Essays on framing, free riding, and punishment

By

Michail Drouvelis

Thesis submitted to the University of Nottingham for the

Degree of Doctor of Philosophy

July 2009

Abstract

This thesis presents an experimental investigation of free riding behavior and, more particularly, individual responses to it using, as a workhorse, the so-called public goods game. This game starkly isolates the conflict between private and collective interest, providing us with a simple measure of the extent of free riding behaviour. The unifying theme of the thesis is elicitation and analysis of different indicators for how subjects perceive free riding under a number of treatment manipulations.

Chapter 2 explores how people judge the morality of free riding in a public goods game by eliciting people's moral evaluations in hypothetical scenarios. The scenarios differed with respect to the framing of the game, the order of moves, and the behaviour of the non-judged player. Our findings suggest that free riding is perceived as morally reprehensible, except when the free rider moves second after observing that the other player free rode as well. We also find that moral judgments depend on others' behaviour, on framing and on the order of moves.

Chapter 3 analyses the effect of framing on social preferences, as measured by self-reported emotions and punishment. Our findings are that, for

a given pattern of contributions, neither punishment nor emotion depends on our framing manipulation.

Chapter 4 assesses the behavioural consequences of unfair punishment. In this experiment, we generate an unfair environment by assigning punishment to all group members, irrespective of their first stage behaviour. We find that, although unfair punishment causes a different time profile of contributions, contributions are, on average, little different from in the standard punishment game; and the assignment of punishment in the latter is unaffected by experience of an environment with unfair punishment. However, a history of unfair punishment causes different reactions to helping behaviour and punishment received, respectively.

Acknowledgements

During the course of my research for this thesis, I have benefited from innumerable discussions and helpful suggestions from my supervisors, Robin Cubitt and Simon Gächter. Without their intermittent assistance and guidance, this thesis would likely not have matured. My special thanks go to them, as their interest in my personal development has been a great encouragement to me.

During the several years in which this endeavour lasted, I have also benefited from many conversations that I had with a number of people, including staff at the University of Nottingham and elsewhere. I would like to thank all my co-authors, with whom I collaborate: John Bone, Maria Montero, Wieland Müller, Alex Possajennikov, Nick Powdthavee, Indra Ray and Martin Sefton. The interaction with all of them has been a beneficial influence on my thesis and my own motivation.

While I was in the final year of my PhD, I was also working as a post-doctoral teaching fellow at the University of York. The discussions with many of the members of staff at York have been very useful for the improvement and the enhancement of my academic knowledge. In particular, I am extremely grateful to John Hey for reading part of this thesis and for his patience during our interesting discussions. I have really benefited from his wisdom. Thanks also go to Andrew Jones and Alan Krause, who have been helpful in other ways.

I would like also to thank a number of friends who helped me to make this project an enjoyable experience.

I gratefully acknowledge financial support for this project consisting of a three-year research studentship from the Economic and Social Research Council (ESRC), as well as a one-year scholarship from the Karelias Foundation.

Last, but certainly not least, I would like to express my gratitude to my family, as without them I would not have been in a position to begin this research. Their unconditional generosity and support have played a decisive role in the fulfillment of my achievements and targets.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 General introduction

This thesis explores an important aspect of human decision making: namely free riding. In particular, we are interested in free riding behaviour in situations in which there is a tension between personal and collective interests. Such situations abound in real-life: warfare, environmental protection, management of commons, tax compliance, voting, participation in collective actions, donations to charities, just to name a few. Inability to solve these collective problems would have important implications for the welfare, and perhaps even survival, of human societies. However, the degree of cooperation observed in the field is often higher than what the hypothesis of self-interest, typically made in economics, would suggest.

The objective of this thesis is to improve our understanding of free riding behavior and its implications, by shedding new light on a number of questions, including: Under which conditions, and how far, does free riding behaviour occur? What do agents think about it, and how do they respond to it behaviourally and emotionally? How far can individuals prevent free riding through private sanctions and rewards?

Most of these questions have been addressed by previous researchers too, so this chapter begins by briefly setting the intellectual landscape. At the most general level, previous work on public goods can be divided into theoretical and empirical literatures. For theoretical arguments, the seminal paper by Samuelson (1954) is a standard starting point. Since then, the theoretical literature on public goods has been further advanced by significant contributions by many prominent social scientists, such as Musgrave, Buchanan and Olson. For a general review, see Cornes and Sandler (1996).

However, the aim of the present thesis is to contribute to the empirical literature of public goods provision, using controlled laboratory experiments.[1] The most common design for exploring public good games in the lab is the so-called "voluntary contributions mechanism". In its standard form, each player receives an identical initial endowment of money. Players have to decide simultaneously what fraction of the endowment they want to contribute to a group account and what fraction they wish to keep. All funds in the group account pay a positive return to each member of the group. The parameters are chosen so that each agent has a dominant strategy to contribute zero to the group account (i.e. to free ride fully) but, in the group

[1] There has also been empirical non-experimental literature investigating the provision of public goods. For an overview of this literature, see Batina and Ihori (2005).

optimum, every agent would contribute all of his endowment to the group account. Thus, this game starkly isolates a conflict between private and collective interest, providing us with a simple measure of the extent of free riding behaviour.

