Cooperation under Peer Sanctioning Institutions

Collective Decisions, Noise, and Endogenous Implementation

Funding for this research was provided by the Netherlands Organization for Scientific Research (NWO, grant 400-09-159).

Nynke van Miltenburg
Cooperation under Peer Sanctioning Institutions
Dissertation, Utrecht University, The Netherlands
Printed by Ridderprint BV
ISBN 978-94-6299-100-2

This book was typesetted using LaTeX by Nynke van Miltenburg

Cooperation under Peer Sanctioning Institutions

Collective Decisions, Noise, and Endogenous Implementation

Instituties met wederzijdse sanctionering ter bevordering van coöperatie

Collectieve beslissingen, ruis, en endogene implementatie

(met een samenvatting in het Nederlands)

Proefschfift

ter verkrijging van de graad van doctor
aan de Universiteit van Utrecht
op gezag van de rector magnificus,
prof. dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op vrijdag 5 juni 2015 des middags te 2.30 uur

door

Nynke van Miltenburg

geboren op 19 april 1987 te Baarn

Promotoren:	Prof. dr. ir. V. Buskens
	Prof. dr. W. Raub
Copromotor:	Dr. D. Barrera

Contents

Chapter 1

The role of peer sanctioning institutions in cooperation problems

1.1 A department's coffee corner

Imagine a large firm or institute in which each department is allocated its own coffee corner. Employees working in the same department jointly benefit from keeping their coffee corner clean. Thus, if one employee makes an effort to clean the coffee corner, the work environment will improve for everyone in the department. Alternatively, this employee could free ride on the cleaning efforts of co-workers. This way, he or she can enjoy a spotless coffee corner without forgoing time and energy that could be dedicated to other tasks. If all employees of the department follow this incentive to free ride, such that no one makes an effort to clean, the coffee corner may rapidly become disorderly, leaving the employees worse off than if they each were to have cleaned.

Now suppose that the employees of a department selectively address one another's cleaning habits. If someone cleans up after using the coffee corner, a co-worker may reward him or her (positive sanction), for example with a compliment. If an employee leaves dirty mugs in the coffee corner, he or she may be punished (negative sanction) by being held responsible for another pesky task. One can imagine that positive and negative sanctions provided by peers encourage employees to complete their share of work. However, there are numerous types of sanctions, and numerous possible ways in which peers may allocate them, and it is not evident how appropriate behavior is successfully promoted. For example, which of the two types of sanctions described above, punishment or reward, best encourages people to maintain a clean coffee corner? And is it sensible to allow all co-workers of a department to individually decide on peer sanctioning? Or would the coffee corner be better maintained when some or all co-workers collectively decide, for example, that someone who did not clean must perform an alternative duty?

The success of a certain type of peer sanction in encouraging co-workers to clean may not mean that employees positively evaluate this type of sanction. For example, employees may find collective sanctioning decisions more fair than individual sanctioning decisions regardless of which of the two results in a cleaner coffee corner. Or the coffee corner might be best maintained through severe punishment (e.g., by neglecting to invite deviant co-workers to department lunch gatherings) than through mild punishment (e.g., frowning at the deviant employee). Still, even if the coffee corner remains spotless under threats of severe punishment, this raises the question whether department members would allow one another to employ severe punishment actions.

Co-workers may not always directly witness one another cleaning or leaving dirty dishes. Rather, it is likely that employees will utilize circumstantial information, which may or may not reflect others' actual behaviors. For example, one employee might observe a co-worker leaving a dirty mug in the sink without knowing that this co-worker already cleaned everyone else's dishes the previous day. Such inaccurate information can lead to misguided sanctioning decisions. How would the possibility of having inaccurate information on others' cleaning habits affect answers to the questions posed above?

The co-workers described in the above example are involved in daily situations where they must decide whether to contribute private resources to a collective endeavor. Similar situations that have major implications for the welfare of the actors include the provision of national public goods and services, environmental protection efforts, and the maintenance of local public infrastructure. Such situations are generally referred to as cooperation problems. While contributing is beneficial to the group of involved actors as a whole, individuals have an incentive to defect on the group by keeping their resources for private use. If every individual follows this incentive, the collective endeavor is not realized, and each actor is worse off than if all individuals had contributed.

In this thesis, the term 'contribution' is reserved for individual decisions and can refer to 'high' as well as 'low' contributions. Contributing less than other group members is referred to as 'free riding,' and neglecting to contribute is referred to as 'defecting.' Contributing the highest possible amount is referred to as 'cooperating.' 'Group contributions' refer to aggregate contributions of all members of a group of actors involved in a cooperation problem.

Cooperation problems are prominent examples of social dilemmas, that is, situations in which individual, goal-directed behaviors lead to collectively suboptimal outcomes (Buskens and Raub, 2013; Dawes, 1980; Kollock, 1998; Raub et al., 2015). The tension between individual and collective interests renders explaining the emergence and maintenance of cooperation, i.e., the problem of social order, one of the fundamental research problems in the field of sociology (Coleman, 1964; Hobbes, 1651; Parsons, 1937). This thesis considers cooperation problems in which the actors involved have an opportunity to allocate positive or negative sanctions to their peers. The use of peer sanctions and the effect of peer sanctioning opportunities on the emergence and maintenance of cooperation are considered, as well as actors' perceptions of peer sanctions and their decision whether to allow an opportunity for peer sanctioning in their group.[1]

1.2 Peer sanctions as institutions

Peer sanctions are defined as positive and negative selective incentives that actors allocate to each other. Thus, the term 'sanction' refers to both punishment and reward. If cooperation problems are supplemented with an opportunity for allocating peer sanctions, this opportunity can be considered an institution that regulates cooperation (cf. Casari and Luini, 2009; Gürerk et al., 2006; Sutter et al., 2010). Hereby, institutions are conceived of as "the humanly devised constraints that shape human interaction" (North, 1990). This includes formal institutions that offer third party enforcement (e.g., courthouses and police departments) and informal institutions, in which actors themselves provide enforcement (e.g., peer sanctioning, cf. Andreoni and Gee, 2012; Kamei et al., 2011; Markussen et al., 2014). Peer sanctioning opportunities thus represent an example of a constraint that regulates behavior in social dilemmas.[2]

The opportunity to allocate peer sanctions, or the implementation of such sanctions, may encourage cooperative behaviors. However, allocating sanctions is typically costly (Axelrod, 1986; Coleman, 1990; Heckathorn, 1989; Voss, 2001, but see Guala, 2012). This implies that a second-order cooperation problem arises, in which each actor has an incentive to refrain from sanctioning while enjoying benefits of increased cooperation resulting from sanctions allocated by others (Oliver, 1980). Still, numerous studies show that costly peer punishments and rewards are often used by human actors, rendering peer sanctions feasible institutions for establishing and maintaining cooperation (see Balliet et al., 2011; Chaudhuri, 2011 for an overview).

In modern societies, cooperation is often enforced by formal authorities such as courts and police. However, such external authorities are typically absent in certain types of cooperation problems, for example in small-scale communities, in relations between nations, and in small groups of actors such as those involved in the coffee corner example (Ostrom, 1990; Decker et al., 2003). In these settings, cooperation may be – or has to be – supported through peer sanctions. Moreover, the emergence of an external authority itself is an outcome of successful cooperation (Fehr and Gintis, 2007; Voss, 1985), presupposing that the second-order cooperation problem has been solved. In sum, more basic mechanisms than external authorities, such as peer sanctions, are relevant in explaining human cooperation (Fehr and Gintis, 2007).

[1] Another interesting topic concerns the effect of peer sanctioning opportunities on actor welfare (e.g., Nikiforakis, 2014). As described in the empirical chapters of this thesis, the settings considered do not allow one to assess welfare effects in a similar manner as previous studies. Therefore, while welfare is addressed in the empirical chapters, the focus of the thesis remains on cooperation.

[2] Note that cooperation problems are typically embedded in a combination of formal and informal institutions. In this dissertation, the effect of peer sanctioning institutions is isolated by ruling out other institutions by design.

A vast amount of previous research addresses the role of peer sanctioning institutions in cooperation problems. Yet, a number of issues that may be especially relevant in settings where peer sanctioning institutions are employed have received relatively little attention. This thesis focuses on three of these topics, with each empirical chapter addressing one or several of the three topics. First, previous research typically considers peer-sanctioning institutions in which every actor individually decides on sanctioning others. In this thesis, such institutions are compared to institutions in which peer sanctions are only implemented after receiving some degree of collective support from actors. Two possible consequences of collective sanctioning decisions relative to individual sanctioning decisions are addressed, namely whether collective sanctioning decisions are better tailored to cooperation enforcement and whether actors perceive collective sanctioning decisions as more fair. Second, the success of peer punishment institutions in supporting cooperation is considered in cooperation problems in which actors receive potentially inaccurate information regarding others' behaviors. Third, in addition to comparing behavior in cooperation problems in settings with and without various peer-sanctioning institutions, this thesis addresses whether actors themselves implement a peer punishment institution in their group when presented with opportunities to do so. As described in the ensuing sections, each of these topics may be particularly relevant to everyday cooperation problems.

1.3 Individual and collective sanctioning decisions

Since the publications of Yamagishi (1986), Ostrom et al. (1992) and Fehr and Gächter (2000, 2002), studies on peer sanctioning have typically considered institutions in which all actors individually decide on sanctioning others. One prominent finding shows that actors differ in principles that they follow when deciding whom to sanction (e.g., Carpenter and Matthews, 2009; Falk et al., 2005; Kamei, 2014). Under individual sanctioning decisions, allocated sanctions resulting from these heterogeneous strategies are implemented simultaneously. Accordingly, actors fail to coordinate on efficient and effective sanctioning levels (Boyd et al., 2010). Moreover, individual sanctioning decisions need not foster cooperation, but they may also be associated with the enforcement of behaviors that hinder cooperation. That is, certain actors allocate 'perverse sanctions' to group members who contribute as much as the others or more (e.g., Cinyabuguma et al., 2006; Ellingsen et al., 2013; Herrmann et al., 2008). This raises questions regarding whether other manners in which sanctioning decisions can be aggregated are better tailored to enforcing cooperation. This also raises questions regarding how actors perceive individual sanctioning decisions and on whether other manners in which sanctions can be implemented are considered more fair. Strictly implementing sanctions that receive a certain degree of collective support from group members may be especially successful in both respects (Casari and Luini, 2009; Strimling and Eriksson, 2014; Hauser et al., 2014).

Ethnographic evidence shows that peer-sanctioning decisions employed in small-scale communities are often collectively made by the group as a whole or by a subgroup (Guala, 2012; Ostrom, 1990; Turnbull, 1961; Veszteg and Narhetali, 2010). Collective sanctioning decisions may be superior or inferior to individual decisions depending on the level of collective agreement required for implementing a sanction. On the one hand, the more collective agreement is required for a sanction to be implemented, the more likely it might be that the agreed-upon sanctions adequately promote group welfare, i.e., high levels of cooperation (Casari and Luini, 2009; Hauser et al., 2014). On the other hand, as a higher degree of collective agreement is required, sanctions become more difficult to realize, rendering collective sanctioning decisions inefficient and reducing their ability to enforce cooperation (Buchanan and Tullock, 1962; Fischer and Nicklisch, 2007). Casari and Luini (2009) show that cooperation is better enforced by punishment that is agreed upon by a small proportion of group members than via individual punishment decisions. This raises questions concerning how enforcement of cooperation depends on levels of collective agreement required for sanctions to be implemented and on whether findings on punishment are generalizable to reward. These two questions are theoretically and empirically addressed in this thesis.

Comparing individual and collective sanctioning decisions is not only relevant to research on cooperative behavior. Rather, sanctioning institutions also affect actors' intrinsic motivation to contribute (Fehr and Falk, 2002; Mulder, 2008; Tenbrunsel and Messick, 1999). The extent to which allocated sanctions promote intrinsic motivation is relevant for upholding cooperation when a punishment institution is no longer in place or when behaviors cannot be sufficiently monitored (Mulder and Nelissen, 2010; Mulder et al., 2006). Relatively little is known on how peer sanctioning decisions made by individuals or made by the group affect actor perceptions of the sanctions that are allocated. However, one reason why collective sanctioning decisions are employed by numerous groups of actors might be that they are valued for non-instrumental reasons. More specifically, actors may differently evaluate the fairness of sanctions allocated through individual and collective decisions (Strimling and Eriksson, 2014). This thesis compares the effect of sanctioning institutions with individual and with collective sanctioning decisions on actor perceptions of the fairness of peer sanctions that they and their peers receive in cooperation problems.

1.4 The accuracy of information on others' contributions

In numerous everyday cooperation problems, the actors involved may not always receive accurate information on one another's behaviors (Bereby-Meyer, 2012). Especially the risk of inaccurately observing others' contribution to the collective good may have severe implications for behavior in cooperation problems with peer sanctioning institutions. Henceforth, the possibility to inaccurately observe contributions of one or several fellow group members is referred to as 'noise'. More precisely, this thesis focuses on settings in which actors either contribute their full endowment or contribute nothing, and in which noise implies a positive probability that an actor observes another actor's cooperation as a defection and vice versa. With noise, actors may make 'misguided' decisions when sanctioning their peers (Ambrus and Greiner, 2012; Grechenig et al., 2010; Wu and Axelrod, 1995). That is, an actor may observe cooperation by a fellow group member as defection and punish accordingly. Alternatively, an actor may observe defection by a fellow group member as cooperation and fail to punish. Misguided sanctioning decisions may hinder cooperation under peer sanctioning institutions. This highlights the importance of examining how peer-sanctioning institutions function in cooperation problems with noise (Bereby-Meyer, 2012).

With noise, some potential sanctioners may correctly observe the contribution or defection of a prospective recipient, while others might observe it incorrectly. An advantage of demanding more collective agreement with noise is thus that the more actors are required to agree on sanctioning, the less likely it is that sanctions that are proposed after an inaccurate observation of another's contribution or defection are implemented. A disadvantage is that the more actors are required to agree on sanctioning decisions, the more the ability to reach collective agreement may be hindered by heterogeneous information between sanctioners on the contribution or defection of a prospective recipient. This raises the question which degree of collective agreement required on sanctioning decisions best supports cooperation under noisy information. In this thesis, the effect of individual and collective sanctioning decisions on cooperation are considered in cooperation problems characterized by noise.

1.5 Endogenous peer sanctioning institutions

The majority of previous studies compare situations in which actors interact under a peer-sanctioning institution to situations in which actors do not have the option to sanction each other. Such studies address the question whether actors react in predictable ways to the opportunities that are created by peer sanctioning institutions (Botelho et al., 2005). However, in numerous small-scale cooperation problems, the involved actors themselves design institutions for maintaining cooperation (Ostrom, 1990). To better understand how endogenous institution formation in cooperation problems proceeds, the deeper theoretical question emerges whether actors involved in cooperation problems anticipate the opportunities created by peer-sanctioning institutions, and permit the implementation of such institutions in their group (Batenburg et al., 2003; Gürerk et al., 2006; Prendergast, 1999). In addressing this question, research should focus on settings in which actors themselves decide whether to implement a sanctioning institution in their group before they engage in cooperation problems.

Available evidence suggests that when actors are given repeated opportunities to decide whether to implement a peer-sanctioning institution with individual sanctioning decisions, actors decide to implement the institution after they acquire experience with the decision situation (e.g., Ertan et al., 2009; Gürerk, 2013; Gürerk et al., 2006, 2009; Markussen et al., 2014; Rockenbach and Milinski, 2006. However, this finding may strongly depend on whether the conditions for cooperation to emerge under peer sanctioning institutions are favorable. As noted above, there are reasons to assume that with noise, peer-sanctioning institutions with individual sanctioning decisions may not be as successful in encouraging cooperation. This raises questions concerning whether the willingness of actors to implement a peer-sanctioning institution with individual sanctioning decisions generalizes to cooperation problems with noise.

Previous studies find group contributions under peer-punishment institutions with individual punishment decisions to be higher when punishments have more severe implications for the recipient, regardless of noise (Ambrus and Greiner, 2012; Egas and Riedl, 2008; Nikiforakis and Normann, 2008). However, when actors voluntarily allow for the implementation of peer punishment institutions in their group, it is unclear whether they prefer an institution in which they can allocate severe punishments, or rather opt for a punishment institution under which punishments are relatively mild. As noted above, punishment allocated through individual decisions may occasionally be directed at cooperators, and this may be especially problematic if punishments are severe. Noise intensifies the risk that cooperators are punished further, as cooperators may receive misguided punishments from group members who observe them as defectors. In this thesis, endogenous institution formation and endogenous punishment severity are considered both with and without noise.

1.6 Theoretical and empirical approach

The three issues described above are addressed in four empirical chapters that employ similar theoretical and empirical approaches. In each empirical chapter, one-shot versions of the n-player Prisoner's Dilemma (PD) or Public Goods Game (PGG) are used for theory formation and empirical testing. The PD and PGG have emerged in the literature as archetypal representations of cooperation problems (e.g., Davis and Holt, 1993; Ledyard, 1995; Sally, 1995). The decision whether to use the PD or PGG is based on practical considerations elaborated on in the respective chapters. Both the PD and PGG employ a contribution stage. In this stage, all members of a group of n actors receive an equal endowment $w > 0$. Actors simultaneously and independently decide on contributing the endowment to a group account. In a PD, an actor i can only choose whether to contribute the entire endowment, i.e., contribution $c_i \in \{0, w\}$, while in a

PGG, actors select a fraction of the endowment to contribute, i.e., $c_i \in [0, w]$. Aggregate contributions of all group members $c = \sum c_i$ are multiplied by a factor m, with $1 < m < n$. Each actor receives an equal share of mc. It can be easily verified that aggregate group earnings $nw - c + mc$ are highest when all actors contribute their entire endowment, as $m > 1$, while individual earnings $w - c_i + mc/n$ decrease with own contributions c_i, because $m < n$. Because collective payoffs are maximized when all actors contribute their full endowment and individuals earn most by contributing nothing, the PD and PGG correspond to the definition of a cooperation problem presented at the start of this chapter.

Peer sanctions are modeled as a second stage in the PD and PGG that starts after actors are informed of the contribution of each fellow group member (Fehr and Gächter, 2000, 2002; Sefton et al., 2007). In the most basic form of a sanctioning stage, each actor i can determine whether to sanction each fellow group member $j \neq i$. If i decides to sanction j, a cost $a > 0$ is deducted from i's earnings. In case of punishment, an amount $b > a$ is deducted from the earnings of actor j, while b is added in case of reward.

Several variants of the PD and PGG are considered in this thesis. The modification of the sanctioning stage that allows actors to collectively decide on sanctioning a group member is outlined in Chapters 2, 3, and 4. The implementation of noise is discussed in Chapters 4 and 5, and the implementation of an endogenous sanctioning stage with endogenous sanction severity is described in Chapter 5.

A similar theoretical approach is used in each empirical chapter where behavior in variants of the PD or PGG is considered.[3] First, behaviors for each variant under standard game-theoretic assumptions are outlined. According to these assumptions, all actors are self-regarding with a sole objective to maximize personal payoffs; all actors act rationally and expect rational behaviors from other actors; and all features of the interaction, including the utility function of the actors, are common knowledge. As noted above, actors maximize their payoffs in the contribution stage by keeping their endowment for themselves. Moreover, because the cost of allocating sanctions exceeds zero, under standard assumptions, actors do not allocate or expect to receive sanctions. Thus, full defection is predicted regardless of whether a peer sanctioning stage is added to the interaction. This is true for every variant of the PD or PGG that is considered in this thesis.

Empirical evidence shows that behaviors of numerous actors in PDs or PGGs do not correspond to predictions that follow from standard game-theoretic assumptions. Instead, robust patterns of alternative contribution and sanctioning are observed in cooperation problems with endogenous and exogenous sanctioning stages (e.g., Chaudhuri, 2011; Gürerk, 2013). More specifically, numerous actors contribute in PDs and PGGs without a sanctioning stage as long as they expect sufficient reciprocation (Fis-

[3]Only Chapter 3 does not consider behavior in variants of the PD or PGG. Here, existing theories on determinants of fairness perceptions are applied to PGGs with a sanctioning stage and with individual or collective punishment decisions.

chbacher et al., 2001; Herrmann and Thöni, 2009; Kamei, 2012; Kocher et al., 2008; Thöni et al., 2012). Furthermore, numerous actors punish free riders or reward high contributors (e.g., Balliet et al., 2011; Fehr and Gächter, 2000, 2002; Sefton et al., 2007), and a small cohort of actors is consistently found to punish high contributors or reward free riders (e.g., Cinyabuguma et al., 2006; Ellingsen et al., 2013; Herrmann et al., 2008; Ostrom et al., 1992). In theorizing on behaviors found in variants of the PD or PGG beyond the standard actor model, it is assumed in each chapter that a proportion of actors in the population acts according to these alternative behaviors. The remaining actors are assumed to be rational and self-regarding and to anticipate that certain proportions of their peers act according to alternative behaviors. The theoretical frameworks describe how the different types of actors might behave in each PD or PGG stage considered, and how experiencing the behavior of other actors may affect subsequent decisions.

This theoretical approach is broadly consistent with behavioral game theory in that it aims to predict how actors actually behave in certain interaction situations (Camerer, 2003). In each empirical chapter, the same alternative behaviors that are empirically validated in previous studies are consistently assumed to occur and are used to predict decisions made in a range of settings (Fehr and Falk, 2002; Gächter and Thöni, 2011; Ockenfels and Raub, 2010).

Several methods have been used in previous studies to examine the effect of (peer) sanctioning institutions on behaviors in cooperation problems. Examples include case studies of cooperation and punishment in naturally occurring cooperation problems (Boehm, 1999; Ostrom, 1990); field experiments on peer punishment in everyday interactions between strangers (Balafoutas and Nikiforakis, 2012; Berger and Hevenstone, 2014; Noussair et al., 2011); lab-in-the-field experiments using subjects of various cultural backgrounds (Baldassarri and Grossman, 2011; Henrich et al., 2006; Herrmann et al., 2008; Marlowe et al., 2008); and laboratory experiments employing the PD, PGG, or related games, predominantly using student subjects (Fehr and Gächter, 2000, 2002; Ostrom, 1990; Yamagishi, 1986). The latter approach is employed in this thesis. In the experiments, decisions are incentivized in that the outcomes of the PDs and PGGs determine subject earnings. Although laboratory experiments are often criticized for artificially representing interaction situations and for lacking a representative subject pool, laboratory experiments allow one to test causal claims by excluding potentially confounding factors. Thus, laboratory experiments are well suited to isolate effects of peer sanctioning institutions on cooperation (see Falk and Heckman, 2009; Jackson and Cox, 2013; Thye, 2007 for a discussion on the use of laboratory experiments in the social sciences).

1.7 Overview of remaining chapters

In Chapters 2 through 5, specific research questions are addressed that relate to the three main topics outlined above. Note that these chapters are written as separate empirical studies, and thus some overlap, especially in descriptions of theoretical frameworks and experimental designs, may be encountered. Chapter 2 considers how the effect of peer sanctioning institutions on cooperation is affected by the extent to which sanctioning decisions are made collectively in cooperation problems where actors have accurate information regarding others' behaviors. A laboratory experiment on the PGG with either a punishment or reward stage is employed. To represent individual and collective sanctioning decisions, three different 'decision rules' are employed, through which punishments or rewards are implemented. Under the individual decision rule (IDR), all sanctions that group members allocate are implemented. Under the collective decision rule (CDR) that requires majority consent, sanctions are only implemented if at least a majority of group members, excluding the prospective recipient, choose to sanction the same actor. Sanctions not achieving majority agreement are not implemented. Under the CDR that requires unanimity consent, all group members with the exception of the prospective recipient must target someone for punishment or reward to be implemented.

Chapter 3 examines the same setting as that explored in Chapter 2; namely a PGG in which peer punishment and reward are allocated through individual, majority, or unanimity decision rules; but focuses on a different outcome. The chapter examines how peer sanctions, and the decision rule through which sanctions are implemented, affect actors' perceptions of the fairness of punishments and rewards that they and fellow group members receive. The same laboratory experiment is employed in Chapters 2 and 3, but Chapter 3 focuses on the perceived fairness of sanctions received by subjects and fellow group members as measured through a questionnaire.

Chapter 2 shows that peer punishments are more successful at promoting cooperation than rewards under all decision rules. Therefore, Chapters 4 and 5 focus on punishment only. In Chapter 4, the enforcement of cooperation through an IDR and through CDRs is revisited. However, in addition to examining cooperation problems in which actors have accurate information regarding one another's contributions or defections, this chapter considers cooperation problems with noise. A laboratory experiment is employed to compare an IDR and two CDRs for allocating punishment in an n-player PD. In Chapter 2, numerous punishments under CDRs do not achieve sufficient agreement to be implemented. Noise may hinder punishment implementation through CDRs even further. Therefore, in Chapter 4, required levels of collective agreement for punishment implementation are lower than in Chapter 2.

Chapter 5 considers cooperation problems in which actors endogenously decide whether to implement a punishment institution with an IDR and endogenously decide whether each punishment has a lower or higher impact on the income of its recipient. Both cooperation problems in which actors possess accurate information on one another's cooperation or defection and cooperation problems with noise are considered. This setting is captured in a third laboratory experiment wherein subjects interact in n-player PDs. Before each interaction commences, subjects decide whether to add a peer punishment stage to the PD, and if so, the severity of punishments that are allocated.

Chapter 6 returns to the three main themes of this thesis, i.e., collective sanctioning decisions, inaccurate information, and the endogenous implementation of peer sanctioning institutions. For each topic, findings of the empirical chapters are summarized, and avenues for future research are identified. The chapter concludes with recommendations based on the results of this thesis for co-workers who wish to keep their shared coffee corner clean.

Chapter 2

Implementing punishment and reward in the Public Goods Game – The effect of individual and collective decision rules*

Abstract

Punishments and rewards are effective means for establishing cooperation in social dilemmas. We compare a setting where actors *individually* decide whom to sanction with a setting where sanctions are only implemented if actors *collectively* agree that a certain group member should be sanctioned. Collective sanctioning decisions are problematic due to the difficulty of reaching consensus. However, when a decision is made collectively, perverse sanctioning (e.g., punishing high contributors) by individual actors is ruled out. Therefore, sanctions implemented through a collective decision rule are more likely to be in the interest of the whole group. We employ a laboratory experiment where subjects play Public Goods Games with opportunities for punishment or reward that is implemented either by an individual, a majority, or unanimously. For both punishment and reward, contribution levels are higher in the individual than the majority condition, and higher under majority than unanimity. Often, majority agreement or unanimity is not reached on punishments or rewards.

2.1 Introduction

A public good is characterized by non-excludability: once it is produced, all actors can enjoy its benefits regardless of their contribution to the provision of the good (Olson,

*This is a slightly different version of Van Miltenburg, Buskens, Barrera, and Raub (2014) that has been published in *International Journal of the Commons*.

1965). Since public good provision is costly, this implies a tension between the individual and collective interest. While full cooperation leads to the best possible group outcome, individuals have an incentive to free ride on the contributions of others.

Contributions to public goods can be supported by positive or negative peer sanctioning institutions, that is, the opportunity for actors to reward or punish each other. Experimental research established that high contributions can be maintained when sanctioning is possible, both in public goods problems and related settings (Balliet et al., 2011; Fehr and Gächter, 2000, 2002; Ostrom et al., 1992; Sefton et al., 2007; Yamagishi, 1986). However, a challenge for research and policy is to design institutions that best enable heterogeneous actors to enforce cooperation (Ostrom, 2010, 2012). In this respect, which method of implementing sanctions is most successful in increasing contributions remains an open question (Gächter and Thöni, 2011). For example, it is unclear whether contributions are higher when the decision of whom to sanction is made individually or when it is made collectively. A sanctioning institution with an individual decision rule (IDR) is an institution in which every actor individually decides whom to sanction and pays the associated costs. A sanctioning institution with a collective decision rule (CDR) is an institution in which sanctions are executed only when multiple actors agree and pay the cost of sanctioning.

In real-life public goods problems, actors often employ a sanctioning institution with a CDR. For example, Ostrom (1990) and Veszteg and Narhetali (2010) describe small communities where group members successfully enforce collective action through collective sanctioning decisions. Typically, members of the community regularly meet to identify free riders and decide upon their punishment, for example in a vote. Also, in international cooperation, nations use collective sanctioning decision rules to ensure provision of global public goods such as international security and economic stability. Sanctioning decisions are usually taken by a variant of majority voting. Unanimity voting is uncommon, because it gives every individual nation the opportunity to veto a sanction, thereby making collective organizations ineffective decision makers (http://www.europa.eu, 2010; http://www.un.org, 2010).

So far, there is limited experimental research comparing the effect of sanctioning through IDRs and CDRs in public goods problems.[1] Casari and Luini (2009) find that, compared to an IDR, contributions to public goods are higher when punishment is only carried out if at least two out of four actors punish the fifth member of their group. Thus, they consider only one CDR, and do not compare positive and negative sanctions.

[1]Numerous experimental studies employ other forms of collective peer sanctioning decisions, but do not compare IDRs with CDRs. Decker et al. (2003) examine the effect of implementing a subset of punishment proposals while actors share punishment costs. In Ertan et al. (2009), Nese and Sbriglia (2009), Sutter et al. (2010), and Botelho et al. (2005) subjects vote on whether to allow individual peer sanctioning. Andreoni and Gee (2012), Feld and Tyran (2002), Guillen et al. (2006), Kamei et al. (2011), Markussen et al. (2014), Putterman et al. (2011), and Tyran and Feld (2006) let subjects collectively decide on implementing various forms of sanctioning institutions.

This leaves a number of unresolved issues, in which the current chapter provides further insight.

First, it is unclear how the effect of a CDR on contribution levels depends upon the proportion of actors required to agree for a sanction to be implemented. On the one hand, the higher the proportion required, the less likely it will be that a sufficient number of actors agrees on the necessity of sanctioning and is willing to incur the associated costs (cf. Buchanan and Tullock, 1962). Thus, while under an IDR all desired sanctions are carried out by definition, under CDRs there is a higher chance that free riders remain unpunished or contributors unrewarded. On the other hand, under an IDR individuals might decide to use sanctions in ways that hurt cooperation and thereby result in decreasing payoffs for the group, i.e.. to reward free riders or to punish contributors (Casari and Luini, 2009; Ellingsen et al., 2013). Consequently, the more actors collectively agree that a certain group member should be sanctioned, the higher the chance that this sanction will be in the collective interest, that is, in accordance with enforcing high contributions to the public good. In the current chapter, we address the effect of the required proportion of consenting actors on contribution levels by comparing contributions under an IDR to a CDR for which majority and a CDR for which unanimity is required. Majority or unanimity refers to the group members other than the prospective target of the sanction. Thus, if in a group of four actors three other actors all want to sanction the fourth, we consider this unanimity. Majority requires that at least two of the three others want to sanction the fourth.

Second, theoretical arguments and empirical results on punishment cannot be straightforwardly generalized to reward. For example, to maintain cooperation rewards have to be repeatedly allocated to contributors. Conversely, the mere threat of punishment can be sufficient to deter free riding (Dari-Mattiacci and De Geest, 2010). This suggests that punishments and rewards may differ in efficiency. Empirically, it has been shown that punishments and rewards might differ also in terms of efficacy (e.g., Sefton et al., 2007; Choi and Ahn, 2013). We therefore study decision rules for assigning both punishment and reward.

The effects of the decision rules on macro-behavior, such as aggregate contribution levels, depend on assumptions about the micro-motives of individual actors (cf. Gächter and Thöni, 2011). For example, these effects depend on which proportion of actors is willing to sanction, who is likely to be targeted, and how sanctions influence contribution decisions. We summarize existing knowledge on individual behavior in the PGG with sanctions. Subsequently, we apply this to predict macro-level behavior in the PGG with different decision rules, and with punishment or reward. We thus assess through which mechanisms our empirical extensions could result in different contribution levels between sanctioning institutions.

The chapter is structured as follows. In the theory section, we review the literature on behavior in public goods problems with opportunities for sanctioning. Subsequently, we develop hypotheses on contribution and sanctioning behavior, and on how this behavior of individuals translates in different contribution levels under IDRs and CDRs. Individual-level and macro-level hypotheses are tested in an experiment where individual, majority, and unanimity decision rules for punishing and rewarding are employed in an incentivized manner.

2.2 Theory

2.2.1 The Public Goods Game

The linear Public Goods Game (PGG; also called Voluntary Contribution Mechanism, e.g., Isaac and Walker, 1988) is used as a model of public goods problems. It is played by n actors. Every actor i receives an endowment w. They simultaneously and independently decide whether to keep this endowment for themselves or contribute an amount $c_i \in [0, w]$ to a 'group account'. The total amount contributed by all n actors together, $c = \sum c_i$, is multiplied by a number m, with $1 < m < n$, and mc is divided equally among all actors. Because $m < n$, the individual return obtained from the amount contributed to the group account is smaller than when it would have been kept to oneself ($mc_i/n < c_i$). Therefore, when the PGG is played once, and when actors are self-regarding in that utility equals own payoff, contributing nothing is a dominant strategy, yielding the highest utility regardless what others do. This results in the unique Nash equilibrium of zero contributions. However, since $m > 1$ the joint group outcome $nw - c + mc$ is maximized when everybody contributes the full endowment. Every player would then be better off compared to when all contribute nothing ($mw > w$). Thus, individually rational behavior in the sense of playing a dominant strategy or, more generally, Nash equilibrium behavior, leads to a Pareto-suboptimal outcome, making the PGG a social dilemma (Dawes, 1980).

2.2.2 Behavior in the PGG

The prediction of complete free riding is typically refuted in experimental research employing the PGG. Instead, contributions averaging 50% of the endowment are consistently observed in one-shot PGGs (Kocher et al., 2008; Walker and Halloran, 2004). Also in repeated PGGs where group composition changes after each round, as in our experiment, subjects initially contribute 50% on average. However, in subsequent rounds contributions gradually decline to very low levels (Ledyard, 1995).

Research explaining this declining contribution pattern focuses on non-standard utility as an alternative behavioral assumption. It has been empirically established that actors in the PGG can be classified in two main preference types (Fehr and Gintis, 2007; Ones and Putterman, 2007; Ostrom, 2000). Actors of the first type are rational and self-regarding free riders who never contribute to the public good. Actors of the second type are conditional cooperators who contribute more, the more they expect others to contribute (see Chaudhuri, 2011; Gächter, 2007 for an overview of empirical evidence). These actors are assumed to derive utility from reciprocating others' expected contribution even in one-shot settings. Conditional cooperators are heterogeneous in the extent to which they match others' contributions. Many are 'imperfect' reciprocators in that they contribute slightly below what they expect others to contribute on average. In an experiment specifically designed to identify preference types, Fischbacher et al. (2001) classify 50% of their subjects as (partial) conditional cooperators and 30% as free riders.[2] Others have roughly replicated this distribution of types in different subject pools (Herrmann and Thöni, 2009; Kamei, 2012; Kocher et al., 2008; Thöni et al., 2012). Conditionally cooperative behavior is consistent with a prosocial orientation (e.g., Van Lange, 1999; Ackermann et al., 2014).

In repeated PGGs, conditional cooperators adapt their expectation of others' contribution on the basis of their experience of the average group contribution in the previous rounds (Fischbacher and Gächter, 2010). The more free riders and imperfect conditional cooperators there are, the lower group contribution will be. Conditional cooperators decrease their contribution accordingly, which causes the average to further decline. This explains the decrease of cooperation over time.

2.2.3 The PGG with peer sanctions

Sanctioning can be modeled by adding a second stage to the standard PGG. After all actors have determined their contribution and observed the contributions of the other group members, they decide for every other group member j whether to pay an amount to punish or reward this actor. Let s_{ij} denote the amount actor i uses to sanction actor j. We assume here that an actor can only choose whether or not to sanction, but not the magnitude of the sanction: s_{ij} is either a fixed amount $a > 0$ or zero. When the amount is used for punishment, a multiple k of a is subtracted from the payoff actor j obtained in the PGG. The same amount is added to the payoff of actor j when s_{ij} is used for reward. Thus, in addition to the payoff from the standard PGG, every actor i loses a total amount $k \sum_j s_{ji}$ of received punishment from all other actors j or gains this amount of received rewards. Moreover, every actor i forfeits $\sum_j s_{ij}$ by assigning sanctions to other actors j. This captures the essential features of how sanctions are implemented in the PGG, denoted here as an IDR.[3]

In sanctioning institutions with a CDR, all actors likewise decide whether to pay an amount to sanction others. Sanctioning under a CDR is different from an IDR in the sense that a sanction is only implemented if at least a proportion x/n of all group

[2] Virtually all remaining subjects were characterized as 'triangle' contributors (Fischbacher et al., 2001). These actors fully reciprocate others' expected contribution at 50% of the endowment, but their contribution declines when they expect others to contribute either more or less than this threshold.

[3] Note that details of this procedure can vary. For example, in many studies the amount s_{ij} used to sanction can be chosen freely by actors between 0 and some positive value.

members, save the prospective recipient, sanctions the same actor. Because we fixed the sanctioning amount s_{ij} to 0 or a, this implies that sanctions under a CDR are more severe than those under an IDR assigned by a smaller number of actors. Thus, every actor j loses an amount $k \sum_i s_{ij}$ under a punishment institution with a CDR when the proportion q_j of actors i for whom $s_{ij} = a$ is larger than or equal to x/n. The same applies to the amount gained under rewards. If $q_j < x/n$ no sanction is executed, that is, actor j does not gain or lose money due to received sanctions. Moreover, the actor(s) who proposed to sanction actor j do not pay the cost of sanctioning if $q_j < x/n$. Thus, every actor i who sanctions j loses an amount $\sum_{j: q_j \geq x/n} s_{ij}$.

We assume one-shot interactions.[4] Thus, actors cannot benefit from group members who increase their contribution in subsequent games after being sanctioned. This implies that long-term incentives for sanctioning, which differ between IDRs and CDRs, are ruled out. Under the IDR, rational self-regarding actors do not allocate costly sanctions in one-shot interactions. Under CDRs, rational self-regarding actors likewise do not sanction if their sanctions are implemented. If sanctions are not implemented, rational self-regarding actors are indifferent between sanctioning or not. Accordingly, the Nash equilibrium of zero contributions for the one-shot PGG with sanctions under standard assumptions of rationality and selfishness remains unchanged, while any amount of punishment or reward that is not implemented under the CDRs might be allocated in equilibrium. Although repeated interactions with sanction opportunities might be more realistic for many applications, we do stick to one-shot interactions also because in repeated interactions actors have alternative sanctioning mechanisms, too. For example, actors can reciprocate others' low contributions by own low contributions in future interactions. This would lead to possible confounding effects of the exogenous sanctioning mechanisms we want to study with the endogenous sanctioning opportunities due to repeated interactions (cf. Fehr and Gächter, 2002).

Given that not all actors are rational and self-regarding, some might sanction despite the prediction that follows from the assumptions of selfish rationality. Under a CDR, non-executed sanctions are costless, and group members are not informed about sanctions that were proposed by others but were not executed. Thus, non-executed sanctions cannot influence behavior of other actors than the ones who proposed the sanction. Therefore, actors have no incentive to take the probability that the sanction is executed into account when deciding whether or not to sanction under a CDR. [5]

[4]In our experiment we employ random matching. Although subjects are likely to interact multiple times, they are not informed on the identity of others. It is common in the experimental literature to treat this as series of one-shot games. However, Botelho et al. (2009) show that under random matching subjects behave slightly different from subjects who play perfect stranger matching. The main difference is that subjects contribute zero more often under random matching, although they do not contribute less on average. We do not expect that these slight differences affect the difference between the experimental conditions we consider.

[5]When a sanction under a CDR is implemented, there are by definition multiple sanctioners. Thus, under a CDR actors know that implemented sanctions are severe for the recipient. Therefore, they

Given these characteristics of the interaction situation, there is no reason to assume that actors make different sanctioning decisions under IDRs and CDRs.

We proceed with a review of empirical evidence and a theoretical account of contributing and sanctioning behavior in the PGG with an IDR. This reveals which actors likely allocate sanctions, and which behaviors are likely to be sanctioned. Multiple individual sanctions for a given behavior imply a high consensus. Thus, given that the decision rule will not directly influence sanctioning decisions, those behaviors that are likely sanctioned individually by many are more likely to be sanctioned when a CDR is used. Behavior in the PGG with an IDR then allows to predict the likelihood that sanctions will be implemented under a CDR.

2.2.4 Behavior in the PGG with peer sanctions under an IDR

Despite the equilibrium prediction, experimental evidence shows that actors frequently use punishment under an IDR in one-shot settings. It is consistently found that punishment is assigned in a pro-social manner, that is, in accordance with enforcing cooperation. Specifically, actors receive more punishment the less they contribute (e.g., Carpenter and Matthews, 2009; Casari and Luini, 2009), and the less they contribute compared to the average contribution of the group (e.g., Carpenter and Matthews, 2009; Ertan et al., 2009; Fehr and Gächter, 2000, 2002; Ones and Putterman, 2007; Sefton et al., 2007). Punishment is mostly executed by high contributors (e.g., Fehr and Gächter, 2002; Sefton et al., 2007). However, it is also observed that low contributors occasionally punish above-average contributors. This 'perverse' punishment is usually carried out by a small number of actors (Casari and Luini, 2009). The extent to which it occurs varies greatly between subject pools, up to 50% of total punishment expenditure (Herrmann et al., 2008), but is typically estimated between 5% and 25% (Casari and Luini, 2009; Cinyabuguma et al., 2006; Ertan et al., 2009; Ones and Putterman, 2007; Ostrom et al., 1992). The effect of pro-social and perverse punishment differs. Below-average contributors increase their contribution in the subsequent round after being punished (e.g. Fehr and Gächter, 2002), but for above-average contributors empirical evidence is mixed. Some studies show that above-average contributors decrease their contribution after being sanctioned (Bochet et al., 2006; Masclet et al., 2003; Ones and Putterman, 2007); others find no effect of receiving perverse punishment on contribution decisions (Denant-Boemont et al., 2007; see also Ellingsen et al., 2013).

Like punishments, rewards are typically used to promote cooperation in one-shot settings. High contributors tend to reward other high contributors (Choi and Ahn, 2013;

might for example be reluctant to sanction mildly deviant behavior, for which a severe sanction might not be deserved. However, the chance that the sanction is implemented will also be smaller for slight deviations. Thus, actors expect a more severe sanction with a lower probability of implementation the more actors are required to agree. There seems to be no reason to assume that this influences their sanctioning decisions.

Ellingsen et al., 2013; Sefton et al., 2007; Sutter et al., 2010; Walker and Halloran, 2004). However, while rewards are mainly allocated to above-average contributors, it is often less clear than for punishment that the amount of rewards received increases with the (positive) deviation from the average group contribution (Choi and Ahn, 2013; Nosenzo and Sefton, 2012; Sefton et al., 2007; Walker and Halloran, 2004, but see Ellingsen et al., 2013). Also, in repeated PGGs where actors can identify each other it is found that rewards are frequently used in every successive interaction (Ellingsen et al., 2013; Milinski and Rockenbach, 2011; Rand et al., 2009), while the use of rewards declines over time in fixed groups when actors cannot infer who rewarded them (Choi and Ahn, 2013; Sefton et al., 2007). As with punishments, the effect of receiving rewards on cooperation differs with the recipients' contribution. Above-average contributors are found to contribute more in the subsequent interaction the more rewards they receive, while below-average contributors contribute less the more they are rewarded (Ellingsen et al., 2013).

In repeated interactions in fixed groups, contributions under reward institutions are sometimes found to be lower than those under punishment institutions (Drouvelis and Jamison, 2012; Milinski and Rockenbach, 2011; Nosenzo and Sefton, 2012; Sutter et al., 2010, low leverage; Wiedemann et al., 2011) although others do not find a difference, at least until the final periods (Choi and Ahn, 2013; Rand et al., 2009; Sefton et al., 2007; Sutter et al., 2010, high leverage; see also Balliet et al., 2011). In repeated one-shot settings, which are most similar to our experiment, it is found that contributions are lower under reward than under punishment (Choi and Ahn, 2013).

2.2.5 Non-selfish utility in the PGG with sanctions

Rational self-regarding free riders never sanction when this is costly and their sanctions are implemented with a non-zero probability. However, anticipation of being sanctioned will induce them to contribute, provided that the loss due to received punishment or gain from rewards offsets the payoff advantage of free riding (Fehr and Fischbacher, 2004). Non-selfish actors could derive utility from sanctioning defectors even in one-shot interactions (Diekmann and Voss, 2003). These pro-social punishers are sometimes classified as a separate type of actor, which partly but not completely overlaps with conditional cooperators in the PGG without punishment (e.g. Ostrom, 2000; Ones and Putterman, 2007).

Empirical evidence is indeed consistent with the assumption that people derive utility from punishing and rewarding in one-shot settings. Fehr and Gächter (2002) already noted that subjects experience anger when they observe free riding in a hypothetical situation. This anger increases the more the free rider deviates from the average contribution of others. Casari and Luini (2009) show that punishment decisions are not influenced by information that others already punished the recipient. Thus, subjects do not care so much about actors being punished, but derive utility from the act of punishing. Fudenberg and Pathak (2010) show that subjects punish even when the recipient is not informed on the punishment, implying that punishment cannot influence future cooperation. In a neurobiological experiment, De Quervain et al. (2004) show that the human reward system is activated in the brain of an actor punishing a defector. Utility from rewarding is addressed by Dawes et al. (2007), who conduct an experiment in which subjects can decide on a costly in- or decrease of a random amount of tokens other subjects had received. They find that subjects who afterwards indicate more anger and annoyance towards those with a high amount also spend more to increase low and reduce high amounts received by others. Yet, despite utility derived from sanctioning, it is found that actors sanction less the higher the costs of sanctioning are (Anderson and Putterman, 2006; Carpenter, 2007; Nikiforakis and Normann, 2008; Sutter et al., 2010; Vyrastekova and Van Soest, 2008). Thus, actors take their own payoff into account in sanctioning decisions (Fehr and Fischbacher, 2004).

As mentioned above, some actors use sanctions 'perversely'. Although they are relatively rare, perverse sanctioners constitute a separate type of actors. These actors free ride in the PGG, and subsequently punish high contributors (e.g. Chaudhuri, 2011; Cinyabuguma et al., 2006; Gächter and Herrmann, 2009; Herrmann et al., 2008). A motive for perverse punishment might be revenge on previous punishment received from high contributors (Denant-Boemont et al., 2007; Fehr and Gächter, 2000; Nikiforakis, 2008; Ostrom et al., 1992), a desire to increase the relative payoff advantage of free riding (Fehr and Gächter, 2000), or a dislike of do-gooders or norm violators (Gächter and Herrmann, 2009; Monin, 2007; Ones and Putterman, 2007). Alternatively, it could be that actors occasionally punish high contributors by mistake (Fehr and Gächter, 2000). Rand et al. (2010) and Rand and Nowak (2011) show that punishment of cooperators can be evolutionary stable, thus providing a potential explanation for the fact that perverse punishment can drive out cooperation. Perverse rewards, i.e., rewards targeted at free riders, just as perverse punishments, increase the payoff discrepancy between high and low contributors. Hence, they are potentially equally detrimental for cooperation (Ellingsen et al., 2013).

Punishment and reward work through different mechanisms. The possibility of being punished might be enough to deter free riding, so that there is no need to actually allocate punishment. However, when an actor makes a high contribution, rewards actually have to be carried out sufficiently often to induce free riders to contribute (Dari-Mattiacci and De Geest, 2010). Thus, when contributions in a population increase due to the availability of a sanctioning institution, more rewards than punishments have to be allocated. In one-shot settings, actors cannot establish a norm of direct mutual rewarding. They are therefore unsure whether the costs of allocating rewards will be offset by reciprocation (Rand et al., 2009). This makes rewards more expensive than punishments in the one-shot PGG. As mentioned, more expensive sanctioning implies that less sanctions are assigned. This explains why, without opportunities for directly reciprocating received rewards, actors initially attempt to reward but eventually give up when others do not continue to reward as well.

2.2.6 Micro-level hypotheses

Before turning to differences in contribution levels between IDRs and CDRs, we capture the framework developed for micro-motives, that is, contributing and sanctioning behavior of individual actors, in a number of hypotheses. These hypotheses are based on empirical regularities observed in previous experiments. The hypotheses will be used as a micro-level framework summarizing which actors are likely to sanction, and how actors react to receiving pro-social or perverse sanctions. When theorizing on the effect of sanctioning decision rules on contribution levels, we assume that actors behave as summarized in this framework.

We first derive hypotheses on sanctioning behavior. Although perverse punishment is sometimes observed, punishment is usually pro-social and allocated by high contributors. Accordingly, we hypothesize that actors are more likely to punish others the more they contributed themselves.

Hypothesis 2.1 *The more an actor contributes, the higher this actor's likelihood to assign punishment.*

Punishment of high contributors is more often targeted at free riders than punishment of low contributors. Thus, the more an actor contributed the more likely he is to punish a free rider. This implies that we expect an interaction between the contribution of the actor allocating punishment and the contribution of the recipient of the punishment on the likelihood to sanction. We argue that actors perceive free riding both in the sense of the recipient contributing a low amount and in the sense of contributing less than the other group members. This means that low as well as below-average contributors are likely to be punished by high contributors.

Hypothesis 2.2a *The more an actor contributes, the more this actor's likelihood of assigning punishment decreases with the contribution of the recipient of the punishment.*

Hypothesis 2.2b *The more an actor contributes, the more this actor's likelihood of assigning punishment increases with the negative deviation of the recipient of the punishment from the group average contribution.*

Pro-social reward is also predominantly allocated by high contributors.

Hypothesis 2.3 *The more an actor contributes, the higher this actor's likelihood to assign reward.*

High contributors are more likely to reward other high contributors. This applies both in an absolute sense, and compared to the average of other group members. Again, we hypothesize an interaction between the contribution of the rewarding actor and the contribution of the recipient of the reward.

Hypothesis 2.4a *The more an actor contributes, the more this actor's likelihood of assigning reward increases with the contribution of the recipient of the reward.*

Hypothesis 2.4b *The more an actor contributes, the more this actor's likelihood of assigning reward increases with the positive deviation of the recipient of the reward from the group average contribution.*

Unlike punishments, rewards have to be allocated repeatedly to high contributors in order to enforce cooperation. Therefore, they are costly to maintain when direct reciprocation is impossible. Accordingly, the likelihood of rewarding decreases over rounds.

Hypothesis 2.5 *The more rounds have already been played, the lower the likelihood that rewards are allocated.*

We now turn to the effect of sanctions on contribution decisions. Receiving punishment leads to conformation to the behavior of other actors, in order to avoid receiving punishment in future interactions. Free riders thus increase and high contributors decrease their contribution the more they are punished. Consequently, their contribution is more in line with others' average.

Hypothesis 2.6 *The more an actor contributing below the average is punished, the more this actor contributes in the subsequent interaction.*

Hypothesis 2.7 *The more an actor contributing above the average is punished, the less this actor contributes in the subsequent interaction.*

Rewards strengthen current deviations from average behavior. Above-average contributors will thus contribute more and below-average contributors less the more they are rewarded, provided they did not already contribute the full endowment or free ride completely, respectively.

Hypothesis 2.8 *The more an actor contributing above the average is rewarded, the more this actor contributes in the subsequent interaction.*

Hypothesis 2.9 *The more an actor contributing below the average is rewarded, the less this actor contributes in the subsequent interaction.*

2.2.7 Macro-level effects of CDRs

Under a CDR, only the sanctions on which required consensus is reached are executed. Given sanctioning behavior as predicted in the micro-level hypotheses, it is likely that there will be more consensus on some sanctions than on others. This gives rise to different contribution levels under IDRs versus CDRs. Macro-level hypotheses differ for punishment and reward.

Under an IDR, all allocated punishments are carried out. This implies that pro-social punishers will frequently punish free riders. Free riders will receive more punishment the less they contribute in an absolute sense and compared to the others. Also, perverse punishers have the opportunity to punish high contributors. The situation is different when only those sanctions are implemented to which a majority of actors consents. A large proportion of actors derives utility from sanctioning. It is therefore likely that majority consent is often reached on punishment of free riders. The more a free rider deviates from the average, the higher the chance that consent is reached. Conversely, when perverse punishment is relatively rare, as is typically found, it will be unlikely that a majority of actors agrees on punishing a high contributor. Thus, a majority sanctioning institution will mitigate perverse punishment while at the same time pro-social punishment is likely to be implemented. We therefore expect a majority decision rule to lead to higher contribution levels than an IDR.

Hypothesis 2.10a *Contributions are higher under a majority than under an individual punishment decision rule.*

Some previous studies indeed found that majority consent is sufficient to rule out perverse punishment, but that pro-social punishment could still be implemented. Casari and Luini (2009) found that punishment was more effective when two out of four actors had to agree on sanctioning a fifth. Perverse punishment was to a large extent ruled out under this decision rule. Likewise, Ertan et al. (2009) let subjects decide by majority vote whether or not to enable punishment of high contributors. While this was sometimes favored by a number of free riders, it was never implemented because a majority opposed the possibility.

Under a unanimity decision rule punishment is only executed when all remaining group members decide to punish an actor. Perverse punishment is therefore even less likely than under a majority decision rule. However, also for pro-social punishment a unanimity decision rule requires a very high proportion of actors willing to punish. Therefore, it will be difficult to implement any punishment at all. Conversely, under an IDR there could be perverse punishment, although the vast majority of punishment should be targeted at below-average contributors. It is therefore likely that contribution levels under a unanimity punishment decision rule are lower than under an individual rule.

Hypothesis 2.10b *Contributions are higher under an individual than under a unanimity punishment decision rule.*

As explained above, continuous need of rewarding makes reciprocating through rewards more expensive than through punishment, which causes the use of rewards to decline (Dari-Mattiacci and De Geest, 2010). This makes sanctioning through punishment more effective. Therefore, we expect that for every decision rule contributions are higher under punishment than under reward.

Hypothesis 2.11 *For every decision rule, contributions are higher under punishment than under reward.*

The more actors are required for a reward to be executed, the more likely it is that too many actors give up on using rewards. Thus, the more actors are required, the more likely it is that consensus cannot be reached anymore. Also, perverse rewards have to be carried out when an actor free rides in anticipation of being rewarded. Perverse rewards are thus likewise costly to maintain. Therefore, while perverse rewards might be occasionally allocated it is unlikely that they are persistently problematic for enforcing cooperation. Thus, rewards under an IDR are not thwarted by perverse sanctions as much as punishment, while it is difficult to raise enough actors to agree on rewards under a CDR. The more actors are required to agree, the more problematic enforcing cooperation becomes. Accordingly, we hypothesize that the more actors are required to agree on rewards, the fewer rewards will be carried out and the lower contribution levels are. Thus, the macro-level hypotheses on rewards are partly different from those on punishment.

