You Are In Charge –
Experimentally Testing the
Motivating Power of Holding
a Judicial Office

Christoph Engel /
Lilia Zhurakhovska

MAX PLANCK SOCIETY

You Are In Charge – Experimentally Testing the Motivating Power of Holding a Judicial Office

Christoph Engel / Lilia Zhurakhovska

October 2016,

revised January 2017

You Are In Charge – Experimentally Testing the Motivating Power of Holding a Judicial Office[*]

Christoph Engel
(Max Planck Institute for Research on Collective Goods, Bonn)

Lilia Zhurakhovska
(University of Duisburg-Essen)

Abstract

Apparently judges' decisions are not motivated by maximizing their own profit. The literature uses two strategies to explain this observation: judges care about the long-term monetary consequences for themselves, or individuals who are more strongly motivated by the common good self-select into the profession. We suggest that there is an additional explanation, the "office motive". In a lab experiment, we rule out both traditional explanations by design. Nonetheless authorities do a reliable job at overcoming a social dilemma. Calling the authorities "public official" or "judge" increases their sensitivity towards the degree by which individuals are selfish, and it reduces the effect of their social value orientation (making them more neutral). This suggests that the socially desirable effect is not driven by anger or sympathy with the victims, but follows from the desire to fulfill the expectations that come with the assigned task. We test three extensions: When given an opportunity to announce an explicit policy, judges become less sensitive to the objective degree of reproach, and more sensitive to their social value orientation. If judges are elected or experienced, they react more intensely to norm violations. Experienced judges are more affected by their social value orientation.

JEL: C91, D03, D63, D73, H11, H41, H83, K41
Keywords: judicial behavior, office motive, public-goods experiment, judicial frame, election, experience

1. Introduction

"What Do Judges and Justices Maximize? The Same Thing Everybody Else Does" (Posner, 1993). The title of Judge Posner's 1993 article has provoked a whole literature. Actually, the body of his paper is much more nuanced, as his later contributions are (they are summarized in Posner, 2010). Yet the original question remains still unresolved. In economic parlance, judges are agents, while the people, the state, or the government are/is their principal(s). A huge literature in economics starts from the assumption that a principal-agent relationship creates a problem (for a summary account see Laffont and Tirole, 1993). If the principal cannot condition the agent's payoff on the agent's action, the agent acts such that she maximizes her own payoff, irrespective of the detriment for the principal. The rules meant to protect judicial independence (almost) completely shield judges from intervention. Why does the principal-agent problem nonetheless seemingly not materialize for judges? The two main answers that have been given are: judges are, indeed, selfish but judicial decision-making is sufficiently contained by constraints; judges hold different preferences and self-select into office.

We claim that this literature overlooks a motive that is highly important for, but not specific to, judges. Judges do not behave (plainly) selfishly simply because they have been assigned a public office. In their majority they live up to the expectations that go with holding the office. To put this claim to the test, we design a lab experiment. In the lab, we are able to completely remove institutional and social constraints. We construct an environment where there is a clear and easily observable conflict between selfish behavior and the expectations that go with the office. We give subjects in the role of authorities the opportunity to address a true social problem, at a cost to themselves. We randomly assign the office, thus excluding any possibility of particularly conscientious individuals self-selecting into the position. We rule out any long-term effects on authorities' profit.

To create a social problem that calls for intervention, we rely on a standard design from experimental economics, a linear public good. This design creates a social problem. Individually, each participant is best off if she does not contribute to a joint project. Yet the group to which this participant is randomly assigned is best off if everybody contributes maximally. This experiment has been replicated hundreds of times (for overviews see Chaudhuri, 2011, Ledyard, 1995, Zelmer, 2003). A very robust pattern results: on average, contributions start somewhere in the middle between nothing and everything, but gradually decay over time. The pattern is reversed, though, if participants are given an opportunity to punish each other, even if this is costly (Fehr and Gächter, 2002, 2000, Nikiforakis and Normann, 2008). The threat of punishment disciplines free-riders. We deviate from the standard design in one respect. We randomly assign to each group of four contributors a fifth participant. This participant alone has power to mete out punishment, which is costly for her. She does not benefit from contributions to the public good. Instead she earns a fixed fee and has an additional endowment that she can use for punishment (or keep). The game is repeated. Throughout the experiment, group composition is held constant.

In the literature, the puzzle is presented in the language of a principal-agent-relationship. Once they are in office, judges could abuse the guarantee of independence to increase their personal profit, e.g. by maximizing leisure. Yet it is a principal-agent problem with a special feature: the principal are the people. The assigned task of the agent is to govern the people,

i.e. the principal. It is this special feature that we claim is critical in behavioral terms. A typical judge feels the urge to fulfill the expectations that come with the office, and not to abuse the powers assigned to her for selfish purposes. The design of the experiment preserves this feature of the real life situation.

An important feature of our design is that we test students with various majors, while we want to understand the choices of judges. For us students are not just a convenience sample. In fact, not testing professional judges is critical for identification. Had we found the same results with judges, we would not have known whether results are driven by the fact that special individuals select into becoming judges, or by their exposure to the judicial professional environment. We would not have been able to isolate the office motive. By contrast, we can do so with students, since we can randomly assign the office to one in five participants. Further note that a situation is frequent in judicial practice that shares this feature with our experiment: individuals on one-time jury duty.

If authorities hold standard preferences, i.e., if they maximize their payoff, we have a clear prediction. There is no punishment. Contributions to the public good decay over time, as they would in the absence of the punishment option. In fact, 28% of the authorities behave this way. But the vast majority of 72% are willing to discipline free-riders, although this means a smaller payoff for them. Punishment is explained by two reasonable and socially desirable factors. Participants are the more likely to be punished the less they have contributed. Authorities, on top, care about the degree by which participants deviate from the group average in the previous period. Authorities thus conjointly apply an absolute and a relative definition of free-riding, and target both. Holding an office indeed strongly motivates one to do what is called for.

Ideally, the legislator would want all wrongdoers to be treated equally, irrespective of the particular authority they are facing. The policy preferences of the authority should be immaterial. To test whether policy preferences of the authority have explanatory power, before the start of the main experiment, using a standard test, we elicit the authorities' social value orientation (Liebrand and McClintock, 1988).[1] We find that punishment behavior is significantly moderated by the authorities' personal social value orientation. Note that in our design an authority does not have power to equal out her own payoff. The fact that her personal social value orientation explains her punishment choices shows that she has a preference for *others* behaving in a way, and for *others'* payoffs being distributed in a way that is consistent with the way she would want to be treated herself. This is why we refer to the authorities' social value orientation as a policy preference.

It is normatively desirable that individuals who have no personal interest in the outcome nonetheless discipline others, even if this comes at a cost. But in and of itself this evidence does not suffice to isolate an office motive. The outsiders could just be angry about the anti-social behavior of some individuals, or they could show sympathy with the victims. Thus, in order to find out whether "being in charge" is relevant, we rerun the experiment, but now add a frame. In one treatment, we call the outsider a "public official".[2] In the other treat-

---

1 Note that feedback about the social value orientation test is only given at the very end of the experiment. That way we can ensure that the authorities' decisions in the public-goods experiment are not influenced by the outcome of the first part of the experiment.
2 "Beamter" in the German original.

ment, we call the outsider a "judge".[3] In both treatments, we refer to active participants as "citizens".[4] Both manipulations make outsiders more sensitive to absolute contributions and, to a lesser extent, to the degree by which participants deviate from the group average in the previous period. These effects are more pronounced if we call the outsider a judge. Both manipulations neutralize the socially undesirable effect of the outsiders' personal policy preferences. This shows that we do not only observe a disciplining effect. The data suggest that the socially desirable effect is driven by the socially desirable motive to live up to the expectations that come with the office.

Lab experiments are not meant to faithfully capture all facets of the social phenomenon that motivates them. The situation is simplified up till the cause of the effect of interest can be identified. In a number of extensions, we investigate whether the socially desirable effect is still present if we add characteristic features of judicial decision making. In all extensions we use the "judge" and "citizen" framing.

In our main experiment norms are implicit. All citizens contributing their entire endowment is the efficient choice. If a citizen contributes less than the remaining group members, she exploits them. This gives a judge two norms to choose from: maximum contributions, or contributions at or above the group average in the previous period. In a first extension, we ask judges to announce a punishment policy before the actual game starts. This manipulation gives us the opportunity to see whether inducing accountability has an effect. We indeed find an effect. Judges become less sensitive to the absolute size of contributions, and to the average contributions in the respective group and the previous period. They, by contrast, become highly sensitive to citizens obeying their norm. If contributions are above this norm, the judge's personal social value orientation becomes immaterial.

In many jurisdictions, judges are not appointed, but elected. The fact that they have an explicit mandate by the electoral body could make them even more responsive to the call of duty. But they might also find it more legitimate to impose their personal preferences on their community. To test whether either effect is present, we implement the second extension of our main experiment. There are two candidates for the role of a judge. As in the previous treatment, both of them announce a policy. Yet two additional participants determine by vote which of the two candidates assumes office. The non-elected candidate earns a considerably lower fee, so that being elected is attractive. Citizens are only informed about the policy announcement of the elected judge. We find that elected judges are less sensitive to the absolute size of contributions than appointed judges. In all remaining respects, their behavior is very similar to the behavior of judges who announce a policy.

Jurisdictions differ in their policy for selecting judges. Some legal orders wait until lawyers are fairly advanced in their professional careers. In other legal orders, individuals become judges right after they have finished legal training. But even in the former legal orders, individual judges have some experience on the job. Judicial experience might bring professionalization. With more pronounced professionalization, it might become more natural for judges to distinguish between personal predilections and the normative expectations of the legal order. But responding to the call of duty might also require self-control. The willingness to do what is socially expected might tire with repetition. To test for either possibility, we run a

---

3   "Richter" in the German original.
4   "Bürger" in the German original.

final extension. In this treatment, an assigned judge first decides over 10 periods on a first group of citizens. These citizens are then replaced by a new group, while the judge stays in office. We find that experienced judges are less sensitive to the absolute size of contributions, and to their relative size, compared with the average contribution of the group in the previous period.

The remainder of the paper is organized as follows: in section 2, we summarize the rich literature that tries to explain why judges seemingly defy rational choice theory. From this literature, we derive the hypotheses we test in the BASELINE. In section 3, we explain the design of the BASELINE experiment. Section 4 relates our paper to the experimental literature. Section 5 reports the results from the BASELINE. Section 6 contrasts the original, unframed experiment with the treatments that put authorities into the position of public officials or of judges. Section 7 checks whether the judges' and citizens' choices change if the judge has announced a policy. Section 8 investigates the behavior of judges that have been elected. In Section 9, we analyze the difference between inexperienced and experienced judges. Section 10 concludes with discussion.

2. Are Judges Not Selfish?

Judicial behavior is only one application of what we call the office motive, but it is the most prominent one. In generating our hypotheses, we focus on this application. We do so since a whole literature has tried to dissolve one important aspect of "the puzzle of judicial behavior" (Baum, 1997). "Self-interest theory reaches its low ebb with judges" (Epstein, 1990). This quote summarizes the starting point of the literature very well. Yet why should that be? Usually, rational choice models start with the assumption that everybody is selfish. Maybe not in a narrow sense. But this modeling strategy assumes that one is able to explain, and even predict, behavior if one knows or infers an individual's utility function. Why should that not be possible with judges? The literature has proposed two strategies for reconciling judicial decision-making with the rational choice model. The first strategy highlights the many constraints. In this perspective, judges hold the same utility functions as everybody else. They only have to act in an opportunity structure that constrains their choices so heavily that one may be misled to believe they had different preferences. The second strategy sees judges as utility maximizers, but claims that the judges' utility functions systematically differ from the utility functions of ordinary people.

A first strand of the literature shows that judges align their decisions with public interest because this is a strategy for advancing their personal career (for a survey of this literature see Shepherd, 2011). For instance, they make decisions they hope will improve their chances for promotion (Cohen, 1992, Sisk et al., 1998), reelection (Brace et al., 1998), or reappointment (Shepherd, 2009). In our experiments, we remove any such shadow of the future. The experiments end after the final announced period. Participants singled out to be authorities know they will never have an opportunity to cash in on their success in disciplining the groups.

A similar effect may also result from more subtle selfish motives. Judges are expected to care about prestige (Cooter, 1983) and reputation (Miceli and Cosgel, 1994), both in the wider public (Baum, 2008) and in the legal community (Schauer, 1999), and in particular

within their panel if the decision is taken by a bench (Posner, 2010). All of these should help align their decisions with the expectations of these communities. We also rule out this motive by design. Authorities are randomly selected. They remain anonymous throughout the experiment. Authorities from different groups neither learn their choices, nor how effective other authorities were at disciplining their respective groups. The authorities' payoff is unrelated to the group's performance. That way, authorities can neither derive pecuniary nor nonpecuniary utility from building a reputation.

The fact that judicial decision-making looks as if judges were not selfish is further explained with features of the institutional framework. If judges make controversial decisions, the parties are likely to appeal and decisions can be reversed (Levy, 2005). Judges have been shown to be averse to this (Randazzo, 2008, Smith, 2006, Songer et al., 1994). A further constraint is the relationship between the legislator, or more broadly speaking the government, and the judiciary. While the individual judge is almost perfectly shielded from state intervention into how she decides a single case, government is able to curb the court, e.g. by curtailing the judiciary's budget (Anderson and Shughart, 1989, Macey, 1994), or by limiting the scope of judicial intervention, through changing the law. From the individualistic perspective of rational choice theory, these threats should translate into peer pressure from other judges who fear for their privileges or powers. In the courtroom, there are powerful institutional safeguards against the patent miscarriage of justice, like criminal sanctions for corrupt judges, or the possibility to recuse a judge who is suspected to be partial (Epstein, 1990). Any exercise of judicial power must be couched in legal doctrine, and must pay due regard to precedent (Graham, 2009). In our experiments, all of these institutional safeguards are removed. If these safeguards are responsible for making judicial decisions appear non-selfish, in our experiments we should see straightforward payoff maximization, which in our design would mean no punishment.

In reality, the income of a judge is not related to the decisions she takes in the courtroom. At least in the short run, selfish judges must find alternative ways to serve their individualistic goals like, most prominently, their striving for leisure (Choi et al., 2012, Schauer, 1999, Yoon, 2006). In the lab, manipulating leisure is difficult. We instead impose a small cost on an authority that decides to punish an active participant. This gives us a direct pecuniary measure for selfishness. If authorities hold the standard preferences of economics textbooks, we should see no punishment. All of the foregoing leads to our null hypothesis:

> **H₀:** Authorities do not use the costly option to punish group members.

With the design of our experiment, we also rule out a potential second explanation. The judiciary might attract individuals who just are different from the ordinary citizen. Judges might self-select into this profession because they are not selfish in the ordinary sense. Yet, in our experiments, authorities are randomly selected.

This leaves us with one explanation, for which we can test. It is said that judges have a "sense of honor and duty" (Epstein, 1990). "We expect judges to try to do the right thing" (Stout, 2001). Being a judge is portrayed as participating in a "tournament of virtue" (Solum, 2004). We expect the mental mechanism behind such behavior to be very simple: if a person has been assigned an office, this person tries to live up to the expectations that go with the office.

In our experiments, we put this claim to the test. We randomly single out individuals to be authorities. We make it costly for them to fulfill these expectations. If they ignore these expectations, they expect neither pecuniary nor social sanctions. The only benefit they may derive from effectively disciplining the group they happen to be assigned to is realizing that they are doing a good job. Their own payoff is totally unrelated to the performance of the group; they are not themselves affected by the policy they implement. Those for whom they secure a higher payoff never learn the identity of their benefactor. We even deprive them of the possibility to compare their own performance with the performance of other authorities. We thereby even rule out a competition motive.

In order to test this claim, we must define what, within the setting of our experiment, is meant by fulfilling the expectations that go with the office. For the reasons we explain in the next paragraphs, in our BASELINE and most of the treatments we have deliberately given authorities more than one possibility for interpreting this expectation. If efficiency is the goal, every member of their group should in every period contribute her entire endowment. Alternatively, the authority could interpret the task as disciplining this one group, given local circumstances. In this interpretation, a participant should be the more likely to be punished the more her contribution falls short of the average contribution in this group. This would imply that the authority is not only sensitive to absolute, but also to relative payoffs in the group she disciplines. This leads to our alternative hypothesis:

> **H₁:** Authorities use costly punishment to discipline the group they are assigned to. They are more likely to punish a participant, and they punish her more severely
> a) the less she contributes;
> b) the further her contribution is below the group average in the previous period.

There is a possible downside of judges being motivated by their sense of duty, which we also want to address. The most prominent explanation of judicial behavior sees judges as policy entrepreneurs (for a survey see Friedman, 2005, Posner, 2010). In the legal literature, this was the main claim of legal realism (for a survey see Graham, 2009), while this perspective is called the attitudinal model in the political science literature (Segal and Spaeth, 1993, 2002). Our experiment is not meant to be a complete test of this rich model. We confine ourselves to one implication. Doctrine is almost always ambiguous. Judges, even if they aim at being good judges and treating the parties fairly, might be influenced by idiosyncratic predilections when dissolving this ambiguity.

To test this hypothesis, we have offered authorities two alternative options for defining the norm. To get a handle on personal predilections, before the main experiment we test authorities on the one personality variable that is most closely related to the definition of the norm they are likely to enforce: social value orientation as a measure for the willingness to give up some individual profit for the sake of others. If our expectation holds, we should find

> **H₂:** The more pronounced the personal social value orientation of an authority, the more she is likely to punish low contributions, and deviations from the group average in the previous period, and the more severe the intervention.

3. Design of the Baseline

To test these hypotheses, in a lab experiment we create a social problem that calls for intervention by an authority. We use a standard design from experimental economics, called public goods. All rules of the game are common knowledge from the beginning on. The experiment is anonymous, i.e., the identities of the subjects are never revealed. Our experiment is repeated 10 times. At the beginning of the experiment, the computer randomly assigns roles to participants (authority or active participant, neutrally labeled "participant B" and "participant A"). Participants are randomly assigned to groups. Each group consists of four active participants and one authority. Groups stay constant over the entire experiment.

In the experiment, we use two different currencies: Euro and an experimental currency unit (ECU).[5] All decisions that are relevant for active participants' earnings are calculated in ECU, while the authority earns a fixed fee of 15€ for participating in the experiment (and additionally a maximum of 40 ECU in the experiment). We deliberately give authorities an income above the expected income of active participants.[6] We do so for reasons of external and internal validity. The typical judge has a higher income than the average defendant. We also want to make it unlikely that authorities compare their own income from the experiment with the income of active participants. That way we aim at minimizing any direct effect of the authority's inequity aversion (Fehr and Schmidt, 1999). This is also why we directly express the flat fee in Euro, rather than in the experimental currency units ECU. At the end of the experiment, 1 ECU is converted into .04€ and all earnings are paid out in cash.