The investigation of the public goods game by experimental economists began to gather pace in the early 80's.[2] These investigations were primarily concerned with the determinants of contributions (e.g., Isaac, et al., 1984; Andreoni, 1988; Isaac and Walker, 1998a, 1988b; Isaac, et al., 1994). For an excellent review for the history of the experimental literature on public goods up to 1995, see Ledyard (1995). Since then, the major new development in the experimental public goods literature has been the analysis of games in which there is a possibility of punishment, following, for example, the seminal contribution by Fehr and Gächter (2000). In sum, the general structure of these games comprises two stages, the first of which is identical to the voluntary contributions game described earlier. In the second stage, individuals are made aware of the profile of contributions of other group members and allowed to punish them. Punishment is costly for the punisher and the recipient. These games are particularly interesting as they shed light both on people's motivations and reactions to free riding, and because they also inform us on whether and if so, how free riding can be prevented. Following Fehr and Gächter's original experiment on

[2] It is worth mentioning that there were also significant contributions by psychologists around this time and earlier (e.g., Dawes, et al, 1977; Marwell and Ames, 1979, 1980, 1981).

punishment, the public goods literature has flourished with an increasing number of contributions being made by many researchers.[3]

The research presented in this thesis has three distinctive characteristics: (i) focus on non-behavioural responses to free riding, not just behavioural ones; (ii) focus on the effects of framing on responses to free riding, not just on contributions; and (iii) analysis of the effects of unfair use of punishment on contributions and on the future use of private sanctions. While there are few precursors to (iii), it is worth mentioning that (i) and (ii) extend earlier work in psychology and economics.

Moral judgment is a prominent notion in psychology (as well, of course, as in philosophy). Other researchers in this field have investigated the process with which people arrive at moral judgments and their importance in various contexts (e.g., Haidt, 2001; Wheatley and Haidt, 2005; Prinz, 2006; Haidt, 2007; Gino, et al, 2008; Croson and Konow, 2009). Others, such as Gächter and Riedl (2005, 2006), who elicit normative judgments of fairness in a bargaining game, have considered the interplay between moral judgment and economic decision making. We extend the analysis of moral judgments to another important form of economic behavior, by eliciting the moral judgments that individuals make of free riders in social dilemma games. In particular, we are concerned with whether these games are perceived as having a moral dimension and if so, how moral judgments of free riding are affected by various aspects of the situation in which it occurs.

[3] For example, see Bowles et al. (2001); Masclet, et al. (2003); Carpenter, et al. (2004); Falk, et al. (2005); Noussair and Tucker (2005); Page, et al., (2005), Anderson and Putterman (2006), Bochet, et al. (2006), Gürerk, et al, (2006); Carpenter (2007a); Carpenter (2007b), Sefton, et al. (2007), Gächter and Herrmann (2007); Herrmann, et al (2008); Gächter and Herrmann (2009).

Our interest on non-behavioural indicators of how free riding is perceived also includes looking at self-reported emotional responses. Emotions constitute an important notion in cognitive and social psychology (e.g. Sonnemans and Frijda, 1994; Sonnemans and Frijda, 1995; Elster, 1998, among others), but also play an important role in the social preferences literature (e.g., Bosman and van Winden, 2002; Kirchsteiger. et al., 2006; Loewenstein, 2000). In our research, we consider emotional responses to the play of others in voluntary contributions games.

In addition, the study of framing effects constitutes an important part of this thesis. Such effects have been studied mainly in relation to individual choice under risk and uncertainty (for starting points, see Kahneman and Tversky, 1979, 1984 and Tversky and Kahneman, 1981, 1986). Nevertheless, there has also been some experimental work investigating the role of framing in the context of public good games (e.g., Andreoni, 1995; Sonnemans, et al., 1998). Findings from these studies suggest that framing matters in the sense that behaviour tends to change if subjects are confronted with alternative wordings for the same objective outcome. In this thesis, our focus is not so much the effect of framing on free riding, but rather on responses to free riding.

1.2 Introduction to each chapter

Chapter 2 investigates people's moral perceptions of free riding behaviour in a public good game. In our experimental design, subjects respond to a questionnaire, in which they are confronted with hypothetical scenarios in a 2-player public good game. These scenarios involve one fictional player who is always a free rider and another fictional player whose behavior varies across scenarios. Subjects participating in this experiment were asked to morally rate (either positively or negatively) the action of free riding in a series of scenarios with these properties. In other words, they act as impartial spectators judging the action of free riding in various endings of these scenarios, without being involved in the decision situation. We manipulated the scenarios with respect to the framing of the public good game; the order in which players make their moves; and the contribution to the public good of the non-judged player.

Our findings from this study suggest that the public good game is indeed perceived as a moral game, with free riding being considered as a morally reprehensible action, except when the free rider moves second after observing that the other player free rode as well. In this case, subjects judge free riding as morally praiseworthy. What is more, as the contribution of the non-judged player rises, the free rider is condemned increasingly strongly, especially when he moves second. We also show that the framing manipulation of the game is a major determinant of moral evaluations. Specifically, subjects tend to condemn a failure to contribute more strongly than a withdrawal although both actions have identical payoff consequences for both players. Our results are robust regardless of whether the game is played simultaneously or sequentially, since framing effects work in the same direction in both conditions. With respect to the order of moves, we find that sequential moves trigger stronger condemnation of free riding than simultaneous moves, provided the non-judged player contributes.