Hypothesis 2.12a *Contributions are higher under an individual than under a majority rewarding decision rule.*

Hypothesis 2.12b *Contributions are higher under a majority than under a unanimity rewarding decision rule.*

2.3 Experimental design

In the experiment, subjects participated in interaction situations based on the PGG as described above with group size $n = 4$; endowment $w = 20$, and multiplier $m = 1.6$. The outcome of the game represented points that subjects earned. After the experiment, subjects received 1 eurocent for every 60 points earned.

The experiment comprised three parts. In the first part, preferences for conditional cooperation were assessed using a measure designed by Fischbacher et al. (2001). First, subjects decided on an unconditional contribution, i.e., how much to contribute in the PGG in a group with three other subjects. Second, subjects made this same decision conditional on others' average contribution. Thus, they decided how much they would contribute for every possible average of the three other group members (strategy method, Selten, 1967). The more conditionally cooperative a subject is, the more contribution should increase with others' average. Subjects were randomly matched in groups of four. For three randomly chosen group members, the payoff was calculated based on the unconditional contribution. For the fourth group member the conditional contribution corresponding to the average unconditional contribution of the three others was used. This makes both decisions incentive-compatible. Note that conditionally cooperative preferences were always assessed at the beginning of a session, prior to playing the actual PGGs. Fischbacher and Gächter (2010) measured conditional cooperativeness using a similar design, administered either at the start or end of the experiment. They did not find a sequence effect, suggesting that measuring preferences does not significantly influence subsequent behavior.

In the second part of the experiment, the standard PGG as described above was played for 10 rounds. Between the rounds, subjects were randomly rematched into different groups. They could not infer their group members' previous decisions. After every round, subjects were informed about the contribution decision of the others in their group and their own payoff. Numerous previous experiments have administered baseline games before the experimental treatments (cf. Casari and Luini, 2009; Sefton et al., 2007). No sequence effects were found in experiments where the order of baseline and punishment treatments was randomized (e.g. Fehr and Gächter, 2002; Herrmann et al., 2008).

In the third part, the PGG with sanctions was employed. In every session, 10 rounds were played with only punishment and 10 rounds with only reward; the order varied between sessions. Both reward and punishment took place in one of three experimental conditions; individual, majority, or unanimity. In all three conditions, subjects first decided on a contribution. Subsequently, they were informed about contributions of their group members and decided for all three others separately whether to sanction this person. If executed, a sanction added or subtracted six points from the earnings of the recipient at a cost of two points. This cost ratio 1:3 is often used in PGG experiments (cf. Fehr and Gächter, 2002). The effect and cost of the sanction were chosen to ensure that receiving a sanction has a severe impact on payoffs. Because the amount by which actors could sanction was fixed, the severity of the sanction is equal to the number of actors sanctioning.[6]

In the individual condition, all assigned rewards and punishments were implemented. Subjects who received multiple sanctions were sanctioned by the cumulative amount while all subjects allocating the sanction paid the cost of two points. The procedure in the majority condition was exactly the same, except that the sanction was only executed when at least two group members wanted to sanction the same recipient.

[6]The 1:3 reward ratio enables increasing group earnings through mutual rewarding. However, note that our random matching scheme excludes direct reciprocity. Subjects are therefore unlikely to unilaterally reward all others for the purpose of increasing efficiency.

Thus, an actor sanctioned by two others lost 12 points, while each of the sanctioning actors lost 2 points. In the unanimity condition, the sanction was only executed when it was requested by all three remaining group members. When the number of subjects who wanted to sanction was insufficient in the majority or unanimity condition, the sanction was not executed and no costs had to be paid. Note that the labels 'majority' and 'unanimity' refer to the level of required consensus among the three other group members, excluding the prospective recipient of a sanction. After each round, subjects were informed about all sanctions that had been executed in their group but could not infer who allocated them. No information was provided about sanctions that were not executed. Again, subjects were randomly rematched between the rounds.

The experiment was programmed using z-Tree (Fischbacher, 2007) and conducted at the ELSE laboratory of Utrecht University. Subjects were recruited using the online recruiting system ORSEE (Greiner, 2004). Twelve sessions were held, four in each experimental condition of which two with reward first and two with punishment first. Instructions were provided on paper. It was made clear that the instructions were always truthful and identical for all subjects in a session. In the first set of instructions, the standard PGG and the first two parts of the experiment were explained. It was announced that there would be further tasks, but not what these tasks entailed. These instructions included a number of control questions, which appeared on the computer screen. When a subject did not answer correctly to a question, the answer was explained on the screen. Additional instructions, adapted for each experimental condition, were provided for the reward as well as for the punishment part. The options in the PGG were labeled in a neutral way: punishment and reward were called 'subtracting' and 'adding' points, respectively.

A total number of 184 student subjects participated in the experiment (32% male; 34% economics students). Both the majority and unanimity sessions comprised 64 subjects in total, while 56 subjects were in a session which was held in the individual condition. Payoffs averaged €12.50, with a minimum of €8.50 and a maximum of €15.

2.4 Methods and results

2.4.1 Descriptive results

All subjects participated first in the baseline. Subsequently, they were assigned to either the individual, majority, or unanimity condition. In this condition they participated in the punishment as well as the reward treatment. A Mann-Whitney test reveals no significant effect of the order in which the punishment and reward treatments were administered on average contribution in either the reward ($z = 1.601$; $p = 0.11$) or punishment ($z = 1.441$; $p = 0.15$) games.[7] Therefore, we combine the first and second

[7]Reported p-values of all nonparametric tests are two-sided. The experimental sessions are used as independent observations.

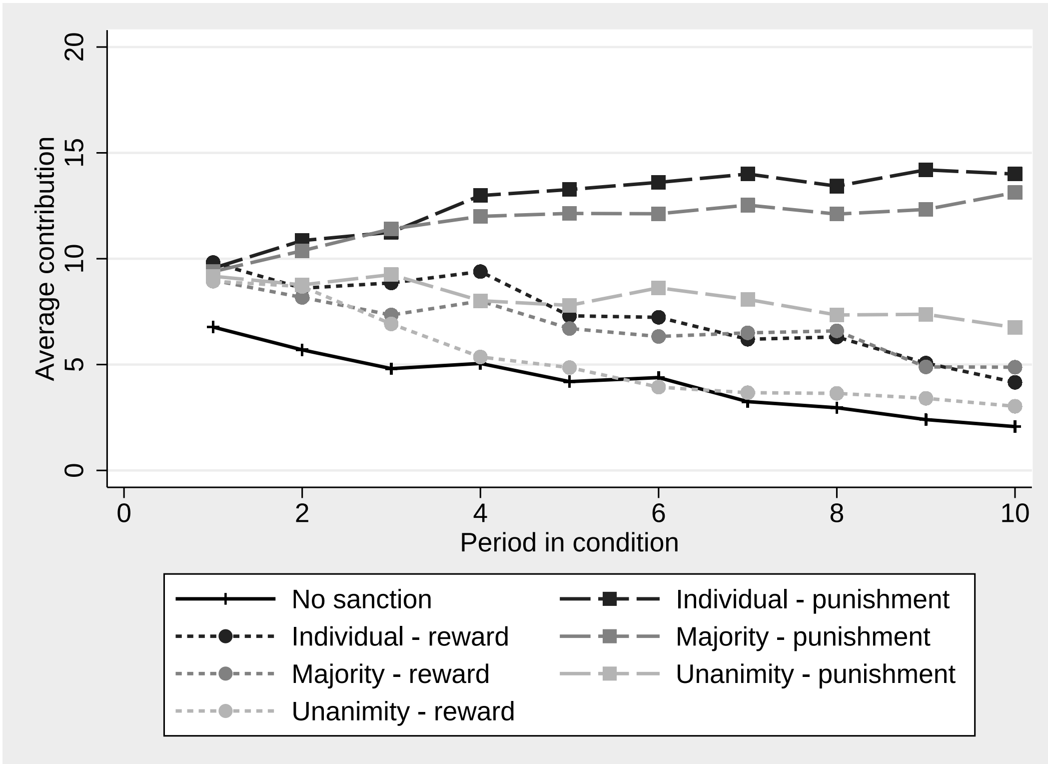


Figure 2.1: Average contribution in the PGGs, separated for each round and experimental condition.

treatment in subsequent analyses. Since the p-values of the order effects are relatively low we check the robustness of our parametric analyses against models in which we include only the first sanctioning treatment that subjects participated in.

Figure 2.1 shows the average contribution levels in the PGGs over the rounds in the baseline and in each experimental condition. Contributions are initially around 50% of the endowment. This is in line with previous findings (Ledyard, 1995). After the first round, Figure 2.1 shows strong differences in contribution levels between the conditions. Contributions in the baseline decline to almost zero. Conversely, individual and majority punishment are the only conditions under which contributions increase over time. A Wilcoxon signed rank test confirms that average contributions are higher in the reward than the baseline ($z = 2.432$; $p = 0.02$) and in the punishment than the reward conditions ($z = 3.059$; $p < 0.01$). For both reward and punishment the individual and majority conditions lead to higher contributions than unanimity, although only the difference between individual and unanimity punishment is significant in a Mann-Whitney test ($z = 2.309$; $p = 0.02$).

Overall average profits are higher in the reward than in both the punishment and baseline treatments (Wilcoxon signed rank test: $z = 2.589$; $p = 0.01$ for baseline vs. reward; $z = 2.981$; $p < 0.01$ for punishment vs. reward; $z = 1.098$; $p = 0.27$ for baseline vs. punishment). However, this is related to our reward technology, which enables earnings to be higher in the reward than the other treatments. The highest possible group earnings are achieved with full contribution in baseline and punishment, and with full contribution and mutual rewarding in the reward treatments. When we consider average earnings as a proportion of the highest possible, this proportion is higher in both the punishment ($z = 3.059$; $p < 0.01$) and baseline ($z = 3.059$; $p < 0.01$) than in the reward treatments. With regard to the decision rules, we find that earnings are higher in the individual than in the unanimity reward condition ($z = 2.021$; $p = 0.04$). Other differences in earnings between the decision rules are insignificant in nonparametric tests.

To test how parametric condition differences in earnings correspond to differences in contribution levels, we replicate Model 1 with earnings as the dependent variable (output not shown). We use a multilevel model with the earnings of a given subject in a given period nested in subjects and sessions. The resulting pattern of condition differences in earnings is the same as that in contributions, except of course that absolute earnings under reward are higher than in the baseline and punishment treatments. Furthermore, only the difference in earnings between the baseline and the unanimity punishment condition is not significant. Thus, the significant increase of contributions under unanimity punishment does not result in a significant increase in earnings. Other differences between the conditions in earnings are significant, although the significance of the differences between the three punishment decision rules drops to the 10%-level in a two-sided test. Because condition differences in contributions and earnings are similar, we report only on contributions in the remainder of the analysis section.

When a subject was targeted for punishment in the majority condition, in 58% of the cases this was by one person only and therefore the punishment was not carried out. Likewise, in 81% of the cases in which a subject was targeted for punishment in the unanimity condition, the required number of three sanctioning subjects was not reached. For reward, in 72% of the cases in which someone was targeted for reward in the majority condition and in 97% of the cases under unanimity, the reward was not implemented. In line with previous research, 25% of punishments were targeted at subjects contributing the average of other group members or more. Of these, 91% and 98% were not implemented in majority and unanimity, respectively. 33% of rewards were targeted at below-average contributors, of which 89% and 100% were not implemented under majority and unanimity.

Figure 2.2 shows the average number of sanctions allocated and average number of sanctions carried out for different deviations of the recipient from the average contribution of the other group members. Note that between one and three other group members can propose to sanction. Figure 2.2 shows a clear trend of more punishment proposed on average the more the recipient negatively deviates from the average contribution of others. Also, more rewards are proposed for above-average contributors, but it is not so clear that more rewards are proposed the further the deviation.

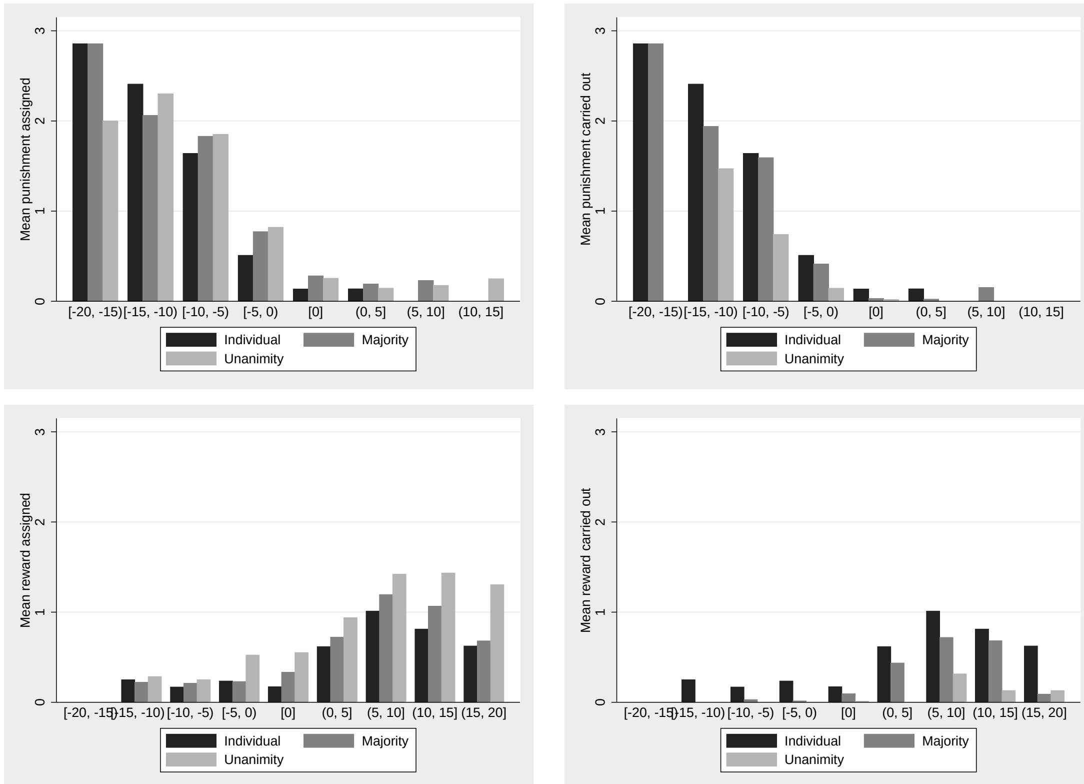


Figure 2.2: Average punishment (above) and reward (below) assigned (left) and carried out (right) for different deviations from the average contribution of other group members, separated for each experimental condition.

2.4.2 Methods – contribution

The first dependent variable, contribution, is measured as the contribution decisions of subjects in the PGG. First, we test the macro-level hypotheses by comparing dummies for the experimental conditions individual, majority, and unanimity punishment and reward. These are less conservative tests for the differences between conditions than the comparisons in the previous subsection, because the interdependencies between the observations are modeled in more detail. Still, the results mainly reconfirm the differences that resulted from the non-parametric tests. Second, we test the micro-level hypotheses *explaining* differences between experimental conditions based on individual decision patterns. The punishment and reward treatments are analyzed separately.

In the micro-level models, sanctions received are measured as the number of others who had sanctioned the subject in the previous round. Only executed sanctions are included. Furthermore, three dichotomous variables indicate whether in the previous round a subject had contributed more than 4 points below the average of other group members, more than 4 points above the average, or did not deviate from the average by more than 4 points. These three dummies for previous deviation are interacted with the number of sanctions received to test whether the effect of being sanctioned is different for above- and below-average contributors.

Previous deviation was measured using dummies for 4 points higher/lower rather than a continuous variable indicating the precise extent of the deviation. This is because a continuous variable interacted with received reward tests if subjects increase (decrease) their contribution more, the higher (lower) the contribution for which they were rewarded. This is unrealistic, since contribution is limited between 0 and 20. The boundaries of 4 points from the average are chosen such that the deviation is substantial enough for subjects to perceive sanctions as clearly norm-enforcing or perverse. Accordingly, log-likelihoods of models with different boundaries are equal to or lower than those of the models presented here. We control for the subject's contribution in the previous round, round number, treatment order, and experimental condition. Furthermore, preference for conditional cooperation is included, measured as the slope of the conditional contribution assessed in the first part of the experiment. The steeper the slope, the more a subject indicated to contribute more when others do so as well.[8]

We use Tobit regression to account for contribution with a limited range, between 0 and 20, with both extremes often chosen. The units of analysis are decisions in the PGGs. Since every subject makes multiple contribution decisions, random effects at the subject level are included to account for decisions nested in subjects. Also, within a session subjects often encounter others with whom they or their group members have interacted previously. Thus, subjects are interdependent within sessions. It is not possible to include both the subject and session level in a three-level Tobit model. Therefore, all models were replicated using multilevel linear regression, in which both subject and session level random effects are included but where contribution is treated as if its range is unlimited. Also, we estimated the models using Tobit regression with random effects at the session level to test if disregarding this level in the models presented below influenced the results, and we ran a Tobit model with robust standard errors adjusted for clustering within sessions. The latter model provides the most conservative way of correcting for the clustering of observations and, therefore, might underestimate the significance of some effects. Given the limited effect of the session level in, e.g., the three-level linear regression model, we have considerable confidence in the estimations of the two-level Tobit models with random effects for subjects reported in the tables. Finally, we examined the possible effects of punishment and reward treatment order in more detail by rerunning all models with only the first treatment that subjects participated in included. Below, effects of treatment order and robustness of the results in alternative analyses are discussed for every model separately.

[8]Two subjects whose slopes are zero, but who do make positive conditional contributions (both unconditionally contribute a certain amount) are excluded from the analysis. A zero slope thus indicates a preference for unconditional free riding. Excluding these subjects did not influence the results.

Table 2.1: Tobit regression on contribution decisions with random effects at subject level (5460 decisions, of which 2376 censored, by 182 subjects).

	Model 1	
	Coeff.	S.e.
Baseline	ref.	
Punishment - individual	13.938**	0.518
Punishment - majority	10.239**	0.464
Punishment - unanimity	5.866**	0.479
Reward - individual	6.184**	0.528
Reward - majority	2.770**	0.474
Reward - unanimity	0.340	0.501
Constant	0.786	0.522
σ_u	6.372**	0.380
σ_e	7.934**	0.113
Log Likelihood	-12773.784	

*Significant at the .05-level. **Significant at the .01-level (2-sided)

2.4.3 Explanatory results – contribution

Table 2.1 shows differences in contribution decisions between the experimental conditions. The baseline condition, in which every subject participated, serves as a reference. Contributions in all experimental conditions except unanimity reward were higher than in the baseline, although the effect of majority reward is insignificant when we adjust for clustering within sessions. Contrary to Hypothesis 2.10a, contribution under punishment is higher in the individual than the majority condition ($\chi^2(1) = 29.51$; $p < 0.01$). The other macro-level hypotheses are confirmed. Contribution under punishment is higher in the individual than the unanimity condition ($\chi^2(1) = 136.58$; $p < 0.01$), confirming Hypothesis 2.10b. As predicted in Hypothesis 2.11, contributions are higher under punishment than reward in the individual ($\chi^2(1) = 228.83$; $p < 0.01$), majority ($\chi^2(1) = 246.01$; $p < 0.01$), and unanimity ($\chi^2(1) = 122.01$; $p < 0.01$) condition. Finally, contribution under reward is higher in the individual than the majority condition ($\chi^2(1) = 23.76$; $p < 0.01$) and higher in the majority than the unanimity condition ($\chi^2(1) = 12.79$; $p < 0.01$). This confirms Hypotheses 12a and 12b. All differences between decision rules are insignificant in the conservative model that accounts for clustering in sessions, but remain highly significant in other model specifications. The differences between punishment and reward remain significant in every alternative specification.

Because we want to exclude that the support for the hypotheses confounds with effects of subjects playing a punishment and reward treatment after each other, we

Table 2.2: Tobit regression on contribution decisions in the punishment conditions with random effects at subject level (1638 decisions, of which 345 censored, by 182 subjects).

	Exp. sign	Hyp.	Model 2		Model 3	
			Coeff.	S.e.	Coeff.	S.e.
Previous punishment received			1.006**	0.170		
Prev. neg. deviation > 4			1.102*	0.440	0.045	0.514
× Punishment received	+	2.6			1.598**	0.226
Prev. dev. ≤ 4			ref.		ref.	
× Punishment received					0.285	0.256
Prev. pos. deviation > 4			-1.905**	0.347	-1.985**	0.353
× Punishment received	–	2.7			0.361	0.829
Previous contribution			0.642**	0.044	0.653**	0.044
Slope conditional contribution			0.986	0.559	0.898	0.543
Period			-0.014	0.043	-0.015	0.043
Individual			ref.		ref.	
Majority			-0.057	0.614	-0.101	0.596
Unanimity			-2.907**	0.650	-2.838**	0.633
Punishment treatment first			-1.328**	0.506	-1.248*	0.492
Constant			10.568**	0.594	10.199**	0.591
σ_u			2.964**	0.263	2.853**	0.262
σ_e			4.189**	0.093	4.180**	0.093
Log Likelihood			-4130.177		-4122.261**	

*Significant at the .05-level. **Significant at the .01-level (2-sided)

also consider effects of the ordering of treatments. Contributions in the punishment conditions are lower when punishment was the first compared to when it was the second treatment. In the reward conditions, contributions are higher when it was the first treatment. Still, when we only consider the first treatments subjects participated in, contributions are higher in the individual than in the majority conditions, although this difference becomes insignificant for reward ($\chi^2(1) = 0.93$; $p = 0.34$). Also, contributions are higher in the majority than in the unanimity conditions. Finally, contributions are higher in the individual and unanimity punishment conditions than in the related reward conditions. Only in the majority condition this difference disappears ($\chi^2(1) = 0.16$; $p = 0.69$). Hence, the confirmation of this part of Hypothesis 2.11 should be interpreted with caution.

The micro-level model for the punishment conditions is presented in Table 2.2. Only main effects are included in Model 2. Several control variables are significant. Contributions are lower in the unanimity compared to the individual condition and when punishment was administered first, and higher the more a subject contributed in the previous round. The difference between the individual and majority condition is not significant in this model. Subjects who contributed 4 points or more below the average increase and subjects who contributed above the average decrease their contribution compared to around-average contributors. Also, contributions are higher the more punishment was received previously.

The interaction effects are included in Model 3. The main effect of punishment is excluded from this model, so the three interactions represent the effect of received punishment for the three groups of subjects belonging to specific deviations from the mean contribution. The model shows that subjects contributing below the average increase their contribution more, the more they are punished. Hypothesis 2.6 is thus confirmed. The insignificant main effect of negative deviation indicates that subjects who contributed below the average but were not punished do not significantly increase their contribution compared to around-average contributors. Subjects who contributed above the average decreased their contribution if they had not been punished, but did not significantly decrease their contribution further after receiving punishment. Thus, no support is found for Hypothesis 2.7. This might be due to the relatively limited amount of sanctioning against high contributors even in the individual condition. The effect remains insignificant in a separate analysis of the individual condition.

The findings in Models 2 and 3 are similar in a multilevel model, with random effects and clustering at session level, and in a model in which only the first treatments are considered. All hypothesis-related effects are robust.

Model 4 in Table 2.3 shows the determinants of contribution decisions in the reward conditions. In this model the differences between the experimental conditions and treatment order are not significant. The other control variables are significant; contributions are higher the more conditionally cooperative a subject is and the more a subject contributed previously, and decreases over rounds. Subjects who previously contributed above the average decrease and those who contributed below the average increase their contribution compared to around-average contributors. Finally, the more rewards a subject had previously received, the higher the contribution.

In Model 5, the interaction effects are included. Again, the three interactions represent the separate main effects. This shows that subjects who had contributed above the average significantly decrease their contribution. However, the decrease is significantly weaker the more they were rewarded. This confirms Hypothesis 2.8. Very few subjects received rewards after a below-average contribution, and virtually all rewards were ruled out by majority and unanimity. Hence, we find no significant effect of being rewarded for around-average or below-average contributors. The effect remains insignificant in a separate analysis of the individual condition. Hypothesis 2.9 is thus not confirmed. Again, findings are similar in a multilevel model, with random effects and clustering at session level, and in a model in which only the first treatments are considered All hypothesis-related effects are robust.

Table 2.3: Tobit regression on contribution decisions in the reward conditions with random effects at subject level (1638 decisions, of which 981 censored, by 182 subjects).

	Exp. sign	Hyp.	Model 4		Model 5	
			Coeff.	S.e.	Coeff.	S.e.
Previous reward received			1.571**	0.511		
Prev. neg. deviation > 4			2.588**	0.869	2.491**	0.884
× Reward received	–	2.9			0.730	1.937
Prev. dev. ≤ 4			ref.		ref.	
× Reward received					0.001	0.952
Prev. pos. deviation > 4			-4.876**	0.994	-5.751**	1.083
× Reward received	+	2.8			2.173**	0.593
Previous contribution			0.882**	0.091	0.901**	0.092
Slope conditional contribution			5.521**	1.563	5.507**	1.571
Period			-0.752**	0.123	-0.753**	0.123
Individual			ref.		ref.	
Majority			0.580	1.703	0.522	1.715
Unanimity			-2.669	1.746	-2.792	1.758
Reward treatment first			2.649	1.384	2.727	1.391
Constant			1.971	1.525	1.910	1.551
σ_u			8.135**	0.774	8.191**	0.775
σ_u			9.593**	0.316	9.560**	0.315
Log Likelihood			-3063.380		-3061.248	

*Significant at the .05-level. **Significant at the .01-level (2-sided)

2.4.4 Methods – sanctioning

The second dependent variable in the analysis of the micro-level framework are the decisions whether or not to sanction. These are three observations for each subject in each period, one for every other group member.

The first independent variable is a subject's own contribution. Second, the contribution of the potential recipient of a sanction is included as a continuous variable. Third, the deviation of the potential recipient from the average contribution of others is measured as the contribution of the potential recipient minus the average of the other group members. This deviation is included as two separate variables. The variable 'positive deviation' includes all positive values of this measure, negative values are set to zero. The variable 'absolute negative deviation' represents the extent of the deviation of all negative values, zero for positive deviations. For punishment, the contribution and absolute *negative* deviation of the potential recipient are interacted with a subject's own contribution to test whether high contributors are more likely to punish the less the potential recipient contributes, and the further she deviates from the average. For reward, contribution and *positive* deviation of the potential recipient are interacted with a subject's own contribution. We control for experimental condition, treatment order, slope of the conditional contribution, and for sanctions assigned and received by a subject in the previous round.

We use logistic regression to analyze the dichotomous sanctioning decisions. Every subject makes three sanctioning decisions, one for every other group member, in all ten periods. Decisions are thus nested within periods and subjects. A multilevel intercept-only model with decisions nested in periods and subjects revealed that variance at the period level is negligible for both punishment and reward decisions. We therefore present multilevel models with decisions nested only in subjects. All models were repeated using only the first treatment subjects participated in. Below, we discuss the sequence effects of all models.

2.4.5 Explanatory results – sanctioning

Models on punishment decisions are displayed in Table 2.4. Model 6 shows that there are no differences between the experimental conditions in the likelihood that a subject decides to punish another. We do find that subjects who have received or have allocated punishment in the previous round are more likely to punish. The likelihood of punishing increases with contribution, confirming Hypothesis 2.1. Also, the more a potential recipient of a punishment negatively deviates from others' contribution, the higher the likelihood that punishment is allocated while no effect is found for positive deviation. Finally, the more a potential recipient contributes, the less likely subjects are to punish this person.

Model 7 shows a significant interaction effect of contribution with the contribution of the potential recipient of a punishment, confirming Hypothesis 2.2a. A significant interaction with the negative deviation of the potential recipient confirms Hypothesis 2.2b. High contributors are thus more likely to punish the less a potential recipient contributes in absolute sense, and relative to the average of others.

The effect that high contributors punish especially others who contribute less than average (Hypothesis 2.2b) is not found if we only consider the first treatment for Model 7. This is probably due to the lower number of observations when only one treatment is included, which makes it more difficult to disentangle the different reasons why high contributors punish others.

Table 2.5 shows the models on reward. Main effects included in Model 8 show that subjects in the unanimity condition are more likely than in the individual condition to allocate rewards. Furthermore, subjects are more likely to reward the more rewards they had allocated in the previous period. The effect of period is significant, confirming Hypothesis 2.5. Also, subjects are more likely to reward the more the potential recipient of the reward contributes, but not the higher the positive deviation from the average.

Table 2.4: Multilevel logistic regression on decisions whether to punish nested in subjects (4914 decisions by 182 subjects).

	Exp. sign	Hyp.	Model 6		Model 7	
			Coeff.	S.e.	Coeff.	S.e.
Contribution	+	2.1	0.091**	0.018	0.121**	0.019
Contribution recipient			-0.220**	0.027	-0.167**	0.029
× Contribution	–	2.2a			-0.020**	0.003
Positive dev. recipient			-0.018	0.032	-0.075*	0.034
Absolute neg. dev. recipient			0.283**	0.029	0.211**	0.033
× Contribution	+	2.2b			0.012**	0.004[1]
Round			0.029	0.020	0.050*	0.021
Individual			ref.		ref.	
Majority			0.203	0.394	0.239	0.444
Unanimity			-0.225	0.410	-0.188	0.458
Slope conditional contribution			0.295	0.357	0.145	0.402
Previous punishment received			0.301**	0.066	0.268**	0.067
Previous punishment assigned			0.291**	0.067	0.255**	0.068
Punishment first treatment			-0.604	0.323	-0.861*	0.364
Constant			-1.500**	0.371	-1.311**	0.409
σ_u			1.930	0.168	2.187	0.190
Log Likelihood			-1505.348		-1444.317**	

*Significant at the .05-level. **Significant at the .01-level (2-sided)

[1]Hypothesized effect not significant when only the first treatment is considered.

Rewarding is less likely the more the potential recipient negatively deviates. Hypothesis 2.3 is supported: subjects who made a higher contribution are more likely to reward.

Model 9 shows the interaction of a subject's own contribution with the contribution and positive deviation of the potential recipient. The significant effects indicate that high contributors are more likely to reward the higher and the further above the average someone contributes, confirming Hypotheses 4a and 4b.

The main effect of contribution (Hypothesis 2.3) and the effect that high contributors reward especially others who contribute much (Hypothesis 2.4a) are not found if we only consider the first treatment for Model 9. Again, this is probably due to the lower number of observations when only one treatment is included, which makes it more difficult to disentangle the different reasons why high contributors reward others.

Table 2.5: Multilevel logistic regression on decisions whether to reward nested in subjects (4914 decisions by 182 subjects).

	Exp. sign	Hyp.	Model 8 Coeff.	Model 8 S.e.	Model 9 Coeff.	Model 9 S.e.
Contribution	+	2.3	0.070**	0.012	0.037**	0.013[1]
Contribution recipient			0.157**	0.019	0.150**	0.021
× Contribution	+	2.4a			0.005**	0.002[1]
Positive dev. recipient			0.032	0.020	0.074**	0.023
× Contribution	+	2.4b			0.007**	0.003
Absolute neg. dev. recipient			-0.115**	0.025	-0.047	0.026
Round	–	2.5	-0.052*	0.024	-0.065**	0.024
Individual			ref.		ref.	
Majority			0.349	0.502	0.361	0.489
Unanimity			1.364**	0.506	1.311**	0.492
Slope conditional contribution			0.745	0.452	0.662	0.439
Previous reward received			-0.179	0.092	-0.187	0.100
Previous reward assigned			0.346**	0.075	0.350**	0.076
Reward first treatment			0.117	0.403	0.126	0.392
Constant			-3.567**	0.442	-3.517**	0.430
σ_u			2.438	0.210	2.355	0.205
Log Likelihood			-1311.265		-1281.752**	

*Significant at the .05-level. **Significant at the .01-level (2-sided)
[1]Hypothesized effect not significant when only the first treatment is considered.

2.5 Conclusion and discussion

We compare the effect of individual, majority, and unanimity decision rules for implementing punishment and reward on actors' ability to enforce cooperation in a Public Goods Game (PGG). For punishment, we conjecture that contributions are higher under a majority than an individual decision rule (Hypothesis 2.10a). We find higher contributions under the individual decision rule instead, although evidence for a difference in earnings is less strong. This is because in the majority condition there are fewer punishments. As expected, we do find that contributions are lower under a unanimity than an individual punishment decision rule (Hypothesis 2.10b). For reward, the hypotheses concerning the effects of decision rules on contribution levels are all confirmed. We find that contributions are higher under an individual than a majority decision rule (Hypothesis 2.12a) and higher under a majority than a unanimity decision rule (Hypothesis 2.12b). In sum, for both punishment and reward contributions are lower, the more actors are required to agree on sanctioning. Also, as hypothesized, contributions are higher under punishment than reward for every decision rule (Hypothesis 2.11), although no difference is found in the majority condition when only the first treatment with sanctions is considered. Condition differences in earnings are similar.

Findings on individual behavior, as captured in the micro-level hypotheses, offer an explanation for the observed differences in contribution between the decision rules. The emerging pattern is very similar for reward and punishment. Hypotheses on the use of pro-social sanctions are all confirmed. High contributors are more likely to punish (Hypothesis 2.1) and to reward (Hypothesis 2.3) than low contributors. These high contributors enforce the norm that others should contribute as well. That is, they are more likely to punish the less a potential recipient contributes (Hypothesis 2.2a) and the lower the contribution of the potential recipient is compared to the other group members (Hypothesis 2.2b). Likewise, high contributors reward group members who also make a high contribution (Hypothesis 2.4a) and who contribute more compared to the others (Hypothesis 2.4b). In other words, there is more consensus on sanctions among high contributors, the more an actor violates or adheres to their cooperative norm. Still, many punishments and rewards under the majority and unanimity decision rules were not executed. This implies that reaching the required number of actors was difficult despite the high consensus on whom to target.

When low contributors are punished, they contribute more in the subsequent interaction (Hypothesis 2.6). Similarly, actors who are rewarded for contributing more than other group members increase their contribution compared to others who are rewarded less (Hypothesis 2.8). Thus, we find strong evidence that pro-social sanctions have a positive effect on contribution levels. Conversely, perverse sanctioning occurred too infrequently to affect contribution levels. We cannot confirm that high contributors decrease their contribution after being punished perversely (Hypothesis 2.7). Likewise, contrary to our expectations, free riders who are rewarded perversely do not decrease their contribution further (Hypothesis 2.9). We did find that almost all perverse sanctions were ruled out under majority and unanimity.

In sum, we find strong evidence for the allocation of pro-social sanctions, and their positive effects on contribution levels. Concurrently, perverse sanctions occur too infrequently to affect cooperation. This makes an individual decision rule (IDR) unproblematic: punishment is mostly targeted at free riders regardless of the possibility for individual actors to sanction perversely. Because more pro-social sanctions are obstructed the more actors are required for the collective decision rule (CDRs), we observe lower contribution levels the more actors are required to agree. The observed micro-level behavior thus explains the macro-level finding of lower contribution levels under unanimity than majority, and lower contributions in the majority than in the individual condition.

The use of rewards decreases over time (Hypothesis 2.5). This provides an additional impediment for CDRs, because it implies that the more actors are required to agree, the sooner consensus cannot be reached anymore. Rewards are therefore even more problematic to enforce than punishment, hence contributions are higher under punishment than reward.

Casari and Luini (2009) find, in groups of five, that punishments on which two out of four actors agree are much more effective than sanctions with an IDR. We use stricter CDRs of two and three actors in groups of four, and find that contributions are highest under an IDR. However, contribution levels in our majority punishment condition (Figure 2.1) and the CDR of Casari and Luini (2009, Figure 2.1) are very similar. The difference between their findings and ours is that contribution in their individual punishment condition is much lower than in our experiment. Herrmann et al. (2008) find such differences in contributions under individual decision rules between subject pools. They attribute this to different levels of perverse punishment. Indeed, Casari and Luini (2009) find that contributions in their individual punishment condition are diminished due to perverse punishments. We find that perverse punishments do not affect contributions even under an IDR.

We started with the observation that actors engaged in real-life public goods problems often use CDRs to successfully enforce cooperation. One possible reason why we find that an IDR is more effective might be that interactions in our experiment are one-shot and anonymous rather than repeated. In many public goods problems outside the laboratory, especially in small communities or between nations, participants interact repeatedly. Moreover, actors can often communicate before deciding whether or not to sanction. Repeated interaction and communication both imply that actors can coordinate on raising the required proportion of agreeing actors.

Furthermore, it is often possible to identify which actors neglected to agree on sanctioning. Therefore, when the required consensus is not reached the actors who did not sanction can be held accountable, for example through second-order punishment (Cinyabuguma et al., 2006; Denant-Boemont et al., 2007; Nikiforakis, 2008). Previous studies found that second-order punishment is not always effective because it is used by defectors to punish first-order punishers. This issue should be alleviated when CDRs are used, because responsibility for punishment is shared by multiple others. Also, when a CDR is used for *implementing* second-order punishment, it may be difficult to reach agreement on punishment of punishers.

Finally, in our experiment actors had complete information about others' contributions. However, in other contexts, some of the actors might make an inaccurate observation of the contributions of some of the others. An IDR might lead to inaccurate sanctioning decisions in such an environment (e.g. Grechenig et al., 2010). However, under a CDR mistaken sanctions caused by a wrong observation of an actor's contribution are likely to be ruled out when inaccurate observations are independent across group members.

Repeated interactions, communication on whom to sanction, public announcement of sanctioning decisions, use of counter-punishment, and noise can be implemented in future experiments to enhance resemblance with public goods problems outside the laboratory. As indicated above, these adaptations might favor CDRs, because coordination of sanctions in CDRs can become easier and mistakes in sanctions can be prevented. Still, the disentangling of sanctions through reciprocal contributions and sanctions through exogenous institutions will remain a challenge in some of these setups. In addition, there might also be some more realistic specifications of the interaction situation which favor IDRs. Most importantly, non-implemented sanctions are costless in our setup. In other contexts, it might be more plausible that people have to invest in sanctioning before knowing whether others will agree. This would make implementation of sanctions under a CDR even more problematic. Future research should further specify conditions under which either CDRs or IDRs are more successful in enforcing cooperation.

Chapter 3

Determinants of perceived fairness of punishment and reward in Public Goods Games*

Abstract

We employ a laboratory experiment on Public Goods Games with opportunities for peer punishment (negative sanction) or reward (positive sanction). We consider equity, collective agreement on sanctioning, and self-serving bias as determinants of subjects' perceived fairness of sanctions that they and fellow group members receive. We find that subjects perceive punishments received by other group members as fairer when punishments result in equitable outcomes. Contrary to our expectations, rewards received by other actors and personally received sanctions are not viewed as fairer when sanctions improve equity levels. Additionally, we cannot confirm our hypothesis that the degree to which sanctioning decisions are imposed through collective agreement among group members systematically influences fairness perceptions. We can confirm that sanctions received by others are perceived as fairer as more of a subject's own proposed sanctions are implemented through collective agreement. Finally, as expected, receiving punishment negatively affects perceived fairness, while receiving reward has a positive effect, suggesting that subject fairness perceptions are biased towards own outcomes.

*This is a slightly different version of Van Miltenburg, Buskens, and Barrera (2014). We thank Werner Raub for providing comments and feedback and Jeroen Weesie for providing suggestions on our analyses.

3.1 Introduction

In daily life, many situations involve decisions on how much one should contribute to a public good. Individuals decide whether to clean a common area, put effort in a team project, or honestly pay their taxes. Though contributing to a public good is costly, every involved actor benefits from each contribution. Consequently, every actor is better off under full cooperation than when no one contributes, while individual actors face an incentive to free ride on the contributions of others, rendering public goods problems a social dilemma (Dawes, 1980; Raub et al., 2015).

One means of encouraging actors to contribute to a public good involves establishing a peer sanctioning institution. In experimental research, peer-sanctioning institutions typically imply that, after observing the contributions of their peers, all actors may decide to reduce or increase the welfare of each fellow group member at a private cost (cf. Fehr and Gächter, 2000, 2002; Ostrom et al., 1992; Yamagishi, 1986). Henceforth, decisions to reduce another actor's welfare are referred to as punishments, and decisions to increase another actor's welfare are referred to as rewards. The general term 'sanction' refers to both reward and punishment decisions.

It is widely recognized that peer punishment and reward institutions are often successful in promoting cooperation (see Balliet et al., 2011; Chaudhuri, 2011 for an overview). Moreover, sanctioning institutions are found to affect actors' intrinsic motivation to contribute to a public good (Fehr and Falk, 2002; Mulder, 2008; Tenbrunsel and Messick, 1999). As it is often found that individuals are more motivated to pursue the collective interest when they perceive an interaction situation as fair (Biel et al., 1997; Eek and Biel, 2003; Fehr and Rockenbach, 2003; Jost and Kay, 2010; Van Prooijen et al., 2008), the perceived fairness of sanctions that are allocated under a peer sanctioning institution may affect actors' intrinsic motivation to contribute. In this chapter, we study determinants of actors' perceptions of the fairness of punishments and rewards allocated through peer-sanctioning institutions in public goods problems. We distinguish between the perceived fairness of sanctions that actors receive personally and of sanctions received by fellow group members.

Perceived fairness as a side effect of peer sanctioning has received relatively little attention in the social dilemma literature (Schroeder et al., 2003; Van Prooijen et al., 2008). However, sanctions that are perceived as unfair may have detrimental effects on cooperative behaviors. To obtain rewards or avoid being punished, actors are forced to make a high contribution if a sanctioning institution is in place and if potential sanctioners observe their actions. If actors are not internally motivated to cooperate as well, cooperation may dissolve when the sanctioning institution is removed (De Cremer and Tyler, 2005; Mulder et al., 2006) or when contributions cannot be monitored sufficiently (Mulder and Nelissen, 2010). Alternatively, actors may react to sanctions that they perceive as unfair by limiting their contribution in another area where enforcement is less strict (Cornelissen et al., 2013). Conversely, it is found that sanctioning institutions that are perceived as fair increase actors' feeling of belonging to the group, thereby motivating them to pursue the group interest (De Cremer and Tyler, 2005). Further insight into determinants of actors' views on the fairness of sanctions that they and their peers receive may help prevent potentially negative side effects of unfair sanctions.

The most basic determinant of fairness perceptions may relate to outcomes attained through peer sanctions, i.e., the manner in which allocated sanctions affect the distribution of earnings between group members (Jost and Kay, 2010). In a public goods problem without a sanctioning institution, those who contribute the least obtain the highest earnings. This negative relation between contributions and earnings becomes weaker and may even be reversed when free riders are punished or when high contributors are rewarded. Sanctions that actors receive themselves and those that other group members receive may be judged as fairer as they alleviate the inverse relation between contributions and earnings.

Fairness perceptions are not only influenced by outcomes but also by procedures through which outcomes are attained (e.g., Aksoy and Weesie, 2009; Dolan et al., 2007; Frey et al., 2004; Jost and Kay, 2010; Sen, 1995). One of the most peculiar features of the procedure through which peer sanctions are typically allocated in experimental research is that each actor can individually decide whether to sanction others (Strimling and Eriksson, 2014). Henceforth, this is referred to as an individual decision rule (IDR). Conversely, in daily life, peer-sanctioning institutions often entail some form of collective decision-making for the allocation of rewards or punishments (Decker et al., 2003; Guala, 2012; Ostrom, 1990). In this chapter, we compare the perceived fairness of sanctions allocated through an IDR to those allocated through collective decision rules (CDRs). Under a CDR, sanctions are only implemented when a certain proportion of group members agree that an actor should be sanctioned. Sanctions that do not achieve sufficient collective agreement are not implemented. When actors are sanctioned themselves, they may perceive the procedure as fairer as more group members are required to agree on sanctioning decisions, because fellow group members are less likely to sanction them erratically as more agreement is required. When fellow group members are sanctioned, actors may perceive procedures as more fair, the more freedom they have in deciding to sanction someone.

Finally, actors may be biased towards their own outcomes in judging the fairness of sanctions that they receive. This may lead actors to perceive personally received punishments as unfair, and rewards as fair, regardless of whether these sanctions resulted in a fairer distribution of resources or were allocated through a fair procedure.

In this chapter, we study outcomes of sanctioning decisions and procedures through which sanctions are allocated as determinants of the perceived fairness of punishments and rewards received by actors themselves and by their peers. In Section 3.2, from a review of previous research, we discuss which established determinants of fairness perceptions are relevant in public goods problems with peer-sanctioning institutions. Hypotheses are formulated accordingly and tested in a laboratory experiment. In the experiment, which is described in Section 3.3, subjects participate in incentivized public goods problems with peer-sanctioning institutions, while fairness perceptions are measured through a questionnaire that originates in organizational research, and that includes measures of perceived outcome and procedural fairness. To the best of our knowledge, we are the first to apply this measure of fairness perceptions to incentivized laboratory cooperation problems. The results are outlined in Section 3.4, and Section 3.5 concludes.

3.2 Theory

3.2.1 The Public Goods Game

In the experiment, public goods problems are modeled through series of one-shot Public Goods Games (PGGs). Each PGG starts with a contribution stage. In this stage, actors interact in groups of size n. Each group member receives an equal endowment and determines how much of this endowment to contribute to a group account. The proportion of the endowment that is not contributed is added to the actors' private resources. The total amount contributed is multiplied by a number m, whereby $1 < m < n$, and is then divided equally among the actors. Group welfare thus increases with every contribution (because $m > 1$), while individual actors earn less the more they contribute (because $m < n$).

Peer punishment and reward are both modeled as a second stage of the PGG. In this stage, each actor observes contributions that every other actor has made to the group account. Subsequently, in a punishment stage, all actors decide for every other group member whether or not to punish, and in a reward stage, all actors decide whether or not to reward their peers. When implemented, both punishments and rewards involve a cost to the actor allocating the sanction. The earnings of the recipient of an implemented sanction are decreased by three times this cost in case of punishment and increased by three times this cost in case of reward. When multiple actors sanction the same group member and the sanction is implemented, all sanctioners pay the cost, and the recipient gains or loses three times the cost for each group member who paid to sanctioned him or her.

We consider three decision rules through which punishments and rewards are implemented. Under the IDR, every sanction is implemented. Under the CDR that requires majority agreement, sanctions are only implemented when a majority of group members, excluding the prospective recipient, sanction the same actor. Under the CDR that requires unanimous agreement, sanctions are only implemented when all remaining group members sanction the same actor. If an actor decides to sanction under a CDR but the sanction is not implemented, no sanctioning costs are paid and the earnings of the intended recipient remain unaffected. After the punishment or reward stage, actors are informed of the implemented sanctions received by each group member and of their own earnings. Actors are not informed of the earnings of the other group members. Non-implemented sanctions are not communicated to others.

In our experiment, subjects are engaged in series of PGGs with either a punishment or reward stage. After the series of PGGs are completed, actors are asked to disclose their views on the overall fairness of the sanctions allocated throughout a series. The structure of each PGG is predefined as described above. Interactions proceed through a computer screen that rules out all forms of interaction between subjects aside from those involved in the PGG. Subjects do not know the identity of their fellow group members. Subjects are randomly matched to different group members after each PGG in a series is completed.

3.2.2 Fairness

Several forms of fairness or justice are identified in the literature (Jost and Kay, 2010), of which two are applicable to our predefined anonymous one-shot PGGs. The first is referred to as distributive fairness, or the fairness of the division of costs and benefits among actors (Adams, 1963, 1965; Deutsch, 1975, 1985). In the contribution stage of the PGG, actors earn more as they contribute less to the public good. Decisions made by actors in the sanctioning stage determine the extent to which the negative relation between contribution and earnings persists. The second type, procedural fairness, implies that irrespective of outcomes, procedures through which allocations are arrived at should be fair (Leventhal, 1980; Thibaut and Walker, 1975). In our experiment, the procedure is represented by decision rules through which sanctions are implemented.

Actors may interpret these two types of fairness differently depending on whether they are judging sanctions that they receive themselves or those that their fellow group members receive. The fairness of sanctions that actors receive themselves may be interpreted in a manner that favors personal outcomes (Messick and Sentis, 1979), while the perceived fairness of sanctions received by fellow group members may be influenced by the fact that some sanctions may have been allocated by the focal actor. Hence, the perceived fairness of sanctions received by actors themselves and by fellow group members is measured and analyzed separately and occasionally differentiated in the theoretical discussion. We proceed by describing these two forms of fairness in the context of our PGGs.

3.2.3 Distributive fairness

After the contribution stage of a PGG, the distribution of earnings across group members is such that actors earn more as they contribute less to the public good. When an actor receives punishment (reward) in the sanctioning stage, the earnings of this actor decrease (increase) relative to the earnings of the other group members, altering the distribution of earnings within the group. For example, suppose group member A made the lowest contribution, and therefore earns more than the others. When group member B decides to punish A, the income of B decreases by the cost of punishment allocation, while the income of A decreases by three times this cost. The earnings of the other group members are unaffected by B's decision to punish A. Thus, the earnings of A decrease relative to those of B, and the earnings of both A and B decrease relative to the rest of the group. In general, the more high contributors are rewarded or the more free riders are punished, the more the negative relation between contribution and earnings is weakened or even reversed. In this way, sanctions alter the distribution of earnings among group members.

Traditionally, individuals are understood to judge distributive fairness based on three principles (Deutsch, 1975). The first is the principle of equality, which implies that everyone should earn the same payoff. The second principle, equity, states that actors who provide the highest input should receive the largest share of the output. In our PGGs, this would imply that actors earn more as they contribute more. Finally, according to the need principle, actors who have a higher need for a resource should receive more than actors with less significant need.

In our anonymous symmetric PGGs, there is no visible heterogeneity between actors in their need for income. Thus, actors cannot apply the need principle in their fairness evaluations. According to the equality principle, sanctions are perceived as fairer when actors who make below-average contributions earn *as much as* above-average contributors. According to the equity principle, sanctions are considered fairer as actors who make below-average contributions earn *less* than above-average contributors. In other words, both equality and equity imply that actors who contribute less than the group average should be punished or that (average and) above-average[1] contributors should be rewarded. However, equity implies that below-average contributors are punished or that (average and) above-average contributors are rewarded as much as possible, while equality implies that these individuals are sanctioned only to the extent that equal payoffs are approached. As subjects in the experiment are not informed of the earnings of their fellow group members, they cannot precisely infer how sanctions affect the distribution of earnings in the group.[2] This implies that subjects cannot distinguish

[1]It depends on the variation in contributions made within a group whether rewarding average contributions increases or decreases equality and equity. For example, if all group members contribute the average amount and only one is rewarded, equality and equity decrease. Conversely, if there are considerable differences between individual contributions made, rewarding an average contributor can enhance equality or equity levels. Theoretically, we find it more appealing to include average contributions among those that 'deserve' rewards in accordance with the fact that they are not included in the contributions that 'deserve' punishment. Empirically, our conclusions on perceived fairness of rewards remain unchanged when 'fair rewards' are classified as rewards for 'above-average' versus 'average and below-average' contributions, or when rewards of average contributions are excluded from the measures.

[2]Subjects observe contributions and the number of received sanctions of all group members. Thus, subjects can calculate fellow group members' earnings from the contribution stage, and by how much each group member's earnings were increased or decreased through received sanctions. However, subjects cannot determine others' overall earnings, as they do not know who paid the cost of allocating

between the extent to which sanctions promote equality or equity. Subjects do observe how much each group member contributes and how many sanctions each group member receives. Therefore, they may perceive the outcome of the sanctioning stage as fairer as more below-average contributors are punished or as more average and above-average contributors are rewarded. That is, actors perceive sanctions as fairer as sanctions promote equity, without knowing the precise extent to which equality is achieved.

Previous experimental findings support the conjecture that actors allocate peer sanctions to increase equity. Dawes et al. (2007) and Klempt (2012) consider settings in which actors experience payoff inequality that does not result from individual differences in cooperative behavior. As outcomes are not a product of subject behavior, all sanctioning motives besides those of restoring equity are ruled out in these studies. Still, it is found that many actors punish those who achieve better outcomes and reward those who achieve lesser outcomes, thus increasing equity. Moreover, Tabibnia and Lieberman (2007) find that the human brain activates a reward system when subjects observe others generating equitable outcomes.

In sum, we define distributive fairness in PGGs as the extent to which actors who contribute less than the group average are punished and the extent to which actors who contribute the same as or more than the group average are rewarded, i.e., the higher the extent to which sanctions promote equitable outcomes. This argument holds both for sanctions that actors receive themselves, and for sanctions received by other group members. Accordingly, we formulate the first hypothesis on reaching equitable outcomes:

Hypothesis 3.1 *Personally received sanctions and sanctions received by others are perceived as fairer the more actors who contribute less than the group average are punished and the more actors who contribute the same as or more than the group average are rewarded.*

First, this implies that actors perceive personally received sanctions as fairer the more the sanctions that they receive increase the equity of the payoff distribution, that is, they more they are punished for below-average or rewarded for average and higher contributions. Second, this implies that actors perceive sanctions received by fellow group members as fairer, the more the sanctions that fellow group members receive result in a more equitable payoff distribution.

3.2.4 Procedural fairness

The procedure through which decisions are made is often recognized as an important determinant of fairness perceptions (Frey, 2003; Jost and Kay, 2010; Leventhal, 1980;

sanctions. Moreover, subjects can examine the screen that presents this information for 30 seconds at most (actors almost always take a few seconds only) before moving on to the next screen. It is therefore highly unlikely that subjects precisely calculate how sanctions affect the distribution of earnings.

Thibaut and Walker, 1975) and is repeatedly found to have a stronger effect than decision outcomes (Schroeder et al., 2003). In our PGGs, procedures through which sanctions are implemented are represented by the decision rules described in Section 3.1.

In a previous study on perceptions of punishment decisions, Strimling and Eriksson (2014) find that individuals perceive punishment as more acceptable as decisions to punish someone are made by the group as a collective. However, respondents in their study are third parties who judge hypothetical interactions between other people, while in our experiment, actors either decide whether to allocate a sanction or are sanctioned themselves. The outcomes of sanctioning decisions affect individual payoffs. Therefore, fairness may be evaluated differently in our setting than in the study of Strimling and Eriksson (2014).