Each period consists of three stages. In the first stage, four active participants receive an endowment of 20 ECU. They are free to keep their endowments or to contribute any number of ECU to a public project of their group. Every ECU that is contributed is multiplied by a marginal per capita rate of .4, and paid out to all active group participants. The payoff π of an active participant is thus given by

$$\pi_i = 20 - c_i + .4 * \sum_{k=1}^{4} c_k$$

In this game, individually each active participant k is best off keeping her endowment since this gives her 1 ECU, instead of only .4 ECU. Yet if all active participants make a contribution of $c_i = 20$, every active group participant earns 0.4*80 = 32 ECU, instead of only 20 ECU. However, if three active participants contribute 20 ECU, while the fourth active participant keeps her endowment, the former only each have 24 ECU, while the latter has 44 ECU. This gives each active participant incentives to freeride by not contributing, which creates the policy problem.

In the second stage, a randomly selected, anonymous fifth participant (the authority) receives information about the individual contributions of the four active participants. The authority has power to inflict punishment (neutrally labeled as "reducing the income") on any active participant. She can at most reduce an active participant's period income by 80 ECU, but only has a punishment power of 80 ECU in total (per period) for the whole group.

---

5   In the instructions, we use the term "Taler", an obsolete German currency.
6   If no active participant contributes in any period to the joint project, they earn 8 €; if all always contribute their entire endowments, they earn 12.80 €.

It would already be interesting if authorities did a good job when this does not matter for their personal profit. But we put the office motive to an even stronger test. Whenever an authority decides to punish an active participant, this costs her 1 ECU (or 0.04 €), irrespective of the severity of punishment inflicted on this active participant. Hence, if an authority in one period does not punish any active participant, she earns an additional 4 ECU. An authority that never punishes thus earns an additional 40 ECU (or 1.6 €), on top of the fixed fee of 15 € for assuming the role of authority.

We have chosen not to condition the cost of punishment on severity. We have done so for reasons of external validity. Arguably dismissing a case quickly saves the judge work and hassle. But usually the judge's (non-pecuniary) cost is not directly related to the severity of the punishment she inflicts. External validity is also the reason why the cost is small. Judges do pay a (non-monetary) price for being diligent, but diligence is not critical for the level of their wealth.

In the third stage, the active participants learn the individual contributions, the punishment assigned, and the period income, of all four members of their group.

Before the main experiment, we measure each authority's individual social value orientation, using a standard procedure, the ring measure test (Liebrand and McClintock, 1988). That way we can investigate whether punishment is exclusively explained by the authorities' policy preferences, whether those preferences change how authorities react to active participants' behavior, or whether they do not explain punishment choices at all.[7]

The experiment was conducted in the Cologne Laboratory for Economic Research in 2016. The experiment is programmed in zTree (Fischbacher, 2007). Participants were invited using the software ORSEE (Greiner, 2015). We had 90 student participants of various majors. We have data from 18 fixed groups of 4 active players and 1 authority each. Participants on average earned 17.86€ (20.04$ on the first day of the experiment). For more information about the conduct of this BASELINE and all five treatments (discussed in sections 6-9), see Table 1. All treatments have been run in the same lab, with only days between treatments. Nobody participated more than once.

---

[7] After the main experiment, we also elicit personality traits, using the well-validated 10-item version of the Big5 test (Rammstedt and John, 2007), but do not use this data for analysis. We also administer the social value orientation test and the personality questionnaire to all other participants, but do not use this information for analysis either.

	Baseline	Public Official	Judge	Announce	Elect	Experience First Group	Experience Second Group
N participants	90	90	90	90	128	85	85
N obs/ authorities	18	18	18	18	16	17	17
N candidates					32		
N voters					32		
Periods	10	10	10	10	10	10	10
Frame	none	public official; citizen	judge; citizen	judge; citizen	judge; citizen	judge; citizen	judge; citizen
Announced policy	no	no	no	yes	yes	no	no
Selection of authority	appointed	appointed	appointed	appointed	elected	appointed	appointed
Authority experienced	no	no	no	no	no	no	yes
Punishment cost	1 ECU per punished active player	1 ECU per punished active player	1 ECU per punished active player	1 ECU per punished active player	1 ECU per punished active player	1 ECU per punished active player	1 ECU per punished active player
Maximum possible punishment per group	80	80	80	80	80	80	80
Mean age	23.87	22.67	22.88	22.52	23.39	23.55	23.05
Percent female	57.78	66.67	64.44	56.67	54.69	60.67	58.82
Mean earnings of authorities	23.16	23.20	23.21	22.96	23.21	41.42	
Mean earnings of active participants	16.53	16.98	16.36	17.15	17.82	16.91	14.73

Table 1
Treatments

4. Related Experimental Literature

Most public-goods games with punishment have used decentralized punishment (more from Herrmann et al., 2008). In Engel and Irlenbusch (2010), an additional player has authority to discipline the group. Yet this player benefits from the success of the group, so that successfully managing the group is in the best pecuniary interest of the authority.

The willingness of third parties with no monetary interest to punish others has been studied in different games. Fehr and Fischbacher (2004) find that third parties are willing to punish dictators who give little, and players who defect in a prisoner's dilemma. They explain their

findings with "strong indirect reciprocity": third parties are willing to punish norm violations, even if they are not personally a victim or can expect potential monetary future benefits from punishment (Carpenter et al., 2004, Putnam, 2001, Zhurakhovska, 2014). Questionnaire data suggest an impulse to punish results, stemming from hurt emotions. Charness et al. (2008) play a trust game where a third player may either sanction the trustee for having sent back little, or may reward the trustor for having sent a lot. They find that many third parties are willing to use either option. They do not specify the third parties' motive. Similarly, Zhurakhovska (2014) gives a third party the opportunity to condition her transfer to a trustee on the trustee's history in a trust game. She finds that trustees are rewarded for high relative return transfers, irrespective of the absolute level of investments they receive. She also finds that third parties' general generosity does not correlate with their reward decisions. Leibbrandt and López-Pérez (2012) have an active player choose between two different allocations of a fixed pie between herself and a passive second player in 10 different games. A third party learns the choices and is allowed to reduce the payoff of either of the two parties, at a cost to herself. A substantial fraction of third parties use that power. The authors exploit the fact that they have multiple punishment choices per individual to classify the distribution norm each punisher adheres to. Almenberg et al. (2011) add a third player to several variants of the dictator game. The third player may either punish or reward players of the dictator game, at a cost to herself. The majority of third parties use one of the options; a substantial minority even use both. The authors explain this result with indirect reciprocity. Our experiment builds on this literature. Nikiforakis and Mitchell (2014) pair a dictator with a passive recipient and an observer. The observer has the possibility to invest some of her endowment in either punishing or rewarding the dictator. Observers use both options.

Our experiment makes a minor and a major contribution to the experimental literature on central altruistic third-party punishment. The minor contribution is our choice of game. Our main game is a repeated four-person public good. The fact that there are more than two players in the main game gives the authority scope for developing a policy (she may, for instance, decide only to target complete free-riders, or to target those who contribute less than the rest of the group, or those who undercut the average contribution from the previous period). The fact that the game is repeated makes it possible to see how punishment behavior develops over time, and how it reacts to the experiences the authority has made in the past.

Our major contribution, however, concerns the identification of the underlying motive. To that end, and in contrast to the previous literature, we have designed our experiment such that we can tell apart personal predilections (measuring the authorities' social value orientation) and the intention to fulfill the expectations that go with the official task (punishment choices).

5. Does an Outsider Discipline Free-riders?

As Figure 1 shows, experimental authorities do a good job overall.[8] Contributions stabilize at a high level. Both the frequency and the severity of punishment are substantial, in particular in early periods. Punishment is effective.

---

8 For descriptive information about all treatments, also see Table 1 in the Appendix A.
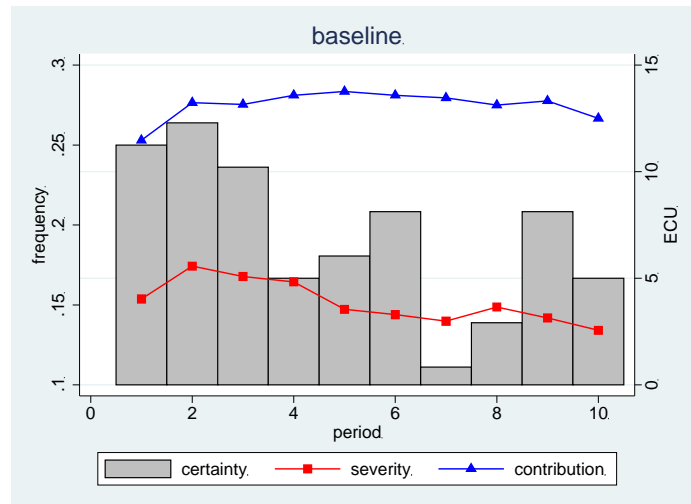
Figure 1
Baseline
certainty: a dummy that is 1 if an active member receives any punishment
severity: number of ECU subtracted from period profit of an active member

These visual impressions are supported by statistical analysis. While 5 authorities behave as predicted by standard theory and never punish, 13 authorities use the punishment option, although this is costly for them. This number of punishing authorities is significantly different from no authority using punishment, as predicted by our null hypothesis **H_0** (one sample sign rank test, N = 18, p = .0003).[9] Active participants contribute significantly more than nothing to the public project (p = .0002)[10] and make a profit significantly higher than 20, which would be the outcome if nobody contributed to the public project (p = .0311).[11] We thus reject the null hypothesis and have

> *Result 1: In a linear public good, outsiders with no pecuniary interest in the performance of active participants effectively discipline them, even if this is costly to the outsider.*

In our alternative hypothesis **H_1**, we further predicted that outsiders would use their punishment power as a technology for enforcing either high absolute contributions or little deviations from the local norm of the group. The latter we capture by the mean contribution, in the respective group, in the previous period. We use the previous period, as for an active participant this is the best information she has if she tries to infer a local norm. We use regressions to test these predictions.

We have two alternative measures for punishment: the probability that an active participant is punished by whatever number of punishment points ("certainty"), and the amount of ECU subtracted from this participant's period profit ("severity"). For parametric analysis, we have challenging data. We capture the data generating process by the following regression equation:

---

9   Our null hypothesis is at the limit of the support, which makes significance testing problematic. We react in the spirit of a confidence interval and additionally report the smallest frequency of punishment that we can reject with our data. It happens to be at 7%, p = .0469.

10 By the same token, we can exclude that average contributions are 10 ECU or less, p = .0294.

11 For profit, the theoretical prediction, assuming standard preferences, is not at the limit of support, which is why we do not need any additional statistical safeguards.

$$pun_{gti}=\beta_0+\beta_1 contr_{gti}+\beta_2\sum_{i=1}^{4} contr_{gi,t\text{-}1} +\varepsilon_g+\varepsilon_{gt}+\varepsilon_{gti}$$

where the dependent variable *pun* refers to the punishment (depending on the model: "certainty" or "severity"), meted out to individual *i* at period *t* in group *g*. This punishment choice by the authority is predicted by a constant, the absolute contribution this individual has currently made (*contr*), the mean contribution her group has made in the previous period ($\sum_{i=1}^{4} contr_{gi,t\text{-}1}$), and three error terms. ε_g captures random variation between groups (and in particular between authorities). ε_{gt} captures the random influence of local circumstances, created by the current choices of the active group members. ε_{gti} is residual error. The three error terms are assumed to be independent of the explanatory variables, and orthogonal to each other. In statistical jargon we thus estimate a mixed effects model (mixing estimates of "fixed effects" [the observed explanatory variables] with random effects). For two reasons we prefer this specification over a model with group fixed effects. First our model perfectly matches the data generating process. Authorities stay the same over the entire experiment. Each period, they simultaneously decide about meting out punishment to 4 different active participants. A fixed effect would lump together both sources of dependence. Moreover fixed effects models are estimated with demeaning, which is why we would only have estimates for the effect of independent variables that change over time, and in particular not of treatment effects, and of social value orientation as a proxy for personal policy preferences. We rely on either or both in all further regressions.

One might further argue that we should adjust functional form. We incentivize the decision to mete out any punishment to an active group member, in the respective period. This dependent variable is binary, which might be captured by a logit model. Many active group members in many periods are not punished at all. This is why our second dependent variable, the number of ECU subtracted from period profit of an active member, is censored from below. This could be captured by a Tobit model. Yet for our story, interaction effects are critical. As is well understood, in such non-linear models, interaction effects (between authorities' punishment policies and their individual social value orientations, in our case) may not be directly interpreted, and their significance cannot be directly derived from the significance of the coefficient (Ai and Norton, 2003). In the paper, we therefore report linear models. For the decision to punish at all ("certainty"), the model predicts the probability. The linear models in particular make it clear which interaction effects are reliable. In Appendix B, we report the corresponding non-linear models.

As Table 2 shows, we find that both the certainty and the severity of punishment are the lower the higher the absolute contribution. However, we do not find a significant effect of average contributions in the previous period. This suggests that authorities strive at enforcing the first norm.

	Certainty	Severity
	Model 1	Model 2
Contribution	-.023***	-.731***
	(.003)	(.088)
Average contribution in t-1	.004	.186
	(.005)	(.143)
Cons	.441***	11.125***
	(.089)	(2.271)
N	648	648

Table 2
Socially Desirable Punishment Motives
linear mixed effects
certainty: a dummy that is 1 if an active member receives any punishment
severity: number of ECU subtracted from period profit of an active member
data from periods 2-10 since one explanatory variable is lagged
standard errors for choices nested in periods nested in groups
standard errors in parenthesis
*** p < .001, ** p < .01, * p < .05

Based on this evidence, we support the alternative hypothesis **H₁** and conclude

> *Result 2: In a linear public good, outsiders with no pecuniary interest in the performance of active participants aim at enforcing high absolute contributions.*

While we were expecting random authorities to discipline the groups to which they are assigned, we were concerned that they might use their powers to impose their personal ideas about group behavior, rather than some neutral standard. To test this hypothesis, we use data from the test for social value orientation that preceded the main experiment. Social value orientation scores do indeed vary considerably. 4 of the 18 authorities hold standard preferences. They only care about personal profit. One authority is rivalistic. She is willing to give up some of her own profit in the interest of increasing the distance in profit between herself and an anonymous counterpart. The remaining 13 authorities are more or less intensely averse to advantageous inequity. If the loss in their own profit is not too pronounced, or if this means a much higher profit for the anonymous counterpart, they are willing to forego the profit-maximizing choice.

The   regressions   in Table 3 show that our concerns have been warranted. The idiosyncratic social value orientation of the respective authority moderates how she reacts to a lower absolute contribution.[12] The interaction effect (Contribution * Authority's SVO) counteracts the main effect of contribution. The more the authority is socially minded herself, the less she is sensitive to the amount contributed by the active participant in question. Model 1 predicts that the authority ignores how much an individual has contributed to the public good if she is mildly socially minded herself,[13] and would even punish those more who have contributed more

---

12 The social value orientation score of authorities is not significantly different from the score of active participants.

13 .042/.003 = 14; social value orientations of authorities in the Baseline range from -5.27 till 26.28, mean 5.83. At a score of 0, the authority is perfectly selfish. A negative score indicates rivalistic preferences: the authority is willing to give up some of her own income in the interest of increasing the payoff difference between herself and others. A positive score indicates social mindedness: the authority is willing to give up some of her own income to reduce the payoff gap between herself and others. A person with a score of 90 would be perfectly altruistic. A person who equalizes payoffs has a score of 45.

themselves if she is very socially minded.[14] This suggests that personal policy preferences partly override the efficiency concern, the more so the more pronounced these policy preferences are.

Interestingly, once we take the personal social value orientation of the authority into account, we also find sensitivity towards relative rather than only towards absolute contributions.[15] Authorities who are not pronouncedly socially minded themselves (have a social value orientation score at or not far from zero) are the more likely to punish the higher the contribution level in the group.[16] Yet the more they are socially minded themselves, the more their own personal attitude towards cooperation dominates (the interaction term partly or fully neutralizes the main effect). Hence the personal social value orientation of the authority moderates the sensitivity of the authority to both the absolute level of contributions, and to relative contributions.[17]

	Certainty	Severity
	Model 1	Model 2
Contribution	-.042***	-1.373***
	(.003)	(.110)
Authority's SVO	-.004	-.453
	(.012)	(.310)
Contribution * Authority's SVO	.003***	.091***
	(.0003)	(.010)
Average contribution in t-1	.019**	.584**
	(.007)	(.191)
Average contribution in t-1 * Authority's SVO	-.002**	-.064**
	(.0008)	(.021)
Cons	.497***	15.250***
	(.122)	(3.233)
N	648	648

Table 3
Socially Undesirable Punishment Motives
linear mixed effects
certainty: a dummy that is 1 if an active member receives any punishment
severity: number of ECU subtracted from period profit of an active member
data from periods 2-10 since one explanatory variable is lagged
standard errors for choices nested in periods nested in groups
standard errors in parenthesis
*** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .1$

To gain a sense of the size of the effects, consider the 51 cases in which an active participant has contributed nothing to the public good. The regressions predict that such individuals run a 38.52% probability of being punished, while this probability is down to 20.74% for those 61

---

14 The net effect, calculated as in the previous footnote, turns positive with a social value orientation score of more than 14.

15 For severity, however, the effect only shows up in the linear, not in the Tobit model; see Table B2 in Appendix B.

16 And the lower the contribution of the active participant in question, since the regression controls for absolute contribution levels.

17 That refers to the relationship between absolute contributions and the contribution level in that group in the previous period.

cases where an individual has contributed the median amount of 15 ECU. The predicted severity of punishment is 8.08 ECU in the first case, and 4.79 ECU in the second case.[18]

Note a further implication of the findings in Table 3. Had authorities with pronounced social value orientation reacted *more* intensely to the absolute level of contributions, and to the average contribution in the previous period, we could not have excluded a self-serving bias (Haisley and Weber, 2010, Loewenstein et al., 1993, Miller and Ross, 1975): the less an authority is socially minded herself, the less she is also willing to incur the cost of punishment, in the interest of improving the payoff of active members.[19] Yet we find that pronounced personal social value orientation makes authorities *less* sensitive to local circumstances. This does not prevent them from punishing free-riders. But their reactions are less fine-tuned to the features of the situation they observe. This does not suggest a self-serving bias.