Motivated by our findings suggesting that self-reported moral judgments are context dependent, Chapter 3 investigates whether a similar effect applies when self-reported emotions are elicited. Emotions have in common with moral judgments that both are non-behavioural indicators of perceptions. However, in this second study, subjects depart from the role of an outside observer and their decisions have a direct monetary impact on their and other group members' earnings. In addition to emotional responses, we study the effects of framing on punishment. This complements (non-behavioral measures of) emotions, by being a key behavioural indicator of attitudes to free riders. Once again, our focus is on the effects of framing on these indicators.

Contrary to our motivating findings from our study on moral judgments, we find that our framing has little effect on our indicators of negative reciprocity in Chapter 3 (namely, punishment and emotional responses). Our findings suggest that the main determinant of punishment and emotional responses is the difference between the contributions of the punisher and the punished group member.

Chapter 4 reports an experiment that studies the impact on contributions and punishment behavior of an institution in which punishment is assigned unfairly, but players can alleviate its effects from each other. To achieve this, we introduce a variant of the standard punishment game which we call as the default punishment game. In this game, all members of a group are punished automatically, regardless of their prior behaviour. The way we implement the automatic penalty is by exogenously imposing on subjects a monetary fine that decreases their payoff. Then subjects are allowed to alleviate the automatic penalty by incurring a cost. However, alleviation benefits its recipients. Our game is motivated by existing evidence that shows punishment to be effective when it is assigned fairly, in the sense that cooperators sanction free riders who in turn accept punishment and as a result, increase their contributions. But how does punishment work when subjects have experienced an environment in which it is assigned unfairly by sanctioning all group members? To test the robustness of the standard punishment game, our experimental subjects participate in both the standard and the default punishment games. We can thus assess whether and, if so, how unfair punishment influences the way punishment is assigned. By this, we mean both how the default punishment game compares with the standard one and how prior experience of the default punishment game affects the play of the standard one.

Our experiment provides evidence that, on average, contributions in the default punishment game are not sensitive with respect to the standard punishment game. However, the time profile of contributions is different: for the default punishment game, contributions follow a hump shaped pattern as they increase at the beginning and decline after the second half of the game; whereas, for the standard punishment game, contributions are rather stable across time. We also find that the assignment of punishment is unaffected after the experience of an environment with unfair punishment.

Yet, it turns out that a history of unfair punishment makes a difference regarding reactions to helping behaviour and punishment received. More specifically, we observe that those subjects who contributed less than the group average increase their contributions per punishment point received in the standard punishment game, but not after a history of the default punishment game, in which case they do not change their contributions significantly per punishment point received.

For each of Chapters 2-4, there are accompanying appendices, giving experimental instructions, tables, and other supplementary material.

Finally, Chapter 5 brings together the main conclusions of the three previous chapters, as well as offering brief suggestions for future research and some final thoughts.

Chapter 2

How do people judge the morality of free riding in social dilemmas?[4]

2.1 Introduction

The purpose of this chapter is to explore experimentally how individuals judge the morality of free riding in a social dilemma game. A social dilemma results from a situation where a group shares a common resource and each individual has to decide whether to contribute to its provision or not. Failure to contribute to the common resource brings about negative side-effects for other individuals; while contributing generates positive side

[4] The experiment reported in this chapter was a collaboration with Robin Cubitt and Simon Gächter. The present chapter is a companion paper to a CeDEx working paper, expected to appear in Summer 2009.

effects. The social dilemma game, thus, isolates the conflict between the personal interest and the collective interest, providing us with a measure for the degree of people's self-interested behaviour. However, in this chapter, our objective is not to measure behaviour but instead to assess how observers judge free riding, from a moral perspective. Our objective is to assess whether social dilemmas are regarded as having a moral dimension; and, if so, its nature.

Although the significance of morality has mostly been emphasised by moral philosophers and psychologists, our study is of interest in the light of several attempts to incorporate the notion of morality in economic situations. For instance, past research by Kahneman, Knetsch and Thaler (1986) highlight its crucial role and investigate individuals' fairness perceptions in understanding when price increases are justifiable. Their findings show that raising prices due to shifts in demand is perceived as unfair, whereas price increases are acceptable only if they are the result of cost increases. Most recently, Gächter and Riedl (2005, 2006) elicit normative judgments of fairness in a bargaining game with infeasible claims, with both studies suggesting that morality is a key ingredient in determining individuals' actual negotiating behaviour.