First, we discuss how decision rules as procedures through which actors receive sanctions from fellow group members affect the perceived fairness of personally received sanctions. Under an IDR, an actor may be sanctioned by every group member who decides to do so. This allows for perverse sanctioning decisions. That is, actors may be punished for cooperating or rewarded for free riding (Casari and Luini, 2009; Ellingsen et al., 2013; Herrmann et al., 2008). Although perverse sanctions may significantly affect cooperative behaviors, they are typically much less frequent than punishments directed at free riders or reward directed at high contributors. Therefore, as more actors must agree on sanctioning decisions, it becomes more likely that only severe free riders are punished and high contributors rewarded. In other words, the more collective agreement is required, the more actors can a priori anticipate behaviors that they may be sanctioned for and the more the decision procedure is free of biases resulting from erratic individual decisions. Dolan et al. (2007) finds neutrality, accuracy, and consistency to be important determinants of fair procedures. Hence, in line with the findings of Strimling and Eriksson (2014), we expect that actors find personally received sanctions fairer as more individuals must agree on sanctioning them.

Hypothesis 3.2 *Personally received sanctions are perceived as fairer as more actors are required to agree on sanctioning decisions.*

Second, we discuss how decision rules as procedures through which sanctions that actors allocate to fellow group members are implemented affect the perceived fairness of sanctions received by other group members. Whereas every actor can independently decide whom to sanction under an IDR, the higher the level of collective agreement required for sanctions to be carried out, the less individual decision-making power each actor has in deciding to sanction someone. Individual influence in the decision procedure forms an important aspect of procedural fairness (Dolan et al., 2007; Frey, 2003). For example, Bies (1993) finds that workers perceive pay cuts as fairer when they are involved in the decision-making process. The ability to voice an opinion, even after a decision has been made, promotes higher levels of experienced procedural fairness (Folger, 1977).

The lower the degree of collective agreement required for sanctions to be implemented, the more influence actors have over sanctioning decisions. Previous research suggests that individuals derive utility from the act of sanctioning (e.g., Dawes et al., 2007; De Quervain et al., 2004). When a CDR is applied, actors perceive themselves to exert more influence as more of their proposed sanctions are implemented by the group. We therefore hypothesize that the proportion of an actor's allocated sanctions implemented under a CDR positively affects the perceived fairness of others' received sanctions.

Hypothesis 3.3 *Sanctions received by other group members are perceived as fairer as more sanctions that the focal actor attempts to allocate under a CDR are implemented.*

3.2.5 Self-serving bias

A widely recognized aspect of perceived fairness is that perceptions are biased towards an actor's personal outcomes (e.g., Messick and Sentis, 1979). This is referred to as self-serving bias. Cognitive dissonance theory is often used to explain this phenomenon (Rode and Le Menestrel, 2011). That is, while individuals are motivated to achieve higher monetary outcomes, they also aspire for a fair division of resources (Eek et al., 1998; Wilke, 1991). Dissonance is reduced by interpreting fairness in a manner that favors personal outcomes. Accordingly, actors may evaluate sanctions in such a way that they interpret personally received punishment as unfair, and personally received reward as fair, regardless of the contribution for which those sanctions are received. This is in line with the results of previous studies that found actors to perceive the same distribution of resources as fairer when in their favor than when they are disadvantaged (e.g., Messick and Sentis, 1979. We formulate the following hypothesis:

Hypothesis 3.4 *Personally received punishments are perceived as less fair as actors receive more punishments, and personally received rewards are perceived as fairer as actors receive more rewards.*

Note that the self-serving bias partly leads to hypotheses that compete with Hypothesis 3.1. Namely, while Hypothesis 3.4 predicts a negative effect of personally received punishments on fairness perceptions and a positive effect of personally received rewards, Hypothesis 3.1 predicts the opposite effect so long as sanctions generate more equitable outcomes.

3.3 Experimental design

We use data from a laboratory experiment in which subjects participated in PGGs with peer-sanctioning institutions as described in Section 3.1. In the experiment, subjects accumulated points that were translated into monetary earnings at the end of the session. We here describe aspects of the experiment that are of relevance to the current chapter. These include a series of ten PGGs with opportunities to punish other group members, a series of ten PGGs with opportunities to reward other group members, and a questionnaire that measures the perceived fairness of sanctions that were implemented. Subjects were rematched to different groups after each PGG in a series was completed. A full description of the experiment is provided in Chapter 2.

Each PGG opened with a contribution stage. In this stage, subjects were divided into groups of four and endowed with 20 points each. The subjects independently and simultaneously decided how much of their endowment to contribute to a 'group account.' All remaining points were held in a 'private account.' All points contributed to the group account were multiplied by 1.6 and then divided equally among all four group members. A subject's earnings made in the contribution stage were composed of points held in his or her private account supplemented with his or her share of the group account.

In the ten PGGs with an opportunity to punish and in the ten PGGs with an opportunity to reward, a sanctioning stage followed after the contribution stage of each PGG. This implied that subjects were informed of the contributions of fellow group members and were asked for each group member whether or not they wanted to sanction this individual. If a subject decided to sanction and the sanction was implemented, in cases of punishment, the recipient lost six points, and in cases of reward, six points were added to the earnings of the recipient. Subjects paid a cost of two points for every sanction that they allocated and that was implemented. After the sanctioning stage, subjects were informed of all sanctions that were implemented in the group and of their own earnings. Subjects were rematched into different groups after every PGG was completed.

In each session, all PGGs with a sanctioning stage were carried out under one of three experimental conditions. Thus, over the course of one session, the same experimental condition was used for the ten PGGs with a punishment stage and for the ten PGGs with a reward stage. In the condition employing an IDR, all allocated sanctions were implemented. In the condition employing a CDR that required majority agreement, sanctions were only implemented when at least two subjects sanctioned the same recipient. In the condition employing a CDR that required unanimous agreement, three subjects needed to sanction the remaining member of the group for sanctions to be implemented. Sanctions not achieving the required level of agreement were not implemented, and subject(s) who attempted to allocate such sanctions did not pay the sanctioning cost of two points. Subjects were not informed of unimplemented sanctions that their fellow group members had attempted to allocate.

Perceived fairness levels were measured through a questionnaire administered after the PGGs to avoid likely confounding effects of eliciting fairness perceptions on subsequent behaviors (Eek and Biel, 2003). The questionnaire included several items for measuring perceived procedural fairness developed by Thibaut and Walker (1975) and items for measuring perceived distributive fairness developed by Leventhal (1976, 1980). Validation of the scales that the items belong to is presented by Colquitt (2001). The scales were originally developed for organizational research. Only items that can be applied to peer-sanctioning institutions in PGGs were included in our questionnaire. For example, an item on the opportunity to appeal to an outcome was excluded, as appeal was not possible by definition. We adapted the remaining items to the context of peer-sanctioning institutions. The six resulting items for punishments received by other group members are: 'have you been able to express your views and feelings in your decrease of other group members' payoffs?'; 'have you had influence over the decrease the other group members have received?' (cf. Thibaut and Walker, 1975); 'has the decrease of the payoff of other group members been applied consistently?'; 'has the decrease of the payoff of other group members upheld ethical and moral standards?' (cf. Leventhal, 1980); 'does the decrease that other group members have received reflect what they have contributed to the group?'; 'is the decrease that other group members have received justified, given their contributions?' (cf. Leventhal, 1976). All questions with the exception of the first were repeated using 'you' rather than 'other group members,' but with otherwise identical phrasing. Likewise, to measure the perceived fairness of received rewards, all questions were repeated with use of the word 'increase' in place of the word 'decrease.' All items were measured on a seven-point scale ranging from 'to a small extent' to 'to a large extent.'

The experiment was programmed using z-Tree (Fischbacher, 2007) and conducted at the ELSE laboratory of Utrecht University. Subjects were recruited through the ORSEE online recruiting system (Greiner, 2004). Separate instructions for the PGGs with an opportunity to punish and for those with an opportunity to reward were provided to subjects on paper before the start of the corresponding part of the experiment. A total of 184 subjects participated in the experiment (32% male; mean age of 23): with 64 participating under the majority CDR, 64 participating under the unanimity CDR, and 56 participating under the IDR. Payoffs averaged at €12.50, with a minimum value of €8.50 and a maximum value of €15.

3.4 Measures and methods

The measures of perceived fairness are constructed using corresponding items from the post-experimental questionnaire. For perceived fairness of punishment, a factor analysis with promax rotation on all items reveals two underlying factors: one for items that measure personally received punishments and one for items that measure punishments received by fellow group members. Accordingly, two corresponding scales are constructed (Cronbach's α own = 0.91; others = 0.90). The correlation between the two scales is strong and significant ($r = 0.61$; $p < 0.01$). Note that both scales include items that measure perceived distributive as well as procedural fairness. Combining distributive and procedural fairness in one scale is not uncommon (Colquitt, 2001).

The underlying factor structure of the perceived fairness of rewards is not as evident as it is for punishment. However, if the number of factors is pre-specified, following promax rotation, items for personally received rewards and for rewards received by others load on two different factors. Only the item 'have you had influence over the increase you have received?' loads on the factor for others rather than that for personally received rewards. Accordingly, this item was removed from the analysis, slightly improving the Cronbach's α of the resulting scale. Two scales were again constructed from the factor scores, with one measuring the perceived fairness of personally received rewards and with the other measuring the perceived fairness of rewards received by others (Cronbach's α own = 0.83; others = 0.80). The correlation between the two scales is strong and significant ($r = 0.72$; $p < 0.01$). The correlation between scales measuring personally received punishments and rewards and those measuring others' received punishments and rewards is significant, but low ($r = 0.29$; $p < 0.01$ for own, $r = 0.33$; $p < 0.01$ for others).

The decision procedures are represented by the three decision rules, i.e., the IDR, the CDR that requires majority consent for sanctions to be implemented, and the CDR that requires unanimous consent for sanctions to be implemented. Each subject experiences only one decision rule that is applied in all PGGs that the subject participates in. Thus, we have one observation for each subject on decision procedures and fairness perceptions. Other explanatory variables, such as sanctions received and the extent to which sanctions promote equity, may vary in each PGG that subjects engage in. Accordingly, remaining explanatory variables are aggregated over the ten PGGs with a punishment stage for an analysis of perceived fairness of punishment and over the ten PGGs with a reward stage for an analysis of perceived fairness of reward. Thus, the measures described below were calculated twice: once for the PGGs with a reward stage and once for the PGGs with a punishment stage.

The proportion of the sanctions a subject allocated that was implemented is measured as the total number of times over all ten PGGs that a sanction a which subject allocated was implemented divided by the total number of sanctions that the subject had attempted to allocate. In each PGG, a subject decides for each of the three fellow group members whether or not to attempt to sanction, resulting in a total of 30 punishment and 30 reward decisions. Subjects who interacted under the IDR, in which all sanctions are implemented by default, and those interacting under CDRs who never attempted to allocate a sanction obtain strictly speaking a missing value, as the proportion cannot be assessed in such cases. For all similar variable constructions, we treat such 'legitimate' missing values in the same way, namely by imputing a value of zero for these missing values and then adding a dummy variable to the regression model indicating whether a zero is imputed.

Sanctions promote equity more effectively as more below-average contributors are punished and as more average and above-average contributors are rewarded. Likewise, sanctions decrease equity when average and above-average contributors are punished or when below-average contributors are rewarded. Because a self-serving bias may cause subjects to judge personally received sanctions differently from sanctions that other group members receive, we consider personally received sanctions and sanctions received by others separately. For sanctions received by other group members, equitability of rewards is measured as the proportion of other group members who contributed at least as much as[3] the group average and who received at least one reward, and the equitability of punishments is measured as the proportion of other group members who contributed less than the group average and who received at least one punishment. Proportions are averaged over all PGGs that a subject participated in, and in which at least one of the group members contributed at least as much as (for rewards) or less than (for punishments) the group average. Similar variables are constructed by measuring sanctions that go against reaching more equitable outcomes, i.e., the proportion of other group members receiving rewards for below-average contributions and receiving punishments for contributing the group average at a minimum.

As in the case of sanctions received by others, personally received sanctions are distinguished among punishments and rewards received for contributing at least as much as the average contribution of the group and contributions falling below the group average. This is measured as the total number of PGGs (between one and 10) in which a subject received at least one sanction when contributing at least as much as the group average divided by the total number of PGGs in which the subject contributed at least as much as the group average. The same procedure was used to measure the proportion of PGGs in which subjects were punished or rewarded after making below-average contributions. Subjects who never contributed at least as much as the group average or below the average contribution are assigned a value of zero. Dummy variables that indicate when a value zero is imputed are included.

In some PGGs, no group members are sanctioned. In addition to the effect of an absence of sanctions on equity as measured in the previous variables, these interactions are more equitable and may be perceived as fairer the more group the members contribute (and earn) the same amount. As heterogeneity among contributions increases, high contributors earn progressively less than low contributors. Thus, PGGs in which no group members receive sanctions are less equitable as contributions differ to a higher degree.[4] Variation is measured as the standard deviation of contributions made by the

[3]Again, our conclusions are not affected by whether rewards received for average contributions are incorporated with above- or below-average contributions or excluded from the measure.

[4]Our measure assumes that in PGGs where at least one group member was sanctioned, the variation in contributions is irrelevant for fairness perceptions. Alternatively, we may argue that the effect of variation in contributions on fairness perceptions gradually weakens the more the sanctions that were implemented restored equity. This would imply an interaction between the variation in contributions and the proportion of others who were sanctioned for above- and below-average contributions. The

four group members. The measure is constructed as the average of the standard deviations of contributions made over all ten PGGs that a subject participated in, and in which no group members received sanctions. Subjects who had not encountered a PGG in which no sanctions were implemented are assigned a value of zero. Dummy variables indicating whether a zero is assigned are included to account for this imputation.

Finally, we control for the mean contribution of the subject and of the other group members, both averaged over the ten PGGs. Experimental subjects are employed as the unit of analysis. We apply multilevel regression with subjects nested in sessions to account for possible dependency at the session level. Two subjects who expressed difficulties understanding the questionnaire, which was offered exclusively in English, are excluded from the analysis.

3.5 Results

3.5.1 Descriptive results

Comparing in general how subjects perceived received sanctions, received punishments are perceived as slightly fairer than received rewards, though little difference is found between the perceived fairness personally received sanctions and sanctions received by others. Among the seven-point items used for the scale measuring the perceived fairness of personally received punishments, subjects score 3.67 (*sd.* = 1.51) on average. On the perceived fairness of punishments received by others, subjects score 3.44 (*sd.* = 1.54) on average. On the perceived fairness of personally received rewards, subjects score 2.41 (*sd.* = 1.30) on average. Finally, on the perceived fairness of rewards received by others, subjects score 2.59 (*sd.* = 1.51) on average.

Table 3.1 presents the descriptive characteristics of variables used in the regression models. Variables for which some subjects are assigned a value of zero are presented as both pre- and post-imputation values. Table 3.1 shows that, on average, slightly more equity-increasing sanctions are implemented in the PGGs with a punishment stage than in the PGGs with a reward stage, i.e., a higher proportion of subjects (themselves or fellow group members) are punished for below-average contributions than those who are rewarded for average and above-average contributions. Conversely, more equity-decreasing sanctions occur under reward, i.e., we observe a higher proportion of subjects who are rewarded for below-average contributions than those who are punished for average and above-average contributions. There is substantially more variation among contributions made in the PGGs with a reward stage than among those made in the PGGs with a punishment stage. Roughly half of all attempted punishments, and only 30% of attempted rewards, are implemented on average. Subjects contribute more

conclusions presented below do not change when such an interaction is employed rather than the current measure. Also, this interaction does not result in better model fit. Accordingly, we present the simpler measure described here.

Table 3.1: Descriptive characteristics of variables used in the regression analysis.

	Punishment				Reward			
Variable	M	S.d.	Min.	Max.	M	S.d.	Min.	Max.
Perceived fairness personally received sanctions	0.000	0.962	-2.169	1.600	0.000	0.913	-1.600	2.123
Perceived fairness sanctions received by others	0.000	0.957	-2.365	1.462	0.000	0.899	-1.654	2.068
Prop. sanctioned others, contr. ≥ av. group	0.050	0.092	0.000	0.526	0.242	0.263	0.000	1.000
Prop. sanctioned others, contr. < av. group	0.386	0.245	0.000	1.000	0.067	0.123	0.000	0.571
Prop. sanctioned self, contr. ≥ av. group[1]	0.061	0.159	0.000	1.000	0.191	0.267	0.000	1.000
Prop. sanctioned self, contr. ≥ av. group[2]	0.060	0.158	0.000	1.000	0.181	0.264	0.000	1.000
Own contribution never at least average[3]	0.028		0.000	1.000	0.051		0.000	1.000
Prop. sanctioned self, contr. < av. group[1]	0.385	0.327	0.000	1.000	0.084	0.182	0.000	1.000
Prop. sanctioned self, contr. < av. group[2]	0.361	0.330	0.000	1.000	0.078	0.178	0.000	1.000
Own contribution never below-average[3]	0.063		0.000	1.000	0.063		0.000	1.000
Var. contr. if no-one in group is sanctioned[1]	2.957	1.919	0.000	8.452	5.805	1.959	0.833	11.033
Var. contr. if no-one in group is sanctioned[2]	2.821	1.974	0.000	8.452	5.673	2.122	0.000	11.033
Never no-one sanctioned[3]	0.046		0.000	1.000	0.023		0.000	1.000
Prop. of own allocated sanctions executed[1]	0.561	0.317	0.000	1.000	0.310	0.334	0.000	1.000
Prop. of own allocated sanctions executed[2]	0.330	0.368	0.000	1.000	0.161	0.286	0.000	1.000
Never allocated sanctions/IDR[3]	0.411		0.000	1.000	0.480		0.000	1.000
Av. contr. others	10.863	3.090	3.733	17.567	6.461	2.166	2.333	11.600
Av. contr. subject	10.745	4.544	0.000	20.000	6.267	5.174	0.000	20.000

[1]Prior to the zero imputation of legitimate missing values.
[2]Following the zero imputation of legitimate missing values.
[3]Dummy variable where a value of one indicates that a zero was imputed for legitimate missing values in the variable directly above.

Table 3.2: Multilevel linear regression on the perceived fairness of personally received punishments (175 subjects nested in 12 sessions).

	Exp. sign	Hyp.	Model 1 Coeff.	Model 1 S.e.	Model 2 Coeff.	Model 2 S.e.
Majority	3.2	+	-0.014	0.263	-0.514	0.344
Unanimity	3.2	+	-0.186	0.262	-0.100	0.415
Prop. others punished when contributing at least group average					-0.231	0.952
Prop. others punished when contributing below group average					0.056	0.479
Variation contributions when no-one in group is punished	3.1	-			0.005	0.057
Prop. punished when contributing at least group average	3.1/3.4	−/−			-0.963*	0.458
Prop. punished when contributing below group average	3.1/3.4	+/−			-0.554*	0.251
Prop. allocated pun. executed					0.604	0.327
Av. contribution subject					0.037	0.024
Av. contribution others					0.083*	0.034
Constant			0.062	0.188	-1.084	0.575
σ_u			0.279**	0.099	0.00	0.00
Log Likelihood			-237.761		-215.756	

*Significant at the .05-level. **Significant at the .01-level (2-sided)
Controlled for dummy variables that indicate whether a value of zero was assigned to legitimate missing values.

on average under punishment than under reward, though substantial heterogeneity is found between subjects in both cases.

3.5.2 Perceived fairness of personally received sanctions

We first discuss our results concerning the hypotheses on personally received punishments and rewards. Table 3.2 shows the results of the regression analysis on the perceived fairness of punishments that subjects received themselves. In Model 1, only experimental conditions are included. There is no significant difference between the conditions, implying that Hypothesis 3.2 is not confirmed in the case of punishment. Other independent variables are included in Model 2. Receiving punishment for below-average contributions, which promotes equity, negatively affects fairness perceptions. Moreover, the effect of variations in contributions when no group member is punished

Table 3.3: Multilevel linear regression on the perceived fairness of personally received rewards (175 subjects nested in 12 sessions).

	Exp. sign	Hyp.	Model 3 Coeff.	Model 3 S.e.	Model 4 Coeff.	Model 4 S.e.
Majority	3.2	+	0.083	0.212	0.622*	0.294
Unanimity	3.2	+	-0.332	0.212	0.375	0.313
Prop. others rewarded when contributing at least group average					0.358	0.556
Prop. others rewarded when contributing below group average					-0.110	0.841
Variation contributions when no-one in group is rewarded	3.1	−			-0.043	0.053
Prop. rewarded when contributing at least group average	3.1/3.4	+/+			1.020*	0.461
Prop. rewarded when contributing below group average	3.1/3.4	−/+			0.664	0.461
Prop. allocated rew. executed					0.381	0.325
Av. contribution subject					-0.038	0.021
Av. contribution others					0.026	0.050
Constant			0.084	0.153	-0.587	0.346
σ_u			0.184*	0.098	0.014	0.606
Log Likelihood			-230.109		-213.146	

*Significant at the .05-level. **Significant at the .01-level (2-sided)
Controlled for dummy variables that indicate whether a value of zero was assigned to legitimate missing values.

is not significant. Together, these effects imply that Hypothesis 3.1 is not confirmed in the case of personally received punishments. Rather, subjects perceive personally received punishments as less fair the more they are punished, regardless whether they are punished for average, above-average, or below-average contributions, confirming that fairness perceptions are biased toward personal outcomes as stated in Hypothesis 3.4. As expected, punishments received by others do not affect fairness perceptions of personally received punishment. We still include these variables to generate comparable models that describe all fairness evaluations. This also illustrates that subjects really distinguish between the four fairness evaluations and that they do not merely report a generalized feeling related to the entire treatment. Finally, subjects find the punishments that they receive to be fairer as others contribute more on average.

Table 3.3 presents the perceived fairness of personally received rewards. Again, Model 3 includes decision rules only, which have no significant effect, indicating that

Hypothesis 3.2 is not confirmed in the case of rewards. Model 4 includes the other independent variables. Equity, as measured by the variation in contributions, does not significantly affect fairness perceptions, thus again providing no support for Hypothesis 3.1. We do find that the number of received rewards for average and above-average contributions positively affects fairness perceptions, which is both consistent with Hypothesis 3.1 on equity and Hypothesis 3.4 on self-serving bias. The effect of receiving rewards for below-average contributions is not significant, supporting neither Hypothesis 3.1 nor Hypothesis 3.4. However, this may be attributable to the fact that few subjects who contributed below the group average were rewarded. The fact that the effect size is similar to that of being rewarded for average and above-average contributions indeed suggests a power problem, while the direction of the effect is consistent with self-serving bias. As for punishment, rewards received by others do not affect fairness perceptions on personal rewards. In contrast with those of punishment, fairness perceptions of personal rewards are not affected by the average contribution of other group members.

3.5.3 Perceived fairness of sanctions received by other group members

Table 3.4 presents the perceived fairness of punishments received by other group members. Model 5 shows that the experimental conditions do not significantly affect fairness perceptions. Model 6 includes the other independent variables. Although the insignificant effect of variation in contributions does not support Hypothesis 3.1, subjects perceive punishments of others who contribute at least the group average as fair, while they perceive punishments of others who contribute less than the group average as unfair, supporting Hypothesis 3.1. Model 6 also shows that subjects perceive punishments received by others as fairer as larger proportions of their own proposed punishments are implemented, suggesting that subjects view having decision-making power in deciding whom to sanction as fair. This confirms Hypothesis 3.3 in the case of punishment. Moreover, we find that subjects perceive punishments of others as less fair the more they are personally punished for below-average contributions, suggesting that subjects still hold negative feelings toward the institution overall when they are punished themselves. Finally, punishments received by others are perceived as fairer as more others contribute on average. This indicates that punishments are perceived as fair if they are successful in promoting high contributions.

Table 3.5 presents an analysis of the perceived fairness of rewards received by other group members. Model 7, with only experimental conditions, shows that rewards received by others are perceived as less fair under a CDR that requires unanimity than in the other conditions. The effect of the unanimity decision rule disappears when the other independent variables are included in Model 8, suggesting that this effect is attributable to experiences subjects have with rewards. Moreover, Model 8 shows that

Table 3.4: Multilevel linear regression on the perceived fairness of punishments received by others (175 subjects nested in 12 sessions).

	Exp. sign	Hyp.	Model 5 Coeff.	Model 5 S.e.	Model 6 Coeff.	Model 6 S.e.
Majority			0.034	0.295	-0.866**	0.301
Unanimity			-0.431	0.294	-0.334	0.363
Prop. others punished when contributing at least group average	3.1	-			-1.712*	0.834
Prop. others punished when contributing below group average	3.1	+			0.833*	0.420
Variation contributions when no-one in group is punished	3.1	–			-0.054	0.050
Prop. punished when contributing at least group average					-0.510	0.402
Prop. punished when contributing below group average					-0.473*	0.220
Prop. allocated pun. executed	3.3	+			1.145**	0.286
Av. contribution subject					0.012	0.021
Av. contribution others					0.076*	0.030
Constant			0.112	0.210	-0.724	0.503
σ_u			0.345**	0.102	0.00	0.00
Log Likelihood			-230.550		-192.583	

*Significant at the .05-level. **Significant at the .01-level (2-sided)
Controlled for dummy variables that indicate whether a value of zero was assigned to legitimate missing values.

subjects perceive rewards received by others as fairer as more of their own proposed rewards are implemented, confirming Hypothesis 3.3 in the case of rewards as well. The insignificant effects of equity as measured by the proportion of others who were rewarded for below-average, or for average and above-average contributions and by the variation in contributions when no one was rewarded indicate that Hypothesis 3.1 is not confirmed for rewards received by others, in contrast with the findings for punishments received by others.

3.6 Conclusion and discussion

The objective of this chapter is to identify determinants of perceived fairness of peer punishments and rewards (i.e., positive and negative sanctions) that actors receive themselves and that are received by their peers. As likely determinants, we consider the

Table 3.5: Multilevel linear regression on the perceived fairness of rewards received by others (175 subjects nested in 12 sessions).

	Exp. sign	Hyp.	Model 7 Coeff.	Model 7 S.e.	Model 8 Coeff.	Model 8 S.e.
Majority			0.171	0.175	0.518	0.277
Unanimity			-0.468**	0.175	0.177	0.294
Prop. others rewarded when contributing at least group average	3.1	+			0.920	0.523
Prop. others rewarded when contributing below group average	3.1	-			0.238	0.791
Variation contributions when no-one in group is rewarded	3.1	-			-0.047	0.050
Prop. rewarded when contributing at least group average					0.614	0.433
Prop. rewarded when contributing below group average					0.541	0.433
Prop. allocated rew. executed	3.3	+			0.619*	0.306
Av. contribution subject					-0.009	0.020
Av. contribution others					-0.018	0.047
Constant			0.103	0.127	-0.320	0.325
σ_u			0.091*	0.139	0.00	0.00
Log Likelihood			-223.375		-202.405	

*Significant at the .05-level. **Significant at the .01-level (2-sided)
Controlled for dummy variables that indicate whether a value of zero was assigned to legitimate missing values.

extent to which sanctions increase the equity of the distribution of payoffs across group members, whether sanctions are allocated through a decision-making procedure that requires individual, majority, or unanimous agreement and self-serving biases towards personal outcomes.

The analysis provides mixed evidence for our expectation (Hypothesis 3.1) that subjects perceive their own and others' received punishments and rewards as fairer as sanctions promote more equitable payoffs. That is, we find no consistent support that sanctions subjects receive themselves are perceived as fairer as more of these sanctions increase equity. Additionally, rewards received by others are not perceived as fairer as equity is increased. However, we do find that subjects perceive punishments received by other group members as fairer, the more the punishments that others received promote equitable outcomes.

With regard to decision-making procedures, we do not find support for our expectation (Hypothesis 3.2) that subjects perceive personally received sanctions as fairer as more collective agreement is required to implement sanctions. However, we confirm Hypothesis 3.3 that subjects perceive others' received punishment and rewards as fairer, the more of the focal subject's own proposed sanctions are implemented under procedures that require collective agreement. This suggests that subjects perceive it as unfair if they do not exert influence over the decision to sanction someone. Finally, as expected based on the self-serving bias (Hypothesis 3.4), we find that subjects perceive personally received punishments as less fair the more they are punished and personally received rewards as fairer the more they are rewarded.

Overall, our results suggest that subjects perceive sanctioning decisions as fair when sanctions that are implemented support their own wishes: when subjects themselves are punished as little as possible and rewarded as much as possible, and when their own proposed sanctions are implemented. Apart from this, only punishments received by others are perceived as fairer, the more they result in equitable outcomes. In contrast to our findings, Strimling and Eriksson (2014) do find individuals to perceive allocated punishments as more acceptable when the level of collective agreement required in decision-making procedures is higher. In their study, respondents are asked to make judgments on hypothetical interactions between other people. Their results may differ from those of the current chapter, as in our study, actors evaluate decision-making procedures in situations that they participate in themselves and that affect their own earnings. This is likely to generate stronger self-serving bias effects.

Future research may further identify the role of individual versus collective sanctioning decisions in fairness perceptions. Such studies may attempt to model procedures through which collective sanctioning decisions can be made in greater detail. Our collective decision rules may have failed to capture important factors that may render collective sanctioning decisions fairer than individual decisions. For example, a setting may be considered in which actors have the opportunity to converse on appropriate sanctioning behaviors before interacting in the Public Goods Game (PGG). Actors may then be better equipped to anticipate behaviors that will be sanctioned, especially under a collective decision rule. Additionally, when communication is possible, actors may better understand why some of the sanctions that they attempt to allocate do not receive sufficient collective agreement for implementation. In this way, more elaborate possibilities for communication may enhance the perceived fairness of sanctions allocated through collective rather than individual decisions (Guala, 2012).

To the best of our knowledge, this study is the first to employ a questionnaire for measuring perceived fairness of punishments and rewards implemented through various decision rules in an incentivized behavioral experiment. This implies that our design and measures are not premised on previous studies. While our measure of fairness perceptions has been extensively validated in organizational research (Colquitt, 2001), to our knowledge, it has never been applied to peer-sanctioning institutions. Thus, we cannot infer the extent to which the items are suitable for this setting. For example, subjects may have had difficulty applying concepts such as 'views and feelings' and 'consistency' to the abstract PGGs that they experienced. Future research efforts may attempt to improve the validation of scales for measuring the perceived fairness of peer sanctions.

To prevent our fairness measure from influencing behaviors observed in the experiment, after all interactions were completed, we asked subjects to disclose their overall fairness perceptions of sanctions that were implemented throughout the entire series of cooperation problems. Therefore, it is not possible to assess how fairness perceptions develop over the course of the interactions. Moreover, subjects may not have remembered every detail of the relevant PGGs when answering the fairness questions, especially when other treatments were administered in between. Future research efforts may employ more complex designs to tackle these two issues without disturbing behavior in the PGGs. For example, in a post-experimental questionnaire, subjects could be presented with scenarios that occurred in each PGG they participated in and asked to judge the perceived fairness of each PGG separately.

Despite these limitations, findings that actors display a self-serving bias in their perceptions of punishment fairness and perceive it as fair to exert influence over decisions to punish or reward someone have some interesting implications. For example, questions arise regarding the extent to which these findings are limited to peer sanctioning scenarios. Under many other sanctioning institutions, for example if sanctions are allocated through formal rules enforced by external authorities, actors have no sanctioning power at all. Our findings suggest that actors may perceive such formal sanctioning institutions as fairer, the less they are punished and the more they are rewarded themselves, and the more the authority sanctions group members whom an actor would also have sanctioned him- or herself.

Chapter 4

Collective decision rules for implementing punishment in n-person Prisoner's Dilemmas with noise in the display of contributions*

Abstract

We examine the effect of various punishment institutions on cooperation and earnings in an experiment of six-person Prisoner's Dilemma (PD) games in which subjects observe others' contribution decisions with some noise. We compare standard peer punishment institution in which each subject individually determines whether to punish another, with institutions in which punishments are only implemented if subjects collectively agree to punish a particular group member. In conditions without noise, we find that contributions and earnings are higher when more subjects must reach collective agreement on a punishment. In conditions with noise, contributions are lower as more subjects are required to agree. Moreover, with noise, earnings are lower under all punishment institutions than in the control condition that does not include punishment opportunities.

4.1 Introduction

Many cases of human social interaction are characterized by a conflict between individual and collective interests. A prominent example are cooperation problems, in which

*This is a slightly different version of Van Miltenburg, Przepiorka, and Buskens (2014) that has been submitted to an international journal. We thank Werner Raub for providing comments and feedback.

actors decide on contributing their private resources to a collective endeavor. While full cooperation generates the best possible outcomes for the collective, individual actors have an incentive to free ride on the contributions of others. This constitutes a social dilemma (Dawes, 1980).

Cooperation problems have been researched extensively in laboratory experiments. Several studies consider settings in which, after observing the contributions of their peers, actors can individually decide to reduce peer payoffs at a cost to themselves. We will call this a peer punishment institution that employs an 'individual decision rule' (henceforth IDR). Typically, peer punishment institutions with an IDR generate high cooperation rates (e.g., Fehr and Gächter, 2000, 2002; Ostrom et al., 1992). However, most studies on peer punishment consider cooperation problems in which actors always accurately observe one another's contribution decisions. Conversely, in numerous real world settings, individuals are provided with imprecise (i.e., noisy) information about the cooperation or defection of others. To study the effect of such noise on cooperation under peer punishment institutions, we conduct a laboratory experiment using six-person Prisoner's Dilemma (PD) games, in which each actor observes another group member's cooperation as a defection or vice versa with positive probability. In the remainder of the chapter, this probability is referred to as noise.

Noise can lead to misguided punishment decisions and thus limit the efficiency of an IDR in sustaining cooperation (Fischer et al., 2013; Grechenig et al., 2010).[1] For example, let us assume that co-workers with different expertise work on a common project. While one of them, person A, may put little effort in the project, her co-workers, who know little about A's field, may believe that she made a sincere contribution. Conversely, suppose person B offers valuable input that appears insignificant, but which requires considerable work behind the scenes. Other group members who did not observe B's real effort may come to believe that B did not contribute his fair share. As the value of a common project depends on the value of actual contributions, all of the co-workers benefit from B's efforts but not from A's work. However, as some will misperceive A and/or B's contributions, B may be criticized for shirking, or A may not be sufficiently reprimanded for not contributing enough. Such feedback might discourage B from putting in the same amount of effort again, and A may be encouraged to continue shirking. In both cases, noise hinders cooperation.

An alternative peer punishment institution that might better support cooperation with noise is one that employs a collective decision rule (henceforth CDR), in which punishments are only implemented if a certain proportion of actors agrees to punish a particular group member. Put differently, under a CDR, if the required agreement is not reached, punishment is not carried out. Numerous social groups that face cooperation problems employ collective punishment decisions (Decker et al., 2003; Guala,

[1] We study one-shot interactions. This implies that theoretical complexities involved in strategic punishments with noise in repeated interactions (Green and Porter, 1984; Wu and Axelrod, 1995; Fudenberg et al., 2012) do not apply to our setting.

2012; Ostrom, 1990; Veszteg and Narhetali, 2010). Under noisy conditions, some peers correctly observe a cooperation, while others observe it as a defection and may apply punishment accordingly. When more actors are required to agree, it becomes less likely that such 'misguided' punishments aimed at cooperators are implemented. However, experimental research on the effect of CDRs on cooperation without noise shows that it is difficult to achieve sufficient collective agreement to punish defectors (Chapter 2). With noise, achieving agreement should be even more difficult, as one or several potential punishers may observe defectors as cooperators. When more actors are required to come to an agreement, defectors are less likely to be punished.

In sum, CDRs present both an advantage and a disadvantage with regards to implementing punishment in noisy environments. The magnitude of both effects depends on the required degree of actor consensus. We here aim to identify an optimal decision rule for encouraging cooperation under noisy conditions – a rule that enables actors to identify and punish defectors, while cooperators are unlikely to be punished erroneously. In our experiment, we compare an IDR with two CDRs for series of one-shot six-person PD games with noise. The first CDR places minor restrictions on collective agreement, and the second requires majority consensus.

In the remainder of this chapter, we first summarize previous findings on behavior from related experimental studies. We then describe our experimental games and derive our hypotheses. In our hypotheses, we focus on noise effects on cooperation and earnings for the three punishment decision rules. Sections 4.4 and 4.5 describe the experimental procedure and results, respectively. Section 4.6 discusses our findings and provides a conclusion.

4.2 Related literature

In the abundant experimental literature on cooperation and peer punishment, a number of findings have been frequently replicated. When punishment options are not available, cooperation rates of 50% are typically found in one-shot interactions. If one-shot interactions are finitely repeated with different partners, cooperation generally declines to lower levels over time (Camerer, 2003; Ledyard, 1995). Behaviors change considerably when a punishment institution with an IDR is employed. Numerous subjects punish defectors, and high levels of cooperation are usually maintained (e.g., Balliet et al., 2011; Chaudhuri, 2011; Fehr and Gächter, 2000, 2002). Many studies also report a small proportion of defectors who punish cooperators, which negatively affects cooperation (e.g., Cinyabuguma et al., 2006; Herrmann et al., 2008; Ostrom et al., 1992). Henceforth, we refer to punishment directed at defectors as pro-social punishment and to punishment directed at cooperators as perverse punishment. Finally, numerous iterations are typically required before the benefits of increased cooperation outweigh punishment costs such that the net effect of a punishment opportunity on earnings becomes positive (Ambrus and Greiner, 2012; Gächter et al., 2008).

As there are typically fewer perverse punishers relative to pro-social punishers in a population, a CDR should generate higher cooperation rates than an IDR if it rules out perverse punishment, while sufficient agreement to implement pro-social punishments can still be achieved. Two recent studies that address CDR punishment appear to be consistent with this intuition. Casari and Luini (2009) study groups of five subjects. Under their CDR, punishment is only implemented if at least two subjects punish the same group member. The authors find that cooperation and earnings are higher under the CDR than under the IDR, as perverse punishments hinder cooperation under the IDR, though such punishment are not typically implemented under the CDR. Chapter 2 examines groups of four actors. Two CDRs are employed: one in which two group members must punish the same recipient, and one in which three members must punish the same recipient. In Chapter 2, perverse punishment is not found to affect cooperation. Moreover, numerous pro-social punishments attempted under CDRs are not implemented. The IDR thus outperformed the CDRs. Both contributions and earnings are lower when a broader consensus is required to enact punishments. Required consensus between three actors was found to be especially problematic to maintaining cooperation.

Both previous studies on CDRs consider settings in which subjects, if they decide to contribute to the public good, also determine how much they will contribute (i.e., these studies employ a Public Goods Game). We are the first to consider CDRs when subjects determine whether or not to contribute their full endowment (i.e., we consider a PD). Relative to continuous contributions, our clear distinction between cooperation and defection may facilitate collective agreement on punishing defectors. However, when contributions are continuous, collective agreement to punish the lowest contributor is often achieved (Casari and Luini, 2009), which is not possible in our setting.

Grechenig et al. (2010) and Fischer et al. (2013) experimentally examine how noise in the display of contributions affects the extent to which cooperation can be maintained through punishment institutions with an IDR. Both studies again consider continuous contributions, while noise refers to an either 10% or 50% chance of a contribution being observed by another actor as a randomly determined higher or lower amount. This renders noise more ambiguous than it is in the current setting, in which a misrepresented contribution is always shown as defection and vice versa. Grechenig et al. (2010) and Fischer et al. (2013) find that subjects do not refrain from employing punishment in the presence of noise, such that two types of error occur in pro-social punishment. First, some punishments are misguidedly directed toward cooperators. Grechenig et al. (2010) find this to be detrimental to their future cooperation, though this is not replicated by Fischer et al. (2013). Second, defectors avoid punishment from others who observe them as cooperators and are thus less strongly discouraged from free riding. Moreover, resources are 'wasted' on misguided costly punishments of cooperators. As a result, both studies find that in the presence of noise, an IDR cannot promote cooperation and earnings as effectively. This latter finding is supported in studies that examine means of implementing noise or inaccurate contribution information that are less related to our setup (Ambrus and Greiner, 2012; Bornstein and Weisel, 2010; Patel et al., 2010).[2] Hence, studies must now identify punishment institutions that are more successful in promoting cooperation and earnings in noisy environments. The current chapter aims to contribute to this endeavor.

4.3 Experimental game and hypotheses

4.3.1 One-shot six-person PD with peer punishment

We theoretically and experimentally consider cooperation problems represented in series of one-shot six-person PDs. The PD model is employed due to the straightforward manner in which noise can be incorporated. In a six-person PD, all $n = 6$ actors receive an equal endowment w. Each actor i independently and simultaneously determines whether to contribute the entire endowment w to a group project, i.e., contribution c_i is either 0 (defection) or w (cooperation). All contributions $c = \sum c_i$ are multiplied by m, with $1 < m < n$, and divided equally among all members. As $m < n$, cooperation generates a lower payoff than defection ($wm/n < w$). However, group payoffs $nw - c + mc$ are maximized when each actor cooperates. Moreover, under full cooperation, individuals earn higher payoffs than they do under full defection ($wm > w$). Individually rational and selfish behavior thus leads to Pareto-suboptimal outcomes, rendering the one-shot PD a classic example of a social dilemma (Dawes, 1980). In the experiment, we use common values for endowments and individual returns from contributing by setting $w = 20$ and $m = 2.4$ (cf. Fehr and Gächter, 2000, 2002).

The current chapter focuses on PDs that include peer punishment opportunities. Following the contribution stage described above, each actor i observes the contribution decisions of all other group members $j \neq i$. All actors then individually and simultaneously determine whether to punish each j. If i decides to punish j, and if the punishment is implemented, actor i pays a fixed cost of $a > 0$, while j loses an amount of $b > a$. If i decides not to punish j, actor i pays no cost, and the earnings of j are unaffected. The total number of group members that i allocates punishment to is denoted by f_i; and the total number of group members who punish i is denoted by g_i. In the experiment, we employ $a = 2$ and $b = 6$, which corresponds to the frequently used 1:3 cost-to-impact ratio of punishment.

[2]Likewise, lesser offers are proposed and accepted in ultimatum games when responders receive limited information on the proposed division (Gehrig et al., 2007; Croson, 1996; Rapoport et al., 1996; Rapoport and Sundali, 1996). Not informing subjects about others' punishment decisions in cooperation problems is found to positively affect cooperation (Fudenberg and Pathak, 2010).

Under an IDR, which reflects how peer punishment institutions are typically employed in cooperation experiments, all punishments are implemented. Thus, each actor's earnings are decreased by allocated punishments af_i and received punishments bg_i. Under a CDR, punishment is only implemented if at least a certain proportion x/n of group members $j \neq i$ punishes the same recipient i. If $g_i/n < x/n$, the punishments directed at i are not carried out, i.e., actors j do not incur cost a for punishing i, and i's earnings are not reduced by bg_i. Thus, an actor i only loses an amount of bg_i due to punishments received if $g_i/n \geq x/n$, and only incurs punishment costs of a for each j whom i has attempted to punish *and* for whom $g_j/n \geq x/n$. Actors are not informed of non-implemented punishments that others attempt to allocate.

In our experiment, we employ two different CDRs: one under which punishment is implemented if at least two actors punish the same recipient (CDR2), and one under which at least three punishers are required (CDR3). In our groups of six, five fellow group members may punish each actor. Thus, CDR3 requires a majority of the other group members for punishment to be carried out ($x/n = 0.6$), while CDR2 requires the lowest possible degree of collective agreement ($x/n = 0.4$).

Under an IDR, rational and self-regarding actors who assume that others are also rational and self-regarding will not allocate or expect to receive punishment in (a series of) one-shot interactions, as opportunities for reputation building are ruled out. Under CDRs, rational, self-regarding actors likewise do not punish others if the punishment is implemented. If a punishment is not implemented, actors are indifferent toward punishing or not. The unique Nash equilibrium of zero contributions of the baseline PD remains unchanged, but punishing below the required level of consensus is allowed in equilibrium.

4.3.2 Noise in the display of contributions

In most public goods or PD experiments, subjects receive accurate information on the contribution decisions of all other group members. Here, we compare this standard setup with one in which actors know that there is a 20% probability that a cooperation may be displayed as a defection, or defection as cooperation. Whether a contribution decision is displayed incorrectly is independently determined for each contribution or defection that each actor observes. The 20% noise implies that on average, each actor's decision will be incorrectly perceived by one of the five other group members. Payoffs are based on the real contributions of all group members. Thus, noise does not affect the payoff structure of the PD such that the Nash equilibrium remains unchanged.

Assuming that initial cooperation rates remain close to 50% as is in most experiments the case, initial average observed cooperation is not heavily affected by noise, as both contributions and defections can be perceived incorrectly. Thus, a typical decline in cooperation in the PD without punishment should also occur in the presence of noise. We refer to the baseline PD as a PD without punishment institution, and assume that it does not matter whether noise is present or not. Conversely, in PDs with a punishment institution, the two errors in observations of others' contribution decisions give rise to two ways in which punishment decision outcomes may deviate from decision maker intentions (henceforth referred to as punishment errors). First, pro-social punishers might fail to punish actual defectors if they observe these defectors as cooperators. Second, pro-social punishers may punish actual cooperators if they observe these cooperators as defectors. This latter punishment error is referred to as 'punishment directed at cooperators observed as defectors', to distinguish it from perverse punishment, which is intended to hit an actual cooperator. Of course, the two errors also occur in cases of perverse punishment. However, the effect of errors in pro-social punishment on cooperation should be more significant, as pro-social punishment is more common than perverse punishment.

Through the two errors, noise changes the amount of punishment that cooperators and defectors can expect to receive when a given number of group members attempt to punish observed cooperation or defection. For example, an actual cooperator might face three group members who punish observed defectors and two others who never punish. Without noise, the cooperator will not be punished. Conversely, with noise, each of the three potential punishers might observe the cooperator as a defector and attempt to punish. How noise affects the likelihood that cooperators and defectors are punished depends on the decision rule in effect. In the example case, under CDR3 the cooperator is only punished when all three potential punishers observe the cooperator as a defector and attempt to punish. This is relatively unlikely. If only one or two punishers observe the wrong contribution decision under CDR3, the cooperator is not punished, as the threshold for implementation is not reached.

Table 4.1 depicts punishments that defectors and cooperators receive in our experiment for each possible number of other group members who punish observed defectors. In our groups of six, each actor faces between zero and five potential pro-social punishers. For the noise conditions, the table shows expected values based on 20% of pro-social punishers on average inaccurately observing the focal actor's contribution decision. In what follows, we strictly focus on pro-social punishment only and neglect perverse punishment for the time being.

Values presented in Table 4.1 were determined as follows. Without noise, punishment for defection is simply the number of punishers multiplied by the points that recipients lose for each punishment (i.e., bg_i). In the experiment, subjects lose six points for each punishment received. For example, defectors who are punished by three group members lose $3 \times 6 = 18$ points. This is shown in the three-punisher section of the no-noise division of Table 4.1. Under the CDRs, actors receive no punishment if the number of punishers falls below the threshold for implementation. Accordingly, the no-noise section of Table 4.1 shows that defectors receive no punishment if the group contains one punisher under both CDRs, and if the group contains two punishers under CDR3. Cooperators do not receive pro-social punishment without noise.

Table 4.1: Punishment points for cooperators and defectors, and the difference in punishment points between cooperators and defectors with noise, for each experimental condition. Each row corresponds to a different number of pro-social punishers in a group. Perverse punishment is not considered. For noise conditions, the table shows expected values based on an average of 20% inaccurate observations. Values are based on parameters used in the experiment.

Number of	No Noise						Noise								
pro-social	IDR		CDR2		CDR3		IDR			CDR2			CDR3		
punsihers	D	C	D	C	D	C	D	C	Diff.	D	C	Diff.	D	C	Diff.
1	6	0	0	0	0	0	4.80	1.20	3.60	0	0	0	0	0	0
2	12	0	12	0	0	0	9.60	2.40	7.20	7.68	0.48	7.20	0	0	0
3	18	0	18	0	18	0	14.40[1]	3.60	10.80	13.82[2]	1.30	12.53	9.22	0.14	9.07
4	24	0	24	0	24	0	19.20	4.80	14.40	19.05	2.34	16.70	17.20	0.50	16.70
5	30	0	30	0	30	0	24.00	6.00	18.00	23.98	3.52	20.46	23.37	1.06	22.31

[1] Example calculation: the probability that zero, one, two, or three pro-social punishers observe defection multiplied by associated punishment points gives $0.008 \times 0 + 0.096 \times 6 + 0.384 \times 12 + 0.512 \times 18 = 14.40$.

[2] Example calculation: as above, except that if only one pro-social punisher correctly observes a defection, the threshold for implementing punishment under CDR2 is not reached, and no punishment is enacted, giving $(0.008 + 0.096) \times 0 + 0.384 \times 12 + 0.512 \times 18 = 13.82$.

With noise, punishments that actual defectors receive depend on the number of punishers who correctly observe their decisions. Thus, if a group contains n punishers, the probability that zero, one, ..., n of these punishers observe a defection correctly is used to weigh the corresponding punishment level. For an actual defector under the IDR with three punishers, there is a 0.8% chance that no punisher correctly observes the decision, causing the defector to receive no punishment at all. Likewise, there is a 9.6% chance that only one punisher observes the defector correctly, causing the defector to receive six punishment points, etc. The probability that zero, one, two, or three punishers correctly observe a defection multiplied by the associated punishment points (b) gives a total expected punishment level of 14.4 points. This can be verified in the IDR noise section of Table 4.1 under defection with three punishing group members. The same calculations hold for actual cooperators, but punishments are in this case weighed by the probability that any number of punishers will observe a decision incorrectly.

Under a CDR with noise, even if enough punishers are available, the punishment of an actual defector is only implemented if a sufficient number of them correctly observe the defection. Thus, the same probabilities for any number of punishers correctly observing a defection apply as under the IDR above, though zero punishment points are associated with probabilities under which too few punishers observe the correct decision. In the example case involving three punishers, under CDR2 defectors receive zero punishment if only one punisher makes a correct observation (9.6% chance), as the punishment of this one group member will not be implemented. As shown in the CDR2 noise section of Table 4.1 under defection and three punishers, this generates 13.8 expected punishment points. Likewise, to calculate the expected punishment of actual cooperators as shown in Table 4.1, the probability that too few punishers incorrectly observe their decision is associated with zero punishment.

4.3.3 Hypotheses

We use Table 4.1 to derive hypotheses regarding differences in cooperation rates and earnings across the experimental conditions. We predict that cooperation rates will be higher as defectors receive more punishment relative to cooperators. Thus, the more punishment cooperators receive in a certain experimental condition, the less cooperation we anticipate, while more cooperation is expected as defectors receive more punishment.[3] Note that implicit in our hypotheses are assumptions that punishment directed at cooperators is roughly as detrimental to cooperation as punishment directed at defectors is beneficial and that the effect of punishment on cooperation does not

[3]Another view on punishment that can be used to interpret Table 4.1, is whether or not defectors are punished enough to offset their payoff advantage. However, many studies show that non-deterrent punishment also affects cooperation, as many actors are not self-regarding (e.g., Engel, 2014; Masclet et al., 2003). Therefore, we chose to assume a linear effect of the amount of punishment received.

depend on noise. These assumptions are empirically tested and evaluated in light of observed differences between the experimental conditions.

Without noise, we have no reason to assume that decision rules affect actors' pro-social punishment decisions. As non-implemented punishments are costless and not communicated to others, actors need not shy away from proposing punishments if they anticipate others may not propose to punish the same actor. Though punishments may become more severe for the recipient as more actors are required to agree, we do not expect that this will render actors reluctant to attempt to punish, as high severity is compensated for by a lower probability of implementation. Moreover, recipients may be punished with equal severity under an IDR. Thus, without noise, we do not expect that decision rules influence pro-social punishment decisions.

Previous studies (Ambrus and Greiner, 2012; Grechenig et al., 2010) have shown that while with noise actors are more likely to punish observed defectors, they employ less severe punishments on average. In our experiment, this is not possible, as actors only determine whether to punish and thus cannot choose the severity of their punishment. If anything, we might expect that noise will render actors reluctant to punish observed defectors, as they may punish an actual cooperator (Bornstein and Weisel, 2010; Patel et al., 2010). The more agreement is required, the lower the probability of punishment errors occurring, and the more closely punishment decisions should correspond to the situation without noise. Thus, if pro-social punishment decisions employed differ across our experimental conditions, we expect that actors will be less likely to punish with noise and more likely to punish under noise conditions as more agreement is required.

Table 4.1 shows that without noise, fewer pro-social punishments are implemented when decision rules are stricter. Thus, cooperation rates should be higher when fewer actors are required to agree on punishment decisions. However, this assumes an absence of perverse punishment. Under the IDR, perverse punishment may negatively affect cooperation, while under the CDRs perverse punishment will be largely ruled out (Casari and Luini, 2009; Chapter 2). As predictions are highly sensitive to deviations in perverse punishment, we refrain from formulating hypotheses on differences in cooperation rates between decision rules without noise.

For any number of punishers, Table 4.1 shows that under each decision rule (expected) punishment for cooperation is higher when noise is present than without noise, while (expected) punishment for defection is lower. Both effects should negatively affect cooperation. Moreover, if noise renders actors more reluctant to punish, with noise defectors will receive even less expected punishment than cooperators, and negative noise effects may be even more pronounced. With regards to earnings, if cooperation rates increase as a result of pro-social punishment, without noise, this implies that fewer punishment costs must be paid. Conversely, with noise even when full cooperation is achieved, some actors are observed as defectors and may be punished (Ambrus and Greiner, 2012). Thus, for high cooperation rates, higher punishment costs must be paid when noise is present, and these punishment costs are offset by a smaller increase in cooperation. Accordingly, we expect noise to negatively affect both cooperation and earnings under each decision rule:

Hypothesis 4.1a *In PD's with a punishment institution, less cooperation is achieved with noise than without noise regardless of the punishment decision rule employed.*

Hypothesis 4.1b *In PD's with a punishment institution, lower earnings are achieved with noise than without noise regardless of the punishment decision rule employed.*

The shaded cells shown in Table 4.1 denote decision rule(s) generating the highest difference between expected punishment for cooperation and expected punishment for defection. For example, when five potential punishers are present, the difference between expected punishment for cooperation and for defection is highest under CDR3 (22.3 points), followed by CDR2 (20.5 points) and the IDR (18 points). If, in line with our assumption, punishment of cooperators hinders cooperation to the same degree as punishment of defectors promotes cooperation, the decision rule resulting in the highest punishment of defectors relative to cooperators will generate the highest cooperation rates. Thus, as long as actors are punished by an intermediate number of punishers rather than by one or all others, CDR2 holds an expected advantage over the other rules. If a higher degree of required consensus renders actors more likely to punish such that more pro-social punishments are allocated under CDR2 than under the IDR, the difference in cooperation between CDR2 and the IDR may be even stronger. If actors are even more likely to punish under CDR3 than under CDR2, the difference between CDR2 and CDR3 may be less pronounced. We formulate the following hypothesis:

Hypothesis 4.2a *With noise, cooperation rates are higher under CDR2 than under an IDR and CDR3.*

Under CDR2, for each given number of punishers in a group, lower punishment of both cooperators and defectors is expected than under the IDR. As we hypothesized that higher cooperation rates are achieved under CDR2 than under the IDR, and that fewer punishments are allocated to achieve this, we can also hypothesize that earnings will be higher under CDR2 than under the IDR.