We therefore also support **H₂** and conclude

> *Result 3: In a linear public good, the willingness of outsiders with no pecuniary interest in the performance of active participants to discipline the group is moderated by their personal preference for equalizing outcomes.*

6. Do Public Officials and Judges Behave Differently?

In our main experiment, the mandate for authorities is implicit. They are the only ones who have power to punish active participants. It is only the situation that calls for their intervention. Strictly speaking, we cannot exclude that they use their assigned power because they are angry ("how can this player be so mean with her partners?"), or because they have sympathy with loyal group members ("I must protect those who sacrifice themselves"). This could, in particular, explain why we find that punishment choices are explained by the authority's idiosyncratic policy preferences. In the interest of inducing a motive, we repeat the main experiment with a single change. In the BASELINE, active participants are labeled "participant with role A", and authorities are labeled "participant with role B". In two treatments, active participants are labeled "citizen", and authorities are either labeled "public official" or "judge". Arguably, participants hold a concept about the legal system in which public officials and judges are disinterested. The legal frame triggers this concept (cf. Engel and Reuben, 2015).

Frames can have a pronounced effect (seminal papers include Deutsch, 1958, Tversky and Kahneman, 1981). This, in particular, holds in dilemma situations (Liberman et al., 2004). Participants may be willing to ignore a dominant strategy (Andreoni, 1995, Ellingsen et al., 2012, Fleishman, 1988, Park, 2000) and may contribute to a public good (Cookson, 2000, Cubitt et al., 2011, Rege and Telle, 2004, Sonnemans et al., 1998). The literature distinguishes two types of frames (Chong and Druckman, 2007, Dufwenberg et al., 2011): valence frames (Levin et al., 1998), and labels. Transforming a take game into a give game (by merely

---

18 Note that the statistical model features random effects. Their effect is factored in when calculating the predicted values.
  In Appendix D, we also reproduce a graph that summarizes average marginal effects of a one unit increase in contribution, and in the average contribution in the previous period, separately for the minimum, the median, the mean, the mean + one standard deviation, and the maximum of authorities' social value orientation scores in our sample.
19 We are grateful to an anonymous referee for pointing this out.

changing the payoffs) is a classic illustration of a valence frame (van Dijk and Wilke, 2000). Calling the same game once a "community game" and once a "Wall Street game" is a classic illustration of mere labeling (Ellingsen, Johannesson, Mollerstrom and Munkhammar, 2012, Kay and Ross, 2003, Liberman, Samuels and Ross, 2004).

For framing to affect behavior, two conditions must be fulfilled: a relevant concept must be stored in memory, and the manipulation of the context must activate this concept (Chong and Druckman, 2007). This might affect choices on two different channels: the frame might trigger guilt (Dufwenberg, Gächter and Hennig-Schmidt, 2011), or it might help active participants and authorities coordinate on the normatively desired outcome (Ellingsen, Johannesson, Mollerstrom and Munkhammar, 2012). If the former channel was critical, we should primarily see an effect on punishment policies (and possibly also on contributions). If the latter channel was critical, we should primarily see an effect on contributions (and possibly in anticipation or reaction on punishment choices). Hence, if we only find an effect on punishment choices, we can exclude the latter channel.

While framing is an established method in social psychology, it has an inherent limitation. Calling one participant a judge, and the remaining participants citizens, evokes an entire schema (Juslin et al., 2003, Smith and Zárate, 1992). This schema is unlikely to be the same for all participants. And even to the extent that there is overlap between the individual schemas, one cannot say which (combination of) element(s) of the schema is critical. In order to minimize this problem we run a separate treatment where we do not label the authority a "judge", but only a "public official". That way we investigate whether the judicial role is special, or whether merely making the role of public authority explicit suffices.

Between the BASELINE and either treatment, we only change one thing: the label. This also holds in the comparison between the JUDGE and the PUBLIC OFFICIAL treatment. We can therefore identify treatment effects. We expect to replicate the office motive that we have already shown in the absence of a frame: the authorities will not maximize their own profit and will strive at disciplining their groups. The interest lies in the correlates to hypotheses H_1 and H_2.[20] We test

> H_3:   a)   If the authority is labeled as "public official" or as "judge", certainty and severity of punishment differ significantly from the BASELINE.
>   b)   These effects are more pronounced if the authority is labeled "judge" than if she is labeled "public official".

We do not have a directed hypothesis for punishment. If the main effect of either label is on active participants (they have an easier time coordinating on a socially beneficial outcome; punishment heightens guilt in free-riders), authorities may reduce punishment in anticipation or reaction. If, by contrast, either frame heightens guilt aversion in authorities (for selfishly saving the cost of punishment), authorities would increase punishment.

Public officials and judges are supposed to serve the public, not their personal interest. The label frames should therefore not only help us isolate the office motive as the cause of punishing free-riders. The labels should also activate the legal norms that expect public officials

---

20   Obviously, our treatment manipulations (which we discuss in this section of the paper as well as in sections 7-9) could also have an effect on the active participants' contributions. However, contributions are not the focus of this paper. In fact, as becomes clear already from the descriptives, contributions do not differ significantly between treatments.

to be unbiased, and to do their duty. Arguably this effect should be even more pronounced in judges. We therefore test

H₄: a) If the authority is labeled as "public official" or as "judge", the authority's personal social value orientation has a smaller effect on her punishment choices than in the BASELINE.

b) The effect is yet smaller if the authority is labeled "judge", rather than "public official".

Figure 2 indicates that "public officials" punish a bit less, in particular in later periods, while "judges" punish a bit more. But we neither find a significant treatment effect with non-parametric nor with parametric statistics, neither on the certainty nor on the severity of punishment.[21]
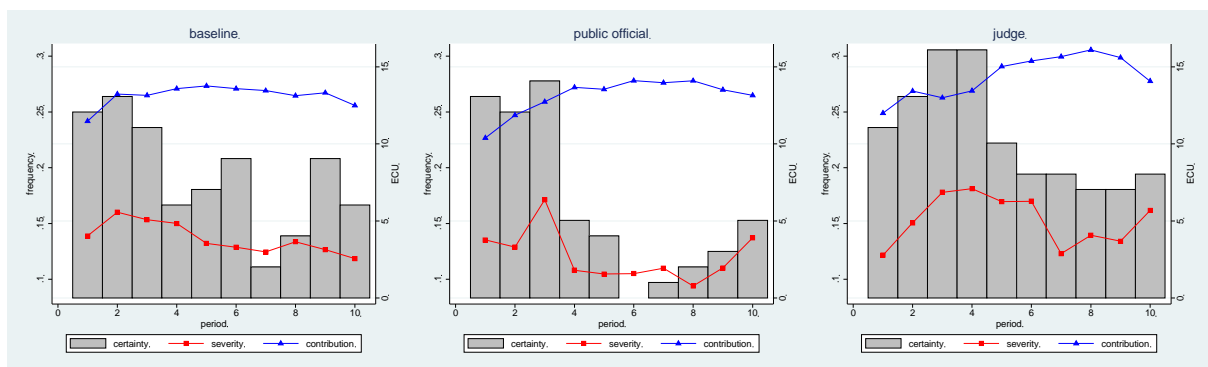

Figure 2
Framing Effects
certainty: a dummy that is 1 if an active member receives any punishment
severity: number of ECU subtracted from period profit of an active member

Yet, as the regressions in Table 4 show, we do find framing effects on the forces driving pun-ishment choices.[22] We begin with treatment effects on socially desirable motives (models 1 and 3). As the highly significant (two-way) interaction effects show, both public officials and judges are substantially more sensitive to the absolute level of contributions. This holds both for the decision to punish at all ("certainty"), as well as for the amount of ECU subtracted from an active participant's period profit ("severity"). For both treatments, we also find sig-nificant treatment effects on the sensitivity towards the contribution level in the respective group.[23] Combining the main effect and the interaction effect, for both dependent variables and in both treatments, we find significant sensitivity towards the contribution level in the respective group,[24] while this is not the case in the BASELINE. Finally, Wald tests show that the judicial frame has an even stronger effect than the public official frame on the sensitivity of

---

21 If we control for the contribution of the respective punishee, and for the average contribution in the group and previ-ous period, we find a significant effect of the "judge" frame, models 1 and 3 of Table 4.

22 In models 1 and 3, we also find significant main effects of the "judge" frame. Yet these effects are insignificant in the non-linear mirror models; see Table B3 in Appendix B.

23 In the Tobit model (see Table B3 in Appendix B), they are only weakly significant.

24 We use a Wald test (on the linear mirror model) to test the hypothesis that the net effect (Average contribution in t-1 + Frame*Average contribution in t-1) is significantly different from 0.

either dependent variable towards the absolute level of contributions, but not towards the contribution level in the respective group.[25]

	Certainty		Severity	
	Model 1	Model 2	Model 3	Model 4
PUBLIC OFFICIAL	.034	-.069	-2.915	-5.760
	(.114)	(.147)	(2.908)	(3.768)
JUDGE	.267*	.114	7.388*	2.493
	(.122)	(.151)	(3.121)	(3.848)
Contribution	-.023***	-.042***	-.727***	-1.340***
	(.003)	(.003)	(.090)	(.116)
PUBLIC OFFICIAL * Contribution	-.015***	.004	-.320***	.152
	(.004)	(.005)	(.128)	(.154)
JUDGE * Contribution	-.029***	-.010*	-.812***	-.287*
	(.003)	(.004)	(.115)	(.142)
Authority's SVO		-.004		-.343
		(.011)		(.274)
PUBLIC OFFICIAL * Authority's SVO		.007		.282
		(.012)		(.308)
JUDGE * Authority's SVO		.020		.640+
		(.013)		(.337)
Contribution * Authority's SVO		.003***		.089***
		(.0003)		(.011)
PUBLIC OFFICIAL * Contribution * Authority's SVO		-.003***		-.060***
		(.0004)		(.015)
JUDGE * Contribution * Authority's SVO		-.003***		-.078***
		(.0004)		(.012)
Average contribution in t-1	.004	.019*	.216	.656**
	(.005)	(.007)	(.145)	(.190)
				(.844)
PUBLIC OFFICIAL * Average contribution in t-1	.010	.003	.460*	.200
	(.007)	(.009)	(.200)	(.250)
JUDGE * Average contribution in t-1	.015*	.006	.462*	.222
	(.007)	(.009)	(.201)	(.245)
Average contribution in t-1 * Authority's SVO		-.002*		-.071***
		(.0007)		(.020)
PUBLIC OFFICIAL * Average contribution in t-1 * Authority's SVO		.0015+		.036
		(.0008)		(.024)
JUDGE * Average contribution in t-1 * Authority's SVO		.0016+		.042
		(.0008)		(.023)
Cons	.441***	.493***	10.673***	13.873***
	(.084)	(.112)	(2.142)	(2.887)
N	1944	1944	1944	1944

Table 4
Framing Effects on Socially Desirable and Undesirable Punishment Motives
linear mixed effects
certainty: a dummy that is 1 if an active member receives any punishment

25 With a Wald test we test the hypothesis that the difference between PUBLIC OFFICIAL*contribution and JUDGE*contribution is significantly different from zero. This is the case for certainty and severity. See also Appendix C for model predictions at the minimum and the median of contributions.

severity: number of ECU subtracted from period profit of an active member
data from periods 2-10 since one explanatory variable is lagged
standard errors for choices nested in periods nested in groups
standard errors in parenthesis
*** $p < .001$, ** $p < .01$, * $p < .05$, [+] $p < .1$

We thus support both claims in hypothesis **H₃** and conclude

> *Result 4: When authorities are called public officials or judges, certainty and severity of punishment are more sensitive to the absolute level of contributions and to the co-operation level in the respective group. Sensitivity to the absolute level of contributions is more pronounced in judges than in public officials.*

In the BASELINE, we have found that punishment choices are moderated by the authorities' personal policy preferences, as measured by their individual social value orientation. Models 2 and 4 of Table 4 investigate whether this socially undesirable moderating effect of personal preferences is also present if authorities are called public officials or judges.[26] In Table 3, we had derived the socially undesirable effect from the fact that the two-way interactions between contributions and the authorities' social value orientation score were significant, as well as the two-way interactions between the level of contributions in the previous period and this score. Consequently the critical coefficients in models 2 and 4 of Table 4 are the three-way interaction terms:[27] does the socially undesirable moderating effect of personal policy preferences go away if the situation is framed?

With sensitivity towards absolute contribution levels, this is straightforwardly true, for both dependent variables. In both regressions, the two-way interaction between contribution and the SVO score of the authority is positive and significant. While in the BASELINE (the reference category) authorities punish those active participants less who have contributed more to the public project (the main effect of contribution is negative and significant), the more the authority is herself socially minded (she has a pronounced SVO score), the more she is likely to punish such active participants regardless (certainty) and the more intensely she punishes them (severity; both two-way interaction terms are positive and significant). Yet, in both regressions, both three-way interactions are significant, have a negative sign and approximately the same size as the two-way interactions. Both public officials and judges are much less influenced by their personal policy preferences. Actually in subsequent Wald tests we can show that they are not influenced at all.[28] In these regressions, we also find that, in the BASELINE, personal policy preferences moderate authorities' sensitivity towards the average contribution level in the respective group (the two-way interaction between Average contribution in the previous period and Authority's SVO is – in the severity regression only weakly – significant and negative: the more the authority is socially minded herself, the less her punishment choices are influenced by the level of cooperation in her assigned group). This effect is partly counteracted if the authorities' role is flagged out as a public or judicial office. However, three-way interactions are not significant at conventional levels.[29]

---

26 As Table 1 in Appendix B shows, on average social value orientation scores of authorities are very similar in the BASELINE and both framing treatments. In a Mann-Whitney test, we do not find a significant difference, neither between the BASELINE and either frame, nor between both frames (N = 36 on each of the three tests).

27 Note that in models 2 and 4 the two-way interactions between frame and contribution, as well as frame and average contribution in t-1, are no longer significant. This suggests that authorities without social preferences (holding an SVO score of 0) are not sensitive to the framing manipulation. Apparently the frame not only holds socially undesirable sensitivity to personal policy in check. Socially minded authorities are also more sensitive to socially desirable determinants of punishment.

28 For the statistical reasons explained when discussing models 1 and 3, we run these tests on the linear mirror models of models 2 and 4. We test whether the net effect of Contribution * Authority's SVO + Frame * Contribution * Authority's SVO is significantly different from zero for either dependent variable and either frame. That is not the case.

29 For certainty, we also find that the net effect of average Contribution in t-1 * Authority's SVO + Frame * Average contribution in t-1 * Authority's SVO is still significantly different from zero. This is not the case with respect to severity.

Higher interactions are genuinely hard to interpret. This is why, in Figure 3, we add a visual representation. [30] Consider the upper left panel: in the BASELINE, authorities are strongly influenced by our proxy for individual policy preferences, i.e., their social value orientation score. As one should have thought, there is no effect of social value orientation for authorities who do not care (have an SVO score of 0). But the more they are socially minded themselves, the less they react to the size of contributions. For authorities with particularly high social value orientation, the effect even reverses the sign: if they contribute more, they are more likely to be punished. Strikingly, this socially undesirable effect of personal policy preferences completely disappears with either frame. For the two framed treatments, the lines are perfectly flat. We just find the socially desirable effect that participants are less likely to be punished the higher their contribution. For severity (lower left panel) framing does not completely erase the effect of the authority's social value orientation (the lines are not completely flat), but the effect is much more moderate. For most levels of social value orientation, we also find a significant difference between the PUBLIC OFFICIAL and the JUDGE frame, in the socially desirable direction.[31] Likewise, the effect of the average contribution of the group in the previous period is considerably less sensitive to personal policy preferences, for both dependent variables (right panels: for the framed treatments slopes are considerably less steep).

---

30 For predicted values at the minimum and the median of contributions of tables 4, 5, 7, and 8 see Appendix C, for average marginal effects of absolute contributions and mean contributions in the previous period, conditional on characteristic values of social value orientation and treatment (for table 3, 5, 7, and 8), see Appendix D, and for power calculations for the 3way interactions (for tables 4, 5, 7, an 8) see Appendix E.
   We refrain from interpreting the remaining significant effects in model 4 since they are not significant in the non-linear mirror model; see Table B4 in Appendix B.
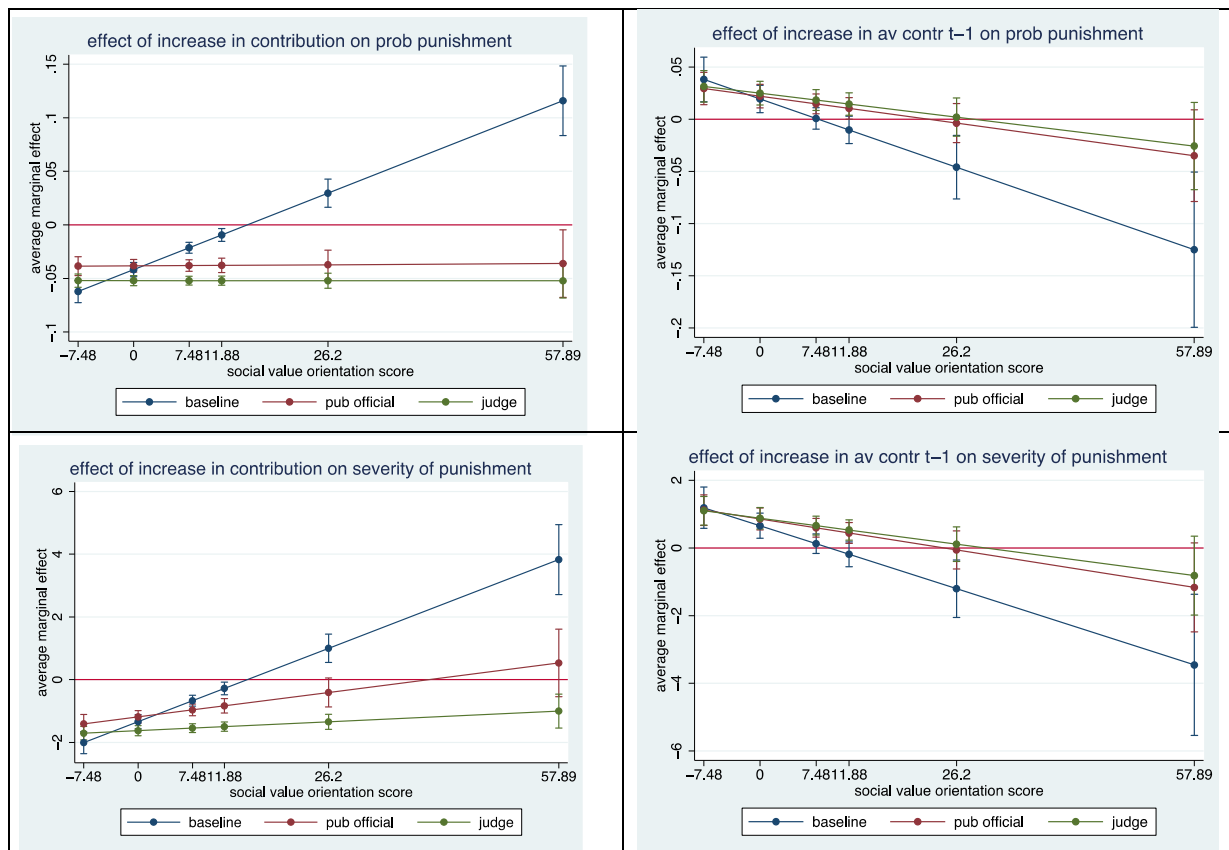31 Confidence intervals do not even overlap.