Drawing on earlier work from experimental psychology, our investigation elicits subjects' moral perceptions of free riding by adapting their techniques to our social dilemma game.[5] Specifically, subjects respond

---

[5] For application of such techniques in social and moral psychology, see, e.g., Spranca, et al., 1991; Baron and Ritov, 1994; Haidt and Baron, 1996; Niedermayer and Chapman, 2001; Wheatley and Haidt, 2005; Cushman, et al., 2006; Kelly, et al., 2007; Gino, et al., 2008.

to a questionnaire, in which they are confronted with hypothetical scenarios, involving a 2-player game. In various endings of these scenarios, one person always free rides; while the other (the non-judged player) contributes different amounts to the common resource. For each scenario separately, subjects are asked to express their positive or negative moral rating towards the free rider without being involved in the decision situation. Thus, subjects merely act as observers, so their judgments should impartial moral evaluations of the free rider.

Our experimental design manipulates three factors with respect to the scenarios. First, as already noted, we examined whether subjects' moral judgments depend on the other player's behaviour. This manipulation enables us to examine whether free riding is maintained as morally (im)permissible regardless of the other player's choice or whether such judgments are sensitive to the other player's actions. Second, we explored whether moral judgments are dependent on the framing of the social dilemma problem. The framing manipulation we look at this study has a Give (contributing to the common resource) vs. Take (withdrawing from the common resource) form, since this manipulation is very common in social psychology studies (e.g. Brewer and Kramer, 1986; Rutte, et al., 1987; McDaniel and Sistrunk, 1991; McCusker and Carnevale, 1995; Sell and Son, 1997; van Dijk and Wilke, 2000). Our third manipulation investigates the extent to which moral judgments depend on the order of moves of the players in the hypothetical scenarios. That is, we explore whether simultaneous moves trigger different moral condemnation of a given action, relative to sequential moves. In particular, we investigate whether the order of moves affects whether and if so, how the behaviour of the non-judged player affects the judgments passed on the free rider.

Our findings demonstrate that free riding is always perceived as a morally blameworthy action except for one case in which it is seen as morally praiseworthy; that is, when the judged free rider moves second after having observed that the other player has not contributed either. Across frames, we find that subjects perceive withdrawing tokens from the public good as being less morally bad than the objectively equivalent action of failing to contribute tokens to the public good, conditional on the other player's behaviour. Finally, we provide evidence that, sequential moves usually trigger higher condemnation of free rider, but irrespective of whether the order of moves are simultaneous or sequential, the higher is a player's contribution, the free rider is condemned increasingly strongly. Interestingly enough, this pattern is observed for a substantial minority in the simultaneous case, but for an overwhelming majority in the sequential case.

The remainder of the chapter is organised as follows. Section 2.2 describes the design, the hypotheses and the procedures of our experiment. Section 2.3 discusses the results and Section 2.4 concludes.

2.2 Design, Hypotheses and Procedures

2.2.1 Design

In our experiment, each subject responded to a questionnaire requiring her to report her moral judgment of a player in hypothetical scenarios. There were four treatments, each defined by a different questionnaire. Each subject responded to the questionnaire for one treatment only. Before explaining the differences between them, we first explain the points which the questionnaires had in common.

Each questionnaire described a decision problem for two fictitious players, named Person A and Person B; and then gave some possible endings, each of which specified players' choices and their payoff consequences. A scenario comprises a description of a decision problem and an ending. Each questionnaire consisted of five scenarios with the same decision problem, but different endings.

In all the scenarios, the players were the two members of a group playing a voluntary contributions game. Within each questionnaire, the behaviour of Person A varied across scenarios, but Person B was always a (complete) free-rider. After each ending, the subject was asked, as a detached observer, to rate the morality of Person B on a scale ranging from -50 (extremely bad) to + 50 (extremely good).[6] Thus, in each treatment, we can test within-subjects for the impact of the behaviour of the non-judged

[6] **Ratings were selected by using a mouse to move a slider on a computer screen. The slider was initially positioned at a rating of zero. Subjects had to click on the slider in order to activate it and were not allowed to proceed unless they had done so. This procedure was intended to prevent subjects from reporting a judgment of zero accidentally, while allowing them to do so after reflection.**

player on the moral rating assigned to the free rider. All other tests are between-subjects and involve comparisons of subjects' responses across treatments.

There were two treatment variables: the framing used to describe the decision problem; and the order of moves in that problem. Each variable had two possible values: "Give" and "Take" for framing; and "Simultaneous" and "Sequential" for order of moves. Each variable was manipulated independently, yielding four treatments: Give-Simultaneous, Take-Simultaneous, Give-Sequential, and Take-Sequential.

To explain the Give versus Take manipulation, we fix for simplicity on the Simultaneous order of moves. In the Give frame, the decision facing each player was how much to contribute to a group project. The description of the decision problem and the first ending for the Give-Simultaneous treatment were as follows:

> ***Imagine a group that consists of two group members, Person A and Person B. Each group member receives an endowment of 20 tokens and has to decide how many tokens to keep for himself and how many to contribute to a group project. Each token he keeps for himself has a value of one pound for him. Each token contributed to the group project has a value of 1.50 pounds to the project. The total value of the project is divided equally between the two group members. So, each token contributed to the project earns both group members 0.75 pounds each. The total income of a group member is the sum earned from tokens kept for himself and his share of the earnings of the group project. Each group member decides simultaneously, that is, without knowing what the other one has done.***
>
> ***A) Assume that Person A contributes 0 tokens to the group project and Person B contributes 0 tokens to the group project. Therefore, the value of the group project is 0 pounds and, thus, as a result of their contributions, Person A's total income is 20 pounds and Person B's total income is 20 pounds. How do you rate Person B's morality?***

As explained above, the scenarios in the Give-Simultaneous questionnaire differed only in respect of Person A's behaviour. Person A's contribution was 0 tokens (as shown) in the first scenario, rising to 20 in increments of 5 over the other four scenarios.