Hypothesis 4.2b *With noise, earnings are higher under CDR2 than under the IDR.*

We refrain from comparing earnings achieved under CDR2 and CDR3, as lower cooperation rates are expected under CDR3, though fewer costly punishments are enacted to achieve this. Thus, the net effect on earnings remains ambiguous.

4.4 Experimental procedure

In our experiment, subjects participated in series of PD games in anonymous groups of six with an endowment $w = 20$ points and multiplier $m = 2.4$ (Section 4.3.1). After the experiment, subjects received €1 for every 160 points earned. Subjects were randomly matched with different partners after each PD.

Experimental sessions were conducted either with or without noise. In sessions without noise, subjects were perfectly informed of the contribution decisions of others and of their own earnings in that period after each game. In sessions with noise, subjects were only informed of noisy contribution decisions and of what their earnings for the period would be if all contribution decisions they observed were correct. The participants were made aware that each contribution decision they observed came with a 20% independent probability of being incorrect. An incorrect observation implied that an actual cooperation was displayed as a defection and vice versa.

In each session, subjects first participated in 15 PD games without the option to punish other group members. After this first series was completed, a punishment stage was added to the PD for two ensuing sequences of 15 interactions. In each period, all subjects received an additional endowment of 10 points at the start of the punishment stage and decided for each of the other group members whether or not to punish them. If a subject decided to punish and the punishment was implemented, six points were deducted from the earnings of the recipient, and two points were deducted from the earnings of the punishing subject. The additional endowment of 10 points thus enabled each subject to punish all five other group members.[4] When multiple subjects targeted the same group member for punishment, and the punishment was implemented, all punishers paid the punishment cost, and the recipient lost the cumulative amount. For example, a participant who was punished by four others lost 24 points. Note that a subject's total accumulated earnings could become negative during this part of the experiment. However, negative total earnings were highly unlikely given the 15 initial baseline periods and the additional endowments. Accordingly, we did not form a procedure for total negative earnings, and negative earnings did not occur in any session.

As noted in Section 4.3.1, three punishment decision rules were employed as experimental conditions. The two series of 15 periods were each conducted under a different decision rule. Under the IDR, all punishments that subjects proposed were implemented. Under CDR2, punishments were only implemented if at least two group members punished the same recipient. Under CDR3, at least three punishers were required for a punishment to be implemented. Further information on punishment implementation through different decision rules is outlined in Section 4.3.1.

After the punishment stage, subjects were shown a screen with others' (noisy) contribution decisions and with punishments that each group member had received. The participants were not informed about who had allocated the punishments. In the ses-

[4]Note that in most previous experiments, subjects pay the cost of punishing from their previous earnings. In the current design, we used a punishment endowment, as in noise conditions, subjects were not informed of their true current earnings. Grechenig et al. (2010) found that assigning a punishment endowment does not significantly affect punishment decisions.

sions with noise, the same noisy contribution decisions were presented as those that were shown after the contribution stage. Again, in the sessions without noise, subjects were informed of their actual earnings after each period, while in the sessions with noise, subjects were informed of the payoff they would have received if their observed contribution decisions were actual decisions. At the end of the experiment, all subjects were informed of their actual aggregate earnings.

The experiment was programmed in z-Tree (Fischbacher, 2007) and conducted in the spring of 2013 at the ELSE laboratory at Utrecht University. Subjects were recruited through ORSEE (Greiner, 2004). Twelve sessions were conducted in total: six with noise and six without noise. Both in sessions with and without noise, different combinations of decision rules were administered for the first and second punishment series of each session (see Table 4.2 for an overview). A total of 252 subjects participated in the experiment (38% male, 86% students, 32% economics students, average age of 22.57). Sessions lasted one hour on average. Payoffs averaged €11, with a minimum of €7 and a maximum of €14.

Table 4.2: Overview of the number of subjects in each experimental session.*

Decision rules (first - second)	# subjects no noise	# subjects noise
IDR - CDR2	18	18
IDR - CDR3	24	24
CDR2 - CDR3	18	18
CDR2 - IDR	24	24
CDR3 - IDR	18	24
CDR3 - CDR2	18	24

*Each session started with 15 periods without punishment

4.5 Methods and results

4.5.1 Order effects

In each experimental session, subjects participated in two sequences of PD games with a punishment stage, each with a different decision rule. However, cooperative behaviors that occurred in the first sequence remained largely consistent through the second sequence despite change in decision rule (see Figure D.1 in Appendix D). A Fisher's exact test confirms that sessions in which cooperation rates are above average in the first sequence tend to generate above-average cooperation rates in the second sequence ($p = 0.08$). In the remainder of this section, we thus only report results of the first sequence of PD games with a punishment stage. The results of the second punishment sequence are available from the authors on request.

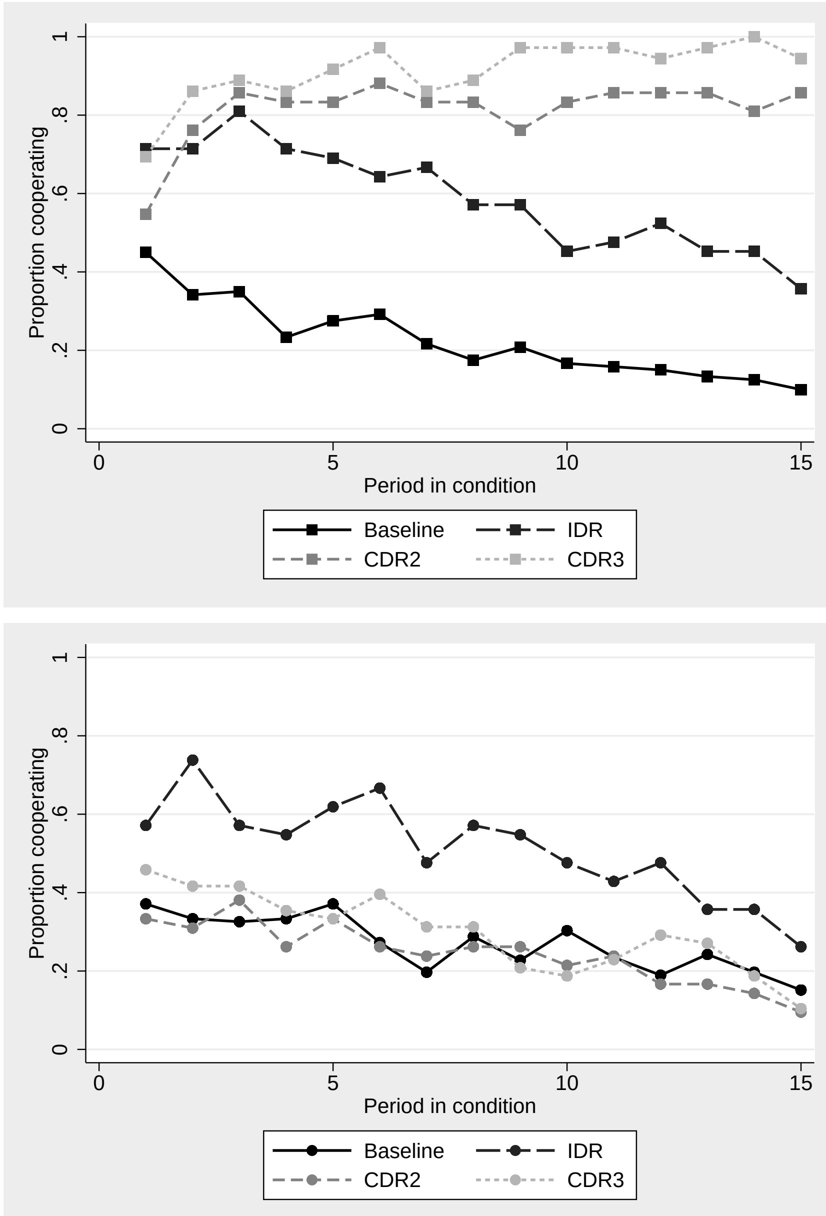


Figure 4.1: Cooperation in each period of the baseline and first punishment sequence by experimental condition without (top) and with (bottom) noise.

4.5.2 Cooperation and earnings

Figure 4.1 presents the proportion of subjects who cooperate in the PD over time and across experimental conditions. In the top panel, it is evident that in the absence of noise, all decision rules lead to an increase in cooperation rates relative to the baseline condition without a punishment stage. Additionally, cooperation rates increase as more group members are required to agree on punishment decisions. Only under the CDRs cooperation is maintained at a high level. The bottom panel of Figure 4.1 presents a very different picture. With noise, only the IDR has a cooperation-enhancing effect relative to the baseline condition, which weakens over time. Cooperation rates for both CDRs are very similar to the baseline condition.

Differences between the conditions shown in Figure 4.1 are confirmed through a regression analysis. Throughout the chapter, we present results based on multilevel regression models with random effects at subject and session levels to control for interdependencies we expect at both levels. We verify the robustness of model estimates with observations clustered at the session level (Tables D4.3-D4.4 in Appendix D). Because only 12 observations were gathered at the session level, clustering at the session level generates very conservative estimates. Nevertheless, these models support our main conclusions. We report below if a hypothesized effect is not robust when a model is estimated with standard errors adjusted strictly for session clustering.

The top panel of Figure 4.2 shows the predicted probability that a subject cooperates in each experimental condition. These predictions are based on the multilevel regression model presented in Table D.2 of Appendix D. Table D.1 presents corresponding descriptive statistics. The first model shown in Table 4.3 presents differences between experimental conditions in the predicted probabilities. For example, the model predicts a 0.014-point higher probability to cooperate with noise than without noise in the baseline games. This difference is not significant. Table 4.3 also shows no significant difference in the predicted probability to cooperate between the noise and the no noise condition under the IDR. Conversely, predicted probabilities to cooperate are lower with noise than without noise under CDR2 and CDR3, supporting Hypothesis 1a for both CDRs but not for the IDR. Without noise, the predicted probability to cooperate is higher under all punishment decision rules than in the no-noise baseline, higher under CDR3 than CDR2 and higher under CDR2 than the IDR. With noise, only the IDR generated a higher predicted cooperation probability than the noise baseline. With noise, the predicted probability to cooperate is higher under the IDR than CDR3, and higher under CDR3 than CDR2. Thus, Hypothesis 2a is not supported. The difference between the two CDRs disappears in the model with clustering at the session level.

The middle panel of Figure 4.2 shows predicted period earnings found for each experimental condition. Descriptive statistics are provided in Table D.1, and predictions are derived from the multilevel regression model presented in Table D.2. The second model shown in Table 4.3 presents differences in the predictions between the experimental conditions. The trends reflect those of predicted cooperation. As shown in Table 4.3, predicted earnings are lower with noise than without noise under all decision rules, supporting Hypothesis 4.1b. Without noise, predicted earnings under all decision rules are above baseline levels, higher under CDR3 than CDR2, and higher under CDR2 than the IDR. With noise, predicted earnings fall below the noise baseline levels under all decision rules. No significant differences with the IDR were detected for both CDR2 and CDR3, however predicted earnings are higher under CDR3 than under CDR2.

Thus, Hypothesis 4.2b is not supported. The difference in predicted earnings between the CDRs disappears if only session clustering is accounted for in the estimation.

To explain effects of the experimental conditions with respect to cooperation and earnings that partly contradict our hypotheses, we now present results on punishment behavior and a detailed analysis of how punishment affects subsequent contribution decisions.

Table 4.3: Differences between experimental conditions of predicted probability to cooperate, predicted period earnings, and predicted probability to punish an observed defector. Based on the fixed parts of multilevel logistic (cooperation and punishment) and linear (earnings) regression models with decisions nested in subjects and sessions. The actual models are displayed in Table D.2 (7,955 punishment decisions, 7,560 PDs, 252 subjects).

	Cooperation		Earnings		Pun. obs. defectors	
	Diff.	S.e.	Diff.	S.e.	Diff.	S.e.
Noise vs. no noise						
Baseline	0.014	0.033	1.146	0.635		
IDR	-0.104	0.066	-3.493**	0.844	0.058	0.137
CDR2	-0.729**	0.035	-14.059**	0.844	-0.737**	0.090
CDR3	-0.716**	0.035	-17.039**	0.848	-0.686**	0.104
Without noise						
IDR - Baseline	0.418**	0.035	2.928**	0.492		
CDR2 - Baseline	0.709**	0.021	12.563**	0.492		
CDR3 - Baseline	0.767**	0.020	16.906**	0.529		
CDR2 - IDR	0.291**	0.042	9.635**	0.682	0.507**	0.128
CDR3 - IDR	0.349**	0.044	13.978**	0.708	0.546**	0.130
CDR3 - CDR2	0.058*	0.023	4.343**	0.708	0.039	0.126
With noise						
IDR - Baseline	0.300**	0.035	-1.711**	0.492		
CDR2 - Baseline	-0.034	0.020	-2.643**	0.492		
CDR3 - Baseline	0.037	0.023	-1.279**	0.462		
CDR2 - IDR	-0.334**	0.042	-0.931	0.682	-0.289**	0.103
CDR3 - IDR	-0.263**	0.042	0.432	0.662	-0.199	0.113
CDR3 - CDR2	0.071*	0.030	1.363*	0.662	0.090	0.050

*Significant at the .05-level. **Significant at the .01-level (2-sided)

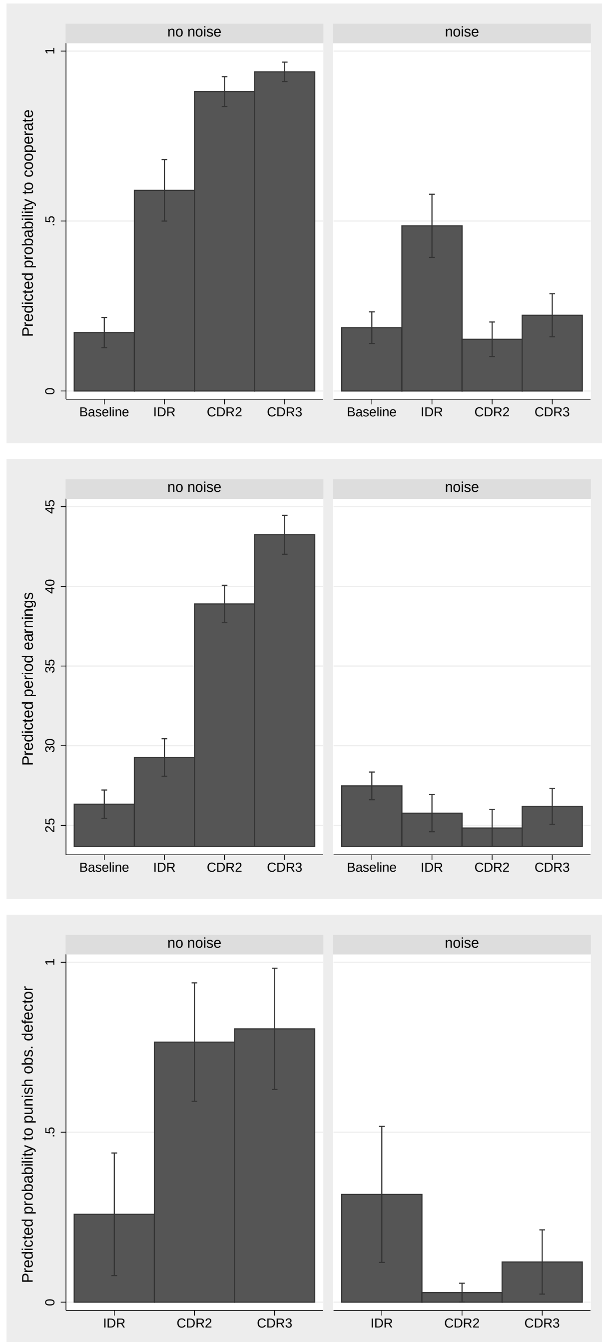


Figure 4.2: Predicted probability to cooperate, predicted earnings, and predicted probability to punish an observed defector for each experimental condition, with 95% confidence intervals. Based on the fixed parts of multilevel regression models (Table D.2) with decisions nested in subjects and sessions. Only the first punishment sequence is considered.

Table 4.4: Number (N) and percentage (%) of subjects who are punished, average number of punishments received among punished subjects (Av. pun), and average number of punishments received by all subjects who made respective contribution decisions (Av. total) by at least one other (punished) and by at least enough others for implementation (implemented), categorized by whether the subject had cooperated or defected and by the experimental condition in the first punishment sequence.

		Actual defectors								Actual cooperators							
		Punished				Implemented				Punished				Implemented			
Noise	Rule	N	%	Av. pun.	Av. total	N	%	Av. pun.	Av. total	N	%	Av. pun.	Av. total	N	%	Av. pun.	Av. total
Yes	IDR	244	79	1.95	1.55	244	79	1.95	1.55	151	47	1.26	0.59	151	47	1.26	0.59
	CDR2	234	49	1.25	0.61	51	11	2.14	0.23	66	43	1.27	0.54	16	10	2.13	0.22
	CDR3	364	72	1.63	1.18	57	11	3.14	0.35	99	46	1.35	0.62	7	3	3.14	0.10
No	IDR	222	85	2.20	1.88	222	85	2.20	1.88	58	16	1.02	0.16	58	16	1.02	0.16
	CDR2	113	97	3.21	3.10	102	87	3.45	3.01	60	12	1.07	0.12	4	1	2	0.02
	CDR3	41	89	3.44	3.07	37	80	3.62	2.91	173	35	1.12	0.39	0	0	0	0

4.5.3 Punishment

Table 4.4 presents allocated and implemented punishment for all decision rules and noise conditions. Overall, defectors are punished more often than cooperators, and when defectors are punished, that is done by a larger number of punishers. Interestingly, when cooperators are punished, the average number of punishers is close to one in all conditions. This suggests that CDR2, for which more than one punisher is required, would have been sufficient to rule out most punishments directed at cooperators. With noise, slightly fewer punishments are aimed at defectors and more are directed at cooperators than without noise, which is consistent with pro-social punishment errors.

Trends of attempted punishment against cooperators and defectors were similar across decision rules. Two exceptions include punishment of defectors under CDR2 with noise, which is relatively infrequent, and punishment of cooperators under CDR3 without noise, which is quite frequent.[5] Without noise, sufficient agreement is often reached under the CDRs on punishing defectors. Additionally, when more subjects are required to agree on punishments, fewer defectors emerge and more are punished on average. Conversely, noise appears to significantly hinder agreement on punishing defectors. Numerous punishments are ruled out, even under CDR2.

Note that defectors earn 12 more points than cooperators in the contribution stage of the experiment. Only under CDR2 and CDR3 without noise, the number of punishment points that subjects receive for defecting is on average 12 points above punishment for cooperation. Thus, under CDR2 and CDR3 punishments offset the payoff advantage of defecting (see columns 'av. total': $3.01 \times 6 - 0.02 \times 6 = 17.94$ punishment points under CDR2; $2.91 \times 6 = 17.46$ punishment points under CDR3). This may explain why CDR2 and CDR3 without noise are the only two conditions in which cooperation increases over time.

The lower panel of Figure 4.2 shows the predicted probability to punish an observed defector for each experimental condition.[6] Descriptive statistics are again listed in Table D.1, while predictions derived from the multilevel regression model are presented in Table D.2. In each period, subjects observe between zero and five defectors in their group, for whom they decide whether or not to punish. The third model shown in Table 4.3 presents condition differences between the predicted probabilities. We expected that decision rules do not affect punishment decisions without noise, that actors might be less likely to punish with noise than without noise, and that actors would more likely punish with noise as more agreement is required. Table 4.3 shows that predicted punishment probabilities are indeed lower with noise than without noise under CDR2 and CDR3. However, without noise, relative to the IDR, the predicted probability to

[5]The two sessions conducted with noise and CDR2 generated the lowest and second-lowest rates of attempted punishment of true defectors, respectively. High perverse punishment rates in CDR3 without noise are mostly attributable to four subjects who punish numerous cooperators. However, these are evenly dispersed over the two sessions such that they never achieve a majority.

[6]We do not find significant condition differences in the likelihood to punish observed cooperators.

punish is higher under CDR2 and CDR3, while with noise, the predicted probability to punish is lower under CDR2 than under the IDR. Thus, with noise, actors are less likely to punish when the likelihood that pro-social punishment errors are implemented is lower, and the effect of required collective agreement on the likelihood to punish differs between the two noise conditions. This is inconsistent with our assumptions. To examine how condition differences in punishment translate into condition differences in cooperation, we examine the effect of receiving punishment after contributing and defecting on subsequent contribution decisions.

4.5.4 Previous game effects on contribution decisions

Table 4.5 presents a regression model with characteristics of the previous interaction as determinants of current contribution decisions. Subjects' own previous contribution decision and received punishments, the number of other group members that the focal subject observed as cooperators in the previous period, and noise interactions for these variables are included. Punishments received for contributing and for defecting are distinguished as two separate effects. Both are specified as a dichotomous variable that states whether the subject was punished or not.[7] Coefficients presented in Table 4.5 are similar when the model is examined for each decision rule separately (output not shown), though the significance of most effects disappears due to the reduced sample size, and it is not possible to estimate the effect of receiving punishment for cooperation under the CDRs.

Table 4.5 shows that subjects are more likely to cooperate when they had cooperated previously. With noise, this effect is still significant but half as strong, reflecting higher fluctuations in contribution decisions over the periods. Receiving punishment for defection has a positive effect on subsequent cooperation. The net effect is insignificant with noise, although the difference between the noise and no-noise conditions is also not significant. Receiving punishment for cooperating negatively affects subsequent cooperation in the no-noise conditions only, while this effect completely disappears with noise.[8] Thus, our assumption that punishment of defectors is as beneficial to cooperation as punishment of cooperators is harmful is not confirmed for the noise conditions. Finally, the more other group members were observed as cooperators in the previous interaction, the higher the likelihood of subsequent cooperation. This effect is significant but much weaker in the noise conditions.

[7] We also estimated a model with a continuous variable indicating by how many others a subject was punished. However, this model suffers from multi-collinearity issues, as the number of punishers is highly correlated with the contribution decisions of the punished subject and of other group members.

[8] Punishment received for defection does have a highly significant effect with noise if specified as a continuous variable. The effect of punishment received for cooperation is insignificant regardless of precise specifications.

Table 4.5: Multilevel logistic regression on contribution decisions for period t, with decisions nested in subjects and sessions in the first punishment sequence (3528 PDs, 252 subjects).

	Coeff.	S.e.
Noise	1.157	0.652
CDR2	1.234**	0.441
× noise	-2.784**	0.636
CDR 3	1.954**	0.502
× noise	-2.986**	0.660
Own contribution $t-1$	2.766**	0.450
× noise	-1.924**	0.533
Punished while defecting $t-1$	1.276*	0.511
× noise	-0.699	0.561
Punished while cooperating $t-1$	-0.863*	0.365
× noise	1.074*	0.449
Obs. n other cooperators $t-1$	0.416**	0.085
× noise	-0.208*	0.107
Period	-0.041*	0.020
× noise	-0.098**	0.027
Constant	-2.471**	0.612
σ_u	0.000	0.353
σ_e	1.687**	0.143
Log Likelihood	-1407.135	

*Significant at the .05-level. **Significant at the .01-level (2-sided)

4.6 Conclusion and discussion

In this chapter, we experimentally compared the effect of different decision rules for implementing punishment on cooperation and earnings in series of one-shot six-person PDs with and without noise in the display of contribution decisions. We hypothesize that cooperation rates are lower with noise than without noise under each decision rule (Hypothesis 4.1a). This hypothesis is supported for the collective decision rules (CDRs), but not for the individual decision rule (IDR). Hereby, it has to be noted that cooperation under the IDR without noise is surprisingly low, increasing the likelihood that similar cooperation rates are achieved with noise. We do find earnings under all decision rule conditions to be negatively affected by noise, supporting Hypothesis 4.1b. Furthermore, we hypothesize that cooperation rates under noise conditions are higher under the least restrictive CDR (CDR2) than under a decision rule that requires majority consensus (CDR3) and the IDR (Hypothesis 4.2a). Earnings are also predicted to be higher under CDR2 than under the IDR (Hypothesis 4.2b). These hypotheses are not supported. Instead, with noise cooperation rates are higher under the IDR and CDR3 than CDR2. We find no differences in earnings between the IDR and CDR2. Moreover, with noise earnings under all decision rules are lower than when punishment is not possible at all.

Noise effects and differences between decision rules that we find at the macro-level may to some extent be explained by noise effects on micro-level behaviors. First, with noise, we find that subjects are less likely to punish observed defectors under the CDRs than the IDR, especially CDR2. Second, punishments directed at cooperators is detrimental to cooperation without noise, but does not significantly affect cooperation with noise. Punishment directed at defectors positively affects cooperation regardless of noise. These two results imply that when noise is present, only under the IDR punishment aimed at defectors was sufficient for maintaining cooperation, while the considerable punishment aimed at cooperators was not problematic. Under both CDRs, too few defectors were punished to enforce cooperation. Yet, under the IDR resources were wasted on misguided punishments aimed at cooperators, and hence in the end also the IDR is unable to maintain profits above baseline level with noise. Ambrus and Greiner (2012), Bornstein and Weisel (2010), Fischer et al. (2013), Grechenig et al. (2010), and Patel et al. (2010) find similar results for related settings.

Thus far, experimental evidence on the effect of noise on individual decision making remains inconclusive. First, we find that subjects are less likely to punish observed defectors with noise than without noise for the CDRs but not the IDR. Grechenig et al. (2010) and Ambrus and Greiner (2012) both employ an IDR, and find that noise makes punishment of both observed cooperators and defectors more frequent, but less severe. Conversely, less related studies find that punishment of defectors through an IDR is less common when information on others' contribution decisions is inaccurate (Bornstein and Weisel, 2010; Patel et al., 2010). Second, like Fischer et al. (2013), we find no significant effect of punishment aimed at cooperators on cooperation with noise, while Grechenig et al. (2010) and Ambrus and Greiner (2012) do find punishment aimed at cooperators to negatively affect subsequent cooperation regardless of noise. Future research may further dissect determinants of individual behaviors under imperfect information. Such research may also address CDR effects on punishment decisions in the presence of noise.

It is interesting to note that in conditions without noise, punishment through an IDR fails to maintain high cooperation rates over time. This contradicts numerous previous findings (e.g., Fehr and Gächter, 2000, 2002). This finding may be attributable to the fact that the subjects in our experiment only decided whether to punish or not. Additionally, each punishment reduced recipient payoffs by six points, which may not have sufficiently discouraged defection. We also find that subjects become more willing to cooperate without noise as more group members are required to agree on punishment decisions. Indeed, most perverse punishments are ruled out while most pro-social punishments are implemented. This suggests that both decision rules were tailored to the proportion of pro-social and perverse punishers in the population. Conversely, with noise cooperation and earnings were not increased by implementing punishment through CDRs. The CDR conditions employed were too restrictive in the noise games, as pro-social punishment rates were low under CDRs with noise, since many pro-social punishments were ruled out.

Chapter 5

n-Person Prisoner's Dilemmas with endogenous peer punishment institutions – The role of noise*

Abstract

We experimentally study six-person Prisoner's Dilemmas (PDs) in which subjects endogenously decide 1) whether to implement a peer punishment institution in their group and 2) whether the punishment institution, if implemented, implies more or less severe punishments. We consider PDs with perfect information, and PDs in which subjects observe each other's cooperation or defection with noise. We expect subjects to be less inclined to implement a punishment institution and less likely to prefer severe punishments with noise than without noise. Indeed, without noise, the majority of groups choose a punishment institution with severe punishments, while with noise most groups do not implement a punishment institution. In both noise conditions, cooperation and earnings increase when a punishment institution is implemented, especially with severe punishments. However, subjects in the noise condition perceive lower earnings under severe punishments than under the other options. Also, with noise, observing that cooperators are punished discourages subjects from implementing a punishment institution in subsequent interactions.

5.1 Introduction

Social groups, such as neighborhoods, work teams, and religious communities, thrive when members cooperate in pursuing group interests. However, cooperation problems arise if defecting on the group benefits individual members (Raub et al., 2015). In

*This is a slightly different version of Van Miltenburg, Buskens, and Raub (2014) that has been submitted to an international journal.

such situations, group members may encourage their peers to cooperate, for example through punishment. Henceforth, the opportunity for all actors to punish their fellow group members, i.e., for peer punishment, is referred to as a 'punishment institution' (cf. North, 1990).

Experimental research demonstrates that cooperation and long-term welfare increase when a punishment institution is exogenously (by the experimenter) implemented in cooperation problems (see Balliet et al., 2011; Chaudhuri, 2011 for an overview). This raises the deeper theoretical question whether actors involved in cooperation problems anticipate the consequences of punishment institutions, and voluntarily allow for the use of peer punishment in their group (cf. Batenburg et al., 2003; Prendergast, 1999). Several studies consider settings in which actors decide whether to interact under a punishment institution. They find that most actors initially prefer to interact without a punishment institution, but that punishment institutions gain popularity over time (e.g., Ertan et al., 2009; Gürerk, 2013; Gürerk et al., 2006, 2009; Markussen et al., 2014; Rockenbach and Milinski, 2006).

We build upon these findings in two ways. First, when actors involved in everyday cooperation problems decide to implement a punishment institution, it is likely that they will also select details of the punishment institution. In our experiment, in addition to deciding whether to implement a punishment institution in their group, actors determine punishment effectiveness, i.e., the extent to which each punishment will reduce its recipient's earnings (cf. Egas and Riedl, 2008; Nikiforakis and Normann, 2008). High punishment effectiveness more efficiently deters defection. Accordingly, cooperation rates are found to increase with exogenous punishment effectiveness (Ambrus and Greiner, 2012; Egas and Riedl, 2008; Nikiforakis and Normann, 2008). However, actors may be reluctant to allow punishment with high effectiveness, as it implies a risk that they may be punished severely themselves (Buchanan and Tullock, 1962). The occurrence of perverse punishment places even cooperators at risk of being punished (Cinyabuguma et al., 2006; Herrmann et al., 2008).

Second, studies on endogenous punishment institutions focus on cooperation problems in which actors have accurate information regarding the decisions of other group members. While this may drive results on institution formation (Nikiforakis, 2014), it does not always reflect everyday cooperation problems (Bereby-Meyer, 2012). We study cooperation problems in which actors are accurately informed about the decisions of their peers and cooperation problems that are characterized by noise. When noise is present, an actor may observe another actor's cooperation as a defection, and vice versa.[1] Noise implies that defectors may remain undetected by one or several potential punishers, while cooperators may receive 'misguided' punishment if other group members observe them as defectors. Accordingly, noise impedes the capacity for exoge-

[1]We consider one-shot interactions. This implies that theoretical complexities involved in strategic punishment with noise in repeated interactions (Green and Porter, 1984; Wu and Axelrod, 1995; Fudenberg et al., 2012) do not apply to our setting.

nous punishment institutions to maintain cooperation (Fischer et al., 2013; Grechenig et al., 2010; Chapter 4). While high effectiveness may better enforce cooperation when noise is present than low effectiveness (Ambrus and Greiner, 2012), misguided punishments aimed at cooperators will be more severe. Thus, it is unclear whether findings on endogenous punishment effectiveness are generalizable to noisy environments.

In our experiment, actors individually indicate (henceforth vote) whether they wish to interact under a punishment institution and which level of effectiveness they prefer. The option that the majority of the group votes for is implemented. Many iterations may be required before voting outcomes and behaviors under each outcome converge to a stable pattern (cf. Ambrus and Greiner, 2012; Gächter et al., 2008). Moreover, while repeated interactions in fixed groups provide incentives for reputation formation that may be anticipated in votes, such incentives are ruled out in one-shot interactions. We thus study endogenous implementation of punishment institutions and effectiveness over a long sequence of one-shot interactions.

The chapter is organized as follows. In Section 5.2, we review related experimental literature. Section 5.3 outlines the theoretical model and hypotheses. The laboratory experiment is presented in Section 5.4, results are reported in Section 5.5, and Section 5.6 presents conclusions.

5.2 Related literature

The majority of experimental research that compares behaviors in Prisoner's Dilemmas or Public Goods Games with and without exogenous punishment institutions has revealed a number of consistent behavioral patterns (Chaudhuri, 2011). Typically, in the absence of a punishment institution, cooperation rates begin at approximately 50% and decline to almost complete defection if one-shot interactions are repeated with different partners (Camerer, 2003; Ledyard, 1995). Several subjects in these interactions can be classified as conditional cooperators who cooperate as long as they expect that others will do the same (e.g., Fischbacher et al., 2001).

Under exogenous punishment institutions, several subjects punish defectors, even in one-shot interactions (e.g., Fehr and Gächter, 2002), and cooperation is typically maintained at a high level (Balliet et al., 2011; Chaudhuri, 2011; Nikiforakis, 2014). Moreover, cooperators often receive some punishment, too (e.g., Cinyabuguma et al., 2006; Herrmann et al., 2008; Ostrom et al., 1992), hindering their subsequent cooperation. Henceforth, punishment directed at defectors is deemed 'pro-social,' while punishment directed at cooperators is deemed 'perverse.' In most experiments, every punishment depletes recipient income by three times the cost of punishment allocation (1:3 effectiveness, e.g., Fehr and Gächter, 2002). If punishment effectiveness is varied exogenously, cooperation rates increase with effectiveness (Ambrus and Greiner, 2012; Egas and Riedl, 2008; Nikiforakis and Normann, 2008). However, more resources are destroyed by punishment when more effective punishments are employed. Accordingly, Egas and Riedl (2008) find that group earnings decrease with effectiveness in series of one-shot interactions.

Experiments employing noise similar to the current study in cooperation problems with exogenous punishment institutions and 1:3 punishment effectiveness find that a 20% or higher probability of inaccurately observing contribution decisions causes a decrease in cooperation rates and earnings (Fischer et al., 2013; Grechenig et al., 2010; Chapter 4). Studies examining other means of implementing noise also find noise to negatively affect earnings in cooperation problems with a punishment institution (Ambrus and Greiner, 2012; Bornstein and Weisel, 2010; Patel et al., 2010). Initial evidence suggests that with noise, cooperation rates and earnings increase with punishment effectiveness in repeated interactions in fixed groups (Ambrus and Greiner, 2012).

A growing number of experimental studies examine the implementation of peer punishment institutions through voting procedures. To our knowledge, none of these studies have addressed noise or endogenous punishment effectiveness. Sutter et al. (2010) allowed subjects to vote on the implementation of a punishment institution, a reward institution, or neither, before the first interaction. The authors consider two exogenous levels of effectiveness. They find that the majority of groups do not opt for a punishment or reward institution under 1:1 effectiveness, and that most groups select a reward institution under 1:3 effectiveness. Botelho et al. (2005) allowed subjects to vote on whether to implement a punishment institution a final round after a series of cooperation problems with and without a punishment institution. The authors find that only one group voted in favor of the punishment institution. However, in studies where fixed groups are offered multiple opportunities to vote, and the outcome applies to several ensuing interactions, punishment institutions gain popularity over time. Ertan et al. (2009) find that most groups allow for punishment of below-average contributors after a number of voting rounds. Kamei et al. (2011) and Markussen et al. (2014) find that after several votes, many subjects prefer peer punishment institutions over institutions that automatically punish defectors if the cost of implementing a peer punishment institution is low relative to automatic punishment. Finally, studies in which subjects can migrate between groups with and without a punishment institution find that many subjects opt for punishment institutions over time (Gürerk, 2013; Gürerk et al., 2009; Rockenbach and Milinski, 2006; Fehr and Williams, 2013).

5.3 Experimental game and hypotheses

5.3.1 A cooperation problem with noise, a punishment stage, and a voting stage

Before describing the voting procedure, we outline how cooperation and punishment are defined in the experiment. Cooperation problems are modeled as one-shot six-person Prisoner's Dilemmas (PDs).[2] Each actor i receives an endowment $w > 0$. Subsequently, all actors independently and simultaneously decide whether to contribute their entire endowment to a 'group account,' i.e., actor i's contribution c_i is either w (cooperation) or 0 (defection). Joint contributions $c = \sum c_i$ are multiplied by m, with $1 < m < 6$, and divided equally among group members. Thus, individual actors gain more from defecting than from cooperating (since $m < 6$), while the group payoff is maximized if everyone cooperates (since $m > 1$). Each actor is better off under full cooperation than if all defect (since $mw > w$). Still, full defection constitutes the unique Pareto-suboptimal Nash equilibrium of the PD under the assumption that actors are rational and self-regarding. Henceforth, we refer to actors who are rational and self-regarding as payoff-maximizing actors. In the experiment, we comply with conventional values of endowment and individual return from cooperation by using $w = 20$ and $m = 2.4$, respectively (cf. Fehr and Gächter, 2000, 2002).

After the contribution stage, each actor i receives a signal o_{ij} for each other group member $j \neq i$ on j's contribution decision. Two variants of the PD are considered, that differ in the accuracy of the signal. In the PD without noise, the signal is always correct (i.e., $o_{ij} = c_j$). In the PD with noise, the signal may be inaccurate, that is, $o_{ij} \neq c_j$ with probability p and $o_{ij} = c_j$ with probability $1 - p$. If a signal is inaccurate, i observes j's cooperation as a defection and vice versa. Whether o_{ij} is accurate is determined independently for each observing actor i of each choice c_j. Payoffs are based on actual contributions, though actors cannot infer actual payoffs from the noisy contributions that they observe. Subjects in the experiment only observe their actual payoffs at the end of a session. Still, noise does not affect the payoff structure of the PD such that the Nash equilibrium of zero contributions remains unchanged for payoff-maximizing actors. In the experiment, $p = 0.2$, and thus each actors' contribution decision is incorrectly observed, on average, by one of the five others in a six-person group.

In both the PD with and without noise, if a punishment institution is implemented, once actors observe signals o_{ij}, each i can decide for each j whether to punish. If i decides to punish j, actor i pays a cost of $a > 0$ for punishing, and the payoff for j is reduced by $b > a$. Thus, i's earnings from the contribution stage of the PD are reduced by an amount of a for each j whom i decides to punish and by an amount of b for

[2] Group size and punishment effectiveness are set such that a different number of punishers is required to deter defectors under each effectiveness and noise condition and such that not too many group members are required to punish to achieve deterrence, such that enforcing cooperation is feasible.

each j who decides to punish i. As punishment is costly, payoff-maximizing actors do not punish in one-shot interactions. Thus, when all actors are payoff maximizing and expect others to be as well, no actor enacts or expects to receive punishment, and the Nash equilibrium of full defection remains unchanged. In the experiment, two levels of effectiveness a:b are considered. The punishment cost remains constant at $a = 2$. With low effectiveness, $b = 6$ so that a:b = 2:6 = 1:3. With high effectiveness, $b = 12$ so that a:b = 2:12 = 1:6. Thus, recipients lose three or six times the amount that actors spend on punishment allocation.

Whether a punishment stage is added to the PD and, if so, the level of effectiveness employed, is decided by vote. The voting stage takes place before actors make contribution decisions. Voting is compulsory and costless. When actors vote, they know whether the PD they are participating in includes or does not include noise. First, actors vote on whether to add a punishment stage to the PD. Second, actors vote for high or low effectiveness of punishment, regardless whether they voted in favor of a punishment stage, and without knowing whether a punishment stage will be added. If a majority votes against the punishment stage, this stage is omitted from the PD. This outcome is referred to as No Punishment (NP). If a majority votes in favor of a punishment stage, this stage is added to the PD with the corresponding punishment effectiveness voted for. These outcomes are referred to as Low Punishment (LP) and High Punishment (HP). If exactly three of the six group members vote in favor of implementing a punishment institution, or for a certain effectiveness, the outcome of the respective vote is randomly determined.[3]

5.3.2 Costs and benefits of interacting under a punishment institution

As stated above, if all actors are payoff maximizing and expect others to be as well, irrespective of noise no punishment is allocated in a sub-game perfect equilibrium, and all group members defect regardless of the presence of a punishment institution. Voting outcomes are then irrelevant for behavior and earnings, and actors are indifferent in the vote for a punishment institution and for effectiveness. However, from the literature cited in Section 5.2, it is evident that some actors in cooperation problems behave in a manner that is inconsistent with the payoff-maximizing actor model. First, conditional cooperators cooperate as long as they expect others to do so as well, also under NP. Second, pro-social and perverse punishers punish group members whom they observe as defectors and cooperators, respectively.

[3]Theoretically, this procedure cannot exclude strategic votes. Actors who prefer HP (LP) over NP and NP over LP (HP) and who expect at least three others to vote for the punishment effectiveness that they least prefer vote against implementing a punishment institution, i.e., their second preference, to avoid ending up under their least preferred outcome, if they expect to be pivotal. We expect the combination of these preference patterns and expectations of other votes to be unlikely. We employed the present voting procedure, as it leads to clear outcomes and is easy for subjects to understand.

The presence of conditional cooperators and punishers alters the implication of the voting outcome for payoff-maximizing actor i. First, while the presence of conditional cooperators can imply a degree of cooperation of i's fellow group members under NP, pro-social punishers are likely to enforce even more cooperation among fellow group members, increasing i's earnings under LP and HP. Second, while actor i always defects and cannot be punished under NP, i may be punished for certain behaviors or may cooperate to avoid receiving punishment under LP and HP. Thus, interacting under LP or HP can 'benefit' i compared to interacting under NP given the higher cooperation rates of other group members. Conversely, interacting under LP or HP instead of NP can have a 'cost' for i, as he or she may have to cooperate and may be punished. Payoff-maximizing actors who anticipate the presence of conditional cooperators and punishers may opt for LP or for HP if they expect the benefit to outweigh the cost.

Table 5.1, as described further below, shows how costs and benefits of interacting under LP or HP rather than NP for payoff-maximizing actors depend on noise and punishment effectiveness. The table is used to develop hypotheses on voting decisions. The far-left column comprises each possible number of pro-social punishers in a group, that is, group members who punish others whom they observe as defectors. Corresponding rows outline how the earnings of payoff-maximizing actors are affected through interacting under LP or HP instead of NP. For the noise conditions, the table shows expected values based on an average of 20% of pro-social punishers inaccurately observing the focal actor's contribution decision. We assume that actors behave as if they use Table 5.1 to maximize their expected payoffs, given their expected number of pro-social punishers present in the group. Perverse punishment is not accounted for in Table 5.1, but we later hypothesize effects of experiencing punishment directed at (observed) cooperators on subsequent votes.

Columns entitled 'Pun D' and 'Pun C' in Table 5.1 show (expected) punishments that actors receive under LP and HP after defecting and cooperating, respectively. Actors are punished by six points under LP, and by twelve points under HP, by every pro-social punisher who observes them as defector (Section 5.3.1). Accordingly, the no noise section of Table 5.1 shows that if actors defect, they receive six punishment points under LP and twelve punishment points under HP for each pro-social punisher. Actors are not punished for cooperating. With noise, an average of 20% of pro-social punishments are targeted at actual cooperators and 80% at actual defectors. Accordingly, the noise section of Table 5.1 shows that under LP actors expect to receive $0.8 \times 6 = 4.8$ punishment points for defecting and $0.2 \times 6 = 1.2$ punishment points for cooperating for each pro-social punisher. Likewise, under HP actors expect to receive $0.8 \times 12 = 9.6$ punishment points for defecting and $0.2 \times 12 = 2.4$ punishment points for cooperating for each pro-social punisher. For example, if actors under LP with noise expect their group to contain four pro-social punishers, they should expect to receive $4 \times (0.8 \times 6) = 19.2$ punishment points for defecting and $4 \times (0.2 \times 6) = 4.8$ punishment points for cooperating.

Table 5.1: Punishment received for defection and cooperation (Pun D, Pun C), payoff-maximizing contribution decisions (Dec); payoff loss compared to defecting under NP given the indicated decisions (Cost); required increases in others' cooperation rates relative to NP to offset indicated costs (Incr) for each voting outcome and number of pro-social punishers in a group. For the noise condition, the table shows expected values based on an average of 20% inaccurate observations. Payoffs correspond to those in the experiment.

Number of pro-social punishers	No Noise										Noise									
	LP					HP					LP					HP				
	Pun D	Pun C	Dec	Cost	Incr	Pun D	Pun C	Dec	Cost	Incr	Pun D	Pun C	Dec	Cost	Incr	Pun D	Pun C	Dec	Cost	Incr
0	0	0	D	0	0	0	0	D	0	0	0	0	D	0	0	0	0	D	0	0
1	6	0	D	6	1	12	0	C	12	2	4.8	1.2	D	4.8	1	9.6	2.4	D	9.6	2
2	12	0	C	12	2	24	0	C	12	2	9.6	2.4	D	9.6	2	19.2	4.8	C	16.8	3
3	18	0	C	12	2	36	0	C	12	2	14.4	3.6	D	14.4	2	28.8	7.2	C	19.2	3
4	24	0	C	12	2	48	0	C	12	2	19.2	4.8	C	16.8	3	38.4	9.6	C	21.6	3
5	30	0	C	12	2	60	0	C	12	2	24	6	C	18	3	48	12	C	24	3

'Dec' columns in Table 5.1 show payoff-maximizing contribution decisions. In our experiment, as cooperators receive only $20 \times 2.4/6 = 8$ points in return for their own contributions, cooperation yields $20 - 8 = 12$ less points than defection. Thus, actors maximize (expected) payoff by cooperating if (expected) punishment for defection ('Pun D') is at least 12 points more severe than (expected) punishment for cooperation ('Pun C'). For example, in groups with four pro-social punishers under LP with noise, expected punishment for defection will be more than twelve points more severe than expected punishment for cooperation ($19.2 - 4.8 = 14.4 > 12$). Accordingly, payoff-maximizing actors in the PD with noise who anticipate the presence of four pro-social punishers will cooperate under LP.

'Cost' columns in Table 5.1 specify the (expected) cost of interacting under LP or HP, i.e., the earnings that payoff-maximizing actors (expect to) lose when they interact under LP or HP rather than NP. The (expected) cost comprises two elements. First, defection is payoff maximizing under NP, while cooperation may maximize (expected) payoffs under punishment institutions. Second, actors cannot be punished under NP, but might receive punishment under LP or HP. Accordingly, if the 'Dec' column shows that defection maximizes (expected) payoffs, 'Cost' is equal to (expected) punishment for defection ('Pun D'). If the 'Dec' column shows that cooperation maximizes (expected) payoffs, 'Cost' is equal to (expected) punishment for cooperation ('Pun C') plus 12, i.e., the cost of cooperation rather than defection. For example, in a group with four pro-social punishers under LP with noise, payoff-maximizing actors cooperate and expect to receive 4.8 punishment points for doing so. Hence, the expected cost when they cooperate under LP rather than defect under NP is $4.8 + 12 = 16.8$.

As noted above, interacting under LP or HP might be advantageous if other group members who would have defected under NP cooperate under LP or HP. In the experiment, cooperation generates an eight-point increase in the earnings of each fellow group member. Accordingly, each additional cooperator under LP or HP compared to NP secures eight more points for everyone else in the group. 'Incr' columns in Table 5.1 outline increases in others' cooperation rates required to offset (expected) cost of interacting under LP and HP rather than NP. This is equal to 'Cost' divided by eight and rounded up to the next integer value. For example, in a group with four pro-social punishers under LP with noise, when three more fellow group members cooperate under LP than under NP, earnings from others' cooperation increase by $3 \times 8 = 24$ points, offsetting the focal actors' expected cost of 16.8 points. Thus, if payoff-maximizing actors in the PD with noise expect four pro-social punishers under LP, they expect to earn more under LP than under NP if they expect that at least three more others cooperate under LP than under NP.

5.3.3 Hypotheses – noise and punishment effectiveness

We wish to test two types of hypotheses. First, we develop hypotheses on cooperation under each voting outcome and on noise effects on cooperation and voting. We expect no major differences in perverse punishment (directed at observed cooperators) between the experimental conditions that affect differences in cooperation rates. Additionally, we expect actors to not *a priori* expect perverse punishment when making contribution and voting decisions. Therefore, we ignore perverse punishment in these hypotheses. Second, we present hypotheses on how experiencing certain behaviors of other group members, including punishments that may be perverse, affect subsequent voting.

The more actors allocate pro-social punishment under a noise condition or voting outcome, the more punishment defectors receive relative to cooperators under this outcome, and this may in turn affect cooperation and voting. How noise and punishment effectiveness affect punishment decisions will be tested empirically and evaluated in light of differences between experimental conditions. On one hand, actors may be less inclined to punish under HP than under LP due to a belief that punishment is too severe. On the other hand, actors may be more inclined to punish under HP than under LP because they get more 'value' for punishment expenditures. Previous studies suggest that actors are more inclined to punish, the higher punishment effectiveness, both with and without noise (Ambrus and Greiner, 2012; Egas and Riedl, 2008). We may expect that noise renders actors reluctant to punish observed defectors, as they may target actual cooperators. A previous study that employs 1:3 effectiveness (Chapter 4) finds that noise has no significant effect on decisions to punish observed defectors.

We first derive hypotheses on noise effects on voting. 'Pun' columns in Table 5.1 show that under every number of pro-social punishers, defectors expect to receive less punishment and cooperators expect to receive more punishment with noise than without noise. If noise renders actors reluctant to allocate pro-social punishment, the extent to which defectors expect to receive less punishment relative to cooperators with noise even increases. Accordingly, we expect that actors anticipate more cooperation under LP and HP without noise. Moreover, 'Incr' columns in Table 5.1 show that actors often require more cooperation with noise to offset expected costs of interacting under LP or HP. In sum, with noise actors do not expect as much cooperation under LP or HP as without noise, though more expected cooperation is necessary to offset expected costs. We expect the lower expected benefits, and higher expected costs, to render actors less inclined to vote in favor of implementing a punishment institution with noise.

Hypothesis 5.1 *Actors are less inclined to vote in favor of implementing a punishment institution with noise.*

'Cost' columns in Table 5.1 show that with noise interacting under HP has a higher cost than interacting under LP for each number of pro-social punishers. This difference is even more pronounced when, in line with previous research, actors are more inclined to punish under HP than under LP. Thus, with noise, actors only vote for HP over LP if they expect that more cooperation under the former offsets the additional cost. Without noise, expected costs of interacting under HP and LP are equal, unless actors expect only one pro-social punisher to be present under LP. Thus, without noise, actors typically do not require a promise of higher cooperation under HP than LP to induce them to vote for the former. Therefore, we hypothesize that actors are more reluctant to vote for HP with noise than without noise.

Hypothesis 5.2 *Actors are less inclined to vote for HP rather than LP with noise.*

One effect that is not accounted for in Table 5.1 may further encourage a negative noise effect on voting for (a more effective) punishment institution after actors gain experience with the PD under each voting outcome. With noise, and when there are more cooperators than defectors, more cooperators are observed as defectors on average than the other way around. Thus, the more actual cooperation exceeds 50%, the more cooperation rates that actors *observe* fall below actual cooperation rates. The opposite holds for cooperation rates lower than 50%. This may imply that others' cooperation rates, and therefore own earnings, that actors observe with noise under a punishment institution are lower than actual cooperation and earnings, while observed cooperation and earnings under NP are higher than in reality. We test empirically whether these observation biases substantially misrepresent the effect of punishment institutions on observed cooperation or earnings.

We now hypothesize on cooperation under each voting outcome. Punishment of defectors promotes cooperation, while punishment of cooperators can hinder subsequent cooperation. First, for both noise conditions, 'Pun' columns in Table 5.1 show that for each number of pro-social punishers, defectors receive more punishment relative to cooperators under HP than under LP. The difference is even more pronounced when, in line with previous research, actors are more inclined to punish under HP than under LP. Second, for each number of pro-social punishers, 'Pun' columns in Table 5.1 show that defectors receive less punishment relative to cooperators with noise than without noise. These differences between noise conditions are even more pronounced when noise makes actors reluctant to allocate pro-social punishment. Thus, in both noise conditions we hypothesize that cooperation rates are higher under HP than under LP, and for each level of punishment effectiveness we hypothesize higher cooperation rates without noise than with noise.

Hypothesis 5.3 *In both noise conditions, cooperation rates are higher under HP than under LP and higher under LP than under NP.*

Hypothesis 5.4 *Cooperation rates under LP and HP are higher without noise than with noise.*

5.3.4 Hypotheses – effects of experience

A second set of hypotheses addresses how experiences affect subsequent votes. First, votes may be affected by the cooperation of fellow group members that actors observed in previous PDs under each voting outcome. Regarding votes on whether to implement a punishment institution, the more cooperation actors have experienced under LP and HP rather than NP, the more likely they will be to expect higher earnings under punishment institutions than under NP. Regarding the vote for punishment effectiveness, the more cooperation of other group members actors have experienced under HP relative to LP, the more inclined they will be to expect higher earnings under HP than under LP.

Hypothesis 5.5 *Actors are more inclined to vote for an option, the more cooperation of fellow group members they have experienced under this option, and the less cooperation of fellow group members they have experienced under the alternative option(s).*

'Incr' columns in Table 5.1 show that with noise, actors often require more cooperation under LP or HP than without noise to offset expected costs of interacting under LP or HP rather than NP. 'Cost' columns show that with noise, interacting under HP has a higher expected cost than interacting under LP such that with noise, actors only vote for HP if they expect more cooperation under HP than under LP. The difference in cost is even more pronounced if actors are more inclined to punish under HP than under LP. Conversely, without noise, the cost of interacting under HP and LP is mostly equal. Thus, with noise, more experienced cooperation may be required than without noise to convince actors to vote for a (more effective) punishment institution.