Figure 3
Visual Representation of Three-Way Interactions from Table 4
All four panels collect average marginal effects of a one unit increase in absolute contributions (left panels) and in the average contribution, in the respective group, in the previous period (right panels), conditional on the lowest social value score of an authority in our dataset (-7.48), zero (i.e., an exclusive preference for profit), the median (7.48), the mean (11.88), the mean plus one standard deviation (26.2) and the highest score we observe (57.89).

For our purposes, the critical coefficients are the three-way interactions: is authorities' sensitivity to socially desirable motives moderated by their personal policy preferences? As the left panels of Figure 3 demonstrate, for sensitivity towards absolute contributions this not only holds overall, but also separately for almost all levels of social value orientation. The moderating effect of the individual's own social value orientation on her sensitivity towards the average contribution in the previous period (right panels) is less pronounced. Overall the frame*average contribution in t-1*SVO terms are only significant at the 10% level for certainty, and they are insignificant for severity. The graphs in Figure 3 suggest that there is a treatment effect if SVO is at the median or higher.[32] But one may wonder whether we had sufficient statistical power to detect these effects.

We estimate regressions with multiple random effects. Therefore power cannot be calculated analytically. Moreover the significance of a three-way interaction depends on the distribution of the exogenous variables, and on the coefficients of the remaining effects. We react by running simulations. In the first step, we generate data that precisely matches the data generating process in our sample, including the distribution of the fixed and random varia-

---

32 The mean of one line is outside the confidence interval of the other.

bles.[33] In the second step we run the regressions of Table 4. For each of 1000 simulation runs, we use a new seed so that all random variables are an independent draw from the underlying distribution. We collect the p-values of the coefficients of interest. The frequency of p-values below a certain α-level is our measure of statistical power.[34] Using this procedure we find that, at the 5% level, we had a 34.6 % chance to find a .0016 effect of frame*average contribution in t-1*SVO on certainty. We had a 62.9% chance to find the corresponding effect on severity. It therefore comes as no surprise that the effects are only weakly significant in model 2, and insignificant in model 4.[35]

We thus largely support **H₄** and conclude

> *Result 5: When authorities are called public officials or judges, certainty and severity of punishment are less sensitive to their personal policy preferences as measured by their social value orientation.*

7. Do Judges Behave Differently When Announcing an Explicit Policy?

By design, experiments are only analogous to the phenomenon in legal reality one wants to understand. External validity is always a concern. The following section is chiefly meant to address one such concern. In the courtroom, judges are bound by the law. There may be room for interpretation. But the norm to be applied is not completely implicit. This we address in a new treatment that builds on the treatment with the judge frame. Before citizens start to make contribution choices, we ask the judge to announce her punishment policy. Specifically, we ask her:[36]

- how severely she intends to punish contributions of 20, 15, 10, 5, and 0 ECU
- whether she intends to punish citizens who contribute less than the average in their group in the previous period
- which of these two norms she prefers, or whether she does not intend to punish at all.

Citizens are informed about these answers. We stress that judges are not bound by these statements. We prefer this design over exogenously imposing a norm. The latter would obviously come closer to judicial practice, but it would imply a strong experimenter demand effect (Zizzo, 2010). We are afraid that the demand effect would overshadow all treatment differences.

To the best of our knowledge, the commitment effect of announcements by disinterested parties has not been studied experimentally. But for interested parties, a pronounced commitment effect of promises has been found (Charness and Dufwenberg, 2010, 2006, Ederer

---

33 This includes the fact that, unsurprisingly, there is high and significant positive correlation between an individual's absolute contribution in the current period and the mean contribution in her group in the previous period (r = .7481). The remaining correlations between exogenous variables are so close to 0 that the simulations assume these variables to be orthogonal.

34 The syntax for these simulations is available from the authors upon request.

35 The simulation for certainty shows that, at the 5% level, we had 76.1% chance to identify the 3way interaction with absolute contributions. For severity, we even had a 95.5% chance to identify this 3way interaction.

36 For the exact wording, please see the instructions in Appendix F.

and Stremitzer, 2014, Ismayilov and Potters, 2016, Vanberg, 2008). If the commitment effect extrapolates to disinterested parties, we might find either more or less punishment, depending on the direction of the effect. Explicitly stating a policy might also make the office motive more salient to judges (on the behavioral effect of salience see Bordalo et al., 2012). On this channel, too, we might see an effect on punishment. On both channels, the effect of personal policy preferences might be further tamed. We therefore predict

H₅: a) If judges are asked to announce their punishment policy, this significantly changes their punishment choices.
b) They are less influenced by their personal social value orientation.

The descriptives suggest no level effect of treatment on certainty and severity, but a decline of both punishment indicators over time (Figure 4). This is indeed what we find statistically. There is no treatment difference on the levels of certainty or severity, neither non-parametrically nor parametrically. But parametrically, we find a negative interaction of treatment with period for both punishment variables.[37]
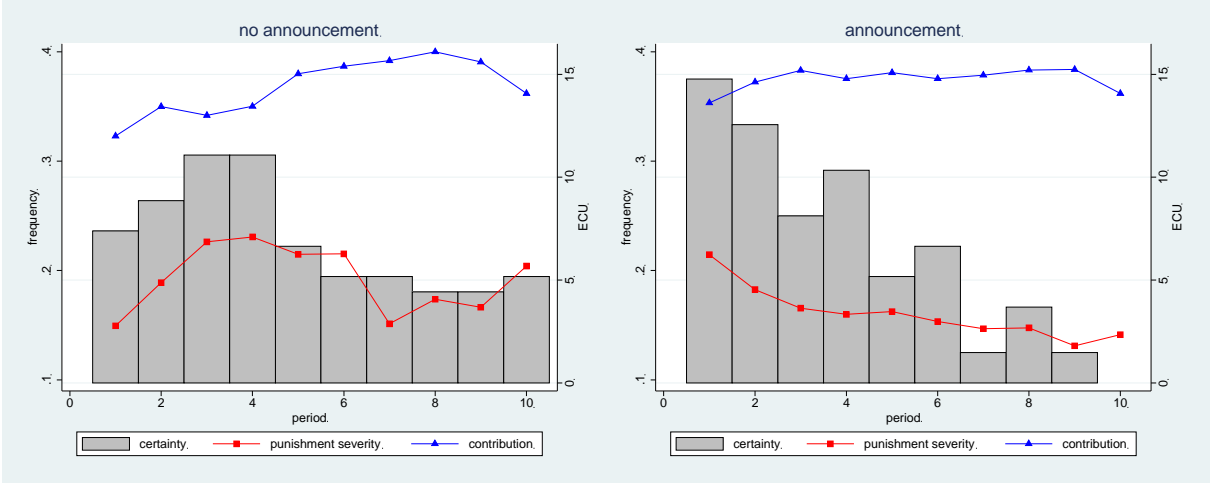

Figure 4
Announcing a Policy
Left-hand side: data from judge frame treatment only
certainty: a dummy that is 1 if an active member receives any punishment
severity: number of ECU subtracted from period profit of an active member

The regressions in Table 5 test in which ways punishment choices are affected by announcing a punishment policy. Models 1 and 3 test in which ways socially desirable motives are affected. As coefficients and significance levels demonstrate, the effect on punishment certainty is more pronounced than on punishment severity. For certainty, the two-way interaction with the absolute contribution level is strong and highly significant. Interestingly the interaction effects partly counteract the respective main effect. This suggests that judges interpret the explicit announcement as empowerment and become less sensitive to what they observe. Nonetheless we still find a significant effect in the normatively desirable direction.[38] The same holds for punishment severity.

---

37 These regressions are available from the authors upon request. In the linear mirror model on severity, the interaction effect is insignificant, though.
38 We run Wald tests of the net effect of Contribution + ANNOUNCE * Contribution.

When comparing the Baseline with the judicial frame, we had found that the frame completely neutralizes the effect of personal policy preferences on sensitivity towards the absolute level of contributions. As model 2 shows, this does not change with the announcement as far as the certainty of punishment is concerned (the three-way interaction is insignificant, as is the net effect of Authority's SVO in this treatment).[39] We do, however, find a significant three-way interaction in the regression explaining punishment severity (model 4). If judges have announced their punishment policy, the severity of their choices is less sensitive to the level of contributions the more they are socially minded themselves.

	Certainty		Severity	
	Model 1	Model 2	Model 3	Model 4
ANNOUNCE	-.121	-.277	-6.382*	-9.354*
	(.133)	(.184)	(3.213)	(4.487)
Contribution	-.052***	-.052***	-1.584***	-1.630***
	(.002)	(.002)	(.068)	(.078)
ANNOUNCE * Contribution	.016***	.020***	.420***	.724***
	(.003)	(.005)	(.108)	(.153)
Judges' SVO		.019*		.301
		(.008)		(.191)
ANNOUNCE * Judges' SVO		-.003		-.011
		(.011)		(.264)
Contribution * Judges' SVO		-.00004		.011*
		(.0002)		(.264)
ANNOUNCE * Contribution * Judges' SVO		-.0002		-.023**
		(.0002)		(.008)
Average contribution in t-1	.021***	.028***	.674***	.887***
	(.006)	(.006)	(.139)	(.156)
ANNOUNCE * Average contribution in t-1	-.009	-.007	-.136	-.349
	(.008)	(.011)	(.198)	(.274)
Average contribution in t-1 * Judges' SVO		-.001*		-.029**
		(.0005)		(.011)
ANNOUNCE * Average contribution in t-1 * Judges' SVO		-.0005		.029[+]
		(.0006)		(.015)
Cons	.691***	.566***	18.250	16.276***
	(.094)	(.102)	(2.801)	(2.528)
N	1296	1296	1296	1296

Table 5
Announcement Effects on Socially Desirable and Undesirable Punishment Motives
linear mixed effects
certainty: a dummy that is 1 if an active member receives any punishment
severity: number of ECU subtracted from period profit of an active member
data from periods 2-10 since one explanatory variable is lagged
standard errors for choices nested in periods nested in groups
standard errors in parenthesis
*** p < .001, ** p < .01, * p < .05, [+] p < .1

---

[39] We find a weakly significant difference in social value orientation scores of judges between the Judge and the Announce treatments (Mann Whitney, N = 36, p = .0596). Note, however, that participants have been randomly selected from a large subject pool, and that the SVO test has been run before the main experiment.

We thus have partial support for **H₅** and conclude

> *Result 6: If judges are asked to announce their policy preference, their punishment choices become less sensitive to the absolute level of contributions and to average contributions in the previous period. Punishment severity becomes sensitive to the judge's personal social value orientation score.*

We have shown that judges are less sensitive to local circumstances when they are asked to announce a policy. The regressions in Table 6 show that this has a very plausible explanation: judges enforce the announced policy. A citizen is much less likely to be punished, and is punished much less severely, if her contribution is above the lowest contribution that this judge has announced not to punish at all. In the regression explaining punishment severity, we also find a main effect and an interaction with the judges' personal social value orientation score. It only matters for the punishment of citizens' who contribute less than the announced minimum.[40] Interestingly, they are punished less severely if the judge is more socially minded herself.

	Certainty		Severity	
	Model 1	Model 2	Model 1	Model 2
Contribution above announced minimum	-.591***	-.627***	-14.485***	-22.994***
	(.037)	(.085)	(.998)	(2.342)
Judges' SVO		.0005		-.340**
		(.004)		(.114)
Contribution above announced minimum * Judges' SVO		.002		.379***
		(.003)		(.094)
Cons	.698***	.715***	15.137**	23.010***
	(.047)	(.098)	(1.145)	(2.663)
N	720	720	720	720

Table 6
Enforcement of Announced Policy
linear mixed effects
certainty: a dummy that is 1 if an active member receives any punishment
severity: number of ECU subtracted from period profit of an active member
standard errors for choices nested in periods nested in groups
standard errors in parenthesis
*** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .1$

8. Do Elected Judges Behave Differently?

Jurisdictions differ in the way how they select judges. While judges are appointed in some jurisdictions, they are elected in others. This difference is not only relevant in terms of external validity. Arguably knowing that she has been elected does also heighten the sense of responsibility in a judge. The electorate has given her the explicit mandate to govern her district. This could create additional accountability. If accountability is made salient, this has been shown to pull individuals towards behavior that is socially expected (Tetlock, 1983, 1985), to make them more sensitive to the possibility of decision errors (Simonson and Nye,

---

40 The interaction effect is also significant in the linear mirror model. A Wald test of the contribution being above the announced minimum + the interaction term is insignificant in this model.

1992) and to tame impulsive reactions (Lerner et al., 1998). Exogenously manipulating accountability has therefore been suggested as a remedy for the governance of regulatory agencies (Seidenfeld, 2001), for reducing the risk of bribery (Serra, 2012), and generally for improving decision quality (Siegel-Jacobs and Yates, 1996). For these reasons, we expect

H₆: a) If judges are elected, this significantly changes their punishment choices.
                b) They are less influenced by their personal social value orientation.

Appointment versus election is not just a framing manipulation. Experimenters therefore have more than words to implement this treatment variation. In the Judge treatment, the authority is appointed by the experimenter. In principle, one could compare this treatment with a second treatment where judges are elected by other (non-active) participants. Yet, to make election meaningful, electors must have something to choose. This is why we do not use the Judge treatment as the BASELINE for this comparison. We rather build on the AN-NOUNCE treatment. We thus hold the announcement procedure constant. But we now have two candidates announce a policy. We also add two voters to each group. The only task of voters is to select one of the two candidates. They are well paid by a fixed fee of 15 €. If votes differ, the judge is randomly selected.[41] The chosen candidate assumes office and – just like the authorities in any other treatment – is equally well paid by 15 €, plus 1 ECU per period for each citizen whom she does not punish. The candidate not chosen only receives a flat fee of 7.50 €. This manipulation is meant to induce serious thought among candidates.

For reasons of external and internal validity, we prefer to distinguish voters from citizens. While elected judges do also rule on their voters, the electorate is much larger than the population of defendants. Even if judges are elected, the individual defendant has to accept a judge on whose selection she had at best a minimal influence. And had we given citizens voting power, we would have tested whether they feel committed to their own democratic choice. This is not what we are interested in. We just want to find out whether the different channel for selecting a judge changes her punishment policy.

Descriptively, both the certainty and the severity of punishment seem to be more pronounced if judges are appointed, in particular in early periods (Figure 5). Yet, statistically, we do not find any significant difference, whether we use non-parametric or parametric statistics. We do not find a significant interaction either between treatment and period on any of these dependent variables.

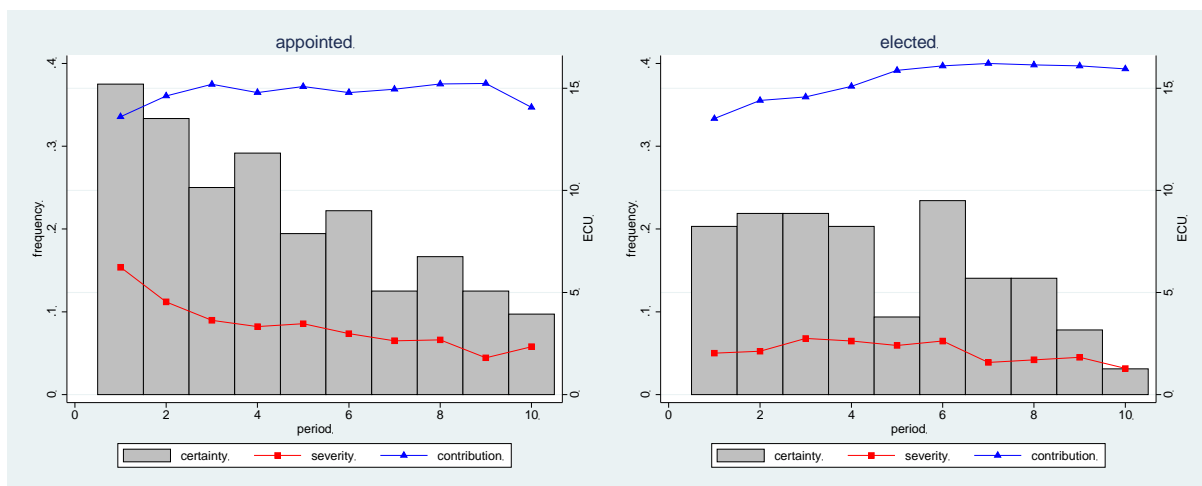41 This happened for 6 of 16 groups.

Figure 5
Appointed vs. Elected Judges
certainty: a dummy that is 1 if an active member receives any punishment
severity: number of ECU subtracted from period profit of an active member

Whether judges are appointed or elected is less important for the sensitivity of authorities to socially desirable and undesirable motives. In model 1 of Table 7, however, we do find that elected judges react more intensely to the absolute level of contributions, in the socially desirable direction (two-way interaction between treatment Eʟᴇᴄᴛ and Contribution). The effect is even more pronounced if the judge is socially minded herself (three-way interaction of treatment Eʟᴇᴄᴛ with Contribution and the Social value orientation score of the judge).[42] This suggests that elected judges feel empowered to act upon their personal (announced) preferences.

---

42 Social value orientation scores of judges do not significantly differ between both treatments. On the non-linear mirror models, we find even more significant effects (see Table B6 in Appendix B), but hesitate to put too much stress on this, given the insignificances in the linear models.

	Certainty		Severity	
	Model 1	Model 2	Model 3	Model 4
ELECT	.076	.027	3.003	3.939
	(.155)	(.218)	(3.085)	(4.471)
Contribution	-.037***	-.032***	-1.128***	-.911***
	(.003)	(.004)	(.068)	(.105)
ELECT * Contribution	-.014***	-.012*	-.021	-.218
	(.004)	(.005)	(.098)	(.139)
Judges' SVO		.017*		.272[+]
		(.008)		(.153)
ELECT * Judges' SVO		.014		.022
		(.012)		(.246)
Contribution * Judges' SVO		-.0003		-.012**
		(.0002)		(.005)
ELECT * Contribution * Judges' SVO		-.0005[+]		.010
		(.0003)		(.008)
Average contribution in t-1	.012*	.021*	.507***	.499**
	(.006)	(.009)	(.117)	(.186)
ELECT * Average contribution in t-1	.007	.010	-.200	-.057
	(.010)	(.014)	(.183)	(.276)
Average contribution in t-1 * Judges' SVO		-.0005		.0006
		(.0004)		(.009)
ELECT * Average contribution in t-1 * Judges' SVO		-.0002		-.009
		(.0007)		(.014)
Cons	.566***	.277[+]	12.311***	7.612*
	(.096)	(.156)	(1.963)	(3.176)
N	1224	1224	1224	1224

Table 7
Election Effects on Socially Desirable and Undesirable Punishment Motives
linear mixed effects
certainty: a dummy that is 1 if an active member receives any punishment
severity: number of ECU subtracted from period profit of an active member
data from periods 2-10 since one explanatory variable is lagged
standard errors for choices nested in periods nested in groups
standard errors in parenthesis
*** p < .001, ** p < .01, * p < .05, [+] p < .1

We thus only find partial support for **H₆** and conclude

Result 7: If judges are elected, the likelihood of punishment is more sensitive to the absolute level of contributions. This, in particular, holds for socially minded judges.