In the Take frame, the decision facing each player was how much to withdraw from a group project. The description of the decision problem and first ending for the Take-Simultaneous treatment were as follows:

Imagine a group that consists of two group members, Person A and Person B. There are 40 tokens in a group project. Each group member has to decide how many, up to a maximum of 20, of these tokens to withdraw for himself and how many to leave in the group project. Each token he withdraws for himself has a value of one pound for him. Each token left in the group project has a value of 1.50 pounds to the project. The total value of the project is divided equally between the two group members. So, each token left in the project earns both group members 0.75 pounds each. The total income of a group member is the sum earned from tokens withdrawn by himself and his share of the earnings of the group project. Each group member decides simultaneously, that is, without knowing what the other one has done.

A) Assume that Person A withdraws 20 tokens from the group project and Person B withdraws 20 tokens from the group project. Therefore, the value of the group project is 0 pounds and, thus, as a result of their withdrawals, Person A's total income is 20 pounds and Person B's total income is 20 pounds. How do you rate ***Person B's*** *morality?*

As with the Give frame, the only difference between the scenarios in a Take frame questionnaire was the behaviour of Person A. Person A's withdrawal was 20 tokens (as shown) in the first scenario, declining to 0 in decrements of 5 over the remaining four scenarios.

It is important to note that the Give and Take frames differ only in respect of the description of the decision problem. There is no difference between the two frames in terms of the feasible sets of monetary outcomes available to a player. In each frame, each player controlled the final destination of 20 tokens, each of which could be allocated either to himself (earning £1 for him) or to the project (earning £0.75 for each player). To emphasise this similarity, we will use the term "effective contribution" below to refer to the tokens allocated by a player to the project, regardless of whether this allocation arises from a failure to contribute or from a withdrawal.

In addition to the Simultaneous treatments, we ran two treatments (one with the Give frame, and one with Take) in which the non-judged player moved first. Each questionnaire for these Sequential treatments was obtained from the corresponding Simultaneous one by replacing the last sentence of the description of the decision situation with "Assume that Person A decides first and Person B observes Person A's choice before making his own decision."[7] In all other respects, Sequential questionnaires were identical to the corresponding Simultaneous ones.

2.2.2 Motivation & Hypotheses

To motivate our design, it is helpful to distinguish two broad accounts of how individuals might arrive at their moral judgments which, for convenience, we call the reason-based model and the emotion-based model, respectively.

The reason-based model or the rationalist approach (Kohlberg, 1969; Piaget, 1932/1965; Turiel, 1983) model sees judgments as being the result of conscious deliberation, using prior moral principles. On this view, an individual's moral judgments arise from application of the moral principles that she endorses to the case in hand. Consequently, hypotheses about how judgements will vary across our scenarios would, according to the reason-

[7] A copy of the instructions for the Give-Simultaneous and Take-Simultaneous treatments can be found in Appendix A. The instructions for the corresponding Sequential treatments differed only in the respect explained.

based model, be conditional on assumptions about subjects' prior moral principles.

In contrast, the emotion-based model sees emotions or gut-instincts as the drivers of moral judgments. On this view, moral judgments are *ex post* rationalisations of individuals' emotions and raw intuitions, not the expression of prior moral principles. Modern examples that emphasise the role of emotions in reaching moral judgments are the social intuitionist model of Haidt (2001) and the sentimentalist theory of Prinz (2006).[8] According to the emotion-based model, any features of our scenarios that trigger different emotions or gut-instincts could give rise to different judgments.

Our experimental design manipulates the framing of the decision problem facing the judged player (Give versus Take); the order of moves in that problem (Simultaneous versus Sequential); and the behaviour of the non-judged player.

The reason-based model delivers a clear prediction in respect of framing, if subjects hold consequentialist principles. For any consequentialist ethical theory, the moral value of an action is determined by a comparison of its consequences with other feasible ones; and so redescribing the decision problem should have no impact on the moral value of an action with given consequences. Conditional on subjects endorsing any form of ethical consequentialism, the reason-based model predicts no difference in moral judgments between otherwise identical Give and Take questionnaires.

[8] These accounts can be seen as modern forms of a much older naturalistic tradition in moral philosophy, going back for example to Hume (1739) and Smith (1759).

If we do observe a difference, the reason-based model would interpret it as evidence of subjects endorsing non-consequentialist ethical principles. However, the emotions-based model suggests a different interpretation, namely that subjects' gut reactions are driven in part by payoff-irrelevant features of the description of the decision problem. For example, emotional response to a player whose effective contribution is zero might differ according to whether this free riding arises from complete failure to contribute to the project or from maximal withdrawal of tokens from it, even though the consequences are the same.