Hypothesis 5.6 *The effect of previously experienced cooperation on the likelihood of voting for a certain option is weaker with noise than without noise.*

Second, votes may be affected by experiencing that observed defectors are punished. Punishment of observed defectors increases expected benefits of interacting under LP or HP, as other's cooperation in future interactions under LP or HP may increase. However, 'Cost' columns in Table 5.1 show that with noise, the expected costs of interacting under punishment institutions also increase with the number of pro-social punishers. Additionally, with noise, the *difference* between expected costs of interacting under HP and of interacting under LP increases with the number of pro-social punishers. Thus, with noise, it remains unclear whether actors expect benefits of experiencing punishment of observed defectors to outweigh costs. Conversely, without noise, the cost of interacting under LP or HP is hardly affected by the number of pro-social punishers. Accordingly, while experiencing that observed defectors are punished may render actors more inclined to opt for a (more effective) punishment institution without noise, we expect that with noise, this likely positive effect will be reduced or become negative.

Hypothesis 5.7 *Noise reduces the effect of experienced punishment of observed defectors under LP and HP on the likelihood to vote in favor of implementing a punishment institution and to vote in favor of HP.*

Finally, votes may be affected by experiencing that observed cooperators are punished. In both noise conditions, punishment of observed cooperators discourages cooperation, especially under HP where punishments are more severe. Thus, actors who observe that cooperators are punished expect lower cooperation rates under punishment institutions in the future and may expect to receive punishment if they cooperate themselves. Accordingly, we hypothesize that punishment aimed at observed cooperators renders actors more reluctant to vote in favor of implementing a punishment institution and for HP.

Hypothesis 5.8 *Actors are less inclined to vote in favor of implementing a punishment institutions and less inclined to vote for HP rather than LP after experiencing punishment of observed cooperators under LP and HP.*

With noise, experiencing that observed cooperators are punished may occur when recipient contribution decisions are wrongly observed by focal actors or punishers. Thus, negative effects of observing cooperators being punished are expected to be smaller with noise than without noise, as without noise punishment of cooperators is perverse by definition.

Hypothesis 5.9 *Effects of experienced punishment of observed cooperators under LP and HP on the likelihood to vote in favor of implementing a punishment institution and to vote for HP are less negative with noise than without noise.*

5.4 Experimental procedure

In the experiment, subjects participated in series of six-person PD games with endogenous punishment institutions and effectiveness described in Section 5.3. In half of the sessions, the PD with noise was employed, while in the other half, subjects participated in the PD without noise. Subjects were informed on whether and how noise was implemented. Payoffs presented in Section 5.3 were represented to subjects as points that were translated into monetary earnings at the end of a session. Throughout each session, subjects were randomly re-matched in different groups of six after every interaction.

At the start of the experiment, subjects received paper instructions that only described the contribution stage of the PD. Subsequently, they answered control questions on the computer to verify that they understood this part of the PD. If a subject did not answer a question correctly, the right answer was presented on the screen. Each session proceeded with five periods of a PD with a contribution stage only to familiarize the subjects with the decision situation. Also these initial PDs were played either with or without noise depending on the experimental condition. After each period, subjects were informed of the (noisy) contribution decisions made by their group members and of their own payoff. In the noise condition, subjects were informed of the payoff that they would earn if contribution decisions that they observed were accurate, instead of their actual earnings. This prevented subjects from inferring actual contribution rates.

After the five initial periods, subjects interacted for 40 periods in the PD with endogenous punishment institutions and effectiveness. Thus, each period opened with a vote on whether or not to implement a punishment institution, and between LP and HP, and the option that the majority of the six-person group voted for was implemented.[4] Subjects received new instructions describing this phase of the experiment. After each voting stage, subjects were informed of the voting outcome in their group, but not on how many group members voted for each alternative. Regardless of voting outcomes, in this part of the experiment subjects received an additional endowment of 10 points after the contribution stage of each PD. Punishment cost was set at two points. Thus, in groups that interacted under LP or HP, the endowment allowed subjects to punish all five fellow group members. Under LP and HP, subjects were informed of punishments that were received by all group members after the punishment stage. Under each voting outcome, subjects were informed of their own earnings after each interaction. Again, in the noise condition, these earnings were based on noisy contribution decisions perceived by subjects. While subjects could theoretically acquire negative aggregate earnings during this part of the experiment, we did not develop a protocol for negative earnings, as they were highly unlikely and indeed never occurred. At the end of the session, subjects were informed of their actual accumulated earnings.

The experiment was programmed using z-Tree (Fischbacher, 2007) and conducted at the ELSE laboratory at Utrecht University. Subjects were recruited using the online recruiting system ORSEE (Greiner, 2004). 156 subjects participated in total: 78 in the noise condition and 78 in the condition without noise (41% male, 85% students, 30% economics students). Subjects earned €12.50 on average (minimum of €7.50, maximum of €15.50).

5.5 Methods and results

5.5.1 Descriptive results

In line with previous findings (Ledyard, 1995), cooperation rates in the five initial PDs without a voting stage steadily decline from roughly 45% to roughly 28% regardless of noise (output not shown). However, marked noise effects are visible in the PDs with a voting stage. Figures 5.1a-5.1d show individual votes, voting outcomes, contribution decisions, and earnings for each PD period with a voting stage and for both noise conditions.

Subjects make two decisions at each voting stage: whether to vote in favor of implementing a punishment institution, and whether to vote for LP or HP. Figure 5.1a

[4] In the experiment, neutral labels were used. Options in the voting stage were referred to as 'system A' (NP), 'system B1' (LP), and 'system B2' (HP).

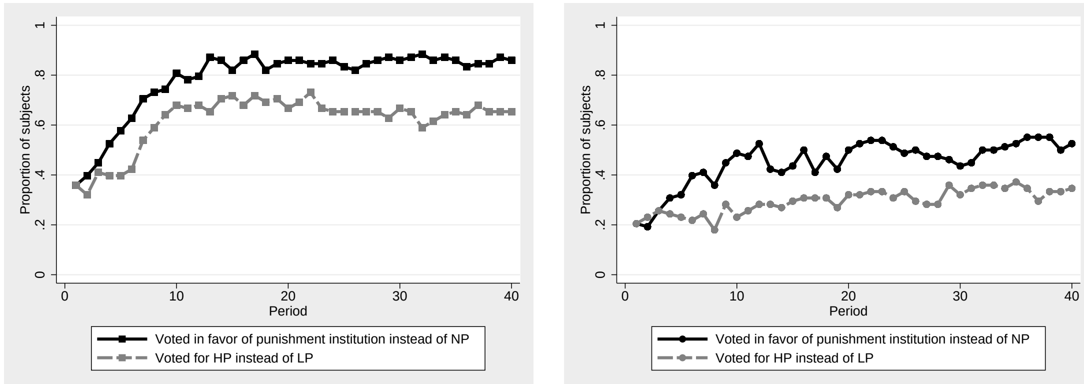


Figure 5.1a: Proportion of subjects voting in favor of implementing a punishment institution rather than NP, and proportion of subjects voting for HP rather than LP without (left) and with (right) noise.

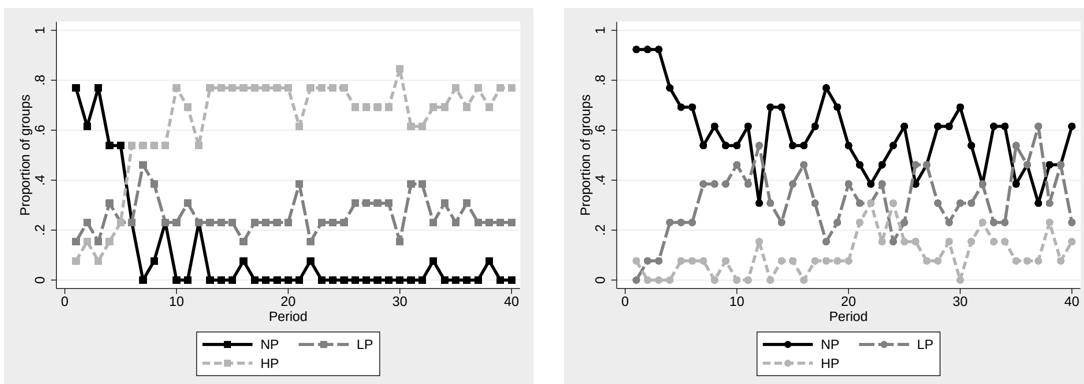


Figure 5.1b: Proportion of groups under each voting outcome without (left) and with (right) noise.

shows the proportion of subjects who vote in favor of implementing a punishment institution rather than NP and the proportion of subjects who vote for HP rather than LP. Note that these are two independent decisions. Without noise, only 35% of all subjects initially vote in favor of implementing a punishment institution. However, punishment institutions quickly increase in popularity, with a stable majority of roughly 85% of subjects voting in favor of implementing a punishment institution after the tenth period. In the vote for effectiveness, 35% of all subjects initially vote for HP. However, also HP increases in popularity over time, with a stable majority of roughly 70% voting for HP after the tenth period. With noise, 20% of all subjects initially vote in favor of implementing a punishment institution, increasing to roughly 50% over the forty periods. In addition, 20% of all subjects initially vote for HP, increasing to 35% over the forty periods. This increasing tendency toward punishment institutions complements previous findings (e.g., Ertan et al., 2009; Gürerk et al., 2006; Markussen et al., 2014).

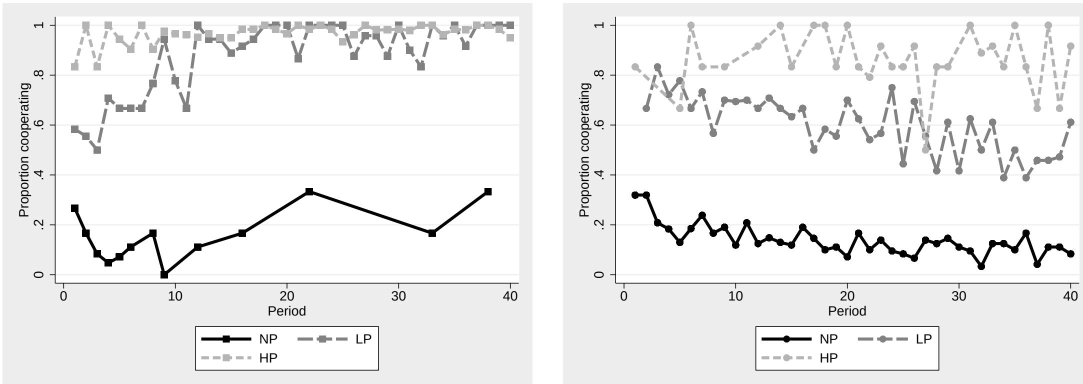


Figure 5.1c: Cooperation rates under each voting outcome without (left) and with (right) noise.

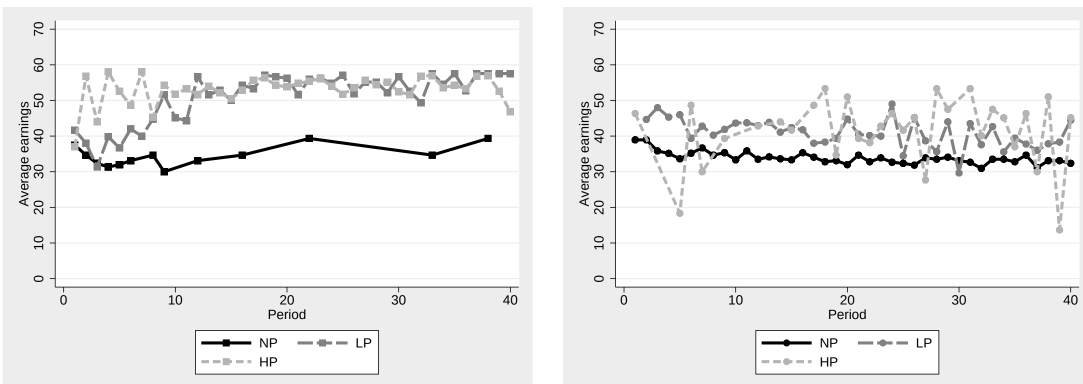


Figure 5.1d: Earnings for each voting outcome without (left) and with (right) noise.

Figure 5.1b presents voting outcomes. Without noise, most groups initially interact under NP, but after a number of rounds, most groups interact under HP, followed by LP. Groups hardly ever interact under NP in later rounds. With noise, almost all groups initially interact under NP. Subsequently, a substantial proportion of groups interact under one of the punishment institutions, but NP remains the most popular in almost every period. Of the two punishment institutions, more groups interact under LP than under HP.

Figure 5.1c shows cooperation rates for each voting outcome. Without noise, cooperation rates under NP are relatively low. Conversely, almost full cooperation is immediately achieved under HP. Initially, roughly 60% of subjects cooperate under LP, but this rapidly increases to nearly full cooperation as well. With noise, cooperation rates under NP likewise remain relatively low. From the first period onward, cooperation rates are considerably higher under both LP and HP. Cooperation rates slowly decrease over time under LP, but remain high under HP. Under both HP and LP, cooperation rates are consistently lower with noise. Previous studies likewise found that cooperation rates increase under endogenous punishment institutions (Kamei et al., 2011; Markussen et al., 2014; Sutter et al., 2010), and that cooperation rates are lower under exogenous punishment institutions with noise (Fischer et al., 2013; Grechenig et al., 2010; Chapter 4) but increase with effectiveness (Ambrus and Greiner, 2012; Egas and Riedl, 2008; Nikiforakis and Normann, 2008).

Figure 5.1d shows that without noise, higher earnings are almost immediately achieved under LP and HP than under NP. With noise, earnings achieved under the LP and HP are somewhat higher than those achieved under NP, but this difference is smaller than without noise. Additionally, while higher cooperation rates are achieved under HP than under LP with noise, this does not appear to translate into higher earnings. Previous studies found that exogenous punishment institutions negatively affect earnings with noise under 1:3 effectiveness (Grechenig et al., 2010; Chapter 4).

Table 5.2 shows punishment received by actual cooperators and defectors. Without noise, 91% of all defectors is punished by at least one fellow group member under LP, and each defector receives punishment under HP. Additionally, the average number of punishments that defectors receive when they are punished is slightly higher under HP than under LP, such that defectors receive more punishments on average under HP than under LP. Conversely, few cooperators are punished, typically by one group member only, such that cooperators on average receive few punishments under both LP and HP. With noise, most actual defectors are punished, though at lower percentages and by fewer punishers than without noise. Accordingly, under LP and HP defectors on average receive fewer punishments with noise than without noise. Several cooperators receive punishment in the noise condition, under HP especially, but only from slightly more than one group member on average, and thus the average number of punishments received by actual cooperators remains low.

Table 5.2 shows that under LP with noise, defectors receive $1.23 \times 6 = 7.38$ punishment points and cooperators received $0.38 \times 6 = 2.28$ punishment points on average. Thus, the difference in average punishment for defection and cooperation ($7.38 - 2.28 = 5.1$ points) did not offset the 12-point payoff advantage of defecting. Rather, under LP with noise, defectors earn more than cooperators on average. Similar calculations reveal that punishment under the other voting outcomes is on average sufficient to deter defection. This may explain why LP with noise is the only punishment institution that does not result in nearly full cooperation and that showed a decline in cooperation over time (Figure 5.1c).

5.5.2 Cooperation, punishment, and earnings

In what follows, multilevel regression models are presented with random effects at the subject level to control for interdependencies within subjects. Too few observations were made at the session level to include random effects that control for interdependencies within sessions. Models with fixed session effects are difficult to interpret, as sessions and noise are perfectly collinear by design. To facilitate a straightforward interpretation

Table 5.2: Number (N) of actual cooperators and defectors, percentage (%) of cooperators and defectors that receive at least one punishment, average number of punishments received by cooperators and defectors who are punished (Av. pun), and average number of punishments received by cooperators and defectors overall (Av. total) under LP and HP with and without noise.

		Actual defectors				Actual cooperators			
Noise	Eff.	N	%	Av. pun.	Av. total	N	%	Av. pun.	Av. total
No	LP	99	91	2.68	2.43	705	17	1.08	0.18
	HP	51	100	3.37	3.37	1929	17	1.09	0.18
Yes	LP	410	74	1.66	1.23	586	33	1.17	0.38
	HP	41	95	2.56	2.44	271	49	1.21	0.59

of noise effects, sessions are not controlled for in the models presented below. We indicate when effects are not robust if session fixed effects are included to control for interdependencies at the session level.

Figures 5.2a-5.2c show the effect of noise condition and voting outcome on the predicted probability to punish an observed defector, predicted cooperation, and predicted period earnings. These predictions are based on regression models shown in Tables E.3 and E.4 in Appendix E. The differences described below are significant according to these models. Table E.1 in Appendix E presents corresponding descriptive statistics.

Figure 5.2a shows the predicted probability that a subject will punish an observed defector.[5] In each period, subjects observe between zero and five defectors in their group, for whom they decide whether to punish. Figure 5.2a shows that in both noise conditions, the predicted probability to punish an observed defector is higher under HP than under LP, and for both for HP and LP the predicted probability is lower with noise. Thus, defectors receive significantly more punishment relative to cooperators under HP than under LP and more without noise than with noise. Note that for deriving our hypotheses, we already anticipated that noise may render actors reluctant to allocate pro-social punishment and that subjects may be more inclined to punish (observed) defectors when punishment effectiveness is high (Section 5.3).

The left panel of Figure 5.2b shows the predicted probability that a subject cooperates. In line with Hypothesis 5.3, Figure 5.2 shows that the predicted probability to cooperate is higher under HP than under LP and higher under LP than under NP regardless of noise. Moreover, under LP and HP the predicted probability to cooperate is lower with noise , confirming Hypothesis 5.4. Thus, as expected, punishment, and

[5]No significant effectiveness or noise differences were found regarding the predicted probability to punish an observed cooperator. Predicted probabilities fall below 0.003 for each level of effectiveness and for each noise condition.

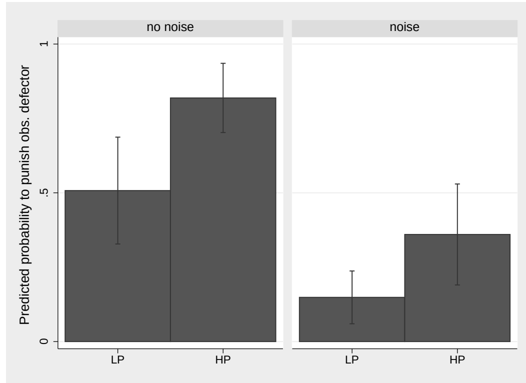


Figure 5.2a: Predicted probability of punishing an observed defector under each effectiveness and noise condition with 95% confidence intervals.

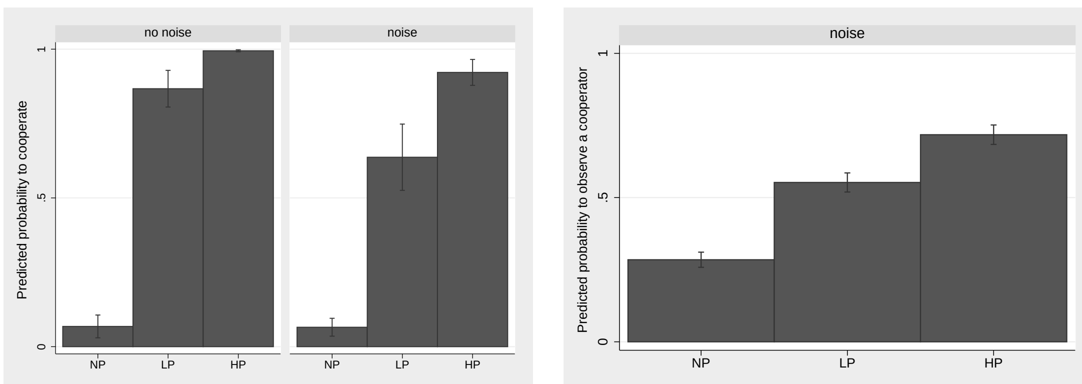


Figure 5.2b: Predicted probability of cooperation under each voting outcome with and without noise, and predicted probability to observe a cooperator with noise with 95% confidence intervals.

especially severe punishment, promotes cooperation, while noise reduces cooperation-enhancing effects of punishment institutions.

We noted that, with noise, if cooperation under LP and HP exceeds 50% while cooperation under NP falls below 50%, cooperation rates that subjects observe under LP and HP may fall below actual cooperation rates, while observed cooperation rates under NP are higher than in reality. The right-hand panel of Figure 5.2b shows the predicted probability to observe a cooperator with noise. These are five observations per subject for each PD, one for the observed contribution decision of each fellow group member. The predicted probability to observe a cooperator follows the same pattern as the predicted actual cooperation, though the difference between NP and the punishment institutions is smaller than in reality. Figure 5.2c shows predicted actual and perceived period earnings with noise. In both noise conditions, predicted actual earnings are higher under HP than under LP and higher under LP than NP. However, with noise, subjects are predicted to observe significantly lower earnings under HP than under NP

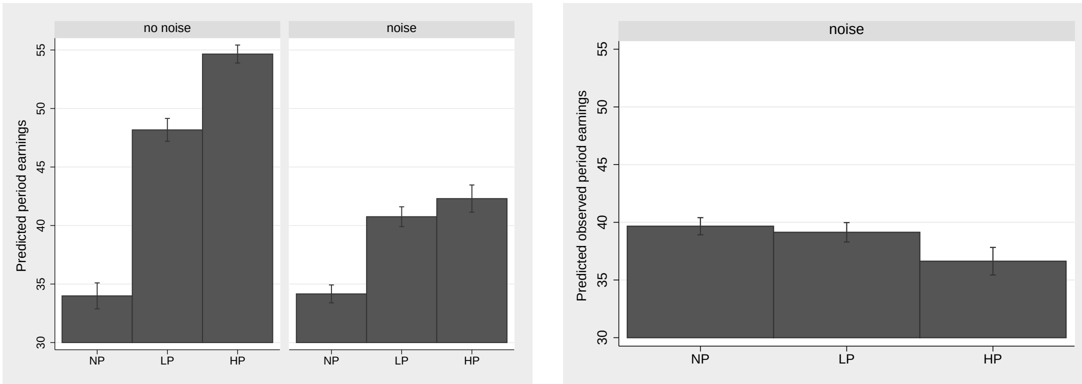


Figure 5.2c: Predicted period earnings for each voting outcome, and predicted observed period earnings with noise with 95% confidence intervals.

and LP. Thus, while the positive effect of interacting under a punishment institution on cooperation rates is partly detected by subjects in the noise condition, the subjects do not observe that this translates into higher earnings.

5.5.3 Votes

We now model the effect of noise and experience on the decision whether to vote in favor of implementing a punishment institution. Experienced cooperation rate was calculated as the average number of other group members that a subject observed as cooperators in all PDs with a voting stage that preceded the current vote, separated by whether the subject interacted under NP, LP, or HP. Experienced punishment of defection is measured as the average number of punishments that observed defectors, including the subject,[6] received in all preceding PDs in which a subject interacted under LP and under HP. The same method was used to measure experienced punishment of cooperation. If subjects in a certain period had never experienced a voting outcome or never observed cooperation or defection under LP or HP, they are assigned a value zero for the corresponding experience variable. Dichotomous variables indicating whether zero was assigned are included to account for this imputation. We control for the period and for whether the subject had voted in favor of implementing a punishment institution in the previous period. Table E.2 in Appendix E presents relevant descriptive statistics.

Model 1 of Table 5.3 shows that subjects are less inclined to vote in favor of implementing a punishment institution with noise than without noise, which is in line with Hypothesis 5.1. However, in Model 2, noise effects are just insignificant, indicating that such effects are partly explained by different experiences actors have in different noise

[6]The number of subjects receiving punishment for defection and for cooperation is insufficient to distinguish between punishments received by subjects and punishments received by other group members. The results presented below do not change when only punishments received by others are considered in the measures.

Table 5.3: Logistic regression on the decision whether to vote in favor of implementing a punishment institution in period t, with random effects at the subject level (6,084 PDs by 156 subjects).

	Hyp.	Exp. sign	Model 1 Coeff.	Model 1 S.e.	Model 2 Coeff.	Model 2 S.e.	Model 3 Coeff.	Model 3 S.e.
Noise	5.1	-	-1.786**	0.313	-0.732	0.419	-2.477	1.379
Experience (observed)								
Cooperation NP	5.5	−			-0.375*	0.181	-0.377	0.310
× noise	5.6	+					0.092	0.392
Cooperation LP	5.5	+			0.419**	0.148	0.278	0.267
× noise	5.6	−					0.269	0.338
Cooperation HP	5.5	+			0.552**	0.157	-1.012	0.690
× noise	5.6	−					1.644*	0.716
Av. pun. def. LP					-0.059	0.130	0.055	0.233
× noise	5.7	−					0.269	0.311
Av. pun. def. HP					0.064	0.110	0.364	0.275
× noise	5.7	−					-0.508	0.314
Av. pun. coop. LP	5.8	−			-0.177	0.375	0.583	0.589
× noise	5.9	+					-2.588**	0.839
Av. pun. coop. HP	5.8	−			-0.743*	0.326	0.223	0.666
× noise	5.9	+					-1.103	0.788
Controls								
Voted for pun. $t-1$			3.499**	0.116	3.277	0.119	3.440**	0.179
× noise							-0.413	0.237
Period			0.033**	0.005	0.009	0.007	0.015	0.011
× noise							0.005	0.016
Constant			-0.779**	0.241	-3.222**	0.844	2.645	3.457
σ_u			2.907**	0.555	3.731**	0.698	3.837**	0.730
Log Likelihood			-1480.390		-1439.326		-1418.964	

*Significant at the .05-level. **Significant at the .01-level (2-sided)
Controlled for dichotomous variables indicating whether a value of zero was assigned to legitimate missing values.

conditions. Furthermore, Model 2 shows that subjects are less inclined to vote in favor of implementing a punishment institution, the more cooperation they experienced under NP, and are more inclined to vote in favor of implementing a punishment institution, the more cooperation they experienced under LP and HP, confirming Hypothesis 5.5. The model further shows that subjects are less inclined to vote in favor of implementing a punishment institution after experiencing more punishment of cooperators under HP, confirming Hypothesis 5.8 for HP but not for LP.

Model 3 presents interactions between experience and noise. The effects of cooperation experienced under NP and LP on the likelihood to vote in favor of implementing a punishment institution are not significantly affected by noise. Moreover, the positive effect of cooperation experienced under HP is stronger with noise than without noise. Thus, there is no evidence that noise weakens the effect of experienced cooperation, implying that Hypothesis 5.6 is not confirmed. Contrary to our conjectures, more experiences of punishment aimed at defectors do not render actors more inclined to vote in favor of implementing a punishment institution without noise. Moreover, the effect of experienced punishment of defection is not significantly reduced with noise, suggesting that Hypothesis 5.7 is not confirmed. Additionally, the effect of experienced punishment of cooperation is not significantly lower with noise than without noise, implying that Hypothesis 5.9 is not confirmed. Instead, only with noise subjects are less inclined to vote in favor of implementing a punishment institution, the more punishment of cooperation they experienced under LP. It may be that too few perverse punishments occurred without noise (Table 5.2) to affect votes. Experienced punishment of cooperation under HP likewise renders actors less inclined to vote in favor of implementing a punishment institution under noise, but the effect is not significant, possibly because few groups interacted under HP.

In sum, subjects are, as expected, more inclined to vote in favor of implementing a punishment institution, the more cooperation they have experienced under punishment institutions relative to NP. However, we find no evidence that the effect of experiences differs between noise conditions in the hypothesized manner (Hypotheses 5.6, 5.7, and 5.9). Instead, observing that cooperators are punished deters subjects from voting in favor of implementing a punishment institution only with noise.

Table 5.4 shows the effect of noise and experience on the decision whether to vote for HP instead of LP. The same experiences are used as in Table 5.3, though experienced cooperation under NP is omitted. All votes are included regardless whether a subject voted in favor of implementing a punishment institution.

Model 4 in Table 5.4 shows that subjects are less inclined to vote for HP with noise, confirming Hypothesis 5.2. Model 5 shows that subjects are more inclined to vote for HP after experiencing lower cooperation rates under LP. However, subjects were not more inclined to vote for HP after experiencing higher cooperation rates under HP, only partly confirming Hypothesis 5.5. The model further shows that experienced punishment of defection under HP positively affects the likelihood to vote for HP. Actors are less inclined to vote for HP after experiencing more punishment of cooperation under HP, confirming Hypothesis 5.8.

Interactions between noise and experience are shown in Model 6. Supporting Hypothesis 5.6, the negative effect of experienced others' cooperation under LP on the likelihood to vote for HP are less significant with noise. However, experienced cooperation under HP does not interact with noise, thus only partly confirming Hypothesis 5.6. The effects of experienced punishment of defection and cooperation do not change

Table 5.4: Logistic regression on the decision whether to vote for HP instead of LP in period t, with random effects at the subject level (6,084 PDs by 156 subjects).

	Hyp.	Exp.	Model 4		Model 5		Model 6	
		sign	Coeff.	S.e.	Coeff.	S.e.	Coeff.	S.e.
Noise	5.2	–	-1.718**	0.299	-0.703*	0.327[1]	-3.357**	1.251
Experience (observed)								
Cooperation LP	5.5	–			-0.506**	0.138	-0.826**	0.227
× noise	5.6	+					0.684*	0.306[1]
Cooperation HP	5.5	+			0.079	0.147	-0.982	0.640
× noise	5.6	–					1.151	0.667
Av. pun. def. LP					0.118	0.127	0.230	0.207
× noise	5.7	–					0.008	0.298
Av. pun. def. HP					0.239*	0.100	0.009	0.233
× noise	5.7	–					0.313	0.282
Av. pun. coop. LP	5.8	–			-0.054	0.352	-0.242	0.525
× noise	5.9	+					-0.135	0.776
Av. pun. coop. HP	5.8	–			-1.403**	0.302	-1.611**	0.554[1]
× noise	5.9	+					0.419	0.680
Controls								
Voted for HP $t-1$			3.579**	0.109	3.421**	0.112	3.307**	0.157
× noise							0.163	0.219
Period			0.017**	0.004	-0.013	0.007	-0.005	0.010
× noise							0.000	0.015
Constant			-1.483**	0.223	0.135	0.772	6.776*	3.144
σ_u			2.595**	0.498	2.227	0.449	2.207	0.457
Log Likelihood			-1612.811		-1559.525		-1552.497	

*Significant at the .05-level. **Significant at the .01-level (2-sided)
[1]Not robust when controlling for session-fixed effects.
Controlled for dichotomous variables indicating whether a value of zero was assigned to legitimate missing values.

with noise, implying that Hypotheses 5.7 and 5.9 are not confirmed. Instead, experiencing punishment of cooperation under HP renders subjects less inclined to vote for HP regardless of noise. Note that this effect is not robust when session fixed effects are considered. This is attributable to the fact that one session without noise has a particularly high rate of perverse punishment, which, in line with our theory, generates more votes for LP than in the other sessions.

In sum, it is partly confirmed that subjects are more inclined to vote for HP after experiencing higher cooperation rates under HP, and lower cooperation rates under LP. However, interactions between noise and experience do not consistently affect voting decisions in the hypothesized manner. Instead, experiencing punishment of cooperators affects subsequent votes regardless of noise.

5.6 Conclusion and discussion

In this chapter, we study endogenous implementation of peer punishment institutions and punishment effectiveness, in cooperation problems where actors accurately observe the behavior of their peers, and in cooperation problems with a 20% probability of cooperation being observed as defection and vice versa. In a laboratory experiment, subjects decide by majority vote whether to interact without a punishment institution (NP) or implement a punishment institution. If the latter was chosen, subjects selected low (LP) or high (HP) punishment effectiveness. As expected, subjects are less inclined to vote in favor of implementing a punishment institution, and are less inclined to vote for HP rather than LP, with noise than without noise (Hypotheses 5.1 and 5.2). Without noise, after some initial reluctance, the vast majority of groups interact under HP, while with noise NP remains most popular. In both noise conditions, as expected, cooperation rates are higher under HP than under LP and higher under LP than under NP (Hypothesis 5.3). Moreover, cooperation rates under punishment institutions are lower with noise (Hypothesis 5.4). Interestingly, in both noise conditions, earnings are higher under LP and HP than under NP, but because of the presence of noise, subjects in the noise condition observe lower earnings under HP than under LP and NP.

These results complement prevalent findings that actors choose the implementation of peer punishment institutions after some experience with the decision situation, in environments characterized by perfect information (e.g., Ertan et al., 2009; Gürerk, 2013; Gürerk et al., 2006, 2009; Markussen et al., 2014; Rockenbach and Milinski, 2006). We add that actors support high peer punishment effectiveness, but less so the more punishment is used perversely (Hypothesis 5.8). However, this increasing preference for punishment institutions and for high punishment effectiveness does not extend to noisy environments. With noise, lower observed earnings under LP than under NP may have discouraged subjects from implementing punishment institutions. Moreover, subjects are reluctant to opt for (more effective) punishment institutions when they observe cooperators receiving punishment.

Our noise results limit previous findings on endogenous institution formation, as noise is often present in cooperation problems outside of laboratory settings (Bereby-Meyer, 2012). This raises questions regarding whether other (more realistic) environments inhibit actors' willingness to interact under a peer punishment institution. For example, actors may be less inclined to opt for a punishment institution that allows for counter-punishment (Denant-Boemont et al., 2007; Nikiforakis, 2008; Nikiforakis and Engelmann, 2011) or may be less inclined to implement a punishment institution in populations where punishment is often used perversely (Herrmann et al., 2008).

In addition to noise effects on voting outcomes and cooperation rates, our analysis pinpoints determinants of individual votes. As expected, actors are more inclined to vote for a certain option, the more cooperation they have thus far experienced under that option relative to alternative options (Hypothesis 5.5). However, we cannot confirm that actors account for noise in their experiences in the hypothesized manner. When actors aim to vote for the option in which they receive the highest earnings, and realize that their experiences are noisy, they may react differently to experiencing certain behavior of fellow group members under noise (Hypotheses 5.6, 5.7, and 5.9). For example, with noise, more cooperation is required under punishment institutions than without noise for earnings under punishment institutions to exceed those generated under NP. Therefore, with noise observing cooperation under LP and HP should not encourage subjects to vote for punishment institutions as much as without noise. Such conjectures are not systematically corroborated by the data. Instead, certain experiences affect subsequent votes regardless of noise. This may suggest that while subjects react to previously experienced positive or negative incentives, they fail to consider noise. Future research may assess how actors in cooperation problems aim to maximize their expected earnings when presented with noisy information.

Chapter 6

Summary of findings and suggestions for future research

6.1 Three issues concerning cooperation problems with peer sanctioning institutions

This thesis opened with a description of a scenario in which a group of co-workers share a departmental coffee corner. The co-workers jointly benefit from keeping their coffee corner clean, but each has an incentive to free ride on the cleaning efforts of others. Contributions to cleaning the coffee corner may be encouraged if each co-worker has the opportunity to reward or punish others for their cleaning behaviors. The situation in which actors benefit from mutual cooperation towards a collective endeavor, but wherein each actor has an incentive to keep resources for private use, represents a cooperation problem. The opportunity for the actors to punish (negative sanction) or reward (positive sanction) one another represents a peer sanctioning institution.

This thesis focuses on three issues that concern cooperation problems with peer sanctioning institutions. First, institutions under which each actor individually decides on sanctioning are compared to institutions under which sanctioning decisions are made by the group or by a subgroup as a collective. The effect of individual versus collective sanctioning decisions on cooperation as well as on actors' perceptions of the fairness of the sanctions that are implemented are considered. Second, the effect of peer punishment institutions on cooperation is studied in cooperation problems wherein actors may inaccurately observe others' contributions to a collective endeavor. Third, it is determined whether actors themselves implement a peer punishment institution when given a choice to do so, and if so, whether they implement one in which allocated punishments are relatively severe or one in which allocated punishments are relatively mild.

Chapters 2 through 5 form the empirical and theoretical basis of this thesis, in which specific research questions regarding one or several of the three topics are addressed.

The theoretical approach employed in these chapters is broadly consistent with behavioral game theory (Camerer, 2003). That is, several empirically validated contribution and punishment behaviors are identified in previous studies. The theoretical framework is based on implications of assuming that actors display these behaviors in the scenario examined. Three laboratory experiments wherein cooperation problems are represented by series of one-shot Public Goods Games (PGGs) or n-player Prisoner's Dilemmas (PDs) are employed for empirical testing.

This final chapter reconsiders the three main issues explored in this thesis. Major findings revealed in the relevant empirical chapters regarding each of the three topics are summarized, and emerging directions for future research on each topic are identified. The chapter closes with general conclusions on cooperation under peer sanctioning institutions and with recommendations for the co-workers who share a coffee corner space.

6.2 Terminology

In what follows, resources that an individual allocates to a collective endeavor are referred to as 'contributions.' Contributing less than other actors involved in a cooperation problem is referred to as 'free riding,' and neglecting to contribute is referred to as 'defection.' Contributing the highest possible amount is referred to as 'cooperation.' 'Group contributions' refer to the aggregate contributions of group members under a certain institution.

'Sanctions' include both negative and positive sanctions, i.e., punishments and rewards. Punishments directed at free riders or defectors and rewards directed at above-average contributors or cooperators are referred to as 'pro-social,' as they should promote the group interest. Punishments directed at above-average contributors or cooperators and rewards directed at free riders or defectors are referred to as 'perverse,' as they oppose the group interest. Terminology relating specifically to one of the three topics examined is introduced in the respective subsections.

6.3 Individual and collective sanctioning decisions

The vast majority of previous studies focus on peer sanctioning institutions under which each actor involved in the cooperation problem individually decides on punishing or rewarding fellow group members. Henceforth, this is referred to as an individual decision rule (IDR). In Chapters 2, 3, and 4, peer sanctioning institutions with an IDR are compared to peer sanctioning institutions under which collective decision rules (CDRs) are used to determine whom to sanction. Under a CDR, sanctions are only implemented if a certain proportion of actors agrees to sanction a certain member of their group. Sanctions not achieving the required level of agreement are ruled out. It is anticipated that a decision rule more effectively enforces cooperation as more pro-social sanctions are implemented and as more perverse sanctions are ruled out. Moreover, it is hypothesized that actors perceive personally received sanctions as more fair when more group members are required to agree on sanctioning decisions and that actors find sanctions received by other group members to be more fair when they are afforded more freedom to decide whom to sanction.

Chapters 2 and 4 consider cooperation under peer punishment institutions with individual and collective decision rules. In both chapters, under every decision rule, it is found that numerous actors attempt to allocate pro-social punishment, while relatively few actors attempt to punish perversely. All proposed punishments are implemented under IDRs. As pro-social punishment is common while perverse punishment is relatively scarce, in Chapters 2 and 4, group contributions under the IDR are higher than in cooperation problems without a sanctioning institution. This finding complements the results of numerous previous studies (e.g., Balliet et al., 2011; Fehr and Gächter, 2000, 2002; Ostrom, 1990). The IDR employed in Chapter 2 more effectively maintains high group contributions than the IDR employed in Chapter 4. This may be attributable to the fact that perverse punishment is more common and more heavily affects cooperation in Chapter 4 than in Chapter 2. Alternatively, this may result from the fact that pro-social punishment under the IDR presented in Chapter 4 is, on average, too weak to offset the payoff advantage of defecting.

In addition to IDRs, Chapters 2 and 4 consider group contributions made under CDRs wherein punishments are implemented if at least a majority of remaining group members agrees to punish a prospective recipient. CDRs for which the required consensus level is lower than the majority are also examined. These CDRs successfully maintain group contributions at a high level. Overall, under CDRs that require a majority consensus or less, some attempted pro-social punishments do not receive sufficient consensus for implementation. Still, on average, free riders or defectors are punished at a level similar or higher to that found under IDRs. Moreover, virtually no perverse punishments are implemented under CDRs that require a majority consensus or less. These results are in line with the results of a previous study by Casari and Luini (2009). In Chapter 4, CDRs that require a majority consensus or less generate higher group contributions than those generated under the IDR; in Chapter 2 a CDR that requires majority consensus produces lower group contributions than the IDR. Although the two chapters are difficult to compare due to differences in experimental designs employed, these results may be attributed to the fact that collective agreement required for punishment to be implemented are less vigorous under the CDRs employed in Chapter 4 than those employed in Chapter 2. Additionally, while in Chapter 4, actors only decided whether to cooperate or defect, in Chapter 2, actors also determined the level of their contribution. The clear distinction between cooperators and defectors may have facilitated collective agreement in Chapter 4.

Chapter 2 also considers a CDR in which unanimous agreement between group members (except for the prospective recipient) is required to implement punishment. Under this CDR, numerous attempted pro-social punishments do not receive sufficient agreement for implementation. This results in considerably less severe punishment of free riders than under the other decision rules. Accordingly, the CDR that requires unanimous agreement maintains group contributions above those of cooperation problems without a sanctioning institution, but below those of other decision rules.

Finally, Chapter 2 compares group contributions under different decision rules for implementing reward. Reward institutions always generate lower group contributions than punishment institutions with the same level of required agreement. Still, in the case of rewards, it is found that numerous actors attempt to allocate pro-social rewards, while few attempt to reward perversely. Under the IDR, wherein every allocated reward is implemented, group contributions are higher than those in cooperation problems without a sanctioning institution. This complements previous findings (e.g., Balliet et al., 2011; Sefton et al., 2007). Under a CDR that requires majority consent, some pro-social rewards are not implemented because the required consensus level is not achieved. Nevertheless, under majority agreement above-average contributors are rewarded by an amount that is on average not much lower than under the IDR. Group contributions made under a CDR that requires majority consent remain higher than those made under cooperation problems without a sanctioning institution, but lower than those made under the IDR. Under a CDR that requires unanimous consent, numerous rewards are not implemented because the required consensus level is not reached. As a result, this CDR produces lower group contributions than the other decision rules and does not increase cooperation levels to above those of a cooperation problem without a sanctioning institution.

In sum, with respect to both punishments and rewards, it is found that CDRs that require majority consensus or less are suitable for increasing group contributions and at the same time rule out perverse sanctions. Consensus requirements that exceed majority levels rule out too many pro-social sanctions to maintain high contributions. A traditional IDR is also found to promote cooperation in the present studies and remains as the most successful decision rule in Chapter 2.

Future experimental research on collective sanctioning decisions may consider the fact that the CDRs employed in this thesis are a stylized representation of how collective sanctioning decisions might proceed outside of laboratory settings. This thesis focused on levels of consensus required for implementing a sanction and on comparisons between reward and punishment. As a next step, other aspects of the process through which collective sanctioning decisions may proceed in cooperation problems outside of laboratory settings may be modeled in further detail. For example, in many cooperation problems, actors may be presented with opportunities to communicate with one another while making peer-sanctioning decisions (Bochet et al., 2006; Ostrom et al., 1992). Alternatively, before making their contributions to the collective good, actors might have the opportunity to agree on a certain contribution level that group members should adhere to in order to obtain rewards or avoid receiving punishment (cf. Ertan et al., 2009; Fehr and Williams, 2013; Kamei et al., 2011; Putterman et al., 2011. If such opportunities are available in addition to sanctioning institutions with a CDR, actors might be better able to coordinate sanctioning certain group members such that cooperation is better maintained through CDRs than in the present setting.

Counter-punishment opportunities, that is, punishments directed towards actors for their previous punishment decisions, serve as another realistic setting in which peer sanctioning institutions with a CDR might be particularly suitable for maintaining cooperation. Under an IDR, it is found that actors use counter-punishment strategies to avenge previously received punishments, and this is found to render punishment institutions with an IDR ineffective at promoting cooperation and earnings (Denant-Boemont et al., 2007; Nikiforakis, 2008; Nikiforakis and Engelmann, 2011). Still, everyday cooperation problems may often include counter-punishment opportunities, highlighting the importance of identifying punishment institutions that are more suitable for maintaining cooperation when counter-punishment is possible (Chaudhuri, 2011; Guala, 2012; Nikiforakis, 2014). Employing a CDR diffuses the responsibility for punishing someone across several actors (Duch et al., 2015), rendering revenge through counter-punishment less feasible. Thus, CDRs may be more successful than IDRs in maintaining cooperation under threats of counter-punishment.

Chapter 3 considers actors' perceived fairness of peer sanctions that are allocated through various decision rules. It cannot be confirmed that the decision rule through which sanctions are allocated systematically affects actors' perceived fairness of the sanctions that they receive themselves. However, it is found that actors perceive sanctions that other group members receive as less fair as more sanctions that the focal actor attempts to allocate are not implemented under CDRs. This suggests that actors prefer decision-making procedures in which they hold decision-making power in sanctioning certain actors. Moreover, actors evaluate personally received punishments as less fair if they had received more punishments and personally received rewards as more fair if they had received more rewards, suggesting that actors are biased towards their own outcomes in their fairness evaluations.

The finding that procedures through which sanctioning decisions are made (i.e., the decision rules) do not significantly affect fairness perceptions of personally received sanctions contradicts the findings of several previous studies. Typically, procedures through which outcomes are arrived at are found to be a strong determinant of fairness perceptions (e.g., Dolan et al., 2007; Messick and Sentis, 1979; Schroeder et al., 2003). One possible explanation for this lack of evidence for a procedural fairness effect on personally received sanctions may be that the CDRs failed to capture important factors that render collective sanctioning decisions fairer than individual decisions. As noted above, in numerous cooperation problems, actors may be afforded opportunities to exchange information on appropriate sanctioning decisions. When such opportunities exist, actors may be better able to anticipate the contribution level that would solicit sanctioning. This is especially true in the case of CDRs, as under an IDR, some fellow group members may not adhere to agreed-upon sanctioning behaviors. Additionally, when communication is possible, actors may better understand why some of the sanctions that they attempt to allocate may not receive sufficient collective agreement for implementation. In this way, more elaborate opportunities for communication might enhance the perceived fairness of sanctions allocated through CDRs relative to those allocated through IDRs (Guala, 2012).

6.4 Noisy information on others' contributions

The second topic of this thesis addresses the effect of peer punishment institutions on group contributions in cooperation problems wherein one or several fellow group members might wrongly observe an actor's contribution decision, i.e., observe cooperation as defection or defection as cooperation. This possibility of wrongly observing others' behaviors is referred to as 'noise.' With noise, actors who wish to allocate pro-social punishment may punish an actual cooperator whom they observe as a defector or neglect to punish an actual defector whom they observe as a cooperator.[1] An advantage of CDRs over IDRs with noise is that when more actors are required to agree on punishment decisions, it is less likely that punishments are implemented that are proposed after observing another actor's cooperation as a defection. A disadvantage of CDRs compared to IDRs with noise is that the more actors are required to agree, the more difficult it might be to reach collective agreement on punishing an actual defector, as potential punishers might observe a defector as a cooperator. Group contributions are expected to be highest under a CDR that results in the most favorable ratio of the expected advantage and disadvantage of CDRs.

Chapter 4 considers group contributions made under peer punishment institutions with different decision rules in cooperation problems with noise. It is found that with noise, a punishment institution with an IDR increases group contributions relative to cooperation problems without a sanctioning institution. This cooperation-enhancing effect of the IDR diminishes over time. Subjects waste resources on punishing actual cooperators whom they observe as defectors such that earnings under the IDR are lower than when no sanctioning is possible. These results complement those of previous studies (e.g., Ambrus and Greiner, 2012; Grechenig et al., 2010). Contrary to expectations, with noise, two CDRs that require majority or lower consensus cannot increase group contributions above cooperation problems without a sanctioning institution. A large proportion of punishments directed at actual cooperators are ruled out under these

[1]Misguided *perverse* punishment decisions may occur as well. However, these are less relevant in explaining cooperation with noise, as perverse punishment is typically less common than pro-social punishment such that there are very few cases in which 'misguided' perverse punishment is assigned to defectors.

CDRs, but at the same time, numerous punishments directed at actual defectors are not implemented. Interestingly, the same CDRs are particularly successful at promoting cooperation without noise.

Chapter 5 shows that peer punishment institutions with an IDR that actors voluntarily implement in their group do increase cooperation and earnings with noise relative to when sanctioning is not possible, and especially when actors implement a punishment institution wherein allocated punishments are severe. Still, also voluntarily implemented peer punishment institutions with an IDR are not as successful at maintaining cooperation with noise as without noise.

Noise may be present in numerous cooperation problems outside of laboratory settings (Bereby-Meyer, 2012). This thesis shows that the possibility to observe cooperation as defection or vice versa clearly disrupts the ability of peer punishment institutions, both with an IDR and with a CDR, to maintain high group contributions and earnings. However, noise can operate in different ways. For example, in some settings, cooperation may more frequently be observed as defection than the other way around, or inaccurate observations of a contribution decision may not be independent across group members. Probabilities of inaccurate observations might vary across settings as well. This raises questions regarding whether other types of noise are less detrimental to the performance of peer punishment institutions with different decision rules. Previous research employing lower noise levels than those used in this thesis likewise find that noise reduces earnings under peer punishment institutions with an IDR (Grechenig et al., 2010). Additionally, a study wherein cooperation is either observed correctly or incorrectly by all group members, and in which defections are always accurately observed, reports that noise has detrimental effects on the performance of peer punishment institutions (Ambrus and Greiner, 2012). However, more detailed empirical evidence is required to better understand how different peer punishment institutions function under different types of noise.

To identify which types of (sanctioning) institutions maintain cooperation with noise, it may be useful to first determine how and why noise affects contribution and punishment behaviors that are empirically validated in cooperation problems without noise. Thus far, mixed results have been found regarding the effect of noise on these behaviors (e.g., Fischer et al., 2013; Grechenig et al., 2010; Chapter 4). Identifying mechanisms that drive potential noise effects on behaviors in cooperation problems may assist in identifying suitable (sanctioning) institutions for noisy environments.

6.5 Endogenous peer punishment institutions

The two aforementioned topics exclusively consider the effects of different peer sanctioning institutions on actor behaviors in cooperation problems and on perceptions of implemented sanctions. This final topic addresses whether actors foresee these effects and voluntarily implement a peer punishment institution in their group when given a choice whether or not to do so. In Chapter 5, a setting is considered in which actors, before engaging in a cooperation problem, indicate whether they wish to implement a peer punishment institution with an IDR, and if so, whether they wish to allow for relatively severe or relatively mild punishments. The option that the majority prefers is implemented. Cooperation problems with and without noise are considered. It is hypothesized that with noise, actors are less likely to implement a punishment institution, and are less inclined to prefer severe punishments, than without noise.

Findings presented in Chapter 5 indicate that without noise, after some experience with the cooperation problem, almost all choose to implement a peer punishment institution. Group contributions and earnings are much higher in groups that implement a punishment institution than in groups that choose to interact without a punishment institution. These results complement those of several previous studies (e.g., Ertan et al., 2009; Gürerk, 2013; Gürerk et al., 2006, 2009; Markussen et al., 2014; Rockenbach and Milinski, 2006). Moreover, most groups implement institutions under which punishments are severe, though observing perverse punishment predisposes actors toward institutions with mild punishments.

Chapter 5 shows that findings on endogenous institution formation do not generalize to noisy environments. With noise, the majority of groups choose to interact without a punishment institution, and if groups do implement a punishment institution, it is most often an institution under which punishments are relatively mild. Group contributions and earnings are higher in groups that implement a peer punishment institution than in groups that do not. However, due to noise, subjects in the experiment observe only slightly higher group contributions and lower earnings under a punishment institution. Moreover, with noise, actors are discouraged more from implementing a peer punishment institution as they witness more group members whom they observe as cooperators being punished.

The finding that actors support severe sanctions to ensure public good provision so long as such sanctions are used to punish defectors may be exploited in future studies that attempt to identify sustainable endogenous institutions that successfully promote cooperation. However, questions remain regarding the extent to which this finding generalizes beyond peer punishment institutions. Previous studies show that actors likewise prefer severe over mild punishments if they can choose whether to implement a mechanism that automatically punishes free riders (Kamei et al., 2011; Kingsley and Brown, 2015). However, aside from this mechanism, punishment may also be allocated by an external authority (e.g., a police officer) who may follow a less accurate and predictable punishment strategy than an automatic mechanism. External authorities may be better able to promote cooperation through severe punishment than via mild punishment, but actors may be more reluctant to permit the use of (severe) punishment by external authorities.

The strong effect of noise on endogenous institution formation again highlights the importance of testing the robustness of established experimental findings in cooperation problems that reflect crucial aspects of daily situations (cf. Nikiforakis, 2014). This raises questions concerning how the endogenous implementation of peer punishment institutions and punishment severity proceeds in other environments that may be less conducive for cooperation to emerge under peer punishment institutions. For example, it may be interesting to examine the voluntary implementation of peer punishment institutions with the option to use counter-punishment. Likewise, the possibility for counter-punishment may render actors less likely to implement a peer punishment institution or to prefer severe punishments relative to the standard IDR. As another example, actors may be less likely to implement a peer punishment institution and severe punishments in populations where numerous actors punish perversely (Herrmann et al., 2008).

To develop a more complete understanding of actors' preference for different institutions, it is important to look beyond the realm of peer punishment. Settings may be considered in which actors can choose between different types of institutions (Rockenbach and Wolff, 2009). Aside from peer punishment mechanisms, actors may prefer institutions with built-in mechanisms that disable perverse punishment, or those that successfully maintain cooperation without resorting to punishment threats (e.g., Markussen et al., 2014; Traulsen et al., 2012).

6.6 The coffee corner and peer sanctioning institutions

Now that the main findings regarding peer-sanctioning institutions in this thesis are summarized, what advice could be given to co-workers who aim to maintain a clean coffee corner? More generally, how can cooperation be maintained through peer sanctions? The results of this thesis are derived from series of one-shot cooperation problems in laboratory experiments that differ in numerous respects from everyday cooperation problems such as the coffee corner scenario. While this constitutes a limitation, some general conclusions regarding means of promoting cooperation under peer sanctioning institutions can be presented.

Suppose that all department members always accurately observe each co-worker's cleaning behaviors. In this case, as long as most co-workers encourage cleaning, a clean coffee corner can be maintained if all department members individually determine punishing or rewarding the cleaning, or lack thereof, of their co-workers. Punishments promote cleaning more effectively than rewards. Department members may voluntarily submit to the scenario wherein all can punish individually and may even prefer to allow severe punishments, that in turn successfully promote cleaning. Still, when everyone is allowed to sanction individually, at least some department members may discourage cleaning by punishing or rewarding co-workers who do not 'deserve' to be sanctioned. To address this issue, an arrangement could be established wherein department members may only sanction someone if they can find other co-workers who agree that the person should be sanctioned. While sanctions that co-workers agree upon might result in a clean coffee corner, they are not always as successful as individual sanctioning decisions. Collective sanctioning decisions should not require too much agreement, as this renders sanctioning inefficient. Additionally, collective agreement arrangements may cause co-workers who cannot find others who agree with their sanctions to perceive sanctions that are allocated as unfair.