9. Do Experienced Judges Behave Differently?

Some judges are new to the bench; others have a lot of professional experience. On the one hand, experience could make judges more sensitive to the normatively desirable considerations in deciding on punishment, and experience could help them tame the socially undesirable influence of personal policy preferences (Langerock, 1915); (on the effect of professionalization on the behavior of legal officers see Podgor, 1999, Zacharias, 2000). On the other hand, being sensitive to socially desirable, and not being sensitive to socially undesirable influences, could require self-control. Self-control has been argued to be an exhaustible mental resource (Baumeister and Juola Exline, 1999, Muraven and Baumeister, 2000); (but see the recent debate on the effect Carter et al., 2015, Hagger et al., 2016, Inzlicht et al., 2015). Since both potential effects point in opposite directions, we do not have a directed hypothesis, but investigate

> **H₇:**   a) If judges are experienced, this significantly changes their punishment choices.
> b) Experience affects how they are affected by their personal social value orientation.

To test this hypothesis, we proceed in two steps. In the first step, we repeat the design of the experiment with the judge and citizen framing. Authorities are invited for an experiment of longer duration. After the first group of active participants has finished the experiment and received their payments, a new group of citizens enter the lab. The judge and the new citizens are now informed that the exact same experiment will be repeated.[43] The new citizens learn that they are assigned to a judge who has already had the same role for another group of four citizens for another 10 periods.[44]

Descriptively with experienced judges, the frequency of punishment decays over time, but this effect is only weakly significant, Figure 6.[45] We do not find a significant effect of experience on punishment severity either.

---

43   Specifically we repeat the main experiment, not the test for social value orientation.

44   Contrary to the previous sections, here we have a within-subject comparison. We compare data of judges in the first phase and in the second phase. This dependence is captured by judge (group) random effects. Table B7 in the Appendix B reports the respective Logit (for certainty) and Tobit models (for severity).

45   In a regression that interacts judges phase of the experiment with the period of each part of the experiment; the interaction term reaches $p = .060$ in the linear model.
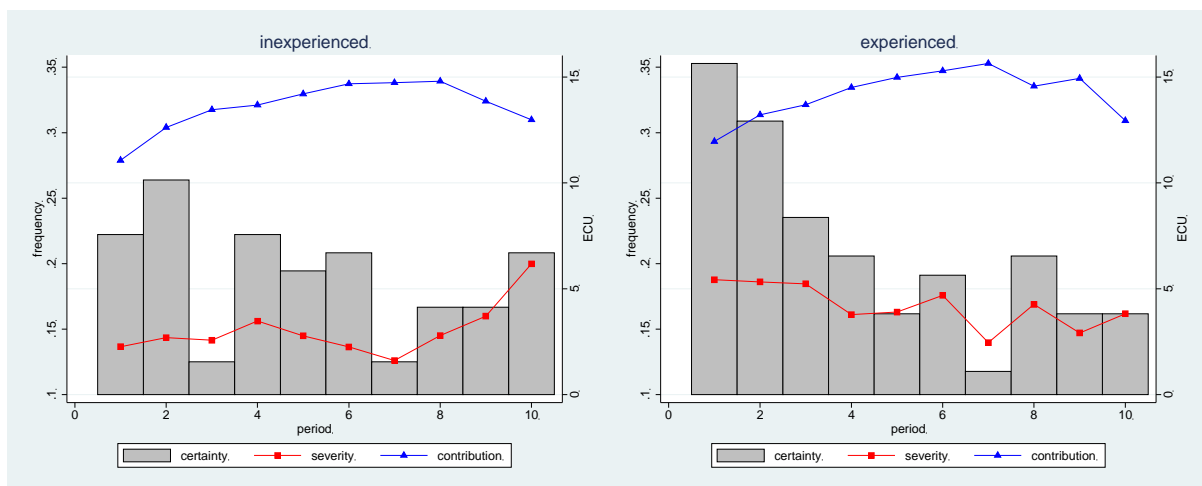
Figure 6
Judicial Experience
certainty: a dummy that is 1 if an active member receives any punishment
severity: number of ECU subtracted from period profit of an active member

We do, however, find clear effects of experience on the sensitivity of judges towards the choices of the subjects they govern. Experienced judges punish more severely in groups that contribute substantially. In the linear model (model 3 of Table 8) we also find that they react more intensely to the absolute size of the contribution.[46] Effects become even more pronounced if we further control for and interact with judges' individual social value orientation (models 2 and 4). For certainty, we now even find a main effect of experience: experienced judges are less likely to punish.[47] More importantly for our story, experienced judges react much more intensely to the absolute size of contributions (two-way interactions with Contribution, for both dependent variables). This normatively desirable effect is, however, the weaker the more the judge is socially minded herself (the significant three-way interactions have opposite sign). We find a similar pattern for reactions to local circumstances, i.e., to average contributions in the previous period. Experienced judges are more sensitive to this information (two-way interactions between Average contributions in the previous period and Experience), but the effect is weakened the more they are socially minded themselves (three-way interactions). This suggests a potential downside. Answering the call of duty might require self-control, which diminishes with repetition.

---

46 However, this effect is insignificant in the Tobit model; see Table B7 in Appendix B.
47 However, this effect is insignificant in the Logit model; see Table B7 in Appendix B.

	Certainty		Severity	
	Model 1	Model 2	Model 3	Model 4
EXPERIENCE	-.109	-.297**	1.935	-5.166
	(.078)	(.111)	(2.224)	(3.216)
Contribution	-.033***	-.022***	-.656***	-.658***
	(.003)	(.004)	(.088)	(.123)
EXPERIENCE * Contribution	.001	-.016**	-.481***	-.872***
	(.004)	(.006)	(.131)	(.183)
Judges' SVO		.004		-.056
		(.004)		(.100)
EXPERIENCE * Judges' SVO		.010**		.173
		(.004)		(.115)
Contribution * Judges' SVO		-.001***		-.0001
		(.0001)		(.004)
EXPERIENCE * Contribution * Judges' SVO		.001***		.023**
		(.0002)		(.007)
Average contribution in t-1	.011*	.001	.264*	.225
	(.004)	(.006)	(.132)	(.180)
EXPERIENCE * Average contribution in t-1	.008	.036***	.694***	1.301***
	(.006)	(.009)	(.188)	(.256)
Average contribution in t-1 * Judges' SVO		.0003		.001
		(.0002)		(.007)
EXPERIENCE * Average contribution in t-1 * Judges' SVO		-.001***		-.035***
		(.0003)		(.010)
Cons	.494***	.454***	8.622***	10.026***
	(.065)	(.090)	(1.804)	(2.573)
N	1260	1260	1260	1260

Table 8
Experience Effects on Socially Desirable and Undesirable Punishment Motives
linear mixed effects
certainty: a dummy that is 1 if an active member receives any punishment
severity: number of ECU subtracted from period profit of an active member
data from periods 2-10 since one explanatory variable is lagged
EXPERIENCE: data from periods 12-20 of final treatment
standard errors for choices nested in periods nested in groups
standard errors in parenthesis
*** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .1$

We thus have partial support for **H$_7$**. From a policy perspective, we have good news and bad news:

Result 8: If judges have experience, they are more sensitive to the absolute level of contributions and to the level of contributions in the group they govern. Yet both effects are the weaker the more the judge is socially minded herself.

10. Conclusion

The most important message of this paper already follows from the Baseline. In the Baseline, we are testing a person that has nothing to gain herself, and even incurs a small cost if she listens to the call of duty. It suffices to provide such a person with the mere opportunity to discipline free-riders. Even if this person is randomly selected and has no social relation whatsoever with the group, she uses the possibility to discipline free-riders. Such authorities are not only willing to exercise the punishment power given to them. More importantly, they do a fairly good job. They target those active group members whose behavior is socially most detrimental: because they make a low contribution to the public good, and because their contribution is below the contribution level that is customary in the group. Remarkably, we replicate the sensitivity of punishment choices to the absolute level of contributions in all treatments. The effect is moderated by treatment conditions, but it is always strong and significant.

But there is a downside. If participants are randomly chosen to be authorities, their punishment choices are moderated by their individual social value orientation. In the context of the public-goods game, this can be interpreted as their personal policy preference. This is where treatment manipulations are most important. This socially undesirable side effect of giving punishment power to a third party is much reduced if this third party is referred to as a "public official" or as a "judge". Arguably framing the interaction as exercising legally contained powers makes individuals more sensitive to their social duties.

In a series of extensions, we step by step bring the situation even closer to judicial decision-making. If the judges get a chance to announce a punishment policy, they become less sensitive to local circumstances. They care less about the absolute level of contributions and the average contributions in the group to which they have been assigned. Announcements also increase the effect of personal policy preferences on punishment policies. Taken together, this suggests that the option to announce their policy is interpreted as empowerment. These effects are counteracted if announcements are made by candidates for judicial office who are selected by external voters. From a normative angle, the effect of experience is mixed. The punishment choices of experienced judges are more targeted. They punish those more who contribute little, and they punish more if the group is in principle behaving well (so that deviations are the more reproachable). But both socially desirable effects are the more diluted the more pronouncedly the judge is socially minded herself.

This paper has been motivated by the long-standing debate over the apparent disconnect between the predictions of rational choice theory and observations from judicial practice. It just does not seem right to assume that judges are plain selfish. By the design of our experiment we exclude the two most prominent attempts at reconciling judicial behavior with rational choice theory. Authorities have truly nothing to gain from doing a good job; actually they have a small amount of money to lose. And participants are randomly assigned to the role of authority, so that results cannot be driven by selection either. In multiple ways and from multiple angles we aim at isolating what we have called the office motive. Being in charge, we argue, has a strong motivating force. We replicate the key findings from the Baseline if we frame the third party as a public official or as a judge. This makes it unlikely that punishment choices are driven by anger or sympathy with the victims of free riding. This inference gains further support from the fact that socially desirable motives get stronger

with these frames. Interestingly, and importantly from a policy perspective, the legal frame has the additional desirable effect of reducing sensitivity towards personal policy preferences. When they are assigned an office, people do not overhear the call of duty, and they do a fairly good job at fulfilling the social expectations that go with the position.

In our experiment, we have investigated the bright side of being in charge. As the (in)famous Stanford prison experiment has made obvious, there is also a dark side. Power may be abused. This may even be enjoyed by those in power (Haney et al., 1973). The legal order is full of safeguards to make this unlikely. To an extent a mere legal frame has already been shown to have this effect (Engel and Reuben, 2015). But a systematic experimental investigation of such safeguards is a topic that must be left to further research.

References

Ai, Chunrong and Edward C. Norton. 2003. "Interaction Terms in Logit and Probit Models." *Economics Letters*, 80, 123-29.

Almenberg, Johan; Anna Dreber; Coren L. Apicella and David Rand. 2011. "Third Party Reward and Punishment. Group Size, Efficiency and Public Goods," N. Palmetti and J. P. Russo, *Psychology of Punishment.* New York: Nova, 73-92.

Anderson, Gary M. and William F. Shughart. 1989. "On the Incentives of Judges to Enforce Legislative Wealth Transfers." *Journal of Law and Economics*, 32, 215-28.

Andreoni, James. 1995. "Warm-Glow Versus Cold-Prickle. The Effects of Positive and Negative Framing on Cooperation in Experiments." *Quarterly Journal of Economics*, 110, 1-17.

Baum, Lawrence. 2008. *Judges and Their Audiences. A Perspective on Judicial Behaviour*. Princeton Univ Pr.

____. 1997. *The Puzzle of Judicial Behavior*. Ann Arbor: University of Michigan Press.

Baumeister, Roy F and Julie Juola Exline. 1999. "Virtue, Personality, and Social Relations. Self-Control as the Moral Muscle." *Journal of Personality*, 67(6), 1165-94.

Bordalo, Pedro; Nicola Gennaioli and Andrei Shleifer. 2012. "Salience Theory of Choice under Risk." *Quarterly Journal of Economics*, 127(3), 1243-85.

Brace, Paul; Melinda Gann Hall and Laura Langer. 1998. "Judicial Choices and the Politics of Abortion. Institutions, Context, and the Autonomy of Courts." *Alabama Law Review*, 62, 1265-304.

Carpenter, Jeffrey P.; Peter Hanns Matthews and Okomboli Ong'Ong'a. 2004. "Why Punish? Social Reciprocity and the Enforcement of Prosocial Norms." *Journal of Evolutionary Economics*, 14(4), 407-29.

Carter, Evan C; Lilly M Kofler; Daniel E Forster and Michael E McCullough. 2015. "A Series of Meta-Analytic Tests of the Depletion Effect. Self-Control Does Not Seem to Rely on a Limited Resource." *Journal of Experimental Psychology: General*, 144(4), 796-815.

Charness, Gary; Ramón Cobo-Reyes and Natalia Jiménez. 2008. "An Investment Game with Third-Party Intervention." *Journal of Economic Behavior & Organization*, 68(1), 18-28.

Charness, Gary and Martin Dufwenberg. 2010. "Bare Promises. An Experiment." *Economics Letters*, 107(2), 281-83.

____. 2006. "Promises and Partnership." *Econometrica*, 74(6), 1579-601.

Chaudhuri, Ananish. 2011. "Sustaining Cooperation in Laboratory Public Goods Experiments. A Selective Survey of the Literature." *Experimental Economics*, 14, 47-83.

Choi, Stephen J.; Mitu Gulati and Eric A. Posner. 2012. "What Do Federal District Judges Want? An Analysis of Publications, Citations, and Reversals." *Journal of Law, Economics, and Organization*, 28, 518-49.

Chong, Dennis and James N Druckman. 2007. "Framing Theory." *Annual Review of Political Science*, 10, 103-26.

Cohen, Marc A. 1992. "The Motives of Judges. Empirical Evidence from Anti-Trust Sentencing." *International Review of Law and Economics*, 12(1), 13-30.

Cookson, Richard. 2000. "Framing Effects in Public Goods Experiments." *Experimental Economics*, 3(1), 55-79.

Cooter, Robert. 1983. "The Objectives of Private and Public Judges." *Public Choice*, 41, 107-32.

Cubitt, Robin P; Michalis Drouvelis and Simon Gächter. 2011. "Framing and Free Riding. Emotional Responses and Punishment in Social Dilemma Games." *Experimental Economics*, 14(2), 254-72.

Deutsch, Morton. 1958. "Trust and Suspicion." *Journal of Conflict Resolution*, 2(4), 265-79.

Dufwenberg, Martin; Simon Gächter and Heike Hennig-Schmidt. 2011. "The Framing of Games and the Psychology of Play." *Games and Economic Behavior*, 73, 459-78.

Ederer, Florian and Alexander Stremitzer. 2014. "Promises and Expectations,"

Ellingsen, Tore; Magnus Johannesson; Johanna Mollerstrom and Sara Munkhammar. 2012. "Social Framing Effects. Preferences or Beliefs?" *Games and Economic Behavior*, 76, 117-30.

Engel, Christoph and Bernd Irlenbusch. 2010. "Turning the Lab into Jeremy Bentham's Panopticon. The Effect of Punishment on Offenders and Non-Offenders," *MPI Collective Goods Preprint 2010/06.*

Engel, Christoph and Alicja Reuben. 2015. "The People's Hired Guns? Experimentally Testing the Motivating Force of a Legal Frame." *International Review of Law and Economics*, 43, 67-82.

Epstein, Richard Allen. 1990. "The Independence of Judges. The Uses and Limitations of Public Choice Theory." *Brigham Young University Law Review*, 827-55.

Fehr, Ernst and Urs Fischbacher. 2004. "Third-Party Punishment and Social Norms." *Evolution and Human Behavior*, 25, 63-87.

Fehr, Ernst and Simon Gächter. 2002. "Altruistic Punishment in Humans." *Nature*, 415, 137-40.

____. 2000. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review*, 90, 980-94.

Fehr, Ernst and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114, 817-68.

Fischbacher, Urs. 2007. "Z-Tree. Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics*, 10, 171-78.

Fleishman, John A. 1988. "The Effects of Decision Framing and Others' Behavior on Cooperation in a Social Dilemma." *Journal of Conflict Resolution*, 32(1), 162-80.

Friedman, Barry. 2005. "The Politics of Judicial Review." *Texas Law Review*, 84, 257-337.

Graham, Randal N.M. 2009. "What Judges Want. Judicial Self-Interest and Statutory Interpretation." *Statute Law Review*, 30(1), 38-72.

Greiner, Ben. 2015. "Subject Pool Recruitment Procedures. Organizing Experiments with Orsee." *Journal of the Economic Science Association*, 1(1), 114-25.

Hagger, Martin S; Nikos LD Chatzisarantis; Hugo Alberts; Calvin Octavianus Anggono; C Batailler; A Birt and M Zwienenberg. 2016. "A Multi-Lab Pre-Registered Replication of the Ego-Depletion Effect." *Perspectives on Psychological Science*, 11, 546-73.

Haisley, Emily C. and Roberto A. Weber. 2010. "Self-Serving Interpretations of Ambiguity in Other-Regarding Behavior." *Games and Economic Behavior*, 68(2), 614-25.

Haney, Craig; W Curtis Banks and Philip G Zimbardo. 1973. "A Study of Prisoners and Guards in a Simulated Prison." *Naval Research Reviews*, 9(1-17).

Herrmann, Benedikt; Christian Thöni and Simon Gächter. 2008. "Antisocial Punishment across Societies." *Science*, 319, 1362-67.

Inzlicht, Michael; Will Gervais and Elliot Berkman. 2015. "News of Ego Depletion's Demise Is Premature. Commentary on Carter, Kofler, Forster, & Mccullough, 2015," *Kofler, Forster, & Mccullough.*

Ismayilov, Huseyn and Jan Potters. 2016. "Why Do Promises Affect Trustworthiness, or Do They?" *Experimental Economics*, 1-12.

