Preface

In economic theory, a public-good game is an abstract description of a class of situations in which the individual pursuit of one's private interest by all actors leads to an outcome leaving everyone worse off than had they cooperated. While the game is a theoretic construct, the class of situation it describes is very real, in the sense that there is an abundant number of instances in everyday life displaying the characteristics of a public-good game. Examples can be found at every level from the small-group problem of doing the dishes in a shared flat over the community-level question of who volunteers to take on an honorary office, to the grand scale, when it comes to the exertion of effort to reduce global warming by reducing one's energy usage and litter production.

Both anecdotal evidence and scholarly research show that these problems are not to be solved easily.[1] Therefore, a number of institutions have been proposed that may contribute to their solution. One of the most prominent candidates is certainly the introduction of punishment opportunities, so that players may sanction others' misbehaviour.[2] However, the mechanisms' ability to foster the collective interest has been challenged recently from a number of different angles. The present collection of articles takes on some of these challenges, trying to give some hints at the robustness of the punishment mechanism.

One challenge directly pertains to the level of cooperation achieved. In two experimental studies conducted independently with different subject pools, Denant-Boemont et al. (2007) and Nikiforakis (2008) show that a peer-punishment mechanism is no longer able to enhance cooperation levels if punished players are given an opportunity to retaliate. Later studies such as Nikiforakis and Engelmann (2009) or the third study in this tome conducted by Andreas Nicklisch and myself provide evidence that this does not need to be the case. The second article in the present collection provides an explanation for the pattern of findings, showing in a model grounded in evolutionary game theory that the breakdown of cooperation – as reported

[1] For an overview of experimental studies on the public-good problem, cf., e.g., Ledyard (1995).

[2] Seminal papers include Yamagishi (1986), Ostrom, Walker, and Gardner (1992), and Fehr and Gächter (2000, 2002).

in the earlier studies – is, in fact, a comparatively likely outcome in a situation with exactly one retaliation stage – as employed by both studies. On the other hand, for situations with multiple retaliation stages, such as in the later two studies and in one treatment of Denant-Boemont et al. (2007), our model predicts an increase in the likelihood of a cooperative outcome when compared to the situation with a single retaliation stage. Assuming that retaliation opportunities, once they exist, rarely are going to be limited to a single stage, the punishment mechanism can be said to have passed this first test of robustness.

A second challenge does not question the mechanism's ability to foster cooperation, but points to the fact that the increase in cooperation levels comes at a cost to both the punishing and the punished players, and therefore, to society as a whole. There has been a considerable number of studies pointing out that in the presence of punishment opportunities, groups often perform worse in terms of their aggregate payoffs than groups that do not have access to this mechanism, unless the time-horizon is sufficiently long.[3] Moreover, the destruction power of the mechanism is often such that in early periods of an experiment, players in a sanctioning environment even perform worse than in the non-cooperative equilibrium.

This gives rise to two questions. First, many public-good games are played repeatedly, as can be easily seen in case of the examples provided above. In some of these games, an actor's cababilities to contribute to the public good are influenced by earlier decisions. If a very cooperative player already holds several honorary offices at his local club, he might not have the resources necessary to contribute to the club's annual festivity. In a situation with punishment opportunities, the resources destroyed through punishment actions may further reduce the actors' contribution capabilities. If, furthermore, early punishment is strong and therefore, wasteful, societies in such environments may be caught up in a poverty-trap situation that leads to falling wealth levels in spite of positive cooperation rates. This is the focus of the first study contained in this dissertation, conducted by Özgür Gürerk, Bettina Rockenbach, and your author. By and large, the punishment mechanism also passes this second test of robustness: on average, an existing early-round disadvantage of the groups in the punishment treatment when compared to those in the corresponding control treatment is made up by the second half of the experiment, and there is considerable evidence that it would turn into a significant advantage, were the time-horizon extended by a few additional rounds.

[3] Gürerk et al. (2006), Nikiforakis and Normann (2008), Gächter, Renner, and Sefton (2008).

Second, if institutions are subject to evolution, would a mechanism that weakens its implementing society temporarily but strongly survive the selection process? The last article in this collection, written by Bettina Rockenbach and your author, deals with a closely related question. If law-makers are given full discretion over the institutions within a society, what institutions would they choose and – with respect to our main topic – would they choose to employ a punishment mechanism? How does the answer change when these law-makers may successively change and improve their institutions? This latter question is even more interesting in light of studies such as Gürerk, Irlenbusch, and Rockenbach (2006) or Gürerk (2009). These studies show that many people instinctively shy away from environments providing punishment opportunities; in the latter study, they do so even though they know that the mechanism is successful in fostering cooperation. What we find in the study contained in this dissertation is rather surprising in this respect. Rather than shying away from sanctions, all law-making groups make use of a punishment mechanism in the initial round. After that, a split occurs: some of the groups try to improve their sanctions mechanism economising on wasted resources through a number of avenues, others abandon this avenue altogether. In the final period, only half of all groups choose to make use of sanctions. However, those who do, do so successfully. In other words, punishment may be an adequate solution to public-good problems, but only if its rules are well-designed. The success of groups in the peer-punishment treatments of studies with a sufficiently long time horizon as in Gürerk, Irlenbusch, and Rockenbach (2006) or Gächter, Renner, and Sefton (2009), on the other hand, suggests that laboratory subjects tend to be apt at administering punishment well.