Our other experimental manipulations can also be analysed in a similar way. In particular, conditional on endorsement of the principle that an agent can only be praised or condemned for her action on the basis of features of the situation that she could have known, the reason-based model suggests no difference in moral judgments across the five scenarios of a Simultaneous questionnaire, since there is no difference between them in what Person B knows at the moment of choice. If subjects endorse the principle that the morality of Person B's action does not depend on what Person A does, even if Person B knows it, the reason-based model also suggests no difference between the judgments passed on Person B across the scenarios of the Simultaneous and Sequential questionnaires with a given framing.

Recall that, although Person B's effective contribution is always 0 tokens, the effective contribution of Person A rises across the successive endings of each questionnaire, leading to outcomes that are progressively less favourable to Person A and more favourable to Person B. Thus, if we observe any difference between the judgments passed on Person B across the five scenarios of a given questionnaire, it seems likely that this will take the form that the higher is Person A's effective contribution, the more negative is the moral rating assigned to Person B. We refer to this hypothesis as the "increasing condemnation hypothesis".

If results from our Simultaneous treatments conform to the increasing condemnation hypothesis, this would support the emotion-based model and be hard to explain using the reason-based model. It is easy to see how Person A's choice might affect a subject's emotional response to Person B's action. For example, the subject might be more angered, or disgusted, by Person B's free riding when it brings about an unequal outcome than when it does not do so, even though Person B could not have predicted these effects. In contrast, it is quite difficult to think of a plausible ethical theory that would rationalise the increasing condemnation hypothesis in Simultaneous treatments as, to do so, it would have to condemn him on the basis of something he could not have known.

Finally, either model could account for results that conform to the increasing condemnation hypothesis in Sequential treatments. The reason-based model would interpret such a finding as evidence that subjects endorse moral principles which license reciprocation by Player B of Player A's behaviour. On this account, it would not be a moral transgression for Player B to free ride when he knows Player A has already done so, but it would be a transgression for Player B to free ride when he knows Player A has not. The emotions-based model would interpret evidence of increasing condemnation in Sequential treatments as an indication that emotional response is more negative to someone who free rides on a known cooperator than it is to someone who "rats on a rat" (i.e. free rides on a known free rider).

2.2.3 Procedures

We recruited subjects from the University of Nottingham pool of students using the ORSEE software (Greiner, 2004). In total, we sent 2,718 email invitations, resulting in participation by 538 subjects. Once a subject registered to take part in the experiment, they were directed to the experiment's website. Subjects were allocated automatically to the four treatments, in a rotating sequence by time of registration for the experiment. After assigning their moral ratings, subjects were asked to give a brief verbal explanation of them.

Each subject saw only the questionnaire for the treatment they were assigned to. They could either respond to the questionnaire immediately, or exit and return to it before the closing date of the experiment (which was one week after invitations were sent out). Subjects returning later could only see the questionnaire they were assigned to initially. Subjects were omitted from the data analysis if they failed to complete a questionnaire by the closing date. To counter the possibility of multiple submissions from the same subject, only one registration was permitted from a given invitation. Our use of ORSEE recruitment software, rather than an open internet experiment[9], enabled us to build in this safeguard, as well as giving us the demographic information on participants held in the ORSEE database.

It is an inherent feature of our study that we could not incentivise task-responses, but we could incentivise participation. We comment on these features in turn.

Our objective was to study subjects' moral attitudes. The use of questionnaire methods without task-related rewards is standard in the study of social attitudes. It is appropriate for our purposes because any means of tying payments to subjects' responses would introduce a confound. In particular, we wished to elicit the judgments that subjects would give in the role of disinterested observer. This precluded having subjects be participants in the voluntary contributions game: hence our use of hypothetical scenarios. Allowing subjects to assign financial penalties or rewards to the players in the scenarios, even hypothetically, would have confounded moral attitudes with attempts to bring about particular distributional consequences: hence our use of pure judgment tasks rather than – say – reward or punishment tasks. As our judgment tasks are moral judgment tasks, as opposed – say – to mathematical puzzles or judgments of distance, there are no objectively "right" or "wrong" answers to them. So, we could not reward subjects for judging correctly. Finally, rewarding subjects for making judgments that conform to particular ethical theories, or to our own ethical views, or to average opinion, would all have introduced obvious biases, relative to the motivation for the experiment.

[9] For a discussion on the advantages and disadvantages of open internet experiments, see Appendix A.2. The aim of this appendix is to offer some general reflections on internet experiments.

We were concerned that, without task-related incentives, it might be difficult to generate a sufficient number of participants. On the other hand, having a substantial reward for participation might have attracted subjects unwilling to give considered responses and only willing to do the minimum necessary to obtain the reward. In the light of these considerations, we used two approaches in parallel. Prior to issuing invitations, we divided our potential subject pool into two equal sub-groups: one for which there would be no payments at all ("No-Payment experiment") and one in which a random participation fee was provided ("Payment" experiment), in the form of entry to a lottery. The latter provided some protection against low participation, while conducting both experiments enabled us to check for any effect of the participation incentive on task-responses.