Juslin, Peter; Henrik Olsson and Anna-Carin Olsson. 2003. "Exemplar Effects in Categorization and Multiple-Cue Judgment." *Journal of Experimental Psychology: General*, 132, 133-56.

Kay, Aaron C and Lee Ross. 2003. "The Perceptual Push. The Interplay of Implicit Cues and Explicit Situational Construals on Behavioral Intentions in the Prisoner's Dilemma." *Journal of Experimental Social Psychology*, 39(6), 634-43.

Laffont, Jean-Jacques and Jean Tirole. 1993. *A Theory of Incentives in Procurement and Regulation*. Cambridge, Mass.: MIT Press.

Langerock, Hubert. 1915. "Professionalism. A Study in Professional Deformation." *American Journal of Sociology*, 21, 30-44.

Ledyard, John O. 1995. "Public Goods. A Survey of Experimental Research," J. H. Kagel and A. E. Roth, *The Handbook of Experimental Economics.* Princeton, NJ: Princeton University Press, 111-94.

Leibbrandt, Andreas and Raúl López-Pérez. 2012. "An Exploration of Third and Second Party Punishment in Ten Simple Games." *Journal of Economic Behavior & Organization*, 84, 753-66.

Lerner, Jennifer S.; Julie H. Goldberg and Philip E. Tetlock. 1998. "Sober Second Thought. The Effects of Accountability, Anger, and Authoritarianism on Attributions of Responsibility." *Personality and Social Psychology Bulletin*, 24, 563-74.

Levin, Irwin P; Sandra L Schneider and Gary J Gaeth. 1998. "All Frames Are Not Created Equal. A Typology and Critical Analysis of Framing Effects." *Organizational Behavior and Human Decision Processes*, 76(2), 149-88.

Levy, Gilat. 2005. "Careerist Judges and the Appeals Process." *Rand Journal of Economics*, 36, 275-97.

Liberman, Varda; Steven M Samuels and Lee Ross. 2004. "The Name of the Game. Predictive Power of Reputations Versus Situational Labels in Determining Prisoner's Dilemma Game Moves." *Personality and Social Psychology Bulletin*, 30(9), 1175-85.

Liebrand, Wim B. and Charles G. McClintock. 1988. "The Ring Measure of Social Values. A Computerized Procedure for Assessing Individual Differences in Information Processing and Social Value Orientation." *European Journal of Personality*, 2, 217-30.

Loewenstein, George; Samuel Issacharoff; Colin F. Camerer and Linda Babcock. 1993. "Self-Serving Assessments of Fairness and Pretrial Bargaining." *Journal of Legal Studies*, 22, 135-59.

Macey, Jonathan R. 1994. "Judicial Preferences, Public Choice, and the Rules of Procedure." *Journal of Legal Studies*, 23, 627-46.

Miceli, Thomas J. and Metin M. Cosgel. 1994. "Reputation and Judicial Decision-Making." *Journal of Economic Behavior & Organization*, 23(1), 31-51.

Miller, Dale T. and Michael Ross. 1975. "Self-Serving Biases in the Attribution of Causality: Fact or Fiction?" *Psychological Bulletin*, 82, 213-25.

Muraven, Mark and Roy F Baumeister. 2000. "Self-Regulation and Depletion of Limited Resources. Does Self-Control Resemble a Muscle?" *Psychological Bulletin*, 126(2), 247-59.

Nikiforakis, Nikos and Helen Mitchell. 2014. "Mixing the Carrots with the Sticks. Third Party Punishment and Reward." *Experimental Economics*, 17(1), 1-23.

Nikiforakis, Nikos S. and Hans-Theo Normann. 2008. "A Comparative Statics Analysis of Punishment in Public Good Experiments." *Experimental Economics*, 11, 358-69.

Park, Eun-Soo. 2000. "Warm-Glow Versus Cold-Prickle. A Further Experimental Study of Framing Effects on Free-Riding." *Journal of Economic Behavior & Organization*, 43, 405-21.

Podgor, Ellen S. 1999. "The Ethics and Professionalism of Prosecutors in Discretionary Decisions." *Fordham Law Review*, 68, 1511-35.

Posner, Richard A. 2010. *How Judges Think*. Harvard Univ Pr.

____. 1993. "What Do Judges and Justices Maximize? (the Same Thing Everybody Else Does)." *Supreme Court Economic Review*, 3, 1-41.

Putnam, Robert D. 2001. *Bowling Alone. The Collapse and Revival of American Community*. Simon and Schuster.

Randazzo, Kirk A. 2008. "Strategic Anticipation and the Hierarchy of Justice in Us District Courts." *American Politics Research*, 36(5), 669-93.

Rege, Mari and Kjetil Telle. 2004. "The Impact of Social Approval and Framing on Cooperation in Public Good Situations." *Journal of Public Economics*, 88(7), 1625-44.

Schauer, Frederick. 1999. "Incentives, Reputation, and the Inglorious Determinants of Judicial Behavior." *University of Cincinnati Law Review*, 68, 615-36.

Segal, Jeffrey Allan and Harold J. Spaeth. 1993. *The Supreme Court and the Attitudinal Model*. Cambridge ; New York, NY: Cambridge University Press.

____. 2002. *The Supreme Court and the Attitudinal Model Revisited*. Cambridge, UK ; New York: Cambridge University Press.

Seidenfeld, Mark. 2001. "The Psychology of Accountability and Political Review of Agency Rules." *Duke Law Journal*, 51, 1059-95.

Serra, Danila. 2012. "Combining Top-Down and Bottom-up Accountability. Evidence from a Bribery Experiment." *Journal of Law, Economics, and Organization*, 28(3), 569-87.

Shepherd, Joanna M. 2009. "Are Appointed Judges Strategic Too?" *Duke Law Journal*, 58, 1589-626.

____. 2011. "Measuring Maximizing Judges. Empirical Legal Studies, Public Choice Theory, and Judicial Behavior." *University of Illinois Law Review*, 1753-66.

Siegel-Jacobs, Karen and J. Frank Yates. 1996. "Effects of Procedural and Outcome Accountability on Judgment Quality." *Organizational Behavior and Human Decision Processes*, 65, 1-17.

Simonson, Itamar and Peter Nye. 1992. "The Effect of Accountability on Susceptibility to Decision Errors." *Organizational Behavior and Human Decision Processes*, 51, 416-46.

Sisk, Gregory C.; Michael Heise and Andrew P. Morriss. 1998. "Charting the Influences on the Judicial Mind. An Empirical Study of Judicial Reasoning." *New York University Law Review*, 73, 1377-2064.

Smith, Eliot R. and Michael A. Zárate. 1992. "Exemplar-Based Model of Social Judgement." *Psychological Review*, 99, 3-21.

Smith, Joseph L. 2006. "Patterns and Consequences of Judicial Reversals. Theoretical Considerations and Data from a District Court." *Justice System Journal*, 27, 28-46.

Solum, Lawrence B. 2004. "A Tournament of Virtue." *Florida State University Law Review*, 32, 1365-400.

Songer, Donald R.; Jeffrey A. Segal and Charles M. Cameron. 1994. "The Hierarchy of Justice. Testing a Principal Agent Model of Supreme Court-Circuit Court Interactions." *American Journal of Political Science*, 38, 673-96.

Sonnemans, Joep; Arthur Schram and Theo Offerman. 1998. "Public Good Provision and Public Bad Provision. The Effect of Framing." *Journal of Economic Behavior & Organization*, 34, 143-61.

Stout, Lynn A. 2001. "Judges as Altruistic Hierarchs." *William and Mary Law Review*, 43, 1605-27.

Tetlock, Philip E. 1983. "Accountability and Complexity of Thought." *Journal of Personality and Social Psychology*, 45, 74-83.

____. 1985. "Accountability. The Neglected Social Context of Judgement and Choice." *Research in Organizational Behaviour*, 7, 297-332.

Tversky, Amos and Daniel Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science*, 211, 453-58.

van Dijk, Eric and Henk Wilke. 2000. "Decision-Induced Focusing in Social Dilemmas: Give-Some, Keep-Some, Take-Some, and Leave-Some Dilemmas." *Journal of Personality and Social Psychology*, 78, 92-104.

Vanberg, Christoph. 2008. "Why Do People Keep Their Promises? An Experimental Test of Two Explanations." *Econometrica*, 76(6), 1467-80.

Yoon, Albert H. 2006. "Pensions, Politics, and Judicial Tenure. An Empirical Study of Federal Judges, 1869-2002." *American Law and Economics Review*, 8(1), 143-80.

Zacharias, Fred C. 2000. "The Professional Discipline of Prosecutors." *North Carolina Law Review*, 79, 721.

Zelmer, Jennifer. 2003. "Linear Public Goods. A Meta-Analysis." *Experimental Economics*, 6, 299-310.

Zhurakhovska, Lilia. 2014. "Strategic Trustworthiness Via Unstrategic Third-Party Reward – an Experiment," *MPI Collective Goods Preprint 2014/6.*

Zizzo, Daniel John. 2010. "Experimenter Demand Effects in Economic Experiments." *Experimental Economics*, 13(1), 75-98.

Appendix A. Descriptive Statistics

	Baseline	Public Official	Judge	Announce	Elect	Experience First Group	Experience Second Group
Mean probability of punishment	19.31	16.53	22.78	21.81	15.63	19.03	21.02
Percentage of authorities who never punish	28	11	22	11	12.5	23.5	
Mean severity of punishment	3.87	2.69	5.04	3.37	2.10	3.03	4.18
Mean contribution	13.12	13.10	14.37	14.76	15.40	13.61	14.18
Mean period profit of active players	24.00	25.17	23.58	25.49	27.14	25.13	24.33

Table A1
Descriptive Statistics
probability of punishment is also referred to as certainty: a dummy that is 1 if an active member receives any punishment
severity: number of ECU subtracted from period profit of an active member

Appendix B. Non-Linear Functional Form

	Certainty	Severity
	Model 1	Model 2
Contribution	-.290***	-2.995***
	(.049)	(.368)
Average contribution in t-1	.049	.441
	(.079)	(.743)
Cons	-.141	-2.957
		(11.784)
N	648	648

Table B1
Socially Desirable Punishment Motives
certainty: a dummy that is 1 if an active member receives any punishment
severity: number of ECU subtracted from period profit of an active member
certainty: logit; severity: Tobit
data from periods 2-10 since one explanatory variable is lagged
standard errors for choices nested in periods nested in groups
standard errors in parenthesis
*** $p < .001$, ** $p < .01$, * $p < .05$

	Certainty	Severity
	Model1	Model 2
Contribution	-.901***	-4.450***
	(.170)	(.433)
Authority's SVO	-.089	-.682
	(.227)	(1.384)
Contribution * Authority's SVO	.045***	.190***
	(.009)	(.028)
Average contribution in t-1	.462*	1.571
	(.184)	(.969)
Average contribution in t-1 * Authority's SVO	-.039*	-.146[+]
	(.016)	(.087)
Cons	1.111	3.901
	(2.395)	(15.503)
N	648	648

Table B2
Socially Undesirable Punishment Motives
certainty: a dummy that is 1 if an active member receives any punishment
severity: number of ECU subtracted from period profit of an active member
certainty: logit; severity: Tobit
data from periods 2-10 since one explanatory variable is lagged
standard errors for choices nested in periods nested in groups
standard errors in parenthesis
*** $p < .001$, ** $p < .01$, * $p < .05$, [+] $p < .1$

	Certainty		Severity	
	Model 1	Model 2	Model 3	Model 4
PUBLIC OFFICIAL	1.716	-1.176	1.293	-8.877
	(2.521)	(4.260)	(14.987)	(19.411)
JUDGE	2.089	.546	5.918	-4.741
	(2.712)	(4.419)	(15.480)	(.407)
Contribution	-.336***	-1.050***	-2.864***	-4.315***
	(.054)	(.166)	(.325)	(.407)
PUBLIC OFFICIAL * Contribution	-.570***	.031	-2.005***	-.247
	(.123)	(.180)	(.500)	(.558)
JUDGE * Contribution	-.706***	-.212	-1.830***	-.335
	(.140)	(.313)	(.431)	(.506)
Authority's SVO		-.103		-.729
		(.313)		(1.267)
PUBLIC OFFICIAL * Authority's SVO		.317		.957
		(.359)		(1.529)
JUDGE * Authority's SVO		.194		1.282
		(.393)		(1.573)
Contribution * Authority's SVO		.051***		.184***
		(.009)		(.028)
PUBLIC OFFICIAL * Contribution * Authority's SVO		-.073***		-.238***
		(.020)		(.064)
JUDGE * Contribution * Authority's SVO		-.061***		-.181***
		(.015)		(.035)
Average contribution in t-1	.021	.505*	.435	1.452[+]
	(.104)	(.199)	(.663)	(.844)
PUBLIC OFFICIAL * Average contribution in t-1	.374*	.170	1.581[+]	1.137
	(.162)	(.271)	(.936)	(1.180)
JUDGE * Average contribution in t-1	.550**	.260	1.607[+]	1.034
	(.175)	(.265)	(.910)	(1.100)
Average contribution in t-1 * Authority's SVO		-.044*		-.135[+]
		(.018)		(.076)
PUBLIC OFFICIAL * Average contribution in t-1 * Authority's SVO		.021		.073
		(.027)		(.118)
JUDGE * Average contribution in t-1 * Authority's SVO		.038[+]		.075
		(.022)		(.089)
Cons	-.578	.539	-1.293	5.450
	(1.750)	(3.055)	(10.747)	(14.134)
N	1944	1944	1944	1944

Table B3
Framing Effects on Socially Desirable and Undesirable Punishment Motives
certainty: a dummy that is 1 if an active member receives any punishment
severity: number of ECU subtracted from period profit of an active member
certainty: logit; severity: Tobit
data from periods 2-10 since one explanatory variable is lagged
standard errors for choices nested in periods nested in groups
standard errors in parenthesis
*** $p < .001$, ** $p < .01$, * $p < .05$, [+] $p < .1$

	Certainty		Severity	
	Model 1	Model 3	Model 1	Model 4
Announce	.019	.656	-2.330	9.866
	(3.296)	(5.087)	(14.280)	(24.585)
Contribution	-1.194***	-1.168***	-4.303***	-4.324***
	(.179)	(.186)	(.252)	(.278)
Announce * Contribution	.636***	.357$^+$	.869*	-1.153$^+$
	(.169)	(.210)	(.356)	(.683)
Judges' SVO		.099		.417
		(.218)		(.851)
Announce * Judges' SVO		-.105		-.816
		(.294)		(1.286)
Contribution * Judges' SVO		-.007		.004
		(.011)		(.018)
Announce * Contribution * Judges' SVO		.019		.087**
		(.013)		(.030)
Average contribution in t-1	.656***	.714	1.873**	2.274***
	(.163)	(.184)	(.538)	(.622)
Announce * Average contribution in t-1	-.587**	-.469	-.781	-.329
	(.206)	(.308)	(.796)	(1.374)
Average contribution in t-1 * Judges' SVO		-.008		-.052
		(.013)		(.041)
Announce * Average contribution in t-1 * Judges' SVO		-.001		.018
		(.016)		(.066)
Cons	1.635	1.101	6.795	3.507
	(2.451)	(2.842)	(9.862)	(11.726)
N	1296	1296	1296	1296

Table B4
Announcement Effects on Socially Desirable and Undesirable Punishment Motives
certainty: a dummy that is 1 if an active member receives any punishment
severity: number of ECU subtracted from period profit of an active member
certainty: logit; severity: Tobit
data from periods 2-10 since one explanatory variable is lagged
standard errors for choices nested in periods nested in groups
standard errors in parenthesis
*** p < .001, ** p < .01, * p < .05, $^+$ p < .1

	Certainty		Severity	
	Model 1	Model 2	Model 1	Model 2
Contribution above announced minimum	-7.069***	-9.126***	-41.628***	-63.743***
	(1.153)	(2.434)	(3.705)	(7.873)
Judges' SVO		-.093		-.941*
		(.090)		(.394)
Contribution above announced minimum * Judges' SVO		.085		.997**
		(.082)		(.300)
Cons	2.565**	4.667*	11.248**	33.169***
	(.937)	(2.280)	(4.152)	(9.031)
N	720	720	720	720

Table B5
Enforcement of Announced Policy
certainty: a dummy that is 1 if an active member receives any punishment
severity: number of ECU subtracted from period profit of an active member
certainty: logit; severity: Tobit
standard errors for choices nested in periods nested in groups
standard errors in parenthesis
*** $p < .001$, ** $p < .01$, * $p < .05$, [+] $p < .1$

	Certainty		Severity	
	Model 1	Model 2	Model 3	Model 4
ELECT	1.190	-.558	12.104	-6.154
	(3.726)	(5.692)	(16.487)	(27.092)
Contribution	-.550***	-.765***	-3.257***	-5.143***
	(.074)	(.148)	(.244)	(.572)
ELECT * Contribution	-.443**	-.060	-.702[+]	1.597*
	(.152)	(.196)	(.397)	(.697)
Judges' SVO		.121		-.120
		(.191)		(.945)
ELECT * Judges' SVO		.228		1.589
		(.333)		(1.506)
Contribution * Judges' SVO		.010[+]		.083***
		(.006)		(.022)
ELECT * Contribution * Judges' SVO		-.030*		-.126**
		(.012)		(.039)
Average contribution in t-1	.100	.383	1.094*	2.134[+]
	(.122)	(.252)	(.554)	(1.186)
ELECT * Average contribution in t-1	.313	.018	-.143	-1.122
	(.250)	(.384)	(1.047)	(1.687)
Average contribution in t-1 * Judges' SVO		-.014		-.044
		(.011)		(.052)
ELECT * Average contribution in t-1 * Judges' SVO		.020		.032
		(.023)		(.093)
Cons	1.757	-.550	5.361	8.791
	(1.995)	(4.019)	(9.272)	(20.436)
N	1224	1224	1224	1224

Table B6
Election Effects on Socially Desirable and Undesirable Punishment Motives
certainty: a dummy that is 1 if an active member receives any punishment
severity: number of ECU subtracted from period profit of an active member
certainty: logit; severity: Tobit
data from periods 2-10 since one explanatory variable is lagged
standard errors for choices nested in periods nested in groups
standard errors in parenthesis
*** $p < .001$, ** $p < .01$, * $p < .05$, [+] $p < .1$