Finally, a different but related debate pertains not to the performance of whole groups in punishment environments, but rather to those engaging in sanctioning the misbehaviour of others. In a recent article, Dreber et al. (2008) show that punishers perform worse than players who do not engage in costly punishment; from an evolutionary perspective, punishment strategies would consequently be maladaptive and should vanish. In other words, the mechanism for 'keeping alive' punishment as a strategy would need to be found in something else than merely its material payoff in public-good settings. An important step on the search for this mechanism is to learn more about the process guiding punishment behaviour.

Past research has mostly implicitly assumed that punishment behaviour is guided by the potentially punished player's deviation from either the group's average or the punisher's own contribution.[4] In a recent article, Carpenter

[4]Examples of studies focusing on the group average are Fehr and Gächter (2000, 2002),

and Matthews (2009) go one step further and statistically infer which of a number of potential reference points is used in experimental subjects' punishment decisions. Their striking result is that an 'absolute norm' in the sense of a player's deviation from a constant number is able to explain the data better than the deviation from any standard pertaining to a group's behaviour. Nevertheless, different decisions seem to be attached to completely different such constants. In the third study of the present collection, Andreas Nicklisch and your author extend this work in several important dimensions: using multiple punishment stages and self-contained episodes of interaction, we disentangle the effects of retaliation, norm-related punishment, and antisocial actions driven by other motivations, such as spite or competitive thinking. An additional treatment provides data on the norms bystanders use in judging punishment actions. We find qualified support of the findings of Carpenter and Matthews: there is strong evidence for the influence of an absolute cooperation norm. This norm is subjects' full endowment which – in contrast to the findings of Carpenter and Matthews (2009) – is consistent over decisions, iterations, and roles (punisher or bystander). At the same time, subjects are prompted to increase both the punishment probability and its severity in the first iteration if the player to be punished has deviated more from the full-contribution norm than the punisher him- or herself. Being present only in the first iteration, this effect suggests a high level of personal investment in the public-good dilemma, whereas in higher iterations, the emotional focus seems to shift to other players' punishment behaviour. Taken together, our results suggest that no single process determines behaviour in all punishment-related decisions. Rather, sanctioning behaviour on the first punishment stage is notably distinct from that in all other decisions. The fact that the full-contribution norm performs best for all decisions suggests that there is, indeed, a common understanding of what constitutes acceptable behaviour among our subjects. This understanding seems to be more general than anything determined by the present interaction.

What are the main lessons to be learnt from this dissertation? First and foremost, punishment mechanisms are a double-edged sword. However, when carefully designed, they can enhance cooperation and efficiency (cf. studies I and IV). This feature of punishment mechanisms is furthermore robust to a number of perturbations. Neither the introduction of dynamics (study I), retaliation and counter-retaliation (study II), nor competing other

Anderson and Putterman (2006), and Sefton et al. (2007), while Herrmann et al. (2008), Egas and Riedl (2008), Sutter et al. (2008), or Reuben and Riedl (2009) focus on the punisher's own contribution.

institutions (study IV) will change this. However, the working mechanism – and thus, the precise conditions under which the mechanism will be effective – is still far from clearly understood. The third study in this dissertation is a first step towards the examination of this question. It analyses the processes giving rise to punishment and suggests that more than a single process is at work. To examine these processes closer should be the aim of future studies. In particular, there are three issues to be addressed. The first two pertain to the processes our third study makes out: (i) the individual processes have to be identified even more clearly, in conjunction with further analyses of which behaviour arises endogenously from the situations examined and which is determined by general norms and behavioural dispositions; (ii) a deeper understanding of the interaction of these processes is needed. The third issue concerns the impact on the punished players' behaviour, leading to the central challenge emanating from this dissertation: (iii) the characteristics of successful punishment (i.e., punishment that induces increases in cooperation levels and efficiency) have to be pinpointed. Meeting this challenge is crucial for the design of punishment institutions in general. As long as we do not properly understand the determinants of successful punishment, we will have to put up with the risk of poverty traps, as our first study readily shows.

As you will note, the order in which the studies appear in the dissertation deviates from the order in which they have been introduced in this preface. The reasons for this are rather trivial: I felt somehow uneasy with starting off a volume that is predominantly experimental in nature with the only non-experimental study contained. And second, I thought that the study with Bettina Rockenbach on institution design would provide a very natural final paper, in light of its exploratory nature that builds a ready bridge to future research. In light of the fact that all four studies are self-contained, I am positive this minor change will not obfuscate the main conclusions conveyed.