More generally, for social dilemmas in which actors are accurately informed of one another's behavior, this thesis finds that cooperation can be maintained through traditional peer sanctioning institutions under which each actor can individually decide on sanctioning. Punishments more effectively promote cooperation than rewards. Peer punishment institutions do not need to be imposed upon actors. Rather, when given a choice whether to implement a peer punishment institution with individual punishment decisions in their group, most actors prefer to do so and to enable severe punishments. Such voluntarily implemented punishment institutions with individual punishment decisions are highly successful at promoting cooperation. Collective sanctioning decisions rule out almost all perverse sanctions and are successful in maintaining cooperation so long as not too much agreement is required to implement sufficient pro-social sanctions. Still, collective sanctioning decisions do not always promote cooperation to the same degree as individual decisions. Actors perceive it as fair when they receive rewards and as unfair when they receive punishments or when punishments that they propose do not receive sufficient collective agreement for implementation.

The situation for the co-workers might be very different when they cannot always accurately observe one another's cleaning behaviors. Strictly assigning punishment when co-workers agree that someone should be punished does not promote cooperation in this setting, as it is difficult to collectively determine whether someone had actually cleaned and whether he or she should be punished such that allocating sufficient punishment to maintain a clean coffee corner becomes infeasible. The coffee corner may be kept clean when every member of the department individually decides on punishing co-workers, especially when punishments are relatively severe. However, this arrangement should be exogenously imposed upon the co-workers. The voluntary implementation of punishment arrangements is unlikely, as the department members dislike the fact that those who clean run a high risk of being punished 'mistakenly' by individuals who inaccurately observe their behavior. Additionally, as information available to each co-worker may be inaccurate, department members may not realize that punishments make their co-workers more inclined to clean the coffee corner.

Generally speaking, in social dilemmas where actors may not accurately observe one another's behaviors, this thesis finds that implementing only those peer punishments that actors collectively agree upon is insufficient to maintain cooperation. Heterogeneous information among potential punishers on an actor's contribution renders collective agreement infeasible. Cooperation is promoted somewhat when every actor can individually decide whom to punish, especially when punishments are severe and when the institution is implemented voluntarily by the group. However, most groups might not implement a punishment institution with individual punishment decisions, as actors are discouraged from implementing a punishment institution when they observe other cooperators being punished and because actors may not realize that their earnings increase under a punishment institution when available information is noisy. Therefore, with noise, a peer punishment institution with an IDR might be exogenously imposed upon actors.

Appendix A

Instructions for the experiment of Chapters 2 and 3

This is the English version of the instructions for the experiment that was used for Chapters 2 and 3. Subjects in the experiment could choose for either a Dutch or an English version of these instructions. The instructions for part 3 and 4 were handed out when the preceding parts were finished. In this version of the instructions, the majority condition was employed in part 3 and 4 with first punishment, then reward. Instructions for the individual and unanimity conditions, or for first reward then punishment are very similar.

- Instructions -

Welcome

Please read the following instructions carefully. These instructions are the same for all participants. The instructions state everything you need to know in order to participate in the experiment. If you have any questions, please raise your hand. One of the experimenters will approach you and answer your question.

You can earn money by means of earning points during the experiment. The number of points that you earn depends on your own choices and the choices of other participants. At the end of the experiment, the total number of points that you earn during the experiment will be exchanged at an exchange rate of:

60 points = 1 Euro

The money you earn will be paid out in cash at the end of the experiment. Other participants will not see how much you have earned. During the experiment you are not allowed to communicate with other participants. Please turn off your mobile phone and put it in your bag. You may only use the functions on the screen that are necessary to carry out the experiment. Thank you very much.

The Decision Situation

First, we introduce the decision situation in which you will interact. You will learn about the procedure of the experiment later.

You are a member of a group of **four participants**. You and the three other members of your group are endowed with **20 points** each. You must all choose how many of these points to keep in your **private account** and how many points to contribute to a **group account**. You can choose any number of points to contribute to the group account, from zero to 20 points. The remaining points go into your private account.

When all group members have made their contribution to the group account, the total number of points in the group account is **multiplied by 1.6**. The multiplied amount will then be **equally** distributed among all four group members. Your earnings from the decision situation are your share of the points from the group account plus the amount of points you kept in your private account. This holds for each participant in your group.

Table 1, which you have received on a separate page, shows how many points you will earn given different levels of your own contribution to the group account and of the total contribution of others in your group. For example, suppose you contribute 5 points to the group account and the others contribute 14, 11 and 0 points respectively. The total contribution of others is thus 14+11+0 = 25 points. Table 1 shows that you then earn 27 points. You can verify this yourself: the total contribution to the group account is 5+14+11+0 = 30 points. Multiplied by 1.6 is 48 points in the group account, divided by four group members is 12 points each. Added to the 20–5 = 15 points of your endowment you kept in your private account, this results in 27 points earned. Note that Table 1 also applies to the other group members; for the group member who contributed

0 points in the example the total contribution of others (including you) was 5+14+11 = 30 points. Table 1 shows that for an own contribution of 0 points and a total other's contribution of 30 points (s)he earned 32 points.

The Experiment

This experiment consists of **four parts**. Each part includes the decision situation described above. The first part comprises two types of decisions, the other three are series of 10 decision rounds. These instructions are for part 1 and part 2; the instructions for part 3 and for part 4 will follow later.

Before you continue reading the instructions, please choose the language in which you want to do the experiment and click "OK" on your computer screen. You will proceed to a number of control questions. Answer the questions to check whether you have understood the decision situation. Your answers to the control questions have no influence on your payment.

Table 1: Your earnings at different contribution levels (note that the actual contributions are not necessarily a multiple of five)

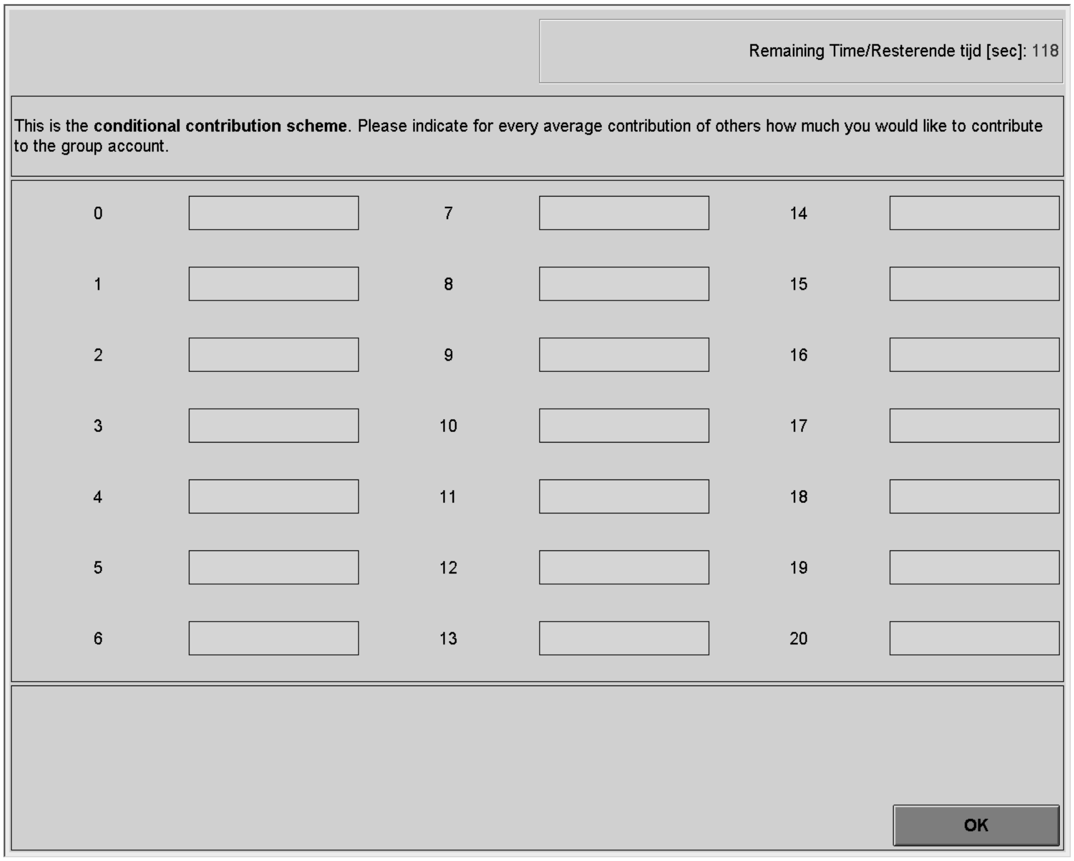
Your contribution to the group account	Total contribution to the group account of others in your group												
	0	**5**	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**
0	20	22	24	26	28	30	32	34	36	38	40	42	44
5	17	19	21	23	25	27	29	31	33	35	37	39	41
10	14	16	18	20	22	24	26	28	30	32	34	36	38
15	11	13	15	17	19	21	23	25	27	29	31	33	35
20	8	10	12	14	16	18	20	22	24	26	28	30	32

Instructions part 1

In the first part of the experiment you are asked to make two types of decisions. First, you decide how much of your endowment of 20 points you would like to contribute to the group account given that you participate **once** in the decision situation described above with three unknown other participants. This is your **unconditional contribution**.

You will see a screen like the one shown above. You type your unconditional contribution to the group account in the box as a number between 0 and 20. The remaining points will be transferred to your private account. Once you have entered a number, click the “OK” button to confirm your decision.

Second, you fill in a **conditional contribution scheme**. Given that you participate once in the decision situation described above you decide how much of your endowment you would like to contribute to the group account **for each possible average contribution of the three others** in your group. Note that this is not the same as the *total* contribution of others shown in Table 1. The three others can contribute on average between 0 and 20 points to the group account. For example, given that the three others contribute 5, 10 and 15 points, respectively, this results in an average of (5+10+15)/3 = 10 points contributed to the group account. Given that the three others all contribute their full endowment of 20 points this results in (20+20+20)/3 = 20 points on average contributed to the group account. You thus make 21 decisions: given that others contribute on average 0, 1, 2, 3, … 20 points, you decide how much of your endowment of 20 points you would contribute to the group account. You will see the following screen:



For every average contribution of others you type the number of points you want to contribute to the group account in the corresponding box.

In this part of the experiment you are randomly matched to three other participants. Your allocation to the group account is determined as follows. Of three randomly chosen players in your group the allocation to the group account is the unconditional allocation. These three unconditional allocations are averaged, and the allocation of the fourth member is the conditional allocation corresponding to the average unconditional allocation of the others. It might thus be that it is your unconditional contribution that is used to calculate your payoff, but it might also be your conditional contribution. Based on the four allocations your payoff is calculated as described above. After this round you will be informed about your allocation, the amount of points in the group account and the number of points that you have earned.

Instructions part 2

The remaining parts of the experiment consist of **10 decision rounds**. In part 2, you are again randomly matched to three other participants. In each round, you receive an endowment of 20 points and determine the number of points you want to contribute to the group account. This is the only decision you have to make in every round, and will determine your earnings.

You will see a screen like the one shown above. You type the number of points you want to contribute to the group account in the box as a number. The amount allocated to your private account is the part of your endowment of 20 points that is left. Once you have entered a number, click the "OK" button to confirm your decision.

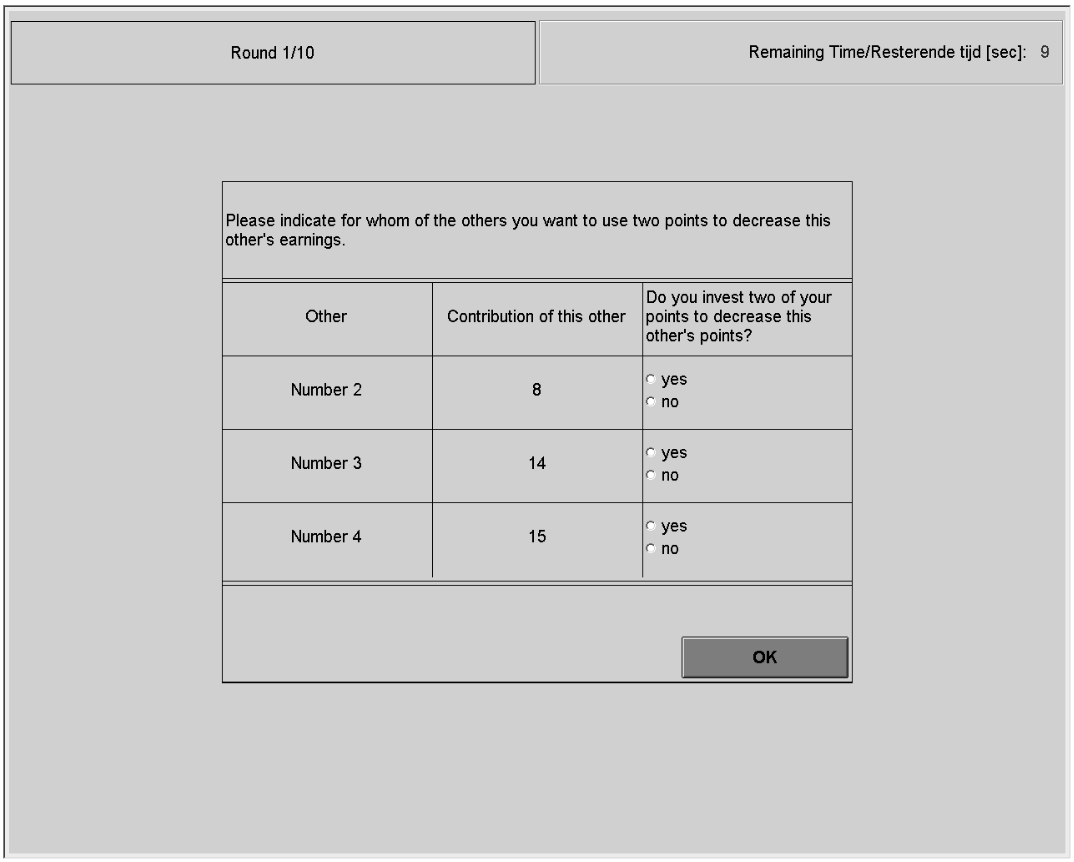
After all participants have made their decisions for the round, the computer will calculate the payoffs and tabulate the results. You will be informed about the contribution every member of your group has made to the group account and about the number of points that you have earned.

This procedure is repeated for 10 rounds. The participants with whom you are together in your group will **change randomly after every round**. You will not be told each other's identities at any time during or after the experiment.

When you have finished the first and second part of the experiment you will receive new instructions for part 3 and finally for part 4.

Instructions part 3

These instructions are for the third part of the experiment, which is again a series of 10 decision rounds. The decision situation as explained in the first instructions remains unchanged. You will thus in every round be asked how much of your endowment of 20 points you want to contribute to the group account. However, every round now has a second stage that follows after every member of your group has decided how much they want to contribute to the group account. In this second stage all group members have a chance to **decrease** the earnings of the others. You will see the following screen:



In the first column you see the contribution every other member of your group (“Number 2”, “Number 3” and “Number 4”) has made to the group account. In the second column you decide whether you want to decrease the earnings of each other group member by clicking “yes” or “no”. You make this decision for each group member independently. If you decide that you want to decrease the earnings of another member, this will be done only **if at least a majority of participants in your group wants to decrease this person’s earnings**. A majority is reached when two out of three group members indicate that they want to decrease a fourth member’s earnings. Thus, if you want to decrease someone’s earnings this is only accomplished if at least one other member of your group also wants to decrease the earnings of this person.

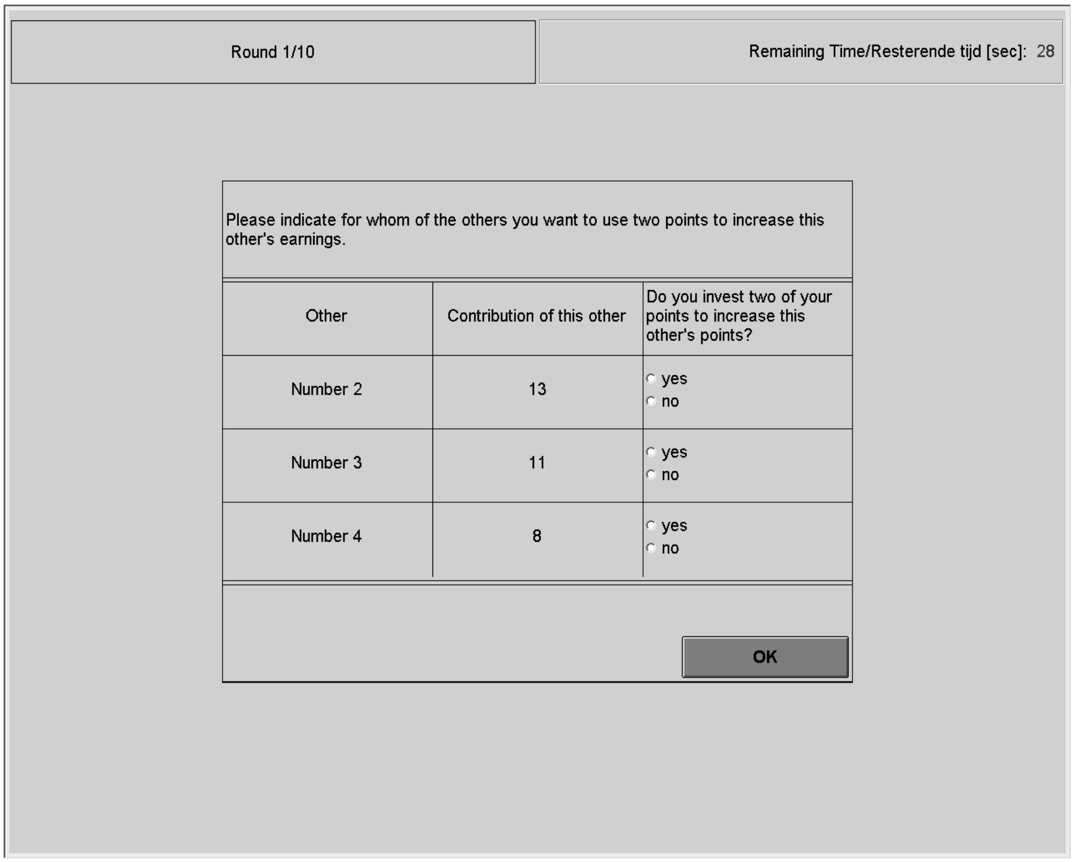
If a majority is reached, for example if you and “Number 2” both want to decrease the earnings of “Number 3”, both you and “Number 2” will lose **2 points**. “Number 3” will lose **6 points** for every group member who decreased his/her earnings, so in this case 2×6 = 12 points. If majority is not reached, for example only “Number 2” wanted to decrease the earnings of “Number 3”, the decrease will not be executed and both “Number 2” and “Number 3” will keep their points.

Your total profit for the round will be the sum of your earnings from the first and the second stage of that round. At the end of the round, you will see a screen showing how much every member of your group contributed to the group account, by how much their earnings have been decreased, and how many points you have earned in this round. If one of the others wanted to decrease your earnings but no majority was reached, this will NOT be shown. Similarly, if you wanted to decrease the earning of a group member, but reached no majority, this will NOT be shown to the group member whose earnings you wanted to decrease or to any of the others. If a majority of others did decrease your points, you will see on this screen that your points have been decreased, it will NOT be shown which participants paid to have your earnings decreased. Similarly, if you and one or more other group member(s) decide to decrease the points of one or more other group member(s), their points will be decreased, but they themselves or any of the others will NOT be informed that you are one of the persons who paid to decrease his/her points.

This procedure is repeated for 10 rounds. The participants with whom you are together in your group again **change randomly after every round**. “Number 2”, “Number 3” and “Number 4” will thus most likely be a different person than in the previous round. You will not be told each other’s identities at any time during or after the experiment. After this third part you will receive new instructions for part four of the experiment.

Instructions part 4

These instructions are for the fourth part of the experiment, which is again a series of 10 decision rounds. The decision situation as explained in the first instructions remains unchanged. You will thus in every round be asked how much of your endowment of 20 points you want to contribute to the group account. However, the second stage that follows after every member of your group has decided how much they want to contribute to the group account has changed. In the second stage all group members now have a chance to **increase** the earnings of the others. You will see the following screen:



In the first column you see the contribution every other member of your group ("Number 2", "Number 3" and "Number 4") has made to the group account. In the second column you decide whether you want to increase the earnings of each other group member by clicking "yes" or "no". You make this decision for each group member independently. If you decide that you want to increase the earnings of another member, this will be done only **if at least a majority of participants in your group wants to increase this person's earnings**. A majority is reached when two out of three group members indicate that they want to increase a fourth member's earnings. Thus, if you want to increase someone's earnings this is only accomplished if at least one other member of your group also wants to increase the earnings of this person.

If a majority is reached, for example if you and "Number 2" both want to increase the earnings of "Number 3", both you and "Number 2" will lose **2 points**. "Number 3" will receive **6 points** for every group member who increased his/her earnings, so in this case

2×6 = 12 points. If majority is not reached, for example only “Number 2” wanted to increase the earnings of “Number 3”, the increase will not be executed and both “Number 2” and “Number 3” will keep their points.

Your total profit for the round will be the sum of your earnings from the first and the second stage of that round. At the end of the round, you will see a screen showing how much every member of your group contributed to the group account, by how much their earnings have been increased, and how many points you have earned in this round. If one of the others wanted to increase your earnings but no majority was reached, this will NOT be shown. Similarly, if you wanted to increase the earning of a group member, but reached no majority, this will NOT be shown to the group member whose earnings you wanted to increase or to any of the others. If a majority of others did increase your points, you will see on this screen that your points have been increased, it will NOT be shown which participants paid to have your earnings increased. Similarly, if you and one or more other group member(s) decide to increase the points of one or more other group member(s), their points will be increased, but they themselves or any of the others will NOT be informed that you are one of the persons who paid to increase his/her points.

This procedure is repeated for 10 rounds. The participants with whom you are together in your group again **change randomly after every round**. “Number 2”, “Number 3” and “Number 4” will thus most likely be a different person than in the previous round. You will not be told each other’s identities at any time during or after the experiment.

Questionnaire

After all decisions have been made you will be asked to fill in a questionnaire. Please take your time to fill in this questionnaire accurately. In the mean time your earnings will be counted. Please remain seated until the payment has taken place.

Appendix B

Instructions for the experiment of Chapter 4

This is the English version of the instructions for part 1 and 2 of the experiment that was used for Chapter 4. Subjects in the experiment could choose for either a Dutch or an English version of these instructions. The instructions for part 2 were handed out after part 1 was finished. In this version of the instructions, CDR2 was employed in part 2. After part 2, instructions for a third series of 15 decision rounds were handed out, in which a different decision rule was employed. Instructions for the third part of the experiment and for the other decision rules are highly similar to the instructions for part 2 shown here. Italics indicate parts of the instructions that were only displayed in the sessions with noise.

- Instructions -

Welcome

Please read the following instructions carefully. These instructions are the same for all participants. The instructions state everything you need to know in order to participate in the experiment. If you have any questions, please raise your hand. One of the experimenters will approach you and answer your question.

You can earn money by means of earning points during the experiment. The number of points that you earn depends on your own choices and the choices of other participants. At the end of the experiment, the total number of points that you earned will be exchanged at an exchange rate of:

160 points = 1 Euro

The money you earn will be paid out in cash at the end of the experiment. Other participants will not see how much you have earned. During the experiment you are not allowed to communicate with other participants. Please turn off your mobile phone and put it in your bag. You may only use the functions on the screen that are necessary to carry out the experiment. Thank you very much.

The decision situation

First, we introduce the decision situation in which you will interact. You will learn about the procedure of the experiment later.

You are a member of a group of **six participants**. You and the five other members of your group are endowed with **20 points** each. You must all choose whether you want to keep your points to yourself in a **private account** or contribute your points to the **group account**. You can only contribute the **entire** amount of 20 points to the group account, or keep the whole endowment to yourself and contribute 0 points.

When all members of your group have decided on their contribution to the group account, the total number of points in the group account is **multiplied by 2.4**. The multiplied amount is then **equally** distributed among all six group members. When you have kept your points in your private account, your earnings from the group account will be added to these points. This holds for each participant in your group.

Table 1 below shows how many points you earn, given the choices you make and the number of others in your group who contribute to the group account. For example, suppose you contribute your 20 points to the group account, and four of the others contribute as well. The total number of people who contribute, including yourself, is five. The sixth group member keeps his/her points on the private account. Table 1 shows that you then earn 40 points. You can verify this yourself: the total contribution to the group account is 5×20 = 100 points. Multiplied by 2.4 gives 240 points in the group account, divided by six group members is 40 points each. Note that Table 1 also applies to the other group members. For the group member who contributed 0 points in the example, five others (including you) have contributed. This group member then receives the 40 points from the group account in addition to the 20 points on his/her private account, and earns 60 points. You can see this in Table 1.

Table 1: the number of points you earn for different own contributions and number of other group members who contribute to the group account.

Your contribution	Number of other group members who contribute 20 points 0	1	2	3	4	5
0	20	28	36	44	52	60
20	8	16	24	32	40	48

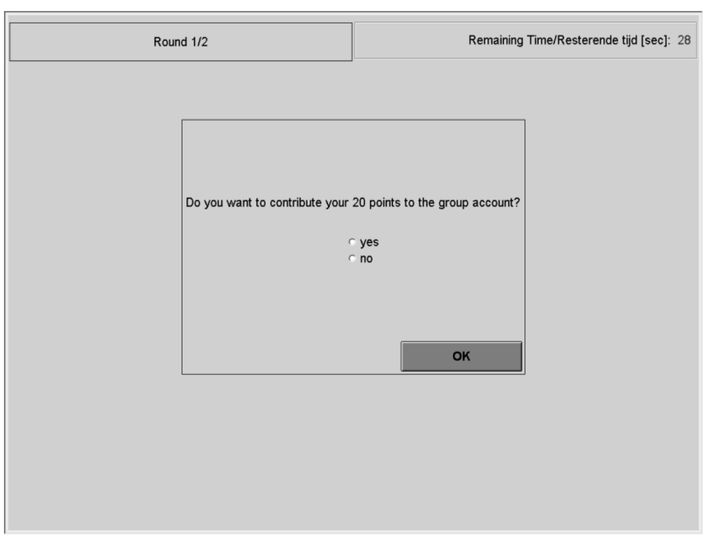
Before you continue reading the instructions, please choose the language in which you want to do the experiment and click "OK" on your computer screen. You will proceed to a number of control questions. Answer the questions to check whether you have understood the decision situation. Your answers to the control questions have no influence on your payment.

The Experiment

This experiment consists of **three parts**. Each part includes the decision situation described above. All three parts are series of 15 decision rounds. These instructions are for part 1; the instructions for part 2 and for part 3 will follow later.

Instructions part 1

You will now participate in the first series of **15 decision rounds**. You are randomly matched to five other participants. In each round, you receive the endowment of 20 points and decide whether you contribute these points to the group account. This is the only decision you have to make in every round, and will determine your earnings. After this decision you are again randomly matched to five other participants.



You will see a screen like the one shown above. You click whether or not you want to contribute your 20 points to the group account. If you choose "no", the 20 points will go to your private account. Once you have made a choice, click the "OK" button to confirm your decision.

After all participants have made their decision for the round, the computer will calculate the payoffs and tabulate the results. You will be informed about the contribution every member of your group has made to the group account.

Please note: this information is not always correct*. For every contribution that you observe, there is a* ***20%*** *chance that this is not the real contribution that this group member made. This means that, for each contribution of 20 that you see, there is a 20% chance that in reality this group member did not contribute to the group account. Likewise, for every contribution of 0 that you see, there is a 20% chance that this group member did contribute. Whether you see the correct contribution will be determined* ***independently*** *for every contribution that you see. It can happen that one, several, or none of the contributions that you see are displayed incorrectly.*

The 20% chance to see a wrong contribution also holds for the others in your group. Again, this chance is independent for every other group member. For example, every group member has a 20% chance to wrongly observe the contribution of the person who is displayed for you as "Number 6". It can happen that for one, several, or none of the other group members the contribution of this person is wrongly displayed. ***This also means that every other group member has a 20% chance to see your contribution incorrectly****. One, several, or none of the other group members can observe your contribution wrongly.*

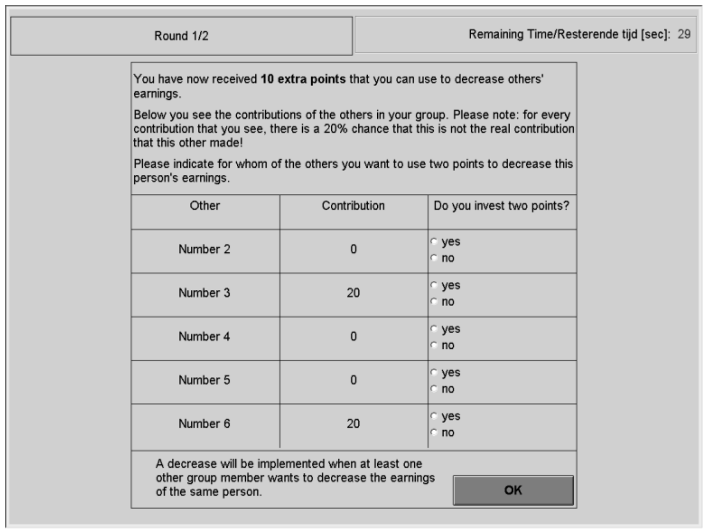
You will not be informed which of the contributions that you observed were displayed incorrectly. However, your earnings will be determined by the ***real*** *contributions of all group members. At the end of every round you are informed about the number of points that you earn in this round if the contributions that you see are the actual contributions. Please note that these might not be your real earnings.*

This procedure is repeated for 15 rounds. The participants with whom you are together in your group will be **randomly selected after every round**. You will not be told each other's identities at any time during or after the experiment.

When you have finished the first part of the experiment you will receive new instructions for part 2 and finally for part 3.

Instructions Part 2

These instructions are for the second series of 15 decision rounds. The decision situation as explained in the first instructions remains unchanged. You will thus in every round be asked whether you want to contribute your endowment of 20 points to the group account. However, every round now has a second part that follows after every member of your group has decided whether they want to contribute. In this second part all group members can **decrease** the earnings of the others. To do this, **every round** again you receive an **additional endowment of 10 points**. You will see the following screen:



In the second column you see the contribution every other member of your group ("Number 2" to "Number 6") has made to the group account. ***Please note****: again, there is a* ***20% chance****, for every contribution that you see here, that this is not the real contribution of this group member. This chance is again independent for everyone, just like in part 1 of the experiment. If you do not remember exactly what this implies, have another look at part 1 of the instructions.*

In the third column, you decide whether or not you want to decrease the earnings of each other group member by clicking "yes" or "no". You make this decision for each group member separately. If you decide that you want to decrease the earnings of another member, this is only possible **if at least one of the other members of your group also wants to decrease this person's earnings**.

When enough people want to decrease the points of the same group member, for example if you and "Number 2" both want to decrease the earnings of "Number 4", you and "Number 2" will lose **2 points** from the 10 extra points that you have both received in this second phase. "Number 4" will lose **6 points** for every group member who decreased his/her earnings, so in this case 2×6 = 12 points. If there are no two people who want to decrease the earnings of the same person, for example only "Number 2" wanted to decrease the earnings of "Number 4", the decrease will not be executed and both "Number 2" and "Number 4" will keep their points.

Your total profit for the round will be the sum of your earnings from part 1 and part 2 of that round. The part of the additional endowment of 10 points that you do not spend on decreasing others' payoff will be added to your earnings. At the end of the round, you will again see a screen showing how much the others have contributed, and with how many points their earnings were decreased. *You will see the same contributions here as you saw in the screen where you could choose whether or not to decrease others' points.* ***Contributions you observed incorrectly in that screen, will thus be displayed incorrectly again****. Hence, you will never be informed about the true contributions. However, your earnings will be determined by the* ***real*** *contributions of all group members. At the end of every round you are informed about the number of points that you earn in this round if the contributions that you see are the actual contributions. Please note that these might not be your real earnings.*

If only one of the others wanted to decrease your earnings, this will NOT be shown to you. Similarly, if you wanted to decrease the earning of another group member, but not at least one other group member wanted this as well, this will NOT be shown to the group member whose earnings you wanted to decrease. If at least two of the others decided to decrease your points, you will see on this screen that your points have been decreased, but you will NOT see which participant paid to have your earnings decreased. Similarly, if you and one or more other group members decide to decrease the points of another group member, (s)he will be shown that his/her points are decreased, but (s)he will NOT be informed that you are one of the persons who paid to decrease his/her points.

This procedure is repeated for 15 rounds. The participants with whom you are together in your group are again **randomly selected after every round**. "Number 2", to "Number 6" will thus most likely be a different person than in the previous round. You will not be told each other's identities at any time during or after the experiment. After this second series you will receive new instructions for the third series.

Appendix C

Instructions for the experiment of Chapter 5

This is the English version of the instructions for the experiment that was used for Chapter 5. Subjects in the experiment could choose for either a Dutch or an English version of these instructions. The instructions for part 2 were handed out after part 1 was finished. Italics indicate parts of the instructions that were only displayed in the sessions with noise. Square brackets indicate parts of the instructions that were only displayed in the sessions without noise.

- Instructions -

Welcome

Please read the following instructions carefully. These instructions are the same for all participants. The instructions state everything you need to know in order to participate in the experiment. If you have any questions, please raise your hand. One of the experimenters will approach you and answer your question.

You can earn money by means of earning points during the experiment. The number of points that you earn depends on your own choices and the choices of other participants. At the end of the experiment, the total number of points that you earned will be exchanged at an exchange rate of:

160 points = 1 Euro

The money you earn will be paid out in cash at the end of the experiment. Other participants will not see how much you have earned. During the experiment you are not allowed to communicate with other participants. Please turn off your mobile phone and put it in your bag. You may only use the functions on the screen that are necessary to carry out the experiment.

The decision situation

First, we introduce the decision situation in which you will interact with other participants. You will learn about the procedure of the experiment later.

You are a member of a group of **six participants**. You and the five other members of your group are endowed with **20 points** each. You choose whether you want to keep your points to yourself in a **private account** or you want to contribute your points to the **group account**. You can only contribute the **entire** amount of 20 points to the group account, or keep the whole endowment to yourself on your private account. The five other members of your group also decide for themselves whether they contribute their 20 points to the group account, or keep them on their private account.

When the members of your group have decided whether or not to contribute 20 points to the group account, the total number of points in the group account is **multiplied by 2.4**. The multiplied amount is then **equally** distributed among all six group members. When you have kept your endowment in your private account, your earnings from the group account will be added to your endowment. This holds for each participant in your group.

Table 1 below shows how many points you earn, depending on whether or not you contribute your endowment to the group account and on the total number group members who contribute. For example, suppose you contribute your 20 points to the group account, and four of the others contribute as well. The total number of people who contribute, including yourself, is five. The sixth group member keeps his/her points on the private account. Table 1 shows that you then earn 40 points. You can verify this yourself: the total contribution to the group account is 5×20 = 100 points. This gives 100 × 2.4 = 240 points in the group account; 240 / 6 = 40 points each. Note that Table 1 also applies to the other group members. The group member who contributed 0 points in the example receives the 40 points from the group account in addition to the 20 points on his/her private account, and earns 60 points. You can see this in Table 1.

Table 1: the number of points you and the other members of your group earn for contributing to the group account and for keeping the points on the private account, given the total number of group members who contribute 20 points.

Number of group members who contribute 20 points	Total amount on the group account (number of contributions × 20)	× 2.4	Earnings for group members who contribute 20 (profit from group account)	Earnings for group members who keep the endowment on their private account
0	0	0	-	20
1	20	48	8	28
2	40	96	16	36
3	60	144	24	44
4	80	192	32	52
5	100	240	40	60
6	120	288	48	-

Before you continue reading the instructions, please choose the language in which you want to do the experiment and click "OK" on your computer screen. You will proceed to a number of control questions. Answer the questions to check whether you have understood the decision situation. Your answers to the control questions have no influence on your payment.

The Experiment

This experiment consists of **three parts**. The first two parts include the decision situation described above. These instructions are for part 1; the instructions for part 2 and 3 will follow later.

Instructions part 1

You will now participate in a series of **5 decision rounds**. You are randomly matched to five other participants. In each round, you receive the endowment of 20 points and decide whether you contribute these points to the group account. This is the only decision you have to make in every round, and that will determine your earnings for this part of the experiment. After this decision you are again randomly matched to five other participants.

You will see a screen like the one shown above. You click whether or not you want to contribute your 20 points to the group account. If you choose "yes", the 20 points will go to the group account. If you choose "no", the 20 points will go to your private account. Once you have made a choice, click the "OK" button to confirm your decision.

After all participants have made their decision for the round, the computer will calculate the payoffs and tabulate the results. You will be informed about the contribution that every member of your group has made to the group account. [Also, you will see how many points you have earned in this round.]

Please note: this information is not always correct*. For every contribution that you observe, there is a* ***20% chance*** *that this is not the real contribution that this group member made. This means that, for each contribution of 20 points that you see, there is a 20% chance that in reality this group member did not contribute to the group account. Likewise, for every contribution of 0 points that you see, there is a 20% chance that this* *group member contributed 20 points. Whether you see the correct contribution will be determined* ***independently*** *for every contribution of 0 or 20 that you see. It can happen that one, several, or none of the contributions that you see are displayed incorrectly.*

The 20% chance to see a wrong contribution also holds for the others in your group. Again, this chance is independent for every group member. For example, everyone has a 20% chance to wrongly observe the contribution of the person who is displayed for you as "Number 6". It can happen that for one, several, or none of the group members the contribution of this person is wrongly displayed. ***This also means that every other group member has a 20% chance to see your contribution incorrectly****. One, several, or none of the other group members can observe your contribution wrongly.*

You will not be informed which of the contributions that you observed were displayed incorrectly. However, your earnings will be determined by the ***real*** *contributions of all group members. At the end of every round, you are informed about the number of points that you would have earned in this round if the contributions that you see would have been the actual contributions. Please note that these might not be your real earnings.*

This procedure is repeated for 5 rounds. The participants with whom you are together in your group will be ***randomly selected after every round****. You will not be told each other's identities at any time during or after the experiment.*

When you have finished the first part of the experiment you will receive new instructions for part 2.

Instructions Part 2

These instructions are for the second part of the experiment. This part consists of a series of **40 decision rounds**. The decision situation as explained in the first instructions remains unchanged. You will thus in every round be asked whether you want to contribute your endowment of 20 points to the group account. However, before you make this decision, a **vote** will take place in **every round**. The vote determines under which **system** the decision round takes place in your group.

Below, you can first read what the different systems entail. Subsequently, we explain the procedure of the decision rounds.

System A

In system A you participate in the decision situation as described in part 1 of the instructions. The only difference is that after the interaction all group members receive an **additional amount of 10 points** regardless of their contribution to the group account. All group members keep this additional amount for themselves, it cannot be contributed it to the group account.

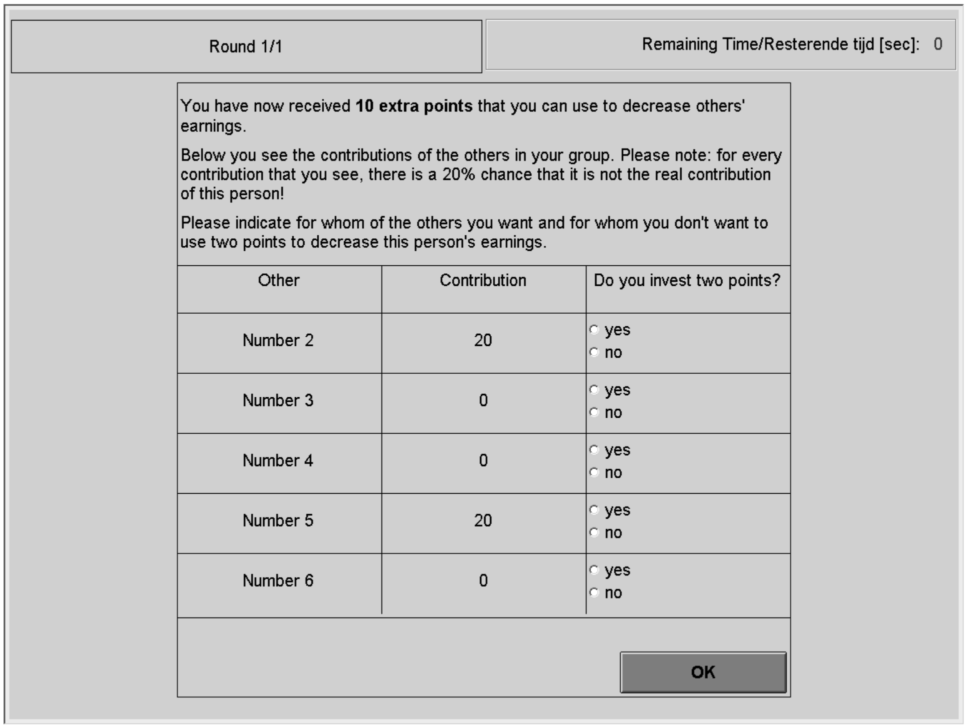
Please note*: again there is a* ***20% chance****, for every contribution that you see, at the end of around, that this is not the real contribution of this group member. This chance is again independent for everyone, just like in part 1 of the experiment. If you do not remember exactly what this implies, have another look at part 1 of the instructions.*

System B

System B has two different versions. Below you can first read the instructions for version B1. Subsequently, the difference with version B2 is explained.

Version B1

In system B, you likewise participate in the decision situation as described in part 1. However, in system B the round has a second stage, that follows after every member of your group has decided whether to contribute 0 or 20 points to the group account. In this second stage all group members can **decrease** the earnings of the others. To do this, after the contribution decisions every group member receives an **additional amount of 10 points**. You will see the following screen:



Round 1/1

Remaining Time/Resterende tijd [sec]: 0

You have now received **10 extra points** that you can use to decrease others' earnings.

Below you see the contributions of the others in your group. Please note: for every contribution that you see, there is a 20% chance that it is not the real contribution of this person!

Please indicate for whom of the others you want and for whom you don't want to use two points to decrease this person's earnings.

Other	Contribution	Do you invest two points?
Number 2	20	yes / no
Number 3	0	yes / no
Number 4	0	yes / no
Number 5	20	yes / no
Number 6	0	yes / no

OK

In the second column you see the contribution every other member of your group ("Number 2" to "Number 6") has made to the group account.

Please note*: again there is a* ***20% chance****, for every contribution that you see here, that this is not the real contribution of this group member. This chance is again independent for everyone, just like in part 1 of the experiment. If you do not remember exactly what this implies, have another look at part 1 of the instructions.*

In the third column, you decide whether or not you want to decrease the earnings of each other group member by clicking "yes" or "no". You make this decision for each group member separately. If you decide to decrease someone's earnings, that person will lose **6 points** while you lose **2 points** from the 10 extra points that you have received in this second stage. If another person in your group decides to decrease your earnings, that person will lose 2 points and you will lose 6 points. If multiple group members decide to decrease the earnings of the same person, those group members will all lose 2 points and the person whose points are decreased will lose 6 points **for every group member who decreased his/her earnings**.

Your total profit for the round will be the sum of your earnings from stage 1 and stage 2 of that round. The part of the additional endowment of 10 points that you do not spend on decreasing others' payoff will be **added to your earnings**. At the end of the round, you will again see a screen showing how much the others have contributed, and with how many points their earnings were decreased. [You will also be informed about the number of points that you have earned in this round.] *You will see the same contributions here as you saw in the screen where you could choose whether or not to decrease others' points.* ***Contributions you observed incorrectly in that screen, will thus be displayed incorrectly again****. Hence, you will never be informed about the true contributions. However, your earnings will be determined by the* ***real*** *contributions of* *all group members. At the end of every round you are informed about the number of points that you would have earned in this round if the contributions that you see would have been the actual contributions. Please note that these might not be your real earnings. NB: information about the number of points by which your earnings or the earnings of other group members are decreased, is always accurate.*

If any of the other group members decides to decrease your earnings, you will see on this screen that your earnings have been decreased, but you will NOT see which participant(s) paid to have your earnings decreased. Similarly, if you decide to decrease the earnings of one or more of the other group members, their earnings are decreased, but they will NOT be informed that you are the person who paid to decrease their earnings.

Version B2
The instructions for version B1 also apply to version B2. The only **exception** is the number of points that group members lose when their earnings are decreased by others.

If you decide to decrease someone's earnings in version B2, that person will lose **12 points** while you lose **2 points** from the 10 extra points that you have received in this second stage. If another person in your group decides to decrease your earnings, that person will lose 2 points and you will lose 12 points. If multiple group members decide to decrease the earnings of the same person, those group members will all lose 2 points and the person whose points are decreased will lose 12 points *for every group member who decreased his/her earnings*.

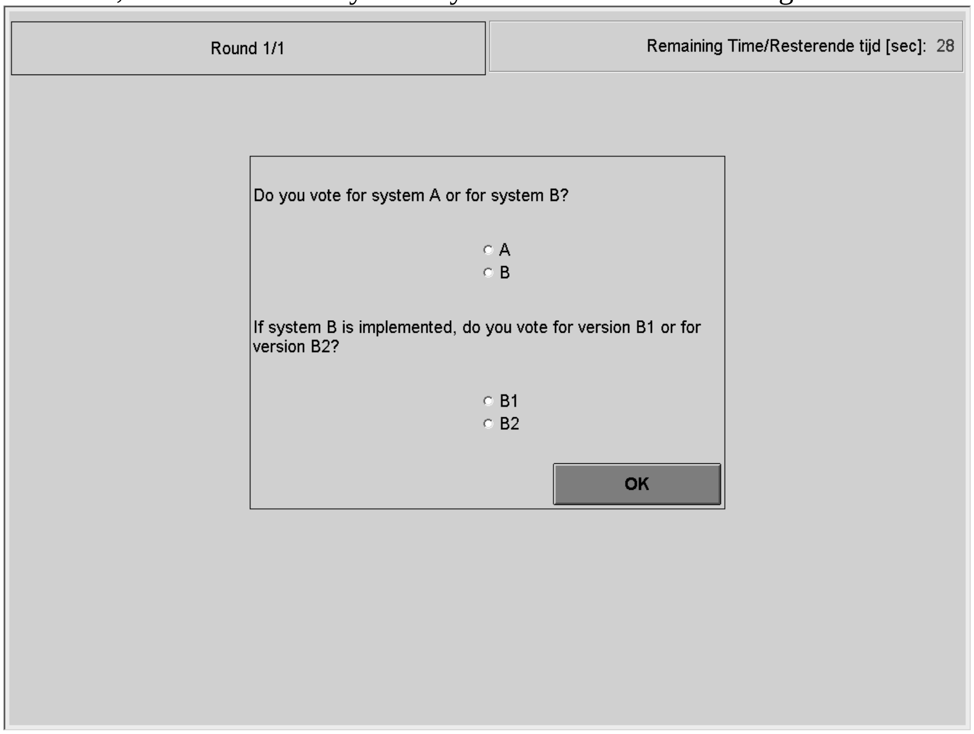
Summary
Table 2 contains a short overview of the content of each system.

Table 2: the systems that you can vote for

System A	Everyone in your group receives an additional amount of 10 points on the private account
System B	Everyone in your group receives an additional amount of 10 points, that can be used to decrease the earnings of other group members. Decreasing someone's earnings costs 2 points. The remaining part of the 10 extra points is added to the private account.
Version B1	A group member whose earnings are decreased, loses 6 points for everyone who decreased his/her earnings.
Version B2	A group member whose earnings are decreased, loses 12 points for everyone who decreased his/her earnings.

Procedure of a round

From now on, at the start of every round you are shown the following screen:



On this screen, you make two decisions. These decisions, and the decisions of the five other members of your group, will determine which system is implemented in your group. It might thus be that different groups in the experiment elect different systems.

First, you vote for system A or B. If more members of your group vote for A than for B, system A will be implemented. If the majority votes for B, a version of system B is implemented. If exactly half of your group votes for system A, and the other half votes for B, the computer will randomly decide whether A or B is implemented.

Second, regardless whether you voted for A or B, you choose between version B1 and B2. If system B is implemented in your group, the version of system B that most people in your group have voted for will be implemented. Hence, the choice between B1 and B2 is only relevant when B is actually implemented. If B is implemented, and exactly half of your group votes for version B1, and the other half votes for B2, the computer will randomly decide whether B1 or B2 is implemented.

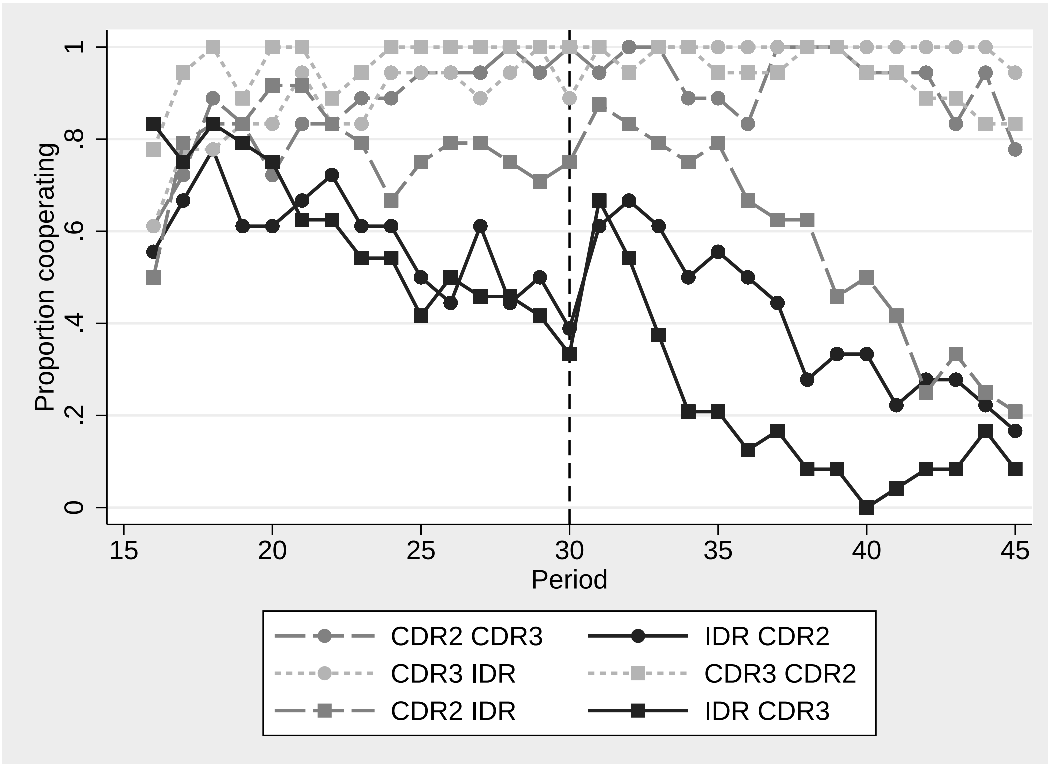
This procedure is repeated for 40 rounds. Thus, every round starts with a vote. After the vote, you are informed about the outcome but not about the number of group members that voted for each option. Subsequently, you participate in the decision situation under the system that your group voted for. The participants with whom you are together in your group are again **randomly selected after every round**. “Number 2”, to “Number 6” will thus most likely be a different person than in the previous round. You will not be told each other’s identities at any time during or after the experiment.

Remaining tasks

After the 40 rounds, you participate in part 3, for which you will receive new instructions. Afterwards, you are asked to fill out a short questionnaire.

Appendix D

Additional analyses for Chapter 4



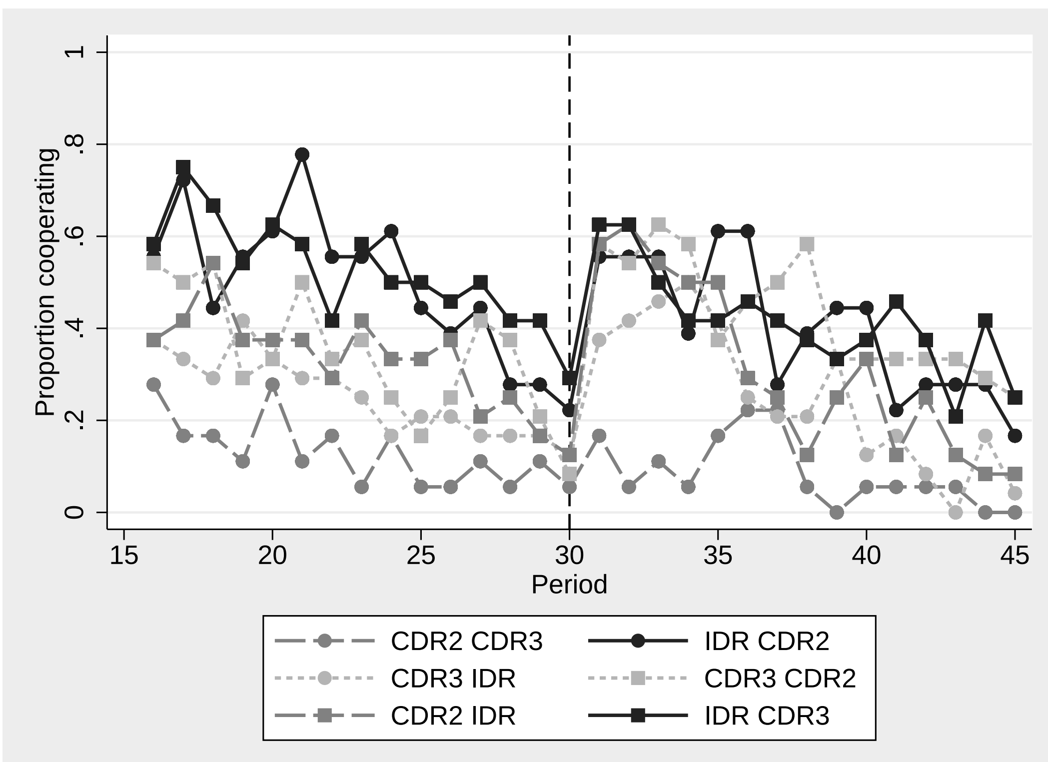


Figure D.1: Cooperation rates in each period of the first and second punishment sequence, separated by sessions, without (top) and with (bottom) noise. Note that the change in experimental condition occurs after period 30 (dashed line).