	Certainty		Severity	
	Model 1	Model 2	Model 3	Model 4
EXPERIENCE	-.318	-1.963	-5.644	-30.428[+]
	(.770)	(1.463)	(8.989)	(15.773)
Contribution	-.303***	-.286***	-3.123***	-3.519***
	(.037)	(.053)	(.355)	(.535)
EXPERIENCE * Contribution	.012	-.186*	-.806	-2.098**
	(.048)	(.083)	(.492)	(.726)
Judges' SVO		.032		.153
		(.046)		(.537)
EXPERIENCE * Judges' SVO		.061		.930[+]
		(.044)		(.490)
Contribution * Judges' SVO		-.001		.014
		(.002)		(.015)
EXPERIENCE * Contribution * Judges' SVO		.009**		.069**
		(.003)		(.024)
Average contribution in t-1	.063	.016	.296	.333
	(.049)	(.073)	(.543)	(.811)
EXPERIENCE * Average contribution in t-1	.040	.344**	1.634*	4.347***
	(.068)	(.120)	(.782)	(1.235)
Average contribution in t-1 * Judges' SVO		.003		.003
		(.003)		(.030)
EXPERIENCE * Average contribution in t-1 * Judges' SVO		-.013**		-.127**
		(.004)		(.041)
Cons	.281	-.328	-2.655	-3.882
	(.769)	(1.212)	(9.432)	(14.061)
N	1260	1260	1260	1260

Table B7
Experience Effects on Socially Desirable and Undesirable Punishment Motives
certainty: a dummy that is 1 if an active member receives any punishment
severity: number of ECU subtracted from period profit of an active member
certainty: logit; severity: Tobit
data from periods 2-10 since one explanatory variable is lagged
EXPERIENCE: data from periods 12-20 of final treatment
standard errors for choices nested in periods nested in groups
standard errors in parenthesis
*** $p < .001$, ** $p < .01$, * $p < .05$, [+] $p < .1$

Appendix C: Predicted Values

treatment	table	min contr	median contr	certainty at min	certainty at median	severity at min	severity at median
BASELINE	Table 4	0 (51)	15 (61)	35.87%	20.37%	8.09	4.61
PUBLIC OFFICIAL	Table 4	0 (40)	15 (73)	35.98%	21.28%	9.50	2.73
JUDGE	Table 4	0 (59)	18 (14)	70.73%	18.65%	20.55	2.49
ANNOUNCE	Table 5	0 (68)	20 (346)	41.20%	5.02%	10.43	0.35
ELECT	Table 7	0 (25)	20 (323)	70.55%	4.24%	15.78	0.08
UNEXPERIENCED	Table 8	0 (27)	15 (84)	57.67%	12.16%	12.15	2.16
EXPERIENCED	Table 8	0 (30)	15 (55)	41.63%	27.54%	10.90	4.12

Table C1
Predicted Reactions to Minimum and Median Contribution
frequency of minimum and median contribution in brackets
min: minimum; contr: contribution;
certainty: predicted probability of punishment;
severity: predicted amount of punishment

Appendix D. Average Marginal Effects

The following graphs report average marginal effects of a one unit increase in contribution and in the average contribution in this group and the previous period, respectively, on either certainty or severity. Means and 95% confidence intervals are reported separately for the minimum, the median, the mean, the mean + one standard deviation, and the maximum, of the social value orientations score of the authority. In the interest of comparability across estimations, we always use the same SVO cutoffs, which are taken from all authorities in the entire sample. The graphs corresponding to Table 4 are reported in the body of the paper as Figure 3.
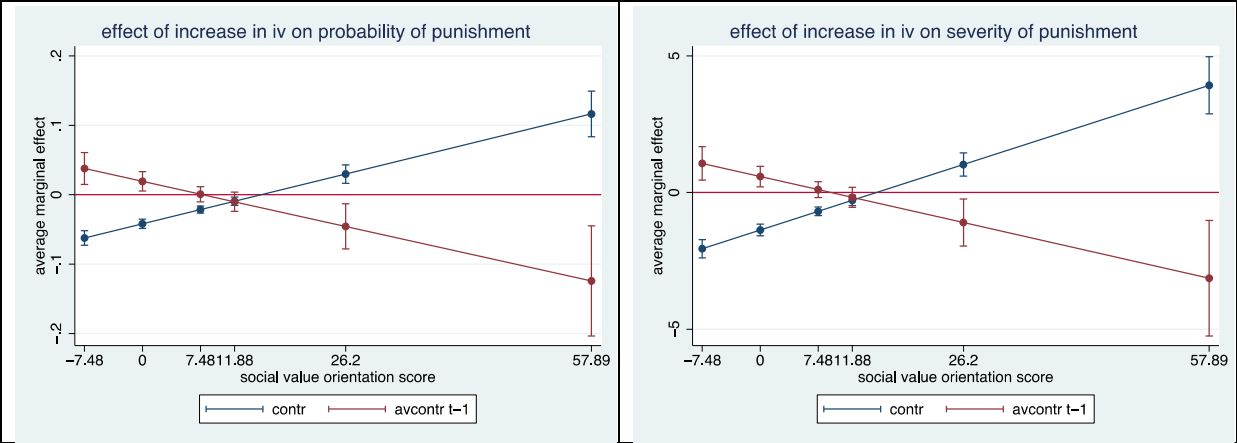
Figure D1
Average Marginal Effects Corresponding to
Table 3

All panels collect average marginal effects of a one unit increase in absolute contributions (left panel) and in the average contribution, in the respective group, in the previous period (right panel), conditional on the lowest social value score of an authority in our dataset (-7.48), zero (i.e., an exclusive preference for profit), the median (7.48), the mean (11.88), the mean plus one standard deviation (26.2) and the highest score we observe (57.89).
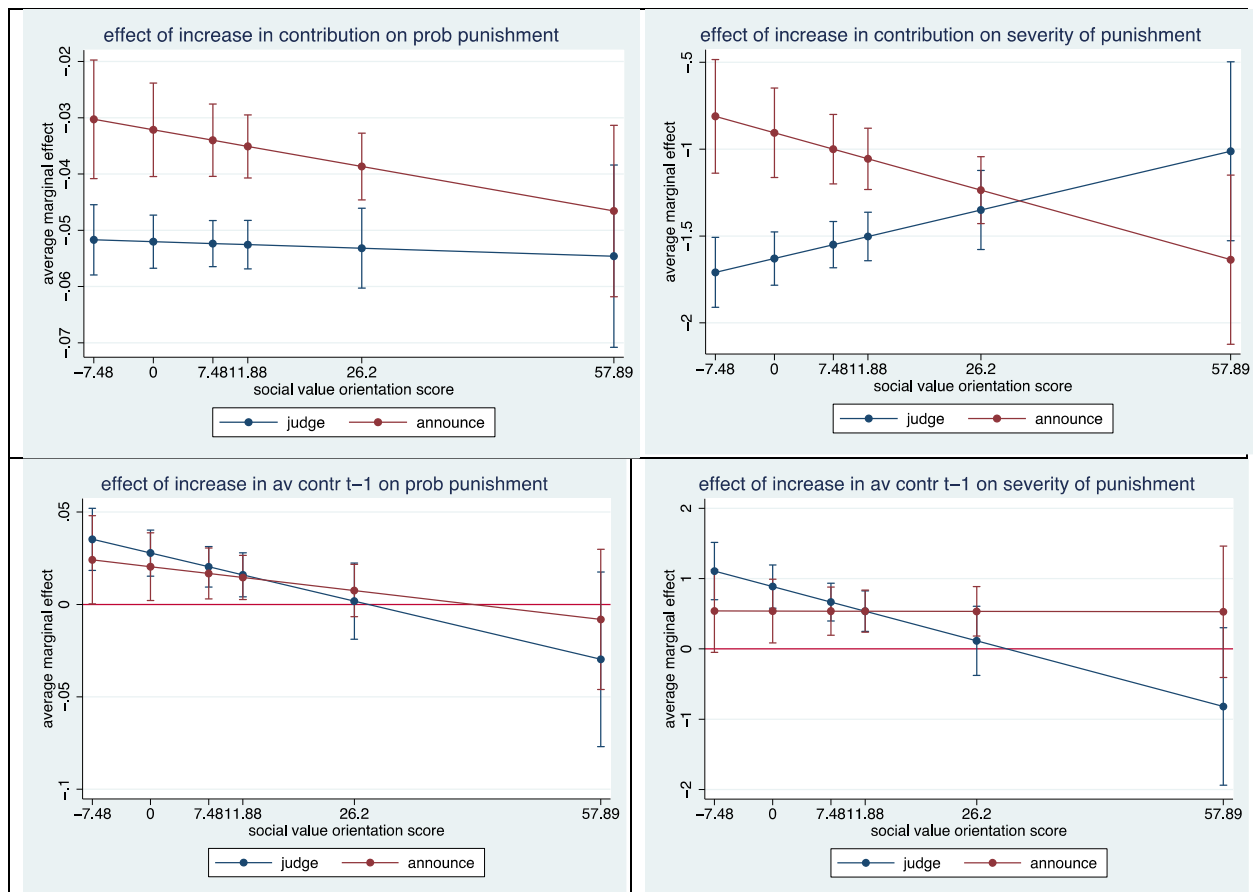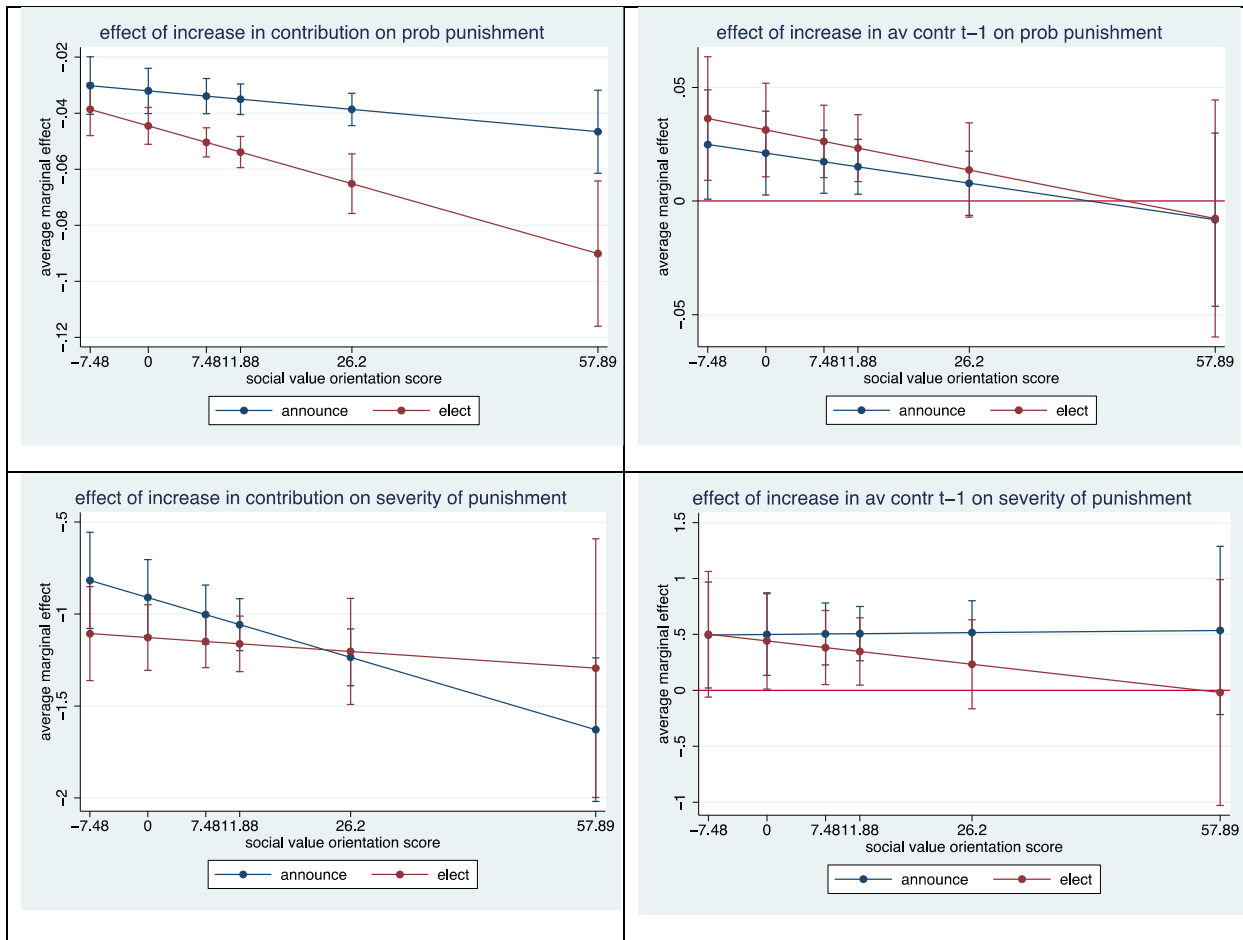
Figure D2
Average Marginal Effects Corresponding to Table 5
All four panels collect average marginal effects of a one unit increase in absolute contributions (left panels) and in the average contribution, in the respective group, in the previous period (right panels), conditional on the lowest social value score of an authority in our dataset (-7.48), zero (i.e., an exclusive preference for profit), the median (7.48), the mean (11.88), the mean plus one standard deviation (26.2) and the highest score we observe (57.89).
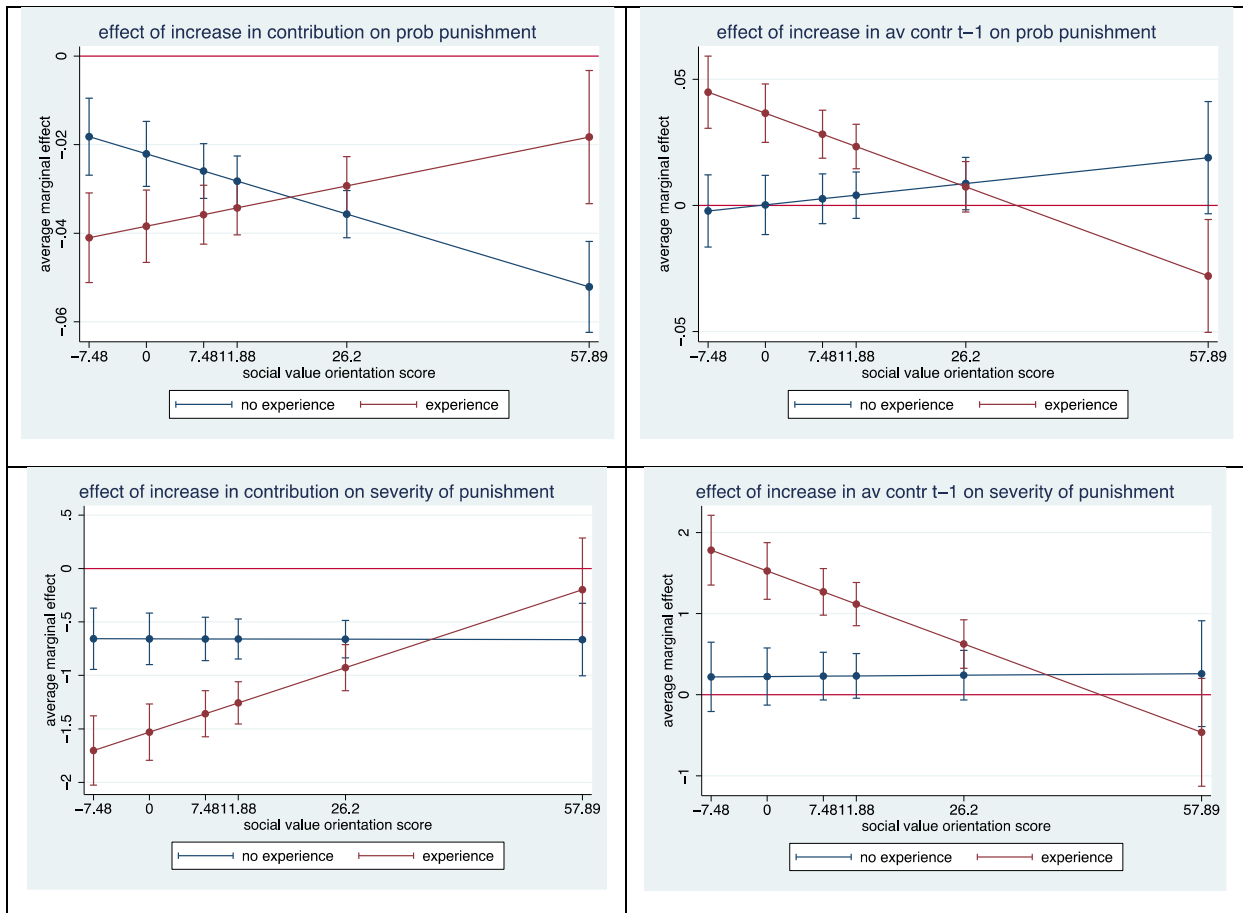
49

Figure D3
Average Marginal Effects Corresponding to Table 7
All four panels collect average marginal effects of a one unit increase in absolute contributions (left panels) and in the average contribution, in the respective group, in the previous period (right panels), conditional on the lowest social value score of an authority in our dataset (-7.48), zero (i.e., an exclusive preference for profit), the median (7.48), the mean (11.88), the mean plus one standard deviation (26.2) and the highest score we observe (57.89).

Figure D4
Average Marginal Effects Corresponding to Table 8
All four panels collect average marginal effects of a one unit increase in absolute contributions (left panels) and in the average contribution, in the respective group, in the previous period (right panels), conditional on the lowest social value score of an authority in our dataset (-7.48), zero (i.e., an exclusive preference for profit), the median (7.48), the mean (11.88), the mean plus one standard deviation (26.2) and the highest score we observe (57.89).

51

Appendix E: Statistical Power for Identifying 3way Interactions

		Certainty				Severity			
		contr		lavcontr		contr		lavcontr	
treatment	table	rep	power	rep	power	rep	power	rep	power
JUDGE	Table 4	<.001	76.1%	.048	34.6%	<.001	95.5%	.098	62.8%
ANNOUNCE	Table 5	.399	9.2%	.432	14.7%	.002	23.8%	.057	41.8%
ELECT	Table 7	.089	13.7%	.817	6.7%	.257	9.8%	.544	10.1%
EXPERIENCED	Table 8	<.001	10.4%	<.001	27.6%	.001	8.3%	<.001	19.4%

Table E1
Statistical Power
certainty: a dummy that is 1 if an active member receives any punishment
severity: number of ECU subtracted from period profit of an active member
contr: absolute level of contributions
lavcontr: average contribution in the group in t-1
rep: reported p-value
power: probability of identifying an effect of the reported size at the 5% level

Appendix F. Instructions

In order to highlight differences between the instructions in the respective treatments, we present all instructions in one file here. Highlights in different colors indicate the treatment differences. Instructions that are not highlighted are seen by participants in all treatments, including the BASELINE. Yellow parts of the instructions are only presented in the BASELINE (no framing), which is discussed in section 5 of the paper. Turquoise parts of the instructions are only presented to subjects in the PUBLIC OFFICIAL treatment, discussed in section 6. Light Green highlights instructions that are specific to all treatments with a judge-frame (discussed in sections 6-9). Pink instructions are only seen by subjects in the ANNOUNCE and in the ELECT treatments (see sections 7 and 8). Only subjects in the ANNOUNCE treatment (section 7) see the Blue-Green instructions, while the Light Grey instructions are presented to subjects in the ELECT treatment only (section 8). Contrary to all other treatments, in the EXPERIENCED treatment (section 9), judges participate in a second part (again a public-goods game with new citizens). The instructions of the second part are highlighted in Dark Grey.