All subjects participating in either experiment were informed about the importance of answering the questionnaire as precisely and honestly as possible and that all responses would remain confidential. Those invited to the Payment experiment were told that those who completed the questionnaire would be entered into a prize draw, conducted publicly with two prizes of £50. Although attending the draw was not compulsory, subjects were given the date, time and the venue of the draw.

2.3. Results

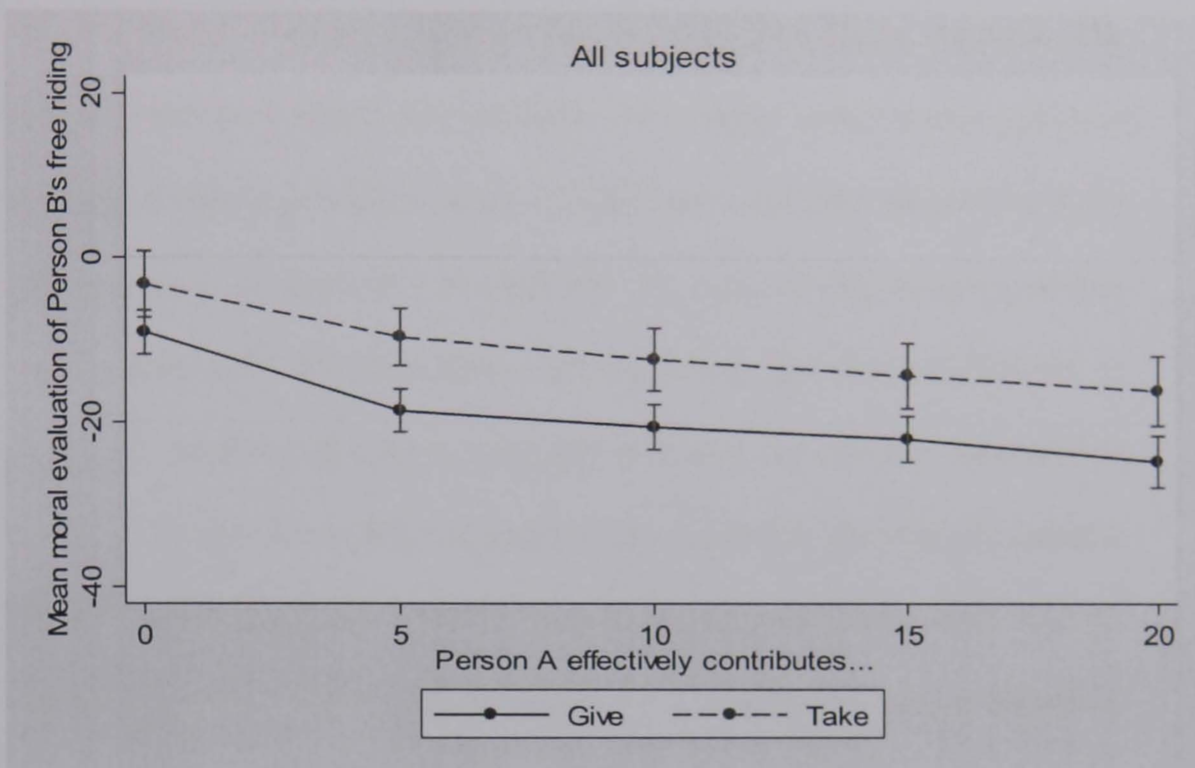
2.3.1 Does paying for participation affect response rates and/or moral evaluations?

We start our data analysis by exploring whether paying a random participation fee is an efficient tool for increasing response rates in our experiment. In the "Payment" experiment, 306 subjects responded a questionnaire, while, in the "No-Payment" experiment 232 subjects responded a questionnaire. That is, when incentives were provided 74 more subjects chose to participate. Performing a Probit regression analysis (see Table B.1 in Appendix B), we find that the probability to participate in the experiment is significantly higher for the subjects receiving the invitation for the "Payment" experiment than those receiving the invitation for the "No-Payment" experiment. We therefore conclude that paying a random participation fee (in our case a random flat fee of £50) is an additional motivating factor for subjects to participate, apart from their intrinsic willingness, increasing significantly the response rates. However, paying a participation fee does not affect subjects' moral evaluations (the corresponding coefficient in Table B.2 in Appendix B is not statistically significant). This implies that there is no selection bias between those who participate in the "Payment" experiment and those who participate in the "No-Payment" one. Since no significant differences in moral evaluations were found, we proceed by pooling the data between the two experiments and thus, not including "Payment" as a separate independent variable in the econometric analysis reported below.

2.3.2 Are moral judgments subject to framing effects?

In this subsection, we investigate the effects of framing on moral judgments in the Simultaneous treatments, that is, when Person A and Person B decide without knowing what the other has done. The tool for our analysis is the mean moral evaluation function, which is an aggregate measure of the moral ratings that subjects assigned to the free rider (Person B), expressed as a function of the five effective contribution levels of his counterpart (Person A). Recall that we refer to effective contribution as the amount of tokens contributed to the public account (in the Give frame) or equally, the amount of tokens left in the public account after the withdrawal (in the Take frame). Figure 2.1 shows the mean moral evaluation function of the free rider for all subjects participated in the corresponding treatments.