Table D.1: Average cooperation, earnings, and punishment of observed defectors in each experimental condition, in the baseline and first punishment sequence (7955 punishment decisions, 7560 PDs, 252 subjects).

	Cooperation		Earnings		Pun. of obs. defectors	
	Mean	S.d.	Mean	S.d.	Mean	S.d.
Baseline no noise	0.225	0.418	26.300	9.139		
Baseline noise	0.269	0.444	27.537	9.649		
IDR no noise	0.587	0.493	29.498	9.799	0.375	0.484
IDR noise	0.511	0.500	25.841	10.716	0.380	0.485
CDR2 no noise	0.814	0.389	38.229	8.790	0.621	0.486
CDR2 noise	0.244	0.430	25.029	9.128	0.146	0.354
CDR3 no noise	0.915	0.279	43.630	6.549	0.613	0.488
CDR3 noise	0.299	0.458	26.128	10.470	0.280	0.449

Table D.2: Logistic regression on the decision whether to cooperate and on the decision whether to punish an observed defector, and linear regression on period earnings, all with decisions nested in subjects and sessions, in the baseline and first punishment sequence (7955 punishment decisions, 7560 PDs, 252 subjects).

	Cooperation		Earnings		Pun. obs. defectors	
	Coeff.	S.e.	Coeff.	S.e.	Coeff.	S.e.
Noise	0.098	0.223	1.146	0.635	0.286	0.672
IDR	1.937**	0.141	2.928**	0.492		
× noise	-0.520**	0.198	-4.639**	0.696		
CDR2	3.573**	0.177	12.563**	0.492	2.235**	0.688
× noise	-3.816**	0.233	-15.205**	0.696	-5.010**	0.987
CDR3	4.306**	0.225	16.906**	0.529	2.466**	0.752
× noise	-4.081**	0.262	-18.185**	0.703	-3.707**	1.004
Constant	-1.572**	0.159	26.333**	0.454	-1.054*	0.480
σ_u	0.130	0.218	0.817	0.267	0.291	0.359
σ_e	1.459	0.087	2.399	0.162	2.649	0.179
Log Likelihood	-3600.076		-27544.172		-3114.541	

*Significant at the .05-level. **Significant at the .01-level (2-sided)

Table D.3: Logistic regression on the decision whether or not to cooperate and on the decision whether or not to punish an observed defector, and a linear regression on period earnings, all with a cluster at session level, in the baseline and first punishment sequence (7955 punishment decisions, 7560 PDs, 252 subjects).

	Cooperation		Earnings		Pun. obs. defectors	
	Coeff.	S.e.	Coeff.	S.e.	Coeff.	S.e.
Noise	0.238	0.169	1.237*	0.868	0.018	0.236
IDR	1.590**	0.138	3.198**	1.444		
× noise	-0.546**	0.164	-4.895**	1.540		
CDR2	2.715**	0.203	11.929**	0.942	1.001**	0.408
× noise	-2.845**	0.419	-14.437**	1.623	-2.274**	0.477
CDR3	3.611**	0.570	17.330**	2.173	0.969**	0.324
× noise	-3.466**	0.581	-18.739**	2.208	-1.422**	0.376
Constant	-1.237**	0.135	26.300**	0.657	-0.509*	0.232
Log Likelihood/R2	-4225.501		0.238		-4603.520	

*Significant at the .05-level. **Significant at the .01-level (2-sided)

Table D.4: Logistic regression on the decision whether to cooperate in period t, with a cluster at session level, in the first punishment sequence (3528 PDs, 252 subjects).

	Coeff.	S.e.
Noise	0.797*	0.391
CDR2	0.731**	0.108
× noise	-1.172**	0.234
CDR 3	1.357**	0.341
× noise	-1.687**	0.346
Own contribution $t-1$	3.598**	0.420
× noise	-1.210*	0.474
Punished while defecting $t-1$	1.239**	0.339
× noise	-0.661	0.395
Punished while cooperating $t-1$	-0.771**	0.148
× noise	0.890*	0.186
Obs. n other cooperators $t-1$	0.245**	0.051
× noise	-0.031	0.076
Period	-0.031	0.020
× noise	-0.035	0.021
Constant	-2.649**	0.407
Log Likelihood	-1520.216	

*Significant at the .05-level. **Significant at the .01-level (2-sided)

Appendix E

Additional analyses for Chapter 5

Table E.1: Average cooperation, observed proportion of others cooperating, earnings, observed earnings, and punishment of observed defectors in each experimental condition (31200 observations of others' decisions, 3442 punishment decisions, 6240 PDs, 156 subjects).

	Cooperation		Obs. coop. others		Earnings		Obs. earnings		Pun. of obs. defectors	
	Mean	S.d.	Mean	S.d.	Mean	S.d.	Mean	S.d.	Mean	S.d.
NP no noise	0.134	0.341			33.75	7.267				
NP noise	0.149	0.356	0.288	0.453	34.172	7.675	39.735	8.854		
LP no noise	0.877	0.329			50.881	9.737			0.487	0.500
LP noise	0.588	0.492	0.547	0.498	40.635	10.239	38.972	10.762	0.299	0.458
HP no noise	0.974	0.158			3.595	8.048			0.674	0.469
HP noise	0.869	0.338	0.721	0.449	42.609	12.587	36.711	14.265	0.543	0.499

Table E.2: Descriptive statistics of average experiences over all PDs preceding the focal interaction (6240 PDs, 156 subjects).

	No noise				Noise			
Variable	Mean	S.d.	Min.	Max.	Mean	S.d.	Min.	Max.
Cooperation NP[1]	0.711	0.417	0.000	3.000	1.574	0.434	0.000	4.000
Cooperation NP[2]	0.706	0.420	0.000	3.000	1.571	0.439	0.000	4.000
Never NP[3]	0.008		0.000	1.000	0.002		0.000	1.000
Cooperation LP[1]	3.342	1.049	0.000	5.000	3.024	0.611	1.000	5.000
Cooperation LP[2]	2.857	1.526	0.000	5.000	2.505	1.268	0.000	5.000
Never LP[3]	0.142		0.000	1.000	0.167		0.000	1.000
Cooperation HP[1]	4.805	0.212	3.000	5.000	3.554	0.769	1.000	5.000
Cooperation HP[2]	4.162	1.648	0.000	5.000	1.861	1.860	0.000	5.000
Never HP[3]	0.130		0.000	1.000	0.464	0.000	1.000	
Av. punishment defectors LP[1]	2.326	1.353	0.000	5.000	1.069	0.715	0.000	1.000
Av. punishment defectors LP[2]	2.300	1.369	0.000	5.000	1.066	0.717	0.000	4.000
Never observed defectors LP[3]	0.012		0.000	1.000	0.003		0.000	1.000
Av. punishment defectors HP[1]	3.065	1.410	0.000	5.000	0.934	1.113	0.000	4.000
Av. punishment defectors HP[2]	2.700	1.656	0.000	5.000	0.913	1.109	0.000	4.000
Never observed defectors HP[3]	0.120		0.000	1.000	0.022		0.000	1.000
Av. punishment cooperators LP	0.206	0.255	0.000	1.250	0.379	0.230	0.000	1.556
Av. punishment cooperators HP	0.208	0.269	0.000	1.333	0.319	0.394	0.000	2.000

[1]Prior to the zero imputation of legitimate missing values.
[2]Following the zero imputation of legitimate missing values.
[3]Dummy variable where a value of one indicates that a zero was imputed for legitimate missing values in the variable directly above.

Table E.3: Multilevel regression or logistic regression on the decision whether to cooperate, period earnings, and decision whether to punish an observed defector, with decisions nested in subjects (3442 punishment decisions in 6240 PDs by 156 subjects).

	Cooperation		Earnings		Pun. obs. defectors	
	Coeff.	S.e.	Coeff.	S.e.	Coeff.	S.e.
Noise	-0.044	0.394	0.168	0.689		
LP	4.492**	0.275	14.183**	0.598		
× noise	-1.270**	0.310	-7.587**	0.697	-1.776**	0.514
HP	7.832**	0.336	20.661**	0.503	1.479**	0.260
× noise	-2.703**	0.410	-12.518**	0.740	-2.084**	0.555
Constant	-2.616**	0.308	33.991**	0.567	0.030	0.366
σ_u	3.896**	0.569	2.933**	0.212	2.840**	0.259
Log Likelihood	-1668.929		-22247.904		-1527.599	

*Significant at the .05-level. **Significant at the .01-level (2-sided)

Table E.4: Multilevel regression or logistic regression on the predicted probability to observe a cooperator, and predicted observed period earnings, with decisions nested in subjects (31200 observations of others' decisions in 6240 PDs by 156 subjects).

	Obs. cooperation		Obs. earnings	
	Coeff.	S.e.	Coeff.	S.e.
Noise	1.085**	0.118	5.688**	0.694
LP	3.406**	0.092	14.449**	0.636
× noise	-2.274**	0.101	-14.975**	0.742
HP	6.033**	0.108	20.583**	0.539
× noise	-4.176**	0.126	-23.617**	0.792
Constant	-2.007**	0.099	33.972**	0.582
σ_u	0.540**	0.041	2.732**	0.212
Log Likelihood	-12993.612		-22668.678	

*Significant at the .05-level. **Significant at the .01-level (2-sided)

References

Ackermann, K. A., J. Fleiß, and R. O. Murphy (2014). Reciprocity as an individual difference. *Journal of Conflict Resolution*, forthcoming.

Adams, J. (1963). Towards an understanding of inequity. *Journal of Abnormal and Social Psychology 67*(5), 422.

Adams, J. (1965). Inequity in social exchange. *Advances in experimental social psychology 2*, 267–299.

Aksoy, O. and J. Weesie (2009). Inequality and procedural justice in social dilemmas. *Journal of Mathematical Sociology 33*(4), 303–322.

Ambrus, A. and B. Greiner (2012). Imperfect public monitoring with costly punishment: An experimental study. *American Economic Review 102*(7), 3317–3332.

Anderson, C. and L. Putterman (2006). Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism. *Games and Economic Behavior 54*(1), 1–24.

Andreoni, J. and L. Gee (2012). Gun for hire: Delegated enforcement and peer punishment in public goods provision. *Journal of Public Economics 96*(11), 1036–1046.

Axelrod, R. (1986). An evolutionary approach to norms. *American Political Science Review 80*(4), 1095–1111.

Balafoutas, L. and N. Nikiforakis (2012). Norm enforcement in the city: A natural field experiment. *European Economic Review 56*(8), 1773–1785.

Baldassarri, D. and G. Grossman (2011). Centralized sanctioning and legitimate authority promote cooperation in humans. *Proceedings of the National Academy of Sciences 108*(27), 11023–11027.

Balliet, D., L. Mulder, and P. Van Lange (2011). Reward, punishment, and cooperation: A meta-analysis. *Psychological Bulletin 137*(4), 594–615.

Batenburg, R. S., W. Raub, and C. Snijders (2003). Contacts and contracts: Dyadic embeddedness and the contractual behavior of firms. *Research in the Sociology of Organizations 20*(1), 135–188.

Bereby-Meyer, Y. (2012). Reciprocity and uncertainty. *Behavioral and Brain Sciences 35*(1), 18–19.

Berger, J. and D. Hevenstone (2014). An international field experiment of norm enforcement in anonymous interactions. How costs and benefits impact altruistic punishment. *Working paper*.

Biel, A., D. Eek, and T. Gärling (1997). Distributive justice and willingness to pay for municipality child care. *Social Justice Research 10*(1), 63–80.

Bies, R. J. (1993). Privacy and procedural justice in organizations. *Social Justice Research 6*(1), 69–86.

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

Boehm, C. (1999). *Hierarchy in the Forest: Egalitarianism and the Evolution of Human Altruism*. Cambridge, MA: Harvard University Press.

Bornstein, G. and O. Weisel (2010). Punishment, cooperation, and cheater detection in "noisy" social exchange. *Games 1*(1), 18–33.

Botelho, A., G. W. Harrison, L. Pinto, and E. E. Rutström (2005). Social norms and social choice. *Working paper*.

Botelho, A., G. W. Harrison, L. M. C. Pinto, and E. E. Rutström (2009). Testing static game theory with dynamic experiments: A case study of public goods. *Games and Economic Behavior 67*(1), 253–265.

Boyd, R., H. Gintis, and S. Bowles (2010). Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science 328*(5978), 617–620.

Buchanan, J. and G. Tullock (1962). *The Calculus of Consent: Logical Foundations of Constitutional Democracy*. Ann Arbor, MI: University of Michigan Press.

Buskens, V. and W. Raub (2013). Rational choice research on social dilemmas: Embeddedness effects on trust. In R. Wittek, T. Snijders, and V. Nee (Eds.), *Handbook of Rational Choice Social Research*. New York, NY: Russell Sage.

Camerer, C. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press.

Carpenter, J. (2007). The demand for punishment. *Journal of Economic Behavior & Organization 62*(4), 522–542.

Carpenter, J. and P. Matthews (2009). What norms trigger punishment? *Experimental Economics 12*(3), 272–288.

Casari, M. and L. Luini (2009). Cooperation under alternative punishment institutions: An experiment. *Journal of Economic Behavior & Organization 71*(2), 273–282.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics 14*(1), 47–83.

Choi, J.-K. and T. Ahn (2013). Strategic reward and altruistic punishment support cooperation in a Public Goods Game experiment. *Journal of Economic Psychology 35*, 17–30.

Cinyabuguma, M., T. Page, and L. Putterman (2006). Can second-order punishment deter perverse punishment? *Experimental Economics 9*(3), 265–279.

Coleman, J. S. (1964). Collective decisions. *Sociological Inquiry 34*(2), 166–181.

Coleman, J. S. (1990). *Foundations of Social Theory*. Cambridge, MA: Harvard University Press.

Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology 86*(3), 386.

Cornelissen, T., O. Himmler, and T. Koenig (2013). Fairness spillovers — The case of taxation. *Journal of Economic Behavior & Organization 90*, 164–180.

Croson, R. T. (1996). Information in ultimatum games: An experimental study. *Journal of Economic Behavior & Organization 30*(2), 197–212.

Dari-Mattiacci, G. and G. De Geest (2010). Carrots, sticks, and the multiplication effect. *Journal of Law, Economics, and Organization 26*(2), 365–384.

Davis, D. D. and C. A. Holt (1993). Experimental economics: Methods, problems, and promise. *Estudios Economicos*, 179–212.

Dawes, C., J. Fowler, T. Johnson, R. McElreath, and O. Smirnov (2007). Egalitarian motives in humans. *Nature 446*(7137), 794–796.

Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology 31*(1), 169–193.

De Cremer, D. and T. R. Tyler (2005). Managing group behavior: The interplay between procedural justice, sense of self, and cooperation. *Advances in Experimental Social Psychology 37*, 151–218.

De Quervain, D., U. Fischbacher, V. Treyer, and M. Schellhammer (2004). The neural basis of altruistic punishment. *Science 305*(5688), 1254–1258.

Decker, T., A. Stiehler, and M. Strobel (2003). A comparison of punishment rules in repeated public good games. *Journal of Conflict Resolution 47*(6), 751–772.

Denant-Boemont, L., D. Masclet, and C. Noussair (2007). Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Economic Theory 33*(1), 145–167.

Deutsch, M. (1975). Equity, equality, and need: What determines which value will be used as the basis of distributive justice? *Journal of Social Issues 31*(3), 137–149.

Deutsch, M. (1985). *Distributive Justice: A Social-Psychological Perspective*. New Haven, CT: Yale University Press.

Diekmann, A. and T. Voss (2003). Social norms and reciprocity. *Arbeitsbericht des Instituts für Soziologie 33*.

Dolan, P., R. Edlin, A. Tsuchiya, and A. Wailoo (2007). It ain't what you do, it's the way that you do it: Characteristics of procedural justice and their importance in social decision-making. *Journal of Economic Behavior & Organization 64*(1), 157–170.

Drouvelis, M. and J. C. Jamison (2012). Selecting public goods institutions: Who likes to punish and reward? *Working Paper*.

Duch, R., W. Przepiorka, and R. Stevenson (2015). Responsibility attribution for collective decision makers. *American Journal of Political Science 59*(2), 372–389.

Eek, D. and A. Biel (2003). The interplay between greed, efficiency, and fairness in public-goods dilemmas. *Social Justice Research 16*(3), 195–215.

Eek, D., A. Biel, and T. Gärling (1998). The effect of distributive justice on willingness to pay for municipality child care: An extension of the gef hypothesis. *Social Justice Research 11*(2), 121–142.

Egas, M. and A. Riedl (2008). The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society B: Biological Sciences 275*(1637), 871–878.

Ellingsen, T., B. Herrmann, M. Nowak, D. Rand, and C. Tarnita (2013). Civic capital in two cultures: The nature of cooperation in Romania and USA. *Working paper*.

Engel, C. (2014). Social preferences can make imperfect sanctions work: Evidence from a public good experiment. *Journal of Economic Behavior & Organization 108*, 343–353.

Ertan, A., T. Page, and L. Putterman (2009). Who to punish? Individual decisions and majority rule in mitigating the free rider problem. *European Economic Review 53*(5), 495–511.

Falk, A., E. Fehr, and U. Fischbacher (2005). Driving forces behind informal sanctions. *Econometrica 73*(6), 2017–2030.

Falk, A. and J. J. Heckman (2009). Lab experiments are a major source of knowledge in the social sciences. *Science 326*(5952), 535–538.

Fehr, E. and A. Falk (2002). Psychological foundations of incentives. *European Economic Review 46*(4), 687–724.

Fehr, E. and U. Fischbacher (2004). Social norms and human cooperation. *Trends in Cognitive Sciences 8*(4), 185–190.

Fehr, E. and S. Gächter (2000). Cooperation and punishment in public goods experiments. *American Economic Review 90*(4), 980–994.

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

Fehr, E. and H. Gintis (2007). Human motivation and social cooperation: Experimental and analytical foundations. *Annual Review of Sociology 33*, 43–64.

Fehr, E. and B. Rockenbach (2003). Detrimental effects of sanctions on human altruism. *Nature 422*(6928), 137–140.

Fehr, E. and T. Williams (2013). Endogenous emergence of institutions to sustain cooperation. *Working paper*.

Feld, L. P. and J.-R. Tyran (2002). Tax evasion and voting: An experimental analysis. *Kyklos 55*(2), 197–221.

Fischbacher, U. (2007). z-Tree: Zürich toolbox for ready-made economic experiments. *Experimental Economics 10*(2), 171–178.

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

Fischbacher, U. and S. Gächter (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review 100*(1), 541–556.

Fischer, S., K. R. Grechenig, and N. Meier (2013). Cooperation under punishment: Imperfect information destroys it and centralizing punishment does not help. *Working paper*.

Fischer, S. and A. Nicklisch (2007). Ex interim voting: An experimental study of referendums for public-good provision. *Journal of Institutional and Theoretical Economics*, 56–74.

Folger, R. (1977). Distributive and procedural justice: Combined impact of voice and improvement on experienced inequity. *Journal of Personality and Social Psychology 35*(2), 108.

Frey, B. S. (2003). Deterrence and tax morale in the European Union. *European Review 11*(3), 385–406.

Frey, B. S., M. Benz, and A. Stutzer (2004). Introducing procedural utility: Not only what, but also how matters. *Journal of Institutional and Theoretical Economics*, 377–401.

Fudenberg, D. and P. A. Pathak (2010). Unobserved punishment supports cooperation. *Journal of Public Economics 94*(1), 78–86.

Fudenberg, D., D. G. Rand, and A. Dreber (2012). Slow to anger and fast to forgive: Cooperation in an uncertain world. *American Economic Review 212*(2), 720–249.

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

Gehrig, T., W. Güth, V. Levati, R. Levinsky, A. Ockenfels, T. Uske, and T. Weiland (2007). Buying a pig in a poke: An experimental study of unconditional veto power. *Journal of Economic Psychology 28*(6), 692–703.

Grechenig, K., A. Nicklisch, and C. Thöni (2010). Punishment despite reasonable doubt — A public goods experiment with sanctions under uncertainty. *Journal of Empirical Legal Studies 7*(4), 847–867.

Green, E. J. and R. H. Porter (1984). Noncooperative collusion under imperfect price information. *Econometrica 52*(1), 87–100.

Greiner, B. (2004). The online recruitment system orsee 2.0 — A guide for the organization of experiments in economics. *Working paper 10*(23), 63–104.

Guala, F. (2012). Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate. *Behavioral and Brain Sciences 35*(1), 1–15.

Guillen, P., C. Schwieren, and G. Staffiero (2006). Why feed the Leviathan? *Public Choice 130*(1-2), 115–128.

Gürerk, Ö. (2013). Social learning increases the acceptance and the efficiency of punishment institutions in social dilemmas. *Journal of Economic Psychology 34*, 229–239.

Gürerk, Ö., B. Irlenbusch, and B. Rockenbach (2006). The competitive advantage of sanctioning institutions. *Science 312*(5770), 108–111.

Gürerk, Ö., B. Irlenbusch, and B. Rockenbach (2009). Voting with feet: Community choice in social dilemmas. *Working paper*.

Gächter, S. (2007). Conditional cooperation. Behavioral regularities from the lab and the field and their policy implications. In B. Frey and A. Stutzer (Eds.), *Economics and Psychology. A Promising New Cross-Disciplinary Field.* Cambridge, MA: MIT Press.

Gächter, S. and B. Herrmann (2009). Reciprocity, culture and human cooperation: Previous insights and a new cross-cultural experiment. *Philosophical Transactions of the Royal Society B: Biological Sciences 364*(1518), 791–806.

Gächter, S. and C. Thöni (2011). Micromotives, microstructure, and macrobehavior: The case of voluntary cooperation. *Journal of Mathematical Sociology 35*(1-3), 26–65.

Hauser, O. P., D. G. Rand, A. Peysakhovich, and M. A. Nowak (2014). Cooperating with the future. *Nature 511*(7508), 220–223.

Heckathorn, D. D. (1989). Collective action and the second-order free-rider problem. *Rationality and Society 1*(1), 78–100.

Henrich, J., R. McElreath, A. Barr, J. Ensminger, C. Barrett, A. Bolyanatz, J. C. Cardenas, M. Gurven, E. Gwako, N. Henrich, C. Lesorogol, F. Marlowe, D. Tracer, and J. Ziker (2006). Costly punishment across human societies. *Science 312*(5781), 1767–1770.

Herrmann, B. and C. Thöni (2009). Measuring conditional cooperation: A replication study in russia. *Experimental Economics 12*(1), 87–92.

Herrmann, B., C. Thöni, and S. Gächter (2008). Antisocial punishment across societies. *Science 319*(5868), 1362–1367.

Hobbes, T. (1651). *Leviathan or the Matter, Forme and Power of a Commonwealth, Ecclesiasticall and Civill.* Bristol: Thoemmes.

http://www.europa.eu (nov. 2010).

http://www.un.org (nov. 2010).

Isaac, R. M. and J. M. Walker (1988). Communication and free-riding behavior: The voluntary contribution mechanism. *Economic inquiry 26*(4), 585–608.

Jackson, M. and D. Cox (2013). The principles of experimental design and their application in sociology. *Annual Review of Sociology 39*, 27–49.

Jost, J. and A. Kay (2010). Social justice. In S. T. Fiske, D. T. Gilbert, and G. Lindzey (Eds.), *Handbook of Social Psychology*. Hoboken, NJ: John Wiley & Sons.

Kamei, K. (2012). From locality to continent: A comment on the generalization of an experimental study. *Journal of Socio-Economics 41*(2), 207–210.

Kamei, K. (2014). Conditional punishment. *Economics Letters 124*(2), 199–202.

Kamei, K., L. Putterman, and J.-R. Tyran (2011). State or nature? Formal vs. informal sanctioning in the voluntary provision of public goods. *Working paper*.

Kingsley, D. C. and T. C. Brown (2015). Endogenous and costly institutional deterrence. *Applied Economics Letters 22*(7), 544–548.

Klempt, C. (2012). Fairness, spite, and intentions: Testing different motives behind punishment in a prisoners' dilemma game. *Economics Letters 116*(3), 429–431.

Kocher, M., P. Martinsson, and M. Visser (2008). Does stake size matter for cooperation and punishment? *Economics Letters 99*(3), 508–511.

Kocher, M. G., T. Cherry, S. Kroll, R. J. Netzer, and M. Sutter (2008). Conditional cooperation on three continents. *Economics Letters 101*(3), 175–178.

Kollock, P. (1998). Social dilemmas: The anatomy of cooperation. *Annual Review of Sociology*, 183–214.

Ledyard, J. O. (1995). Public goods: A survey of experimental research. In J. Kagel and A. Roth (Eds.), *Handbook of Experimental Economics*. Princeton, MA: Princeton University Press.

Leventhal, G. S. (1976). The distribution of rewards and resources in groups and organizations. *Advances in Experimental Social Psychology 9*, 91–131.

Leventhal, G. S. (1980). *What Should Be Done With Equity Theory?* New York, NY: Springer.

Markussen, T., L. Putterman, and J.-R. Tyran (2014). Self-organization for collective action: An experimental study of voting on sanction regimes. *Review of Economic Studies 81*(1), 301–324.

Marlowe, F. W., J. C. Berbesque, A. Barr, C. Barrett, A. Bolyanatz, J. C. Cardenas, J. Ensminger, M. Gurven, E. Gwako, J. Henrich, C. Lesorogol, R. McElreath, and D. Tracer (2008). More 'altruistic' punishment in larger societies. *Proceedings of the Royal Society B: Biological Sciences 275*(1634), 587–592.

Masclet, D., C. Noussair, S. Tucker, and M. Villeval (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *American Economic Review 93*(1), 366–380.

Messick, D. M. and K. P. Sentis (1979). Fairness and preference. *Journal of Experimental Social Psychology 15*(4), 418–434.

Milinski, M. and B. Rockenbach (2011). On the interaction of the stick and the carrot in social dilemmas. *Journal of Theoretical Biology*.

Monin, B. (2007). Holier than me? Threatening social comparison in the moral domain. *International Review of Social Psychology 20*(1), 53–68.

Mulder, L. B. (2008). The difference between punishments and rewards in fostering moral concerns in social decision making. *Journal of Experimental Social Psychology 44*(6), 1436–1443.

Mulder, L. B. and R. M. Nelissen (2010). When rules really make a difference: The effect of cooperation rules and self-sacrificing leadership on moral norms in social dilemmas. *Journal of Business Ethics 95*(1), 57–72.

Mulder, L. B., E. Van Dijk, D. De Cremer, and H. A. Wilke (2006). Undermining trust and cooperation: The paradox of sanctioning systems in social dilemmas. *Journal of Experimental Social Psychology 42*(2), 147–162.

Nese, A. and P. Sbriglia (2009). Individuals' coting choice and cooperation in repeated social dilemma cames. *Working paper*.

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics 92*(1), 91–112.

Nikiforakis, N. (2014). Self-governance through altruistic punishment? In P. A. M. Van Lange, B. Rockenbach, and T. Yamagishi (Eds.), *Reward and Punishment in Social Dilemmas.* Oxford: Oxford University Press.

Nikiforakis, N. and D. Engelmann (2011). Altruistic punishment and the threat of feuds. *Journal of Economic Behavior & Organization 78*(3), 319–332.

Nikiforakis, N. and H.-T. Normann (2008). A comparative statics analysis of punishment in public-good experiments. *Experimental Economics 11*(4), 358–369.

North, D. C. (1990). *Institutions, Institutional Change and Economic Performance.* Cambridge: Cambridge University Press.

Nosenzo, D. and M. Sefton (2012). Promoting cooperation: The distribution of reward and punishment power. *Working paper*.

Noussair, C. N., D. van Soest, and J. Stoop (2011). Punishment, reward, and cooperation in a framed field experiment. *Social Choice and Welfare*, 1–23.

Ockenfels, A. and W. Raub (2010). Rational und fair. In G. Albert and S. Sigmund (Eds.), *Soziologische Theorie kontrovers. Sonderheft 50, Kölner Zeitschrift für Soziologie und Sozialpsychologie.*

Oliver, P. (1980). Rewards and punishments as selective incentives for collective action: Theoretical investigations. *American Journal of Sociology*, 1356–1375.

Olson, M. (1965). *The logic of Collective Action: Public Goods and the Theory of Groups*, Volume 124. Cambridge, MA: Harvard University Press.

Ones, U. and L. Putterman (2007). The ecology of collective action: A public goods and sanctions experiment with controlled group formation. *Journal of Economic Behavior & Organization 62*(4), 495–521.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action.* Cambridge: Cambridge University Press.

Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives 14*(3), 137–158.

Ostrom, E. (2010). Beyond markets and states: Polycentric governance of complex economic systems. *American Economic Review*, 641–672.

Ostrom, E. (2012). Coevolving relationships between political science and economics. *Rationality, Markets and Morals 3*(54).

Ostrom, E., J. Walker, and R. Gardner (1992). Covenants with and without a sword: Self-governance is possible. *American Political Science Review 86*(2), 404–417.

Parsons, T. (1937). *The Structure of Social Action. Volume I: Marshall, Pareto, Durkheim.* New York, NY: Free Press.

Patel, A., E. Cartwright, and M. Van Vugt (2010). Punishment cannot sustain cooperation in a public good game with free-rider anonymity. *Working papers in economics 451*.

Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature*, 7–63.

Putterman, L., J.-R. Tyran, and K. Kamei (2011). Public goods and voting on formal sanction schemes. *Journal of Public Economics 95*(9), 1213–1222.

Rand, D. G., J. J. Armao IV, M. Nakamaru, and H. Ohtsuki (2010). Anti-social punishment can prevent the co-evolution of punishment and cooperation. *Journal of Theoretical Biology 265*(4), 624–632.

Rand, D. G., A. Dreber, T. Ellingsen, D. Fudenberg, and M. A. Nowak (2009). Positive interactions promote public cooperation. *Science 325*(5945), 1272–1275.

Rand, D. G. and M. A. Nowak (2011). The evolution of antisocial punishment in optional Public Goods Games. *Nature Communications 2*, 434.

Rapoport, A. and J. A. Sundali (1996). Ultimatums in two-person bargaining with one-sided uncertainty: Offer games. *International Journal of Game Theory 25*(4), 475–494.

Rapoport, A., J. A. Sundali, and D. A. Seale (1996). Ultimatums in two-person bargaining with one-sided uncertainty: Demand games. *Journal of Economic Behavior & Organization 30*(2), 173–196.

Raub, W., V. Buskens, and R. Corten (2015). Social dilemmas and cooperation. In N. Braun and N. Saam (Eds.), *Handbuch Modellbildung und Simulation in den Sozialwissenschaften*. Wiesbaden: VS-Verlag für Sozialwissenschaften / Springer Fachmedien.

Rockenbach, B. and M. Milinski (2006). The efficient interaction of indirect reciprocity and costly punishment. *Nature 444*(7120), 718–723.

Rockenbach, B. and I. Wolff (2009). Institution design in social dilemmas: How to design if you must? *Working paper*.

Rode, J. and M. Le Menestrel (2011). The influence of decision power on distributive fairness. *Journal of Economic Behavior & Organization 79*(3), 246–255.

Sally, D. (1995). Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. *Rationality and Society 7*(1), 58–92.

Schroeder, D. A., J. E. Steel, A. J. Woodell, and A. F. Bembenek (2003). Justice within social dilemmas. *Personality and Social Psychology Review 7*(4), 374–387.

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

Selten, R. (1967). Die strategiemethode zur erforschung des eingeschräncht rationalen verhaltens in rahmen eines oligopolexperiments. In H. Sauermann (Ed.), *Beiträge zur experimentellen Wirtschaftforschung*. Tübingen: JCB Mohr.

Sen, A. (1995). Rationality and social choice. *American Economic Review 85*(1), 1–24.

Strimling, R. and K. Eriksson (2014). Regulating the regulation: Norms about how people may punish each other. In P. Van Lange, B. Rockenbach, and T. Yamagishi (Eds.), *Reward and Punishment in Social Dilemmas*. Oxford: Oxford University Press.

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

Tabibnia, G. and M. D. Lieberman (2007). Fairness and cooperation are rewarding. *Annals of the New York Academy of Sciences 1118*(1), 90–101.

Tenbrunsel, A. E. and D. M. Messick (1999). Sanctioning systems, decision frames, and cooperation. *Administrative Science Quarterly 44*(4), 684–707.

Thibaut, J. W. and L. Walker (1975). *Procedural justice: A psychological analysis*. Hinsdale, IL: NJ Erlbaum Associates.

Thöni, C., J.-R. Tyran, and E. Wengström (2012). Microfoundations of social capital. *Journal of Public Economics 96*(7), 635–643.

Thye, S. R. (2007). Logical and philosophical foundations of experimental research in the social sciences. *Laboratory experiments in the social sciences*, 57–86.

Traulsen, A., T. Röhl, and M. Milinski (2012). An economic experiment reveals that humans prefer pool punishment to maintain the commons. *Proceedings of the Royal Society B: Biological Sciences*, rspb20120937.

Turnbull, C. (1961). *The Forest People*. New York, NY: Simon and Schuster.

Tyran, J.-R. and L. P. Feld (2006). Achieving compliance when legal sanctions are non-deterrent. *Scandinavian Journal of Economics 108*(1), 135–156.

Van Lange, P. A. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology 77*(2), 337.

Van Miltenburg, N., V. Buskens, and D. Barrera (2014). Determinants of perceived fairness of punishment and reward in Public Goods Games. *Working paper*.

Van Miltenburg, N., V. Buskens, D. Barrera, and W. Raub (2014). Implementing punishment and reward in the Public Goods Game: The effect of individual and collective decision rules. *International Journal of the Commons*, 47–78.

Van Miltenburg, N., V. Buskens, and W. Raub (2014). n-Person Prisoner's Dilemmas with endogenous peer punishment institutions - The role of noise. *Working paper*.

Van Miltenburg, N., W. Przepiorka, and V. Buskens (2014). Collective decision rules for implementing punishment in n-person Prisoner's Dilemmas with noise in the display of contributions. *Working paper*.

Van Prooijen, J.-W., M. Gallucci, and G. Toeset (2008). Procedural justice in punishment systems: Inconsistent punishment procedures have detrimental effects on cooperation. *British Journal of Social Psychology 47*(2), 311–324.

Veszteg, R. and E. Narhetali (2010). Public-good games and the balinese. *International Journal of Social Economics 37*(9), 660–675.

Voss, T. (1985). *Rationale Akteure und soziale Institutionen: Beiträge zu einer endogenen Theorie des sozialen Tauschs*. Munchen: Oldenbourg.

Voss, T. (2001). Game-theoretical perspectives on the emergence of social norms. In M. Hechter and K. Opp (Eds.), *Social Norms*. New York, NY: Russell Sage Foundation.

Vyrastekova, J. and D. Van Soest (2008). On the (in) effectiveness of rewards in sustaining cooperation. *Experimental Economics 11*(1), 53–65.

Walker, J. and M. Halloran (2004). Rewards and sanctions and the provision of public goods in one-shot settings. *Experimental Economics 7*(3), 235–247.

Wiedemann, V., D. Barrera, and V. Buskens (2011). Comparing consequences of carrots and sticks on cooperation in repeated Public Good Games. *Working paper*.

Wilke, H. A. (1991). Greed, efficiency and fairness in resource management situations. *European Review of Social Psychology 2*(1), 165–187.

Wu, J. and R. Axelrod (1995). How to cope with noise in the iterated Prisoner's Dilemma. *Journal of Conflict Resolution*, 183–189.

Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology 51*(1), 110.

Samenvatting

Inleiding

Sociale groepen zoals teams, buurten en organisaties, functioneren vaak het beste wanneer hun leden zich inzetten voor het belang van de groep, door bijvoorbeeld tijd of geld bij te dragen. Coöperatieproblemen ontstaan wanneer groepsleden een reden hebben om hun eigen belang boven het groepsbelang te stellen. In een bedrijf of instituut kunnen de medewerkers van een afdeling bijvoorbeeld samen verantwoordelijk zijn voor het schoonhouden van de koffiehoek. Alle medewerkers hebben belang bij een schone koffiehoek. Zodra een van de medewerkers de koffiehoek schoonmaakt, profiteert daarom iedereen op de afdeling hiervan. Deze medewerker zou echter tijd en moeite besparen, en even zo goed profiteren van een schone koffiehoek, wanneer hij of zij wacht tot iemand anders schoonmaakt. Maar als alle medewerkers wachten tot een ander schoonmaakt, wordt de koffiehoek een puinhoop en zijn de medewerkers slechter af dan wanneer ze zich gezamenlijk hadden ingezet voor een schone koffiehoek.

In dit proefschrift worden coöperatieproblemen onderzocht waarbij de betrokken actoren de mogelijkheid hebben elkaar voor hun bijdrage aan het collectieve belang positief of negatief te sanctioneren, oftewel, te belonen of te straffen. De medewerkers die een koffiehoek delen zouden collega's die nooit schoonmaken kunnen straffen, bijvoorbeeld door ze verantwoordelijk te maken voor een andere vervelende taak, of ze zouden collega's die vaak schoonmaken kunnen belonen met een compliment. Als zulke sancties worden uitgedeeld of als alleen al de mogelijkheid bestaat dat dit gebeurt, zou dat de medewerkers ertoe kunnen aanzetten vaker schoon te maken. De mogelijkheid om anderen te sanctioneren kan daarom worden gezien als een 'institutie' ter bevordering van coöperatie. In dit proefschrift wordt er vanuit gegaan dat aan het uitdelen van sancties persoonlijke kosten verbonden zijn. Hierdoor ontstaat een tweede-orde coöperatieprobleem, waarbij alle betrokken actoren een reden hebben om zelf niet te sanctioneren, maar wel te profiteren van de verhoogde coöperatie die volgt wanneer ander groepsleden sancties uitdelen. Uit voorgaand onderzoek is echter gebleken dat veel mensen in dergelijke situaties bereid zijn te sanctioneren ondanks de daaraan verbonden kosten, en dat de beschikbaarheid van dergelijke instituties bijdraagt aan het bevorderen van coöperatie.

Dit proefschrift behandelt drie thema's rond coöperatieproblemen met wederzijdse sanctioneringsinstituties. Ten eerste worden instituties waarin iedere actor individueel beslist een ander wel of niet te sanctioneren vergeleken met instituties waarin de betrokken actoren gezamenlijk de beslissing nemen om een groepslid te sanctioneren. Hierbij gaat het zowel om het effect van deze verschillende instituties op coöperatie als om de mate waarin geïmplementeerde sancties als eerlijk ervaren worden. Ten tweede worden coöperatieproblemen waarin actoren elkaars bijdrage aan het collectieve belang altijd correct observeren, vergeleken met coöperatieproblemen waarin actoren soms verkeerde informatie ontvangen over de bijdragen van anderen. Verkeerde informatie betekent dat ondanks dat iemand veel aan het collectieve goed bijdraagt, een of meerdere groepsleden dit kunnen observeren alsof deze persoon niets bijdraagt, terwijl iemand die niets bijdraagt kan worden geobserveed alsof hij of zij juist veel bijdraagt. Ten derde wordt onderzocht of actoren die betrokken zijn bij een coöperatieprobleem er zelf voor kiezen om in hun groep de mogelijkheid te hebben elkaar te straffen.

Theoretische en empirische benadering

Hoofdstuk 2 tot en met 5 beschrijven vier empirische studies naar een of meerdere van de drie hoofdthema's. In deze vier hoofdstukken worden coöperatieproblemen gemodelleerd in het eenmalige publiekgoedspel (PGS) of n-personen gevangenendilemma (GD). 'Eenmalig' houdt in dat actoren een spel slechts één keer in een bepaalde groepssamenstelling spelen, en de groepen na deze interactie opnieuw ingedeeld worden. Er is onder andere voor eenmalige interacties gekozen, omdat actoren in herhaalde spellen de mogelijkheid hebben om elkaar op alternatieve manieren te sanctioneren, bijvoorbeeld door meer of minder bij te dragen afhankelijk van het eerdere gedrag van anderen. Door deze mogelijkheid van herhaalde interacties uit te sluiten, wordt het pure effect van de sanctioneringsinstitutie geïsoleerd ten opzichte van deze alternatieve effecten.

In zowel het PGS als het n-personen GD ontvangen alle actoren in een groep van grootte n een startbedrag w. De actoren beslissen vervolgens tegelijkertijd en onafhankelijk van elkaar of ze hun startbedrag storten op 'groepsrekening'. In een GD kunnen de actoren er alleen voor kiezen om ofwel niets ofwel het volledige startbedrag te storten (voor actor i's bijdrage c_i geldt $c_i \in \{0, w\}$), terwijl actoren in een PGD kunnen beslissen om een fractie van hun startbedrag bij te dragen ($ci \in [0, w]$). Nadat iedereen een keuze heeft gemaakt wordt het totaal aan bijdragen $c = \sum c_i$ vermenigvuldigd met een factor m, waarbij $1 < m < n$. De factor m vertegenwoordigt de toegevoegde waarde die individuele bijdragen hebben voor de hele groep. De uitkomst voor de groep als geheel ($nw - c + mc$) stijgt naarmate er meer op de groepsrekening gestort wordt. De uitkomst voor een individu ($w - c_i + mc/n$) is echter hoger, naarmate het individu minder bijdraagt, ongeacht de bijdragen van anderen. Dit maakt het PGS en het GD tot klassieke voorbeelden van coöperatieproblemen.

De mogelijkheid tot wederzijdse sanctionering wordt in dit proefschrift gemodelleerd in een tweede fase van het PGS of GD. In de meest simpele vorm van deze fase observeren actoren de bijdragen c_j van ieder ander groepslid $j \neq i$. Vervolgens kunnen alle actoren voor ieder ander beslissen wel of niet te sanctioneren. Wanneer i besluit om j te sanctioneren, betaalt i hiervoor een bedrag $a > 0$. In het geval van beloning ontvangt j daardoor $b > a$, in het geval van straf verliest j het bedrag b. Verderop wordt beschreven hoe het PGS en GD aangepast kunnen worden om collectieve sanctioneringsbeslissingen, verkeerde observaties, en endogene instituties te implementeren.

In ieder hoofdstuk waarin gedrag in het PGS of GD wordt onderzocht, wordt telkens gebruik gemaakt van dezelfde theoretische benadering. Alleen in hoofdstuk 3 wordt het gedrag in deze spellen niet onderzocht. In hoofdstuk 3 worden bestaande theorieën over welke regels als eerlijk ervaren worden, toegepast op wederzijdse sanctionering. In de overige empirische hoofdstukken wordt ten eerste beschreven hoe actoren zouden handelen onder standaard aannames uit de speltheorie. Deze aannameshouden in dat alle actoren alleen hun eigen belang nastreven (egoïsme), precies weten met welke beslissingen ze hun eigen verwachte uitkomst maximaliseren (rationaliteit) en ook egoïsme en rationaliteit verwachten van hun groepsgenoten. Omdat het steeds om eenmalige interacties gaat, en aan het uitdelen van sancties kosten a verbonden zijn, volgt uit deze aannamesdat actoren geen sancties uitdelen en ook niet verwachten dat anderen dit doen. In eenmalige interacties kunnen actoren hiermee ten slotte niet het gedrag van anderen waarmee ze later weer te maken krijgen beïnvloeden. Actoren verwachten dus geen straffen of beloningen te ontvangen ongeacht hun bijdrage, en maximaliseren daarom hun verwachte uitkomst door niets bij te dragen. Onder deze aannames worden dus geen bijdragen en geen sanctionering verwacht.

Eerder onderzoek wijst uit dat wanneer mensen aan het PGS of GD deelnemen, hun gedrag vaak niet aan de aannames van rationaliteit en egoïsme voldoet. Een groot aantal studies laat zien dat ongeveer 50% van de mensen geclassificeerd kan worden als 'conditionele coöperator', wat inhoudt dat ze meer bijdragen of eerder geneigd zijn bij te dragen, naarmate ze verwachten dat anderen meer bijdragen, ook als ze niet voor hun beslissing gesanctioneerd kunnen worden. Wanneer wederzijdse sanctionering mogelijk is, is een groot deel van de mensen bereid anderen die veel bijdragen te belonen en anderen die weinig bijdragen te straffen, ook als hier kosten aan verbonden zijn. Vanaf nu wordt hiernaar gerefereerd als 'prosociale sanctionering', omdat zulke sancties hoge bijdragen zouden moeten bevorderen. Ten slotte wordt vaak gevonden dat er niet alleen prosociaal wordt gesanctioneerd, maar dat sancties ook op een 'antisociale' manier worden gebruikt. Dit betekent dat actoren anderen belonen die weinig bijdragen, of anderen straffen die veel bijdragen. Typische bevindingen geven aan dat rond de 20% van het totale aantal sancties antisociaal is. Naar aanleiding van deze bevindingen wordt er in de empirische hoofdstukken vanuit gegaan dat een aanzienlijk aantal actoren bereid is te sanctioneren, dat een klein deel antisociaal sanctioneert, en dat veel actoren bereid zijn bij te dragen zolang anderen dit ook doen, zelfs als sanctionering niet mogelijk is.

Vervolgens worden hypothesen opgesteld over wat de aanwezigheid van deze 'typen' actoren impliceert wanneer verschillende sanctioneringsinstituties aan het GD of PGS worden toegevoegd.

Voor de empirische test van hypothesen wordt gebruik gemaakt van laboratoriumexperimenten, waarin proefpersonen in groepen van 4 of 6 personen anoniem deelnemen aan series van varianten van het eenmalige PGS of GD. Laboratoriumexperimenten bieden de mogelijkheid gecontroleerd verschillende condities met elkaar te vergelijken door storende factoren uit te sluiten. Laboratoriumexperimenten zijn daarom bij uitstek geschikt voor het vaststellen van causale relaties. De uitkomsten van het PGS of GD worden aan proefpersonen weergegeven als punten, die aan het einde van een experimentele sessie worden uitbetaald. De proefpersonen hebben er dus belang bij zo veel mogelijk punten te verdienen.

In de volgende paragrafen wordt telkens een van de drie hoofdthema's besproken. Per hoofdthema worden ten eerste kort de theoretische intuïties geschetst. Vervolgens worden de belangrijkste resultaten beschreven. Ten slotte worden aanbevelingen gedaan voor vervolgonderzoek. Deze samenvatting sluit af met een aantal algemene aanbevelingen voor de collega's die samen verantwoordelijk zijn voor het schoonhouden van de koffiehoek.

Collectieve beslisregels

De overgrote meerderheid van het onderzoek naar instituties met wederzijdse sanctionering focust op instituties waarin iedere actor individueel kan beslissen om ieder ander wel of niet te belonen of te straffen (vanaf nu individuele beslisregel, afgekort IBR). In dit proefschrift worden zulke instituties vergeleken met instituties waarin een bepaald percentage van de groepsleden, bijvoorbeeld een meerderheid, gezamenlijk besluit om iemand te belonen of te straffen (vanaf nu collectieve beslisregel, afgekort CBR). In alledaagse coöperatieproblemen in kleinschalige groepen of gemeenschappen wordt de beslissing om iemand wel of niet te sanctioneren vaak gezamenlijk genomen. In het GD of PGS houdt een CBR in dat een sanctie van actor i gericht aan actor j alleen wordt geïmplementeerd wanneer een bepaald aantal groepsleden tegelijkertijd besluit om dezelfde j te sanctioneren. Wanneer niet genoeg actoren dezelfde j sanctioneren, wordt de sanctie niet uitgevoerd. Degenen die sanctioneerden betalen dan niet de kosten a en het beoogde doelwit verliest of verkrijgt niets. Groepsleden worden niet geïnformeerd over sancties die niet zijn uitgevoerd.

CBRs hebben bij het bevorderen van coöperatie een voordeel en een nadeel ten opzichte van IBRs. Aan de ene kant is de kans kleiner dat antisociale sancties, die coöperatie schaden, bij gebruik van een CBR door de groep worden goedgekeurd. Hoe meer collectieve overeenstemming gevraagd wordt, hoe kleiner de kans dat een antisociale sanctie wordt uitgevoerd. Aan de andere kant betekent hogere vereiste overeenstemming dat de kans kleiner wordt dat genoeg actoren tegelijkertijd dezelfde persoon sanctioneren om prosociale sancties te implementeren. Hoe meer overeenstemming gevraagd wordt, hoe groter de kans dat er niet genoeg collectieve overeenstemming is om voldoende prosociale sancties uit te voeren om coöperatie te bevorderen.

Naast het mogelijke effect van CBRs op coöperatie, wordt in dit proefschrift onderzocht hoe actoren de sancties ervaren die in sanctioneringsinstituties met verschillende beslisregels worden geïmplementeerd. Hierbij wordt gekeken naar de mate waarin actoren de sancties die ze zelf ontvangen en die de andere groepsleden ontvangen als eerlijk beoordelen. Voor sancties die groepsleden zelf ontvangen wordt verwacht, dat deze als eerlijker ervaren worden naarmate meer collectieve overeenstemming vereist is. Voor sancties ontvangen door anderen wordt verwacht dat actoren het als eerlijk ervaren wanneer ze zelf een hoge mate van autonomie ervaren in de beslissing om iemand te sanctioneren, dus naarmate een groter aandeel van de door hun voorgestelde sancties onder een CBR wordt geïmplementeerd.

In hoofdstuk 2 en 4 wordt coöperatie bij een IBR met coöperatie bij verschillende CBRs vergeleken, waarbij actoren elkaar ofwel konden belonen ofwel konden straffen. Zoals verwacht laten de resultaten ziendat veel actoren bereid zijn prosociaal te sanctioneren, met andere woorden anderen te straffen die weinig bijdragen of anderen te belonen die veel bijdragen. Slechts een klein aantal actoren sanctioneert antisociaal. Bij een IBR, waarbij al deze sancties worden geïmplementeerd, wordt dan ook aanmerkelijk hogere coöperatie geobserveerd dan wanneer sanctionering niet mogelijk is. Dit geldt zowel voor de interacties met de mogelijkheid tot belonen als voor interacties met de mogelijkheid tot straffen, hoewel het straffen tot hogere coöperatie leidt dan het belonen. In hoofdstuk 2 hebben antisociale sancties geen significante invloed op latere coöperatie, in hoofdstuk 4 hebben antisociale sancties een negatief effect. Dit zou kunnen verklaren waarom coöperatie bij een IBR in hoofdstuk 2 hoger is dan in hoofdstuk 4.

Naast de IBR worden in hoofdstuk 2 en 4 CBRs onderzocht waarbij collectieve sanctionering door een meerderheid of minder dan een meerderheid wordt vereist. Onder deze CBRs worden vrijwel geen antisociale sancties geïmplementeerd, terwijl een groot deel van de prosociale sancties wél wordt uitgevoerd. Deze CBRs zijn dan ook allen succesvol in het bevorderen van coöperatie ten opzichte van de situatie waarin geen sanctionering mogelijk is. Ook voor deze beslisregels leidt de mogelijkheid tot straf tot hogere coöperatie dan de mogelijkheid tot beloning. In hoofdstuk 2 wordt bij CBRs waarvoor een meerderheid vereist is, iets lagere coöperatie geobserveerd dan bij een IBR, terwijl in hoofdstuk 2 CBRs die een meerderheid of minder overeenstemming vereisen tot hogere coöperatie dan de IBR leiden. Hoewel de twee hoofdstukken vanwege verschillen in de experimentele designs moeilijk met elkaar te vergelijken zijn, is een mogelijke verklaring voor deze discrepantie in de resultaten dat in hoofdstuk 4 minder collectieve overeenstemming werd vereist dan in hoofdstuk 2. Een andere verklaring zou kunnen zijn dat straffen onder de IBR in hoofdstuk 4 gemiddeld te zwak waren om hoge bijdragen voor individuen rendabel te maken.

Ten slotte worden in hoofdstuk 2 CBRs beschouwd waarbij de gehele groep – exclusief het beoogde doelwit – unaniem moet besluiten om iemand te sanctioneren voordat een sanctie wordt uitgevoerd. Bij deze CBRs worden aanmerkelijk minder prosociale sancties geïmplementeerd dan bij de andere onderzochte beslisregels. De resultaten laten dan ook zien dat CBRs waarvoor unanimiteit is vereist minder effectief zijn in het bevorderen van coöperatie dan de andere beslisregels. Hoewel een CBR waarbij groepsleden door unanieme beslissingen straffen leidt tot hogere coöperatie dan wanneer geen sanctionering mogelijk is, geldt dit niet voor een CBR waarbij groepsleden door unanieme beslissingen belonen.

De CBRs die in dit proefschrift werden onderzocht vormen een zeer abstract model van collectieve sanctioneringsbeslissingen. In toekomstig onderzoek zouden aspecten van collectieve beslissingen die waarschijnlijk een rol spelen in coöperatieprobelemen buiten het laboratorium in meer detail onderzocht kunnen worden. Actoren zouden bijvoorbeeld de gelegenheid kunnen hebben om met elkaar te communiceren over welke bijdragen voor sanctionering in aanmerking komen, zodat collectieve sanctioneringsbeslissingen beter gecoördineerd kunnen worden. Daarnaast zou het in toekomstig onderzoek interessant zijn om collectieve sanctioneringsbeslissingen te combineren met mogelijkheden voor wederzijdse represailles, ofwel de mogelijkheid om anderen te bestraffen voor de straffen die ze uitdelen. Hoe meer actoren collectief beslissen om iemand te straffen, hoe lastiger het wordt om represailles in te zetten om prosociale straffen te ontmoedigen.

In hoofdstuk 3 wordt onderzocht hoe actoren sancties gericht aan zichzelf en gericht aan anderen ervaren die worden geïmplementeerd door middel van IBRs en CBRs. Er wordt geen systematische evidentie gevonden voor de verwachting dat de beslisregel een directe invloed heeft op de ervaren eerlijkheid van geïmplementeerde sancties. De resultaten tonen echter wel aan dat actoren de sancties die anderen ontvangen als eerlijker ervaren, naarmate een groter aandeel van de sancties die ze zelf voorstellen wordt geïmplementeerd door middel van een CBR. Bovendien ervaren actoren straffen die ze zelf ontvangen als oneerlijker naarmate ze meer gestraft worden, en beloningen als eerlijker naarmate ze meer worden beloond. In toekomstige studies zou kunnen worden onderzocht in hoeverre deze resultaten toe te wijzen zijn aan de operationalisering van collectieve beslissingen als CBRs zonder mogelijkheden tot communicatie, en in hoeverre deze resultaten ook gelden voor andere vormen van sanctioneringsinstituties, waarin bijvoorbeeld een formele instantie verantwoordelijk is voor het uitdelen van sancties.