In the following experiment, you can earn a substantial sum of money, depending on your decisions. It is therefore very important that you read these instructions carefully.

No communication of any kind is allowed during the experiment. If you have any questions, please ask us. Disobeying this rule will automatically lead to exclusion from the experiment and all payments.

You will definitely receive a fee of 4 Euro for taking part in the experiment.

The experiment consists of two parts. We will begin by explaining the first part. You will receive separate instructions for the second part.

In the following, we will ask you to make some decisions. In order to do this, **you will be randomly matched with another test subject**. You may give yourself and the other person points in the course of several distribution decisions. In order to do this, you must repeatedly choose **between two distributions, "X" and "Y"**. The points you give yourself are paid to you at the end of the experiment, at an exchange rate of **500 points = 1.00 Euro**. At the same time, you are also randomly matched with **another test subject** who, in turn, allocates points to you by choosing distributions. This person is **not the same one** as the person to whom you can allocate points. The points allocated to you are also credited to your account. The **sum** of all points that you give yourself and those allocated to you by another test subject is paid to you at the end of the experiment, at an exchange rate of 500 points = 1.00 Euro.

The individual decision tasks will look like this:

Possibility X:		Possibility Y:	
Your points	The points of the test subject you are matched with	Your points	The points of the test subject you are matched with
0	500	304	397

X		Y

In this example, if you clicked on X, you would give yourself 0 points and the other person 500 points. If you clicked on Y, you would give yourself 304 points and the other person 397 points.

Please note that we will inform you of the results of this part of the experiment only at the very end of the complete experiment.

Second Part of the Experiment

In this part of the experiment, we speak not of €, but of Taler. Your income from this part of the experiment is hence initially calculated in Taler. The total number of Taler accumulated by you during the experiment will then be converted into € at the end, at a rate of

1 Taler = 4 Eurocents.

In this part of the experiment, there are **two roles: A Citizens and B Public Officials Judges**. Four Participants A Citizens and one Participant B Public Official Judge together constitute one group. One Participant A Citizen is called "Orange", one is called "Blue", another is called "Brown", and one is called "Green". The colors have no particular meaning. They are merely used as short names in order to simplify comprehension of the study. The computer will randomly draw your role, short name and your group members at the beginning of the experiment. This part of the experiment consists of 10 equal periods. **All Participants Citizens and Judges Public Officials retain their roles and their short names throughout the entire second part of the experiment and remain in one group with the same Participants A Citizens and the same Participants B Public Officials Judges.**

Please note that the test subjects matched with you in this part of the experiment are not the same subjects with whom you collaborated in the previous part of the experiment. You are therefore matched with different test subjects.

The following pages describe the exact proceedings of this part of the experiment.

Information on the Exact Proceedings of the Second Part of the Experiment

This part of the experiment has **two steps in each period**. In the first step, the Participants A Citizens decide on contributions to a project. In the second step, the Participant B Public Official Judge can lower the income of the Participants A Citizens.

Step 1:

In the first step of each period, **only the four Participants A Citizens** of a group make a decision. The decision of each Participant A Citizen is relevant for the income of all other Participants A Citizens from his or her group. The income of the Participant B Public Official Judge is not affected by this decision. At the beginning of this step, each **Participant A Citizen** receives **20 Taler**, which we refer to henceforth as the **endowment**. As a Participant A Citizen, you are required to decide how many of the 20 Taler you wish to contribute to a **project**, and how many you decide to keep for yourself.

If you are a **Participant A Citizen**, **your income from Step 1** is made up of two parts:

(1) the Taler that you have kept for yourself (**"income from endowment retained"**)
(2) the **"income from the project"**. *(The income from the project is calculated as follows: your income from the project = 0.4 times the total sum of the contributions to the project.)*

If you are a **Participant A Citizen**, your **income from Step 1 in each period** is calculated as follows:

(20 Taler) – (your contribution to the project) + 0.4*(total sum of contributions to the project)

      Income from endowment retained   **Your income from the project**

The **income of all Participants A Citizens of a group from the project** is calculated **according to the same formula**, i.e., each Participant A Citizen draws the same income from the project.
(If, for instance, the sum of the contributions from all Participants A Citizens amounts to 60 Taler, you and all other Participants A Citizens receive an income from the project of 0.4*60 = 24 Taler each. If the Participants A Citizens have contributed a total of 9 Taler to the project, you and all other Participants A Citizens will each receive 0.4*9 = 3.6 Taler as an income from the project.)

For every Taler that you keep for yourself, you earn an income of 1 Taler. If instead you contribute one Taler of your endowment to your group's project, the sum of the contributions to the project rises by 1 Taler, and your income from the project rises by 0.4*1 = 0.4 Taler. However, this means that the income of all other Participants A Citizens in your group also rises by 0.4 Taler each, so that the total income of the group rises by 0.4*4 = 1.6 Taler. Through your contributions to the project, the other Participants A Citizens thus also earn something. In turn, you also earn something from the contributions of the other Participants A Citizens to the project. For every Taler that another Participant A Citizen contributes to the project, you earn 0.4*1 = 0.4 Taler.

Please note that the Participant B Public Official Judge cannot contribute to the project and does not draw any income from the project.

Step 2:

At the beginning of the second step, the four ==Participants A== ==Citizens== and the ==Participant B== ==Public Official== ==Judge== are told how much each of the four ==Participants A== ==Citizens== with the short names "==Orange==", "==Blue==", "==Brown==" and "==Green==" has contributed to the project in step 1.

In the second step, **only the ==Participant B== ==Public Official== ==Judge==** decides. As ==Participant B== ==Public Official== ==Judge==, you may **lower the income of each ==Participant A== ==Citizen==, or else leave it as it is**.

As a ==Participant B== ==Public Official== ==Judge==, you are now required to decide, for **each** of the four ==Participants A== ==Citizens==, whether and, if so, by how many Taler you wish to lower the player's income. You have to enter something. If you do not wish to alter the income of a certain ==Participant A== ==Citizen==, please enter 0. If you choose a figure higher than 0, you will reduce the income of the respective ==Participant A== ==Citizen== by this number of Taler.
However, you may only reduce the incomes of all ==Participants A== ==Citizens== by a <u>maximum total</u> of 80 Taler.
(For instance, as a ==Participant B== ==Public Official== ==Judge==, you may reduce the income of a ==Participant A== ==Citizen== by 80 Taler, leaving the incomes of the other ==Participants A== ==Citizens== unchanged. Alternatively, for example, you might lower the income of each ==Participant A== ==Citizen== by 20 Taler.)

As a ==Participant B== ==Public Official== ==Judge==, you may also lower the incomes of the ==Participants A== ==Citizens== by less than 80 Taler.
(For instance, you can reduce the income of a ==Participant A== ==Citizen== by 30 Taler, the income of another ==Participant A== ==Citizen== by 5 Taler, and leave the income of the remaining ==Participants A== ==Citizens== unchanged.)

You may also abstain entirely from reducing the income of the ==Participants A== ==Citizens==.

If you are a ==Participant A== ==Citizen==, your **total income in each period** from both steps is therefore calculated as follows:

(Income from Step 1) – (Reduction from Step 2)

Please note that the income at the end of Step 2 can also be negative for the ==Participants A== ==Citizens==. This can happen when the income reduction is higher than the income from Step 1.

The income of the ==Participant B== ==Public Official== ==Judge== does **not** depend on the income of the ==Participants A== ==Citizens== from Step 1 or on the income from the project. Participant B **receives a fixed payment of 15.00 Euro for taking part in the second part of the experiment.**

Further, the ==Participant B== ==Public Official== ==Judge== is given 4 Taler for every period in which he or she has not reduced the income of any of the ==Participants A== ==Citizens==. For every decision that lowers the income of a ==Participant A== ==Citizen== by at least one Taler, the ==Participant B== ==Public Official== ==Judge== is given one Taler less. If a ==Participant B== ==Public Official== ==Judge== reduces the income of a ==Participant A== ==Citizen== in one period by at least one Taler, it is irrelevant for the

Participant B income Income as Public Official Income as a Judge by how many Taler he or she has reduced the income of this Participant A Citizen in total. **For each period, therefore, he or she can earn a Participant B income Income as Public Official Income as a Judge of 0 Taler, 1 Taler, 2 Taler, 3 Taler, or 4 Taler, in addition to the fixed payment.**

The **Participant B income Income as Public Official Income as a Judge from Step 2 in each period** is therefore calculated as follows:

(4 – [Number of Participants A Citizens whose income he or she reduces]) Taler

---

and

The total income of the **Participant B Public Official Judge** for the **entire second part** is:

15.00 Euro

Once all the Participants A Citizens and B Public Official Judge have reached their decisions, you will be told your period income and your current total income on your screen at the end of each period. Thus, each Participant A Citizen is informed by how many Taler the income of each of the four Participants A Citizens with the short names "Orange", "Blue", "Brown" and "Green" has been reduced by the Participant B Public Official Judge.

This part of the experiment ends after period 10.

Please note that not every player takes part in the second part of the experiment. This preliminary stage determines which candidate will take part, and in which role, in the second part of the experiment.

At the beginning of the preliminary stage, each participant is informed of the role that the computer randomly assigns: **citizen, voter**, or **potential judge**. (In case you are a citizen, you will additionally be informed of the short name that is randomly assigned to you.) For the preliminary stage, the computer randomly assigns to each group 2 voters, 4 citizens, and 12 potential judges.

- The participants who have been assigned the role of **voter take part in the preliminary stage**. In this preliminary stage, they decide by voting, for their group, which of the two potential judges shall be given the role of judge. The voters **do not take part in the second part of the experiment**.
- The participants who have been assigned the role of a **potential judge make an announcement in the preliminary stage** and **participate in the second part of the experiment** in the role of judge after the preliminary stage. **can be elected as a judge by the voters**. If a potential judge is elected as a judge, he or she **will participate as a judge in the second part of the experiment**. If a potential judge **is not elected as a judge**, he or she **will not participate as a judge in the second part of the experiment**.
- The participants who have been assigned the role of **citizen** do not make any decisions in the preliminary stage. They **take part in the second part of the experiment**, in their role as citizens, once the preliminary stage is over.

The preliminary stage proceeds as follows:

1. The **potential judges announce how they intend**, if they are elected as judges, to make use of their opportunity to reduce the citizens' income **in the second part of the experiment in Step 2**.
2. The **voters** are informed of the announcements of the potential judges and then **vote for the judges**.
3. The **citizens are informed of the intention of their elected judge** to make use of his or her opportunity to reduce the citizens' income in the second part of the experiment in Step 2.

Announcement of the Potential Judges:

The announcement of the judges consists of the answers to the following questions:

1. By how many Taler would you reduce, in the second Step, the income of a citizen who has contributed the following to the project in Step 1?

a) The citizen has contributed **20**.	
b) The citizen has contributed **15**.	
c) The citizen has contributed **10**.	
d) The citizen has contributed **5**.	
e) The citizen has contributed **0**.	

f) The citizen's contribution corresponds to the **average sum** that all citizens have contributed **in the previous period**. (*If, for example, we are in period 8, the "previous period" is period 7.*)	

2. Which criterion is the most important one to you in your decision to reduce a citizen's income in Step 2? *Please decide on one option.*

    a. **The absolute number of Taler of the citizen's contribution**: I would reduce a citizen's income if that person contributes less than a particular value.

    b. **The relative number of Taler of the citizen's contribution**: I would reduce a citizen's income if that person has contributed less, on average, than all other citizens in the previous period.
(*If, for example, we are in period 4, the "previous period" is period 3.*)

    c. **I would never reduce a citizen's income.**

Please note: a(n) elected judge is never actually obliged to decide, in the second part of the experiment, as he or she announced.

Voting Mechanism:

Immediately after all potential judges have answered both questions, the voters are informed of the answers that the two potential judges in their group have given.

For each group of 2 voters, 4 citizens, and 2 potential judges, the voters elect one of the two potential judges as judge. If in one group both potential judges have been given the same number of votes, a draw will decide which of them will be allowed to be the judge.

The potential judges who have not been elected receive a flat payment of 7.50 €. They do not participate in the second part of the experiment. They must wait until the second part of the experiment has ended. The voters also receive a flat payment, this time of 15.00 €. After the voters have cast their votes, they may not make any further decisions. They, too, must wait until the second part of the experiment has ended. The citizens and elected The citizens and judges participate in the second part of the experiment and are paid according to the instructions to this part of the experiment.

Please note: judges and citizens retain their roles, and stay in their respective groups, in all ten periods of the second part of the experiment. The preliminary stage is not repeated.

Instructions

In the following experiment, you can earn a substantial sum of money, depending on your decisions. It is therefore very important that you read these instructions carefully.

No communication of any kind is allowed during the experiment. If you have any questions, please ask us. Disobeying this rule will automatically lead to exclusion from the experiment and all payments.

You will definitely receive a fee of 4 Euro for taking part in the experiment.

In this experiment, we speak not of €, but of Taler. Your income from this experiment is hence initially calculated in Taler. The total number of Taler accumulated by you during the experiment will then be converted into € at the end, at a rate of

1 Taler = 4 Eurocents.

In this experiment, there are **two roles: citizens and judges**. Four citizens and one judge together constitute one group. One citizen is called "Orange", one is called "Blue", another is called "Brown", and one is called "Green". The colors have no particular meaning. They are merely used as short names in order to simplify comprehension of the study. **Please note: immediately before this experiment, we already conducted the same experiment in the same lab. The participants who had the role of judge in the previous experiment also have the role of judge in this experiment. The other participants are given the role of citizen.** If you are a citizen, the computer will randomly draw your short name at the beginning of the experiment. It will also match you with the other members of your group. This experiment consists of ten equal periods. **All citizens and judges retain their roles and their short name throughout the entire experiment and remain in one group with the same citizens and the same judge.**

This experiment has **two steps in each period**. In the first step, the citizens decide on contributions to a project. In the second step, the judge can reduce the income of the citizens.

The following pages describe the exact proceedings of the experiment.

Step 1:

In the first step of each period, **only the four citizens** of a group make a decision. The decision of each citizen is relevant for the income of all other citizens from his or her group. The income of the judge is not affected by this decision. At the beginning of this step, each **citizen** receives **20 Taler**, which we refer to henceforth as the **endowment**. As a citizen, you are required to decide how many of the 20 Taler you wish to contribute to a **project**, and how many you decide to keep for yourself.

If you are a **citizen**, **your income from Step 1** is made up of two parts:

(1) the Taler that you have kept for yourself (**"income from endowment retained"**)
(2) the **"income from the project"**. *(The income from the project is calculated as follows: your income from the project = 0.4 times the total sum of the contributions to the project.)*

If you are a **citizen**, your **income from step 1 in each period** is therefore calculated as follows:

(20 Taler) – (your contribution to the project) + 0.4*(total sum of contributions to the project)

    Income from endowment retained   **Your income from the project**

The **income of all citizens of a group from the project** is calculated **according to the same formula**, i.e., each citizen draws the same income from the project.
*(If, for instance, the sum of the contributions from all citizens amounts to 60 Taler, you and all other citizens receive an income from the project of 0.4*60 = 24 Taler each. If the citizens have contributed a total of 9 Taler to the project, you and all other citizens will each receive 0.4*9 = 3.6 Taler as an income from the project.)*

For every Taler that you keep for yourself, you earn an income of 1 Taler. If instead you contribute one Taler of your endowment to your group's project, the sum of the contributions to the project rises by 1 Taler, and your income from the project rises by 0.4*1 = 0.4 Taler. However, this also means that the income of all other citizens in your group rises by 0.4 Taler each, so that the total income of the group rises by 0.4*4 = 1.6 Taler. Through your contributions to the project, the other citizens thus also earn something. In turn, you also earn something from the contributions of the other citizens to the project. For every Taler that another citizen contributes to the project, you earn 0.4*1 = 0.4 Taler.

Please note that the judge cannot contribute to the project and does not draw any income from the project.

Step 2:

At the beginning of the second step, the four citizens and the judge are told how much each of the four citizens with the short names "Orange", "Blue", "Brown" and "Green" has contributed to the project in step 1.

In the second step, **only the judge** decides. As a judge, you may **lower the income of each citizen, or else leave it as it is**.

As a judge, you are now required to decide, for **each** of the four citizens, whether and, if so, by how many Taler you wish to lower the player's income. You have to enter something. If you do not wish to alter the income of a certain citizen, please enter 0. If you choose a figure higher than 0, you will reduce the income of the respective citizen by this number of Taler. **However, you may only reduce the incomes of all citizens by a <u>maximum total</u> of 80 Taler.** *(For instance, as a judge, you may reduce the income of a citizen by 80 Taler, leaving the incomes of the other citizens unchanged. Alternatively, for example, you might lower the income of each citizen by 20 Taler.)*

As a judge, you may also lower the incomes of the citizens by less than 80 Taler. *(For instance, you can reduce the income of a citizen by 30 Taler, the income of another citizen by 5 Taler, and leave the income of the remaining citizens unchanged.)*

You may also abstain entirely from reducing the income of the **citizens**.

If you are a **citizen**, your **total income in each period** from both steps is therefore calculated as follows:

(Income from Step 1) – (Reduction from Step 2)

Please note that the income at the end of step 2 can also be negative for the citizens. This can happen when the income reduction is higher than the income from Step 1.

The income of the **judge** does **not** depend on the income of the citizens from Step 1 or on the income from the project. The judge **receives a fixed payment of 15.00 Euro for taking part in the experiment.**

Further, the judge is given 4 Taler for every period in which he or she has not reduced the income of any of the citizens. For every decision that lowers the income of a citizen by at least one Taler, the judge is given one Taler less. If a judge reduces the income of a citizen in one period by at least one Taler, it is irrelevant for the judge's income by how many Taler he or she has reduced the income of this citizen in total. **For each period, therefore, he or she can earn a judge's income of 0 Taler, 1 Taler, 2 Taler, 3 Taler, or 4 Taler, in addition to the fixed payment.**

The **judge's income from Step 2 in each period** is therefore calculated as follows:

(4 – [Number of citizens whose income he or she reduces]) Taler

and

The total income of the **judge** for the **entire experiment** is:

15.00 Euro

Once all the citizens and the judge have reached their decisions, you will be told your period income and your current total income on your screen at the end of each period. Thus, each citizen is informed by how many Taler the income of each of the four citizens with the short names "Orange", "Blue", "Brown" and "Green" has been reduced by the judge.

This part of the experiment ends after period 10.