Inaccurate informatie

Voor het tweede hoofdthema van dit proefschrift worden coöperatieproblemen onderzocht waarin de betrokken actoren mogelijk verkeede informatie over de bijdragen van anderen ontvangen. Naar de mogelijkheid dat informatie verkeerd is, wordt in het vervolg gerefereerd als 'ruis'. In dit proefschrift houdt ruis in dat één of meerdere actoren in het GD mogelijk observeren dat een groepsgenoot niets bijdroeg, terwijl deze persoon in werkelijkheid zijn of haar volledige startbedrag heeft bijgedragen, of dat één of meerdere actoren observeren dat iemand zijn of haar volledige startbedrag bijdroeg, terwijl deze persoon in werkelijkheid niets heeft bijgedragen. Of actor i de bijdrage van actor j correct observeert, wordt onafhankelijk bepaald voor iedere observatie van j's bijdrage door elke actor i.

Ruis kan de mate waarin instituties met wederzijdse bestraffing coöperatie bevorderen negatief beïnvloeden. Ten eerste kan het met ruis voorkomen dat een actor iemand die heeft bijgedragen aanziet voor iemand die niet bijdroeg, en daarom straf uitdeelt. Ten tweede kan het voorkomen dat een potentiële bestraffer iemand die niets bijdraagt observeert alsof deze persoon wél heeft bijgedragen, zodat het potentiële doelwit zijn of haar straf (deels) ontloopt. Wanneer een IBR wordt toegepast, worden alle straffen uitgevoerd die 'per ongeluk' gericht zijn aan actoren die bijdragen. Wanneer een CBR wordt toegepast is de kans dat zulke straffen worden uitgevoerd kleiner naarmate meer collectieve overeenstemming gevraagd wordt, omdat daarmee de kans kleiner wordt dat genoeg actoren een bijdrage verkeerd observeren om de straf te implementeren. CBRs hebben echter ook tot gevolg dat wanneer een deel van de potentiële prosociale straffers niet correct observeert dat iemand niet bijdraagt, de kans groot is dat collectieve overeenstemming niet bereikt wordt. Hoe meer overeenstemming gevraagd wordt, hoe kleiner de kans dat voldoende prosociale straffers een lage bijdrage correct observeren en bestraffen. CBRs hebben dus een voordeel ten opzichte van een IBR in coöperatieproblemen met ruis, maar ook een nadeel. Zowel het voordeel als het nadeel worden groter, naarmate meer collectieve overeenstemming is vereist.

Hoofdstuk 4 vergelijkt een IBR met een CBR waarvoor overeenstemming van een meerderheid van de overige groepsleden gevraagd wordt om een straf te implementeren, en een CBR waarvoor minder dan meerderheidsovereenstemming vereist is, in coöperatieproblemen met ruis. De resultaten wijzen uit dat de IBR coöperatie enigszins bevordert, terwijl beide CBRs niet in staat zijn om coöperatie tot boven het niveau te brengen dat bereikt wordt als geen sanctionering mogelijk is. Hoewel bij de CBRs vrijwel geen straffen geïmplementeerd worden die gericht zijn aan actoren die in werkelijkheid hebben bijgedragen, wordt ook een groot deel van de straffen gericht aan actoren die niets bijdroegen niet geïmplementeerd. Bij de IBR worden degenen die niet bijdragen wél voldoende gestraft om coöperatie te bevorderen. Het komt echter ook vaak voor dat actoren al dan niet 'per ongeluk' worden bestraft wanneer ze hun volledige startbedrag bijdragen. Hoewel zulke straffen toekomstige coöperatie niet significant negatief beïnvloeden, worden er aanzienlijke opbrengsten mee vernietigd (kosten a en effect b), zodat actoren met betrekking tot het aantal verdiende punten onder een IBR, evenals onder de CBRs, uiteindelijk slechter af zijn dan wanneer sanctionering niet mogelijk is.

In hoofdstuk 5 wordt een institutie met wederzijdse sanctionering door middel van een IBR onderzocht die actoren in coöperatieproblemen met ruis vrijwillig in hun groep implementeren. Hier bevordert de mogelijkheid tot sanctioneren coöperatie wél en verdienen actoren meer als ze voor deze mogelijkheid kiezen dan als ze dit niet doen, vooral wanneer sancties een hoog bedrag b van de opbrengsten van het doelwit vorderen. Desalniettemin zijn ook vrijwillig geïmplementeerde sanctioneringsinstituties in coöperatieproblemen met ruis niet zo succesvol in het bevorderen van coöperatie als in coöperatieproblemen waarin informatie altijd correct is.

Aangezien in alledaagse coöperatieproblemen vaak enige mate van ruis aanwezig is, plaatsen deze bevindingen een belangrijke beperking op de resultaten van voorgaand onderzoek waarin met ruis geen rekening is gehouden. Dit roept de vraag op in hoeverre de huidige resultaten ook gelden wanneer ruis op andere manieren wordt geïmplementeerd, bijvoorbeeld wanneer de kans op een verkeerde observatie hoger of lager is, of wanneer de kans dat één actor een bijdrage verkeerd observeert, samenhangt met de kans dat een ander dezelfde bijdrage verkeerd waarneemt. Ook zou toekomstig onderzoek duidelijk kunnen maken hoe ruis de beslissing van actoren om bij te dragen en om te sanctioneren precies beïnvloedt.

Endogene sanctioneringsinstituties

Evenals het overgrote deel van het voorgaand onderzoek beschouwen de vorige twee hoofdthema's uitsluitend situaties waarin sanctioneringsinstituties exogeen (door de onderzoeker) aan actoren worden opgelegd. Daarmee kan worden onderzocht welke gevolgen bepaalde instituties hebben voor gedrag in coöperatieproblemen. Het laatste hoofdthema van dit proefschrift draait om de vraag of actoren zelf deze gevolgen voorzien en er voor kiezen de mogelijkheid tot wederzijdse bestraffing door middel van een IBR in hun groep toe te laten wanneer ze de keuze hebben dit wel of niet te doen. Hiervoor wordt gekeken naar situaties waarin actoren in een GD voordat ze beslissen wel of niet bij te dragen aangeven (vanaf nu: 'stemmen') of ze de mogelijkheid tot bestraffen in de groep willen toelaten of niet. Tegelijkertijd stemmen actoren over de vraag of iedere uitgedeelde straf een hoog of een laag bedrag b van het doelwit vordert (vanaf nu: 'effectiviteit') in het geval dat de mogelijkheid tot bestraffen wordt toegelaten. De optie waar de meerderheid van de groep voor stemt, wordt geïmplementeerd. Zowel coöperatieproblemen waarin informatie altijd correct is als coöperatieproblemen met ruis worden onderzocht. Omdat met ruis de mogelijkheid bestaat dat actoren 'per ongeluk' gestraft worden wanneer ze bijdragen of straf ontlopen wanneer ze niet bijdragen, wordt verwacht dat actoren in het geval van ruis minder vaak de mogelijkheid tot bestraffen toelaten dan wanneer informatie altijd correct is, en minder vaak kiezen voor de hoge effectiviteit.

De resultaten van hoofdstuk 5 laten zien dat in coöperatieproblemen zonder ruis actoren er na enige ervaring vrijwel altijd voor kiezen om de mogelijkheid tot bestraffen toe te laten. Hierbij kiezen de meeste groepen voor straffen met hoge effectiviteit, en wordt vaker voor de lage effectiviteit gekozen naarmate straffen vaker antisociaal gebruikt worden. Zowel bijdragen als uitkomsten zijn hoger in groepen waar voor de mogelijkheid tot bestraffen gekozen wordt dan in groepen die er voor kiezen om deze mogelijkheid niet te hebben.

Voor coöperatieproblemen met ruis geven de resultaten van hoofdstuk 5 een ander beeld. Hier kiest een meerderheid van de groepen er voor om niet de mogelijkheid tot bestraffen te hebben, en als er toch voor de mogelijkheid tot bestraffen wordt gekozen, kiezen groepen het vaakst voor straffen met een lage effectiviteit. In groepen waar voor de mogelijkheid tot bestraffen wordt gekozen, zijn bijdragen en opbrengsten iets hoger dan in groepen waar niet voor deze mogelijkheid gekozen wordt. Omdat actoren vanwege de ruis echter niet elkaars werkelijke bijdragen te zien krijgen, observeren ze juist dat de mogelijkheid tot bestraffen tot lagere opbrengsten leidt. Bovendien stemmen actoren in coöperatieproblemen met ruis minder vaak voor de mogelijkheid tot bestraffen wanneer ze observeren dat degenen die bijdragen straf ontvangen.

De bevinding dat actoren de mogelijkheid voor het uitdelen van relatief zware straffen vrijwillig implementeren zolang deze straffen op een prosociale manier worden gebruikt, zou verder uitgewerkt kunnen worden in toekomstig onderzoek naar karakteristieken van duurzame instituties ter bevordering van coöperatie. Het blijft echter de vraag in hoeverre deze bevinding gegeneraliseerd kan worden naar andere vormen van sanctioneringsinstituties. Het is bijvoorbeeld niet duidelijk in hoeverre actoren in de huidige setting voor de mogelijkheid tot effectieve bestraffing kiezen omdat ze van mening zijn dat dit coöperatie het beste bevordert, of omdat ze zelf de gelegenheid willen hebben om zware straffen uit te delen.

Het sterke effect van ruis op de endogene vorming van instituties roept opnieuw vragen op naar de mate waarin onderzoeksresultaten die verkregen zijn in coöperatieproblemen met accurate informatie te generaliseren zijn naar situaties buiten het laboratorium. Een andere factor die, zoals ruis, in alledaagse situaties vaak aanwezig is en de endogene formatie van instituties zou kunnen beïnvloeden is bijvoorbeeld de mogelijkheid tot represailles. In toekomstige studies zou het interessant zijn om te onderzoeken of actoren er minder vaak voor kiezen om de mogelijkheid tot wederzijdse bestraffing toe te laten wanneer de mogelijkheid tot represailles bestaat. Ook zou het interessant zijn om te kijken naar endogene institutievorming in populaties waar straffen vaak antisociaal worden gebruikt. Ten slotte zou voor completer beeld van institutievorming gekeken moeten worden naar meerdere soorten instituties dan alleen wederzijdse sanctionering.

Het schoonhouden van de koffiehoek

Ter afsluiting van dit proefschrift worden de belangrijkste conclusies van de vier empirische hoofdstukken toegepast op de situatie van collega's die samen verantwoordelijk zijn voor het schoonhouden van de koffiehoek. De eenmalige coöperatieproblemen die voor de laboratoriumexperimenten zijn gebruikt waar de conclusies van dit proefschrift zich op baseren, verschillen natuurlijk op talloze manieren van de situatie van de medewerkers. Ondanks deze kanttekening lijkt het toch mogelijk om een aantal algemene conclusies te presenteren.

Een eerste set conclusies betreft situaties waarin de medewerkers altijd volledig en correct geïnformeerd zijn over de mate waarin hun collega's er aan hebben bijgedragen de koffiehoek schoon te houden. Zolang de meeste medewerkers sancties gebruiken om coöperatie te bevorderen, blijft de koffiehoek in dit geval schoon als iedereen zelf besluit anderen wel of niet te sanctioneren. Hierbij wordt samenwerking beter bevorderd door slecht gedrag te bestraffen dan door goed gedrag te belonen. De medewerkers zouden individuele sanctionering vrijwillig in de groep toe kunnen staan, en zelfs de voorkeur kunnen geven aan de mogelijkheid tot het uitdelen van zware straffen, die op hun beurt het schoonhouden van de koffiehoek zeer effectief bevorderen. Toch blijft het, wanneer iedereen zelf beslist om wel of niet te sanctioneren, mogelijk dat individuen sancties op een perverse manier gebruiken. Om dit te voorkomen, zou een institutie aangewend kunnen worden waarin medewerkers elkaar alleen mogen sanctioneren wanneer anderen het er mee eens zijn dat iemand gesanctioneerd dient te worden. Collectieve sanctioneringsbeslissingen zijn echter niet altijd even effectief in het bevorderen van coöperatie als individuele beslissingen. Bovendien zou niet te veel overeenstemming gevraagd moeten worden, omdat dit collectieve beslissingen inefficiënt maakt. Bovendien bestaat het risico dat medewerkers die iemand willen sanctioneren, maar niet genoeg collega's vinden die het voorstel ondersteunen, dit als oneerlijk ervaren.

Een tweede set conclusies betreft situaties waarin de medewerkers niet altijd van elkaars beslissing op de hoogte zijn. Collectieve strafbeslissingen zijn in dit geval niet in staat om schoonmaken te bevorderen, omdat het lastig is om collectief vast te stellen dat iemand niet heeft bijgedragen en dient te worden gestraft. De koffiehoek zou schoongehouden kunnen worden wanneer de medewerkers individueel kunnen beslissen om collega's te straffen, vooral wanneer deze institutie vrijwillig door de medewerkers wordt geïmplementeerd en wanneer relatief zware straffen worden toegestaan. Het is echter niet waarschijnlijk dat medewerkers deze institutie vrijwillig implementeren wanneer ze merken dat mensen die bijdragen toch een hoog risico lopen om gestraft te worden. Bovendien zorgt de verkeerde informatie er voor dat medewerkers zelf niet merken dat de mogelijkheid tot bestraffen coöperatie bevordert. Een dergelijke institutie zou dus aan de medewerkers moeten worden opgelegd.

Acknowledgments

Many people have directly or indirectly provided invaluable support during the writing of this thesis. First and foremost my supervisors Vincent Buskens, Werner Raub, and Davide Barrera, who make a great team and who supported me at every opportunity. Vincent, for the past seven years I could not have had a more helpful and inspiring daily supervisor. Your encouragement greatly influenced my decision to pursue a PhD in experimental sociology. Throughout the project, you were always available to help me with problems or questions, and always did so in a very open and cooperative manner. Your remarkable analytical insights often made me see my work in a different light and helped me to take a chapter to the next level. Werner, I highly appreciated your contributions, your encouragement, and your countless efforts to create opportunities for me. Despite your busy schedule, you often found time for providing insightful comments on the parts of the manuscript that probably needed it most. Retrospectively, I also appreciate how you diligently pointed out each flaw, inconsistency, and 'Verschlimmbesserung', which tremendously improved the end result. Davide, you secured the funding for this project and were especially involved in the early stages, where your research ideas and your knowledge of the literature have been invaluable.

Chapter 4 of this thesis was partly written during a research visit at Nuffield College, University of Oxford. I would like to thank the members of the college and of the Nuffield Centre for Experimental Social Science for their hospitality and their feedback. Especially the creative and well thought suggestions of my college supervisor and great colleague Wojtek Przepiorka have much improved the chapter. Sarah, Jorien, Line, Marisa, and Anna, thank you for the great time we spent in Oxford!

This project was embedded in the research group 'Cooperation in Social and Economic Relations'. Many thanks go out to the members and former members of this group. Vincenz, Rense, Ozan, David, Dominik, Charlotte, Nikki, and Bas, thank you for your input and for your help in finalizing and conducting the experiments. I always enjoyed working in the lab with you. Jeroen, thank you for your help in designing the statistical models.

Writing a thesis can be lonely, making it all the more important to be surrounded by a group of nice and cooperative colleagues. Fortunately, the Utrecht Department of Sociology comprises a lively community, where everyone is happy to help others with their work, and where there is always time for a chat, coffee, 'Wie is de Mol'-pool, or an after work game of Werewolves. I want to thank everyone at the department, and especially Mariska, Sarah, Sabrina, Wike, Anja, Jesper, Vincenz, Sanne, Tim, Esther, Zoli, David, Sara, Franciëlla, Marco, and Manja for their support and for the fun times. The administrative staff has been a great help, too, especially Ellen, who took care of the financial side of the project. I also thank the members of the Utrecht University Strategic Theme 'Institutions' for their refreshing suggestions. Finally, this thesis has greatly benefited from the stimulating environment and the outstanding PhD program of the ICS research school.

Natuurlijk is het ook belangrijk om af en toe even niet met het proefschrift bezig te zijn. Gelukkig had ik dan altijd mijn vrienden in de buurt voor een spelletjesavond, spinning, een kopje thee, een etentje, een feest, of een epische vakantie op IJsland. Inge, Cora, Nicole en Marleen, Fimke, Linde, Timothy, Anna, Olga, Alex en Sanne, Sarah, Vera, Jorien, jullie maakten de afgelopen jaren tot een hele leuke tijd. Bedankt voor de afleiding!

De Prangers wil ik graag bedanken voor de gezellige weekends in het rustieke Dedemsvaart, en voor hun geweldige kweek-, brouw- en kookkunsten, die er altijd voor zorgden dat ik weer met hernieuwde energie aan de slag kon.

Dit proefschrift had ik niet kunnen schrijven zonder de geweldige steun van mijn familie. Toos, bedankt voor je enthousiasme en voor je interesse in mijn werk. Sietse, je bent de coolste broer en ik word altijd blij van de gezelligheid die jij en Wieke uitstralen. Theo, je was en blijft het beste voorbeeld dat ik had kunnen hebben. Joke, ik vind het heel fijn om te weten dat je altijd achter me staat. Bedankt voor alles.

Casper, jouw liefde, steun en positiviteit maken mij altijd gelukkig, en ik kijk uit naar alles wat ons nog te wachten staat.

Curriculum Vitae

Nynke van Miltenburg was born in Baarn, The Netherlands on April 19, 1987. In 2005, she enrolled in the Sociology Bachelor's program at Utrecht University. During her Bachelor's, she spent an exchange period at the University of Cologne. In 2008, she proceeded with the Utrecht University Research Master's program Sociology and Social Research, which she graduated cum laude. In September 2010, she started working as a PhD candidate at the Interuniversity Center for Social Science Theory and Methodology (ICS) and the Department of Sociology of Utrecht University, where she wrote this thesis. During her PhD trajectory, Van Miltenburg spent a term as a visiting scholar at Nuffield College, University of Oxford.

ICS dissertation series

The ICS series presents dissertations of the Interuniversity Center for Social Science Theory and Methodology. Each of these studies aims at integrating explicit theory formation with state of the art empirical research or at the development of advanced methods for empirical research. The ICS was founded in 1986 as a cooperative effort of the universities of Groningen and Utrecht. Since 1992, the ICS expanded to the University of Nijmegen. Most of the projects are financed by the participating universities or by the Netherlands Organization for Scientific Research (NWO). The international composition of the ICS graduate students is mirrored in the increasing international orientation of the projects and thus of the ICS series itself.

1. Cornelis van Liere (1990). *Lastige leerlingen. Een empirisch onderzoek naar sociale oorzaken van probleemgedrag op basisscholen.* Amsterdam: Thesis Publishers.
2. Marco H.D. van Leeuwen (1990). *Bijstand in Amsterdam, ca. 1800-1850. Armenzorg als beheersings- en overlevingsstrategie.* ICS-dissertation, Utrecht.
3. Ineke Maas (1990). *Deelname aan podiumkunsten via de podia, de media en actieve beoefening. Substitutie of leereffecten?* Amsterdam: Thesis Publishers.
4. Marjolein I. Broese van Groenou (1991). *Gescheiden netwerken. De relaties met vrienden en verwanten na echtscheiding.* Amsterdam: Thesis Publishers.
5. Jan M.M. van den Bos (1991). *Dutch EC policy making. A model guided approach to coordination and negotiation.* Amsterdam: Thesis Publishers.
6. Karin Sanders (1991). *Vrouwelijke pioniers. Vrouwen en mannen met een 'mannelijke' hogere beroepsopleiding aan het begin van hun loopbaan.* Amsterdam: Thesis Publishers.
7. Sjerp de Vries (1991). *Egoism, altruism, and social justice. Theory and experiments on cooperation in social dilemmas.* Amsterdam: Thesis Publishers.
8. Ronald S. Batenburg (1991). *Automatisering in bedrijf.* Amsterdam: Thesis Publishers.
9. Rudi Wielers (1991). *Selectie en allocatie op de arbeidsmarkt. Een uitwerking voor de informele en geïnstitutionaliseerde kinderopvang.* Amsterdam: Thesis Publishers.
10. Gert P. Westert (1991). *Verschillen in ziekenhuisgebruik.* ICS-dissertation, Groningen.
11. Hanneke Hermsen (1992). *Votes and policy preferences. Equilibria in party systems.* Amsterdam: Thesis Publishers.
12. Cora J.M. Maas (1992). *Probleemleerlingen in het basisonderwijs.* Amsterdam: Thesis Publishers.
13. Ed A.W. Boxman (1992). *Contacten en carrière. Een empirisch theoretisch onderzoek naar de relatie tussen sociale netwerken en arbeidsmarktposities.* Amsterdam: Thesis Publishers.
14. Conny G.J. Taes (1992). *Kijken naar banen. Een onderzoek naar de inschatting van arbeidsmarktkansen bij schoolverlaters uit het middelbaar beroepsonderwijs.* Amsterdam: Thesis Publishers.
15. Peter van Roozendaal (1992). *Cabinets in multi party democracies. The effect of dominant and central parties on cabinet composition and durability.* Amsterdam: Thesis Publishers.
16. Marcel van Dam (1992). *Regio zonder regie. Verschillen in en effectiviteit van gemeentelijk arbeidsmarktbeleid.* Amsterdam: Thesis Publishers.
17. Tanja van der Lippe (1993). *Arbeidsverdeling tussen mannen en vrouwen.* Amsterdam: Thesis Publishers.
18. Marc A. Jacobs (1993). *Software: Kopen of kopiëren? Een sociaal wetenschappelijk onderzoek onder PC gebruikers.* Amsterdam: Thesis Publishers.
19. Peter van der Meer (1993). *Verdringing op de Nederlandse arbeidsmarkt. Sector- en sekseverschillen.* Amsterdam: Thesis Publishers.
20. Gerbert Kraaykamp (1993). *Over lezen gesproken. Een studie naar sociale differentiatie in leesgedrag.* Amsterdam: Thesis Publishers.
21. Evelien Zeggelink (1993). *Strangers into friends. The evolution of friendship networks using an individual oriented modeling approach.* Amsterdam: Thesis Publishers.
22. Jaco Berveling (1994). *Het stempel op de besluitvorming. Macht, invloed en besluitvorming op twee Amsterdamse beleidsterreinen.* Amsterdam: Thesis Publishers.
23. Wim Bernasco (1994). *Coupled careers. The effects of spouse's resources on success at work.* Amsterdam: Thesis Publishers.
24. Liset van Dijk (1994). *Choices in child care. The distribution of child care among mothers, fathers and non parental care providers.* Amsterdam: Thesis Publishers.
25. Jos de Haan (1994). *Research groups in Dutch sociology.* Amsterdam: Thesis Publishers.
26. Kwasi Boahene (1995). *Innovation adoption as a socio economic process. The case of the Ghanaian cocoa industry.* Amsterdam: Thesis Publishers.
27. Paul E.M. Ligthart (1995). *Solidarity in economic transactions. An experimental study of framing effects in bargaining and contracting.* Amsterdam: Thesis Publishers.
28. Roger Th. A.J. Leenders (1995). *Structure and influence. Statistical models for the dynamics of actor attributes, network structure, and their interdependence.* Amsterdam: Thesis Publishers.
29. Beate Völker (1995). *Should auld acquaintance be forgot...? Institutions of communism, the transition to capitalism and personal networks: the case of East Germany.* Amsterdam: Thesis Publishers.
30. Anna M. Cancrinus Matthijsse (1995). *Tussen hulpverlening en ondernemerschap. Beroepsuitoefening en taakopvattingen van openbare apothekers in een aantal West Europese landen.* Amsterdam: Thesis Publishers.

31. Nardi Steverink (1996). *Zo lang mogelijk zelfstandig. Naar een verklaring van verschillen in oriëntatie ten aanzien van opname in een verzorgingstehuis onder fysiek kwetsbare ouderen.* Amsterdam: Thesis Publishers.
32. Ellen Lindeman (1996). *Participatie in vrijwilligerswerk.* Amsterdam: Thesis Publishers.
33. Chris Snijders (1996). *Trust and commitments.* Amsterdam: Thesis Publishers.
34. Koos Postma (1996). *Changing prejudice in Hungary. A study on the collapse of state socialism and its impact on prejudice against gypsies and Jews.* Amsterdam: Thesis Publishers.
35. Jooske T. van Busschbach (1996). *Uit het oog, uit het hart? Stabiliteit en verandering in persoonlijke relaties.* Amsterdam: Thesis Publishers.
36. René Torenvlied (1996). *Besluiten in uitvoering. Theorieën over beleidsuitvoering modelmatig getoetst op sociale vernieuwing in drie gemeenten.* Amsterdam: Thesis Publishers.
37. Andreas Flache (1996). *The double edge of networks. An analysis of the effect of informal networks on cooperation in social dilemmas.* Amsterdam: Thesis Publishers.
38. Kees van Veen (1997). *Inside an internal labor market: Formal rules, flexibility and career lines in a Dutch manufacturing company.* Amsterdam: Thesis Publishers.
39. Lucienne van Eijk (1997). *Activity and well being in the elderly.* Amsterdam: Thesis Publishers.
40. Róbert Gál (1997). *Unreliability. Contract discipline and contract governance under economic transition.* Amsterdam: Thesis Publishers.
41. Anne Geerte van de Goor (1997). *Effects of regulation on disability duration.* ICS-dissertation, Utrecht.
42. Boris Blumberg (1997). *Das Management von Technologiekooperationen. Partnersuche und Verhandlungen mit dem Partner aus empirisch theoretischer Perspektive.* ICS-dissertation, Utrecht.
43. Marijke von Bergh (1997). *Loopbanen van oudere werknemers.* Amsterdam: Thesis Publishers.
44. Anna Petra Nieboer (1997). *Life events and well being: A prospective study on changes in well being of elderly people due to a serious illness event or death of the spouse.* Amsterdam: Thesis Publishers.
45. Jacques Niehof (1997). *Resources and social reproduction: The effects of cultural and material resources on educational and occupational careers in industrial nations at the end of the twentieth century.* ICS-dissertation, Nijmegen.
46. Ariana Need (1997). *The kindred vote. Individual and family effects of social class and religion on electoral change in the Netherlands, 1956-1994.* ICS-dissertation, Nijmegen.
47. Jim Allen (1997). *Sector composition and the effect of education on wages: an international Comparison.* Amsterdam: Thesis Publishers.
48. Jack B.F. Hutten (1998). *Workload and provision of care in general practice. An empirical study of the relation between workload of Dutch general practitioners and the content and quality of their Care.* ICS-dissertation, Utrecht.
49. Per B. Kropp (1998). *Berufserfolg im Transformationsprozeß. Eine theoretisch empirische Studie über die Gewinner und Verlierer der Wende in Ostdeutschland.* ICS-dissertation, Utrecht.
50. Maarten H.J. Wolbers (1998). *Diploma inflatie en verdringing op de arbeidsmarkt. Een studie naar ontwikkelingen in de opbrengsten van diploma's in Nederland.* ICS-dissertation, Nijmegen.
51. Wilma Smeenk (1998). *Opportunity and marriage. The impact of individual resources and marriage market structure on first marriage timing and partner choice in the Netherlands.* ICS-dissertation, Nijmegen.
52. Marinus Spreen (1999). *Sampling personal network structures: Statistical inference in ego-graphs.* ICS-dissertation, Groningen.
53. Vincent Buskens (1999). *Social networks and trust.* ICS-dissertation, Utrecht.
54. Susanne Rijken (1999). *Educational expansion and status attainment. A cross-national and over-time comparison.* ICS-dissertation, Utrecht.
55. Mérove Gijsberts (1999). *The legitimation of inequality in state-socialist and market societies, 1987-1996.* ICS-dissertation, Utrecht.
56. Gerhard G. Van de Bunt (1999). *Friends by choice. An actor-oriented statistical network model for friendship networks through time.* ICS-dissertation, Groningen.
57. Robert Thomson (1999). *The party mandate: Election pledges and government actions in the Netherlands, 1986-1998.* Amsterdam: Thela Thesis.
58. Corine Baarda (1999). *Politieke besluiten en boeren beslissingen. Het draagvlak van het mestbeleid tot 2000.* ICS-dissertation, Groningen.
59. Rafael Wittek (1999). *Interdependence and informal control in organizations.* ICS-dissertation, Groningen.

60. Diane Payne (1999). *Policy making in the European Union: An analysis of the impact of the reform of the structural funds in Ireland.* ICS-dissertation, Groningen.
61. René Veenstra (1999). *Leerlingen-klassen-scholen. Prestaties en vorderingen van leerlingen in het voortgezet onderwijs.* Amsterdam, Thela Thesis.
62. Marjolein Achterkamp (1999). *Influence strategies in collective decision making. A comparison of two models.* ICS-dissertation, Groningen.
63. Peter Mühlau (2000). *The governance of the employment relation. A relational signaling perspective.* ICS-dissertation, Groningen.
64. Agnes Akkerman (2000). *Verdeelde vakbeweging en stakingen. Concurrentie om leden.* ICS-dissertation, Groningen.
65. Sandra van Thiel (2000). *Quangocratization: Trends, causes and consequences.* ICS-dissertation, Utrecht.
66. Rudi Turksema (2000). *Supply of day care.* ICS-dissertation, Utrecht.
67. Sylvia E. Korupp (2000). *Mothers and the process of social stratification.* ICS-dissertation, Utrecht.
68. Bernard A. Nijstad (2000). *How the group affects the mind: Effects of communication in idea generating groups.* ICS-dissertation, Utrecht.
69. Inge F. de Wolf (2000). *Opleidingsspecialisatie en arbeidsmarktsucces van sociale wetenschappers.* ICS-dissertation, Utrecht.
70. Jan Kratzer (2001). *Communication and performance: An empirical study in innovation teams.* ICS-dissertation, Groningen.
71. Madelon Kroneman (2001). *Healthcare systems and hospital bed use*. ICS/NIVEL-dissertation, Utrecht.
72. Herman van de Werfhorst (2001). *Field of study and social inequality. Four types of educational resources in the process of stratification in the Netherlands.* ICS-dissertation, Nijmegen.
73. Tamás Bartus (2001). *Social capital and earnings inequalities. The role of informal job search in Hungary.* ICS-dissertation Groningen.
74. Hester Moerbeek (2001). *Friends and foes in the occupational career. The influence of sweet and sour social capital on the labour market.* ICS-dissertation, Nijmegen.
75. Marcel van Assen (2001). *Essays on actor perspectives in exchange networks and social dilemmas.* ICS-dissertation, Groningen.
76. Inge Sieben (2001). *Sibling similarities and social stratification. The impact of family background across countries and cohorts.* ICS-dissertation, Nijmegen.
77. Alinda van Bruggen (2001). *Individual production of social well being. An exploratory study.* ICS-dissertation, Groningen.
78. Marcel Coenders (2001). *Nationalistic attitudes and ethnic exclusionism in a comparative perspective: An empirical study of attitudes toward the country and ethnic immigrants in 22 countries.* ICS-dissertation, Nijmegen.
79. Marcel Lubbers (2001). *Exclusionistic electorates. Extreme right wing voting in Western Europe.* ICS-dissertation, Nijmegen.
80. Uwe Matzat (2001). *Social networks and cooperation in electronic communities. A theoretical-empirical analysis of academic communication and internet discussion groups.* ICS-dissertation, Groningen.
81. Jacques P.G. Janssen (2002). *Do opposites attract divorce? Dimensions of mixed marriage and the risk of divorce in the Netherlands.* ICS-dissertation, Nijmegen.
82. Miranda Jansen (2002). *Waardenoriëntaties en partnerrelaties. Een panelstudie naar wederzijdse invloeden.* ICS-dissertation, Utrecht.
83. Anne Rigt Poortman (2002). *Socioeconomic causes and consequences of divorce.* ICS-dissertation, Utrecht.
84. Alexander Gattig (2002). *Intertemporal decision making.* ICS-dissertation, Groningen.
85. Gerrit Rooks (2002). *Contract en conflict: Strategisch Management van Inkooptransacties.* ICS-dissertation, Utrecht.
86. Károly Takács (2002). *Social networks and intergroup conflict.* ICS-dissertation, Groningen.
87. Thomas Gautschi (2002). *Trust and exchange, effects of temporal embeddedness and network embeddedness on providing and dividing a surplus.* ICS-dissertation, Utrecht.
88. Hilde Bras (2002). *Zeeuwse meiden. Dienen in de levensloop van vrouwen, ca. 1850-1950.* Aksant Academic Publishers, Amsterdam.
89. Merijn Rengers (2002). *Economic lives of artists. Studies into careers and the labour market in the cultural sector.* ICS-dissertation, Utrecht.

90. Annelies Kassenberg (2002). *Wat scholieren bindt. Sociale gemeenschap in scholen.* ICS-dissertation, Groningen.
91. Marc Verboord (2003). *Moet de meester dalen of de leerling klimmen? De invloed van literatuuronderwijs en ouders op het lezen van boeken tussen 1975 en 2000.* ICS-dissertation, Utrecht.
92. Marcel van Egmond (2003). *Rain falls on all of us (but some manage to get more wet than others): Political context and electoral participation.* ICS-dissertation, Nijmegen.
93. Justine Horgan (2003). *High performance human resource management in Ireland and the Netherlands: Adoption and effectiveness.* ICS-dissertation, Groningen.
94. Corine Hoeben (2003). *LETS' be a community. Community in Local Exchange Trading Systems.* ICS-dissertation, Groningen.
95. Christian Steglich (2003). *The framing of decision situations. Automatic goal selection and rational goal pursuit.* ICS-dissertation, Groningen.
96. Johan van Wilsem (2003). *Crime and context. The impact of individual, neighborhood, city and country characteristics on victimization.* ICS-dissertation, Nijmegen.
97. Christiaan Monden (2003). *Education, inequality and health. The impact of partners and life course.* ICS-dissertation, Nijmegen.
98. Evelyn Hello (2003). *Educational attainment and ethnic attitudes. How to explain their relationship.* ICS-dissertation, Nijmegen.
99. Marnix Croes en Peter Tammes (2004). *Gif laten wij niet voortbestaan. Een onderzoek naar de overlevingskansen van joden in de Nederlandse gemeenten, 1940-1945.* Aksant Academic Publishers, Amsterdam.
100. Ineke Nagel (2004). *Cultuurdeelname in de levensloop*. ICS-dissertation, Utrecht.
101. Marieke van der Wal (2004). *Competencies to participate in life. Measurement and the impact of school.* ICS-dissertation, Groningen.
102. Vivian Meertens (2004). *Depressive symptoms in the general population: a multifactorial social approach.* ICS-dissertation, Nijmegen.
103. Hanneke Schuurmans (2004). *Promoting well-being in frail elderly people. Theory and intervention.* ICS-dissertation, Groningen.
104. Javier Arregui (2004). *Negotiation in legislative decision-making in the European Union.* ICS-dissertation, Groningen.
105. Tamar Fischer (2004). *Parental divorce, conflict and resources. The effects on children's behaviour problems, socioeconomic attainment, and transitions in the demographic career.* ICS-dissertation, Nijmegen.
106. René Bekkers (2004). *Giving and volunteering in the Netherlands: Sociological and psychological perspectives.* ICS-dissertation, Utrecht.
107. Renée van der Hulst (2004). *Gender differences in workplace authority: An empirical study on social networks.* ICS-dissertation, Groningen.
108. Rita Smaniotto (2004). '*You scratch my back and I scratch yours' versus 'Love Thy neighbour'. Two proximate mechanisms of reciprocal altruism.* ICS-dissertation, Groningen.
109. Maurice Gesthuizen (2004). *The life-course of the low-educated in the Netherlands: Social and economic risks.* ICS-dissertation, Nijmegen.
110. Carlijne Philips (2005). *Vakantiegemeenschappen. Kwalitatief en kwantitatief onderzoek naar gelegenheid en refreshergemeenschap tijdens de vakantie.* ICS-dissertation, Groningen.
111. Esther de Ruijter (2005). *Household outsourcing.* ICS-dissertation, Utrecht.
112. Frank van Tubergen (2005). *The integration of immigrants in cross-national perspective: Origin, destination, and community effects.* ICS-dissertation, Utrecht.
113. Ferry Koster (2005). *For the time being. Accounting for inconclusive findings concerning the effects of temporary employment relationships on solidary behavior of employees.* ICS-dissertation, Groningen.
114. Carolien Klein Haarhuis (2005). *Promoting anti-corruption reforms. Evaluating the implementation of a World Bank anti-corruption program in seven African countries (1999-2001).* ICS-dissertation, Utrecht.
115. Martin van der Gaag (2005). *Measurement of individual social capital.* ICS-dissertation, Groningen.
116. Johan Hansen (2005). *Shaping careers of men and women in organizational contexts.* ICS-dissertation, Utrecht.
117. Davide Barrera (2005). *Trust in embedded settings.* ICS-dissertation, Utrecht.
118. Mattijs Lambooij (2005). *Promoting cooperation. Studies into the effects of long-term and short-term rewards on cooperation of employees.* ICS-dissertation, Utrecht.

119. Lotte Vermeij (2006). *What's cooking? Cultural boundaries among Dutch teenagers of different ethnic origins in the context of school.* ICS-dissertation, Utrecht.
120. Mathilde Strating (2006). *Facing the challenge of rheumatoid arthritis. A 13-year prospective study among patients and a cross-sectional study among their partners.* ICS-dissertation, Groningen.
121. Jannes de Vries (2006). *Measurement error in family background variables: The bias in the intergenerational transmission of status, cultural consumption, party preference, and religiosity.* ICS-dissertation, Nijmegen.
122. Stefan Thau (2006). *Workplace deviance: Four studies on employee motives and self-regulation.* ICS-dissertation, Groningen.
123. Mirjam Plantinga (2006). *Employee motivation and employee performance in child care. The effects of the introduction of market forces on employees in the Dutch child-care sector.* ICS-dissertation, Groningen.
124. Helga de Valk (2006). *Pathways into adulthood. A comparative study on family life transitions among migrant and Dutch Youth.* ICS-dissertation, Utrecht.
125. Henrike Elzen (2006). *Self-management for chronically ill older people.* ICS-Dissertation, Groningen.
126. Ayse Güveli (2007). *New social classes within the service class in the Netherlands and Britain. Adjusting the EGP class schema for the technocrats and the social and cultural specialists.* ICS-dissertation, Nijmegen.
127. Willem-Jan Verhoeven (2007). *Income attainment in post-communist societies.* ICS-dissertation, Utrecht.
128. Marieke Voorpostel (2007). *Sibling support: The exchange of help among brothers and sisters in the Netherlands.* ICS-dissertation, Utrecht.
129. Jacob Dijkstra (2007). *The effects of externalities on partner choice and payoffs in exchange networks.* ICS-dissertation, Groningen.
130. Patricia van Echtelt (2007). *Time-greedy employment relationships: Four studies on the time claims of post-Fordist work.* ICS-dissertation, Groningen.
131. Sonja Vogt (2007). *Heterogeneity in social dilemmas: The case of social support.* ICS-dissertation, Utrecht.
132. Michael Schweinberger (2007). *Statistical methods for studying the evolution of networks and behavior.* ICS-dissertation, Groningen.
133. István Back (2007). *Commitment and evolution: Connecting emotion and reason in long-term relationships.* ICS-dissertation, Groningen.
134. Ruben van Gaalen (2007). *Solidarity and ambivalence in parent-child relationships.* ICS-dissertation, Utrecht.
135. Jan Reitsma (2007). *Religiosity and solidarity – Dimensions and relationships disentangled and tested.* ICS-dissertation, Nijmegen.
136. Jan Kornelis Dijkstra (2007). *Status and affection among (pre)adolescents and their relation with antisocial and prosocial behavior.* ICS-dissertation, Groningen.
137. Wouter van Gils (2007). *Full-time working couples in the Netherlands. Causes and consequences.* ICS-dissertation, Nijmegen.
138. Djamila Schans (2007). *Ethnic diversity in intergenerational solidarity.* ICS-dissertation, Utrecht.
139. Ruud van der Meulen (2007). *Brug over woelig water: Lidmaatschap van sportverenigingen, vriendschappen, kennissenkringen en veralgemeend vertrouwen.* ICS-dissertation, Nijmegen.
140. Andrea Knecht (2008). *Friendship selection and friends' influence. Dynamics of networks and actor attributes in early adolescence.* ICS-dissertation, Utrecht.
141. Ingrid Doorten (2008). *The division of unpaid work in the household: A stubborn pattern?* ICS-dissertation, Utrecht.
142. Stijn Ruiter (2008). *Association in context and association as context: Causes and consequences of voluntary association involvement.* ICS-dissertation, Nijmegen.
143. Janneke Joly (2008). *People on our minds: When humanized contexts activate social norms.* ICS-dissertation, Groningen.
144. Margreet Frieling (2008). *'Joint production' als motor voor actief burgerschap in de buurt.* ICS-dissertation, Groningen.
145. Ellen Verbakel (2008). *The partner as resource or restriction? Labour market careers of husbands and wives and the consequences for inequality between couples.* ICS-dissertation, Nijmegen.
146. Gijs van Houten (2008). *Beleidsuitvoering in gelaagde stelsels. De doorwerking van aanbevelingen van de Stichting van de Arbeid in het CAO-overleg.* ICS-dissertation, Utrecht.

147. Eva Jaspers (2008). *Intolerance over time. Macro and micro level questions on attitudes towards euthanasia, homosexuality and ethnic minorities.* ICS-dissertation, Nijmegen.
148. Gijs Weijters (2008). *Youth delinquency in Dutch cities and schools: A multilevel approach.* ICS-dissertation, Nijmegen.
149. Jessica Pass (2009). *The self in social rejection.* ICS-dissertation, Groningen.
150. Gerald Mollenhorst (2009). *Networks in contexts. How meeting opportunities affect personal relationships.* ICS-dissertation, Utrecht.
151. Tom van der Meer (2009). *States of freely associating citizens: Comparative studies into the impact of state institutions on social, civic and political participation.* ICS-dissertation, Nijmegen.
152. Manuela Vieth (2009). *Commitments and reciprocity in trust situations. Experimental studies on obligation, indignation, and self-consistency.* ICS-dissertation, Utrecht.
153. Rense Corten (2009). *Co-evolution of social networks and behavior in social dilemmas: Theoretical and empirical perspectives.* ICS-dissertation, Utrecht.
154. Arieke J. Rijken (2009). *Happy families, high fertility? Childbearing choices in the context of family and partner relationships.* ICS-dissertation, Utrecht.
155. Jochem Tolsma (2009). *Ethnic hostility among ethnic majority and minority groups in the Netherlands. An investigation into the impact of social mobility experiences, the local living environment and educational attainment on ethnic hostility.* ICS-dissertation, Nijmegen.
156. Freek Bucx (2009). *Linked lives: Young adults' life course and relations with parents.* ICS-dissertation, Utrecht.
157. Philip Wotschack (2009). *Household governance and time allocation. Four studies on the combination of work and care.* ICS-dissertation, Groningen.
158. Nienke Moor (2009). *Explaining worldwide religious diversity. The relationship between subsistence technologies and ideas about the unknown in pre-industrial and (post-) industrial societies.* ICS-dissertation, Nijmegen.
159. Lieke ten Brummelhuis (2009). *Family matters at work. Depleting and enriching effects of employees' family lives on work outcomes.* ICS-dissertation, Utrecht.
160. Renske Keizer (2010). *Remaining childless. Causes and consequences from a life Course Perspective.* ICS-dissertation, Utrecht.
161. Miranda Sentse (2010). *Bridging contexts: The interplay between family, child, and peers in explaining problem behavior in early adolescence.* ICS-dissertation, Groningen.
162. Nicole Tieben (2010). *Transitions, tracks and transformations. Social inequality in transitions into, through and out of secondary education in the Netherlands for cohorts born between 1914 and 1985.* ICS-dissertation, Nijmegen.
163. Birgit Pauksztat (2010). *Speaking up in organizations: Four studies on employee voice.* ICS-dissertation, Groningen.
164. Richard Zijdeman (2010). *Status attainment in the Netherlands, 1811-1941. Spatial and temporal variation before and during industrialization.* ICS-dissertation, Utrecht.
165. Rianne Kloosterman (2010). *Social background and children's educational careers. The and secondary effects of social background over transitions and over time in the Netherlands.* ICS-dissertation, Nijmegen.
166. Olav Aarts (2010). *Religious diversity and religious involvement. A study of religious markets in Western societies at the end of the twentieth century.* ICS-dissertation, Nijmegen.
167. Stephanie Wiesmann (2010). *24/7 Negotiation in couples transition to parenthood.* ICS-dissertation, Utrecht.
168. Borja Martinovic (2010). *Interethnic contacts: A dynamic analysis of interaction between immigrants and natives in Western countries.* ICS-dissertation, Utrecht.
169. Anne Roeters (2010). *Family life under pressure? Parents' paid work and the quantity and quality of parent-child and family time.* ICS-dissertation, Utrecht.
170. Jelle Sijtsema (2010). *Adolescent aggressive behavior: Status and stimulation goals in relation to the peer context.* ICS-dissertation, Groningen.
171. Kees Keizer (2010). *The spreading of disorder.* ICS-dissertation, Groningen.
172. Michael Mäs (2010). *The diversity puzzle. Explaining clustering and polarization of opinions.* ICS-dissertation, Groningen.
173. Marie-Louise Damen (2010). *Cultuurdeelname en CKV. Studies naar effecten van kunsteducatie op de cultuurdeelname van leerlingen tijdens en na het voortgezet onderwijs.* ICS-dissertation, Utrecht.
174. Marieke van de Rakt (2011). *Two generations of crime: The intergenerational transmission of convictions over the life course.* ICS-dissertation, Nijmegen.

175. Willem Huijnk (2011). *Family life and ethnic attitudes. The role of the family for attitudes towards intermarriage and acculturation among minority and majority groups.* ICS-dissertation, Utrecht.
176. Tim Huijts (2011). *Social ties and health in Europe. Individual associations, cross-national variations, and contextual explanations.* ICS-dissertation, Nijmegen.
177. Wouter Steenbeek (2011). *Social and physical Disorder. How community, business presence and entrepreneurs influence disorder in Dutch neighborhoods.* ICS-dissertation, Utrecht.
178. Miranda Vervoort (2011). *Living together apart? Ethnic concentration in the neighborhood and ethnic minorities' social contacts and language practices.* ICS-dissertation, Utrecht.
179. Agnieszka Kanas (2011). *The economic performance of immigrants. The role of human and social capital.* ICS-dissertation, Utrecht.
180. Lea Ellwardt (2011). *Gossip in organizations. A social network study.* ICS-dissertation, Groningen.
181. Annemarije Oosterwaal (2011). *The gap between decision and implementation. Decision making, delegation and compliance in governmental and organizational settings.* ICS-dissertation, Utrecht.
182. Natascha Notten (2011). *Parents and the media. Causes and consequences of parental media socialization.* ICS-dissertation, Nijmegen.
183. Tobias Stark (2011). *Integration in schools. A process perspective on students' interethnic attitudes and interpersonal relationships.* ICS-dissertation, Groningen.
184. Giedo Jansen (2011). *Social cleavages and political choices. Large-scale comparisons of social class, religion and voting behavior in Western democracies.* ICS-dissertation, Nijmegen.
185. Ruud van der Horst (2011). *Network effects on treatment results in a closed forensic psychiatric setting.* ICS-dissertation, Groningen.
186. Mark Levels (2011). *Abortion laws in European countries between 1960 and 2010. Legislative developments and their consequences for women's reproductive decision-making.* ICS-dissertation, Nijmegen.
187. Marieke van Londen (2012). *Exclusion of ethnic minorities in the Netherlands. The effects of individual and situational characteristics on opposition to ethnic policy and ethnically mixed neighbourhoods.* ICS-dissertation, Nijmegen.
188. Sigrid M. Mohnen (2012). *Neighborhood context and health: How neighborhood social capital affects individual health.* ICS-dissertation, Utrecht.
189. Asya Zhelyazkova (2012). *Compliance under controversy: analysis of the transposition of European directives and their provisions.* ICS-dissertation, Utrecht.
190. Valeska Korff (2012). *Between cause and control: Management in a humanitarian organization.* ICS-dissertation, Groningen.
191. Maike Gieling (2012). *Dealing with diversity: Adolescents' support for civil liberties and immigrant rights.* ICS-dissertation, Utrecht.
192. Katya Ivanova (2012). *From parents to partners: The impact of family on romantic relationships in adolescence and emerging adulthood.* ICS-dissertation, Groningen.
193. Jelmer Schalk (2012). *The performance of public corporate actors: Essays on effects of institutional and network embeddedness in supranational, national, and local collaborative contexts.* ICS-dissertation, Utrecht.
194. Alona Labun (2012). *Social networks and informal power in organizations.* ICS-dissertation, Groningen.
195. Michal Bojanowski (2012). *Essays on social network formation in heterogeneous populations: Models, methods, and empirical analyses.* ICS-dissertation, Utrecht.
196. Anca Minescu (2012). *Relative group position and intergroup attitudes in Russia.* ICS-dissertation, Utrecht.
197. Marieke van Schellen (2012). *Marriage and crime over the life course. The criminal careers of convicts and their spouses.* ICS-dissertation, Utrecht.
198. Mieke Maliepaard (2012). *Religious trends and social integration: Muslim minorities in the Netherlands.* ICS-dissertation, Utrecht.
199. Fransje Smits (2012). *Turks and Moroccans in the Low Countries around the year 2000: Determinants of religiosity, trend in religiosity and determinants of the trend.* ICS-dissertation, Nijmegen.
200. Roderick Sluiter (2012). *The diffusion of morality policies among Western European countries between 1960 and 2010. A comparison of temporal and spatial diffusion patterns of six morality and eleven non-morality policies.* ICS-dissertation, Nijmegen.
201. Nicoletta Balbo (2012). *Family, friends and fertility.* ICS-dissertation, Groningen.
202. Anke Munniksma (2013). *Crossing ethnic boundaries: Parental resistance to and consequences of adolescents' cross-ethnic peer relations.* ICS-dissertation, Groningen.

203. Anja Abendroth (2013). *Working women in Europe. How the country, workplace, and family context matter.* ICS-dissertation, Utrecht.
204. Katia Begall (2013). *Occupational hazard? The relationship between working conditions and fertility.* ICS-dissertation, Groningen.
205. Hidde Bekhuis (2013). *The popularity of domestic cultural products: Cross-national differences and the relation to globalization.* ICS-dissertation, Utrecht.
206. Lieselotte Blommaert (2013). *Are Joris and Renske more employable than Rashid and Samira? A study on the prevalence and sources of ethnic discrimination in recruitment in the Netherlands using experimental and survey data.* ICS-dissertation, Utrecht.
207. Wiebke Schulz (2013). *Careers of men and women in the 19th and 20th centuries.* ICS-dissertation, Utrecht.
208. Ozan Aksoy (2013). *Essays on social preferences and beliefs in non-embedded social dilemmas.* ICS-dissertation, Utrecht.
209. Dominik Morbitzer (2013). *Limited farsightedness in network formation.* ICS-dissertation, Utrecht.
210. Thomas de Vroome (2013). *Earning your place: The relation between immigrants' economic and psychological integration in the Netherlands.* ICS-dissertation, Utrecht.
211. Marloes de Lange (2013). *Causes and consequences of employment flexibility among young people. Recent developments in the Netherlands and Europe.* ICS-dissertation, Nijmegen.
212. Roza Meuleman (2014). *Consuming the nation. Domestic cultural consumption: its stratification and relation with nationalist attitudes.* ICS-dissertation, Utrecht.
213. Esther Havekes (2014). *Putting interethnic attitudes in context. The relationship between neighbourhood characteristics, interethnic attitudes and residential behaviour.* ICS-dissertation, Utrecht.
214. Zoltán Lippényi (2014). *Transitions toward an open society? Intergenerational occupational mobility in Hungary in the 19th and 20th centuries.* ICS-dissertation, Utrecht.
215. Anouk Smeekes (2014). *The presence of the past: Historical rooting of national identity and current group dynamics.* ICS-dissertation, Utrecht.
216. Michael Savelkoul (2014). *Ethnic diversity and social capital. Testing underlying explanations derived from conflict and contact theories in Europe and the United States.* ICS-dissertation, Nijmegen.
217. Martijn Hogerbrugge (2014). *Misfortune and family: How negative events, family ties, and lives are linked.* ICS-dissertation, Utrecht.
218. Gina Potarca (2014). *Modern love. Comparative insights in online dating preferences and assortative mating.* ICS-dissertation, Groningen.
219. Mariska van der Horst (2014). *Gender, aspirations, and achievements: Relating work and family aspirations to occupational outcomes.* ICS-dissertation, Utrecht.
220. Gijs Huitsing (2014). *A Social network perspective on bullying*. ICS dissertation, Groningen.
221. Thomas Kowalewski (2015). *Personal growth in organizational contexts.* ICS-dissertation, Groningen.
222. Manu Muñoz-Herrera (2015). *The impact of individual differences on network relations: Social exclusion and inequality in productive exchange and coordination games.* ICS-dissertation, Groningen.
223. Tim Immerzeel (2015). *Voting for a change. The democratic lure of populist radical right parties in voting behavior.* ICS-dissertation, Utrecht.
224. Fernando Nieto Morales (2015). *The control imperative: Studies on reorganization in the public and private sectors.* ICS-dissertation, Groningen.
225. Jellie Sierksma (2015). *Bounded helping: How morality and intergroup relations shape children's reasoning about helping.* ICS-dissertation, Utrecht.
226. Tinka Veldhuis (2015). *Captivated by fear. An evaluation of terrorism detention policy.* ICS-dissertation, Groningen.
227. Miranda Visser (2015). *Loyality in humanity. Turnover among expatriate humanitarian aid workers.* ICS-dissertation, Groningen.
228. Sarah Westphal (2015). *Are the kids alright? Essays on postdivorce residence arrangements and children's well-being.* ICS-dissertation, Utrecht.
229. Britta Rüschoff (2015). *Peers in careers*: *Peer relationships in the transition from school to work.* ICS-dissertation, Groningen.
230. Nynke van Miltenburg (2015). *Cooperation under peer sanctioning institutions: Collective decisions, noise, and endogenous implementation.* ICS-dissertation, Utrecht.