Figure 2.1. Give vs. Take in Simultaneous treatments – The moral evaluation function



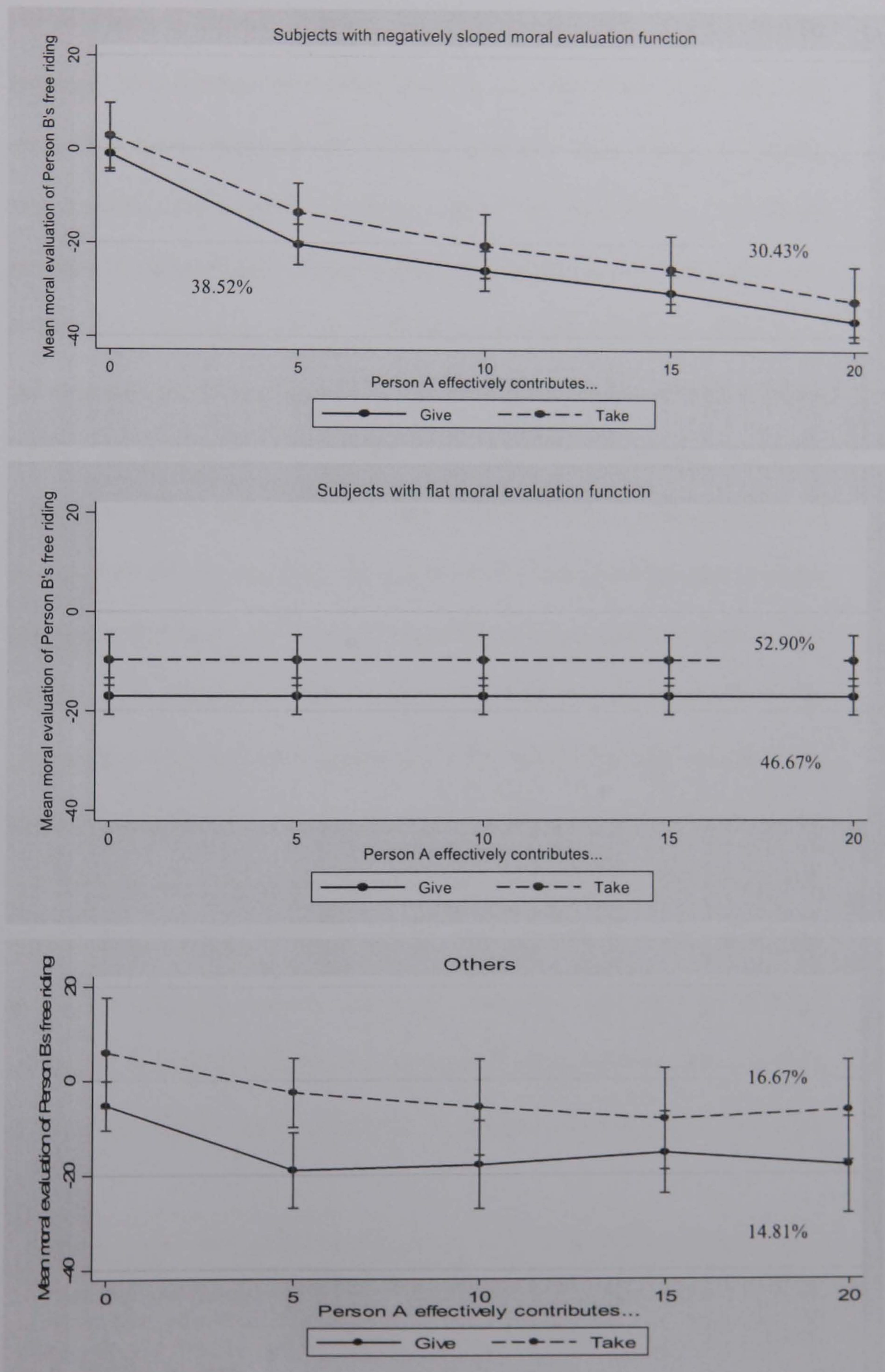
The horizontal axis indicates the amount of tokens Person A effectively contributes. The vertical axis indicates the average moral rating that subjects assign to Person B who is the free rider and is always judged. On this axis, the point 0 denotes that free riding is perceived to be of no moral significance. Ratings below 0 imply that subjects perceive free riding in this scenario as a morally blameworthy action; whereas, ratings above 0 imply that subjects perceive free riding as a morally praiseworthy action. In this graph, the 95% confidence intervals for the mean moral evaluation in each of the five possible scenarios are also shown. With respect to framing, Figure 1 demonstrates that the moral evaluation function in the Give treatment is always below the moral evaluation function in the Take treatment, suggesting that subjects are more condemning in the Give than in the Take frame. Additionally, we observe that under both frames in all five possible scenarios free riding is perceived as a morally blameworthy action, with the moral evaluation function having a negative slope, implying that the more Person A effectively contributes, the higher is the condemnation of the judged free rider. Econometric evidence provides also support for these observations (see Table B.3 in Appendix B). Specifically, we estimated an OLS model, with the dependent variable being the moral evaluation of Person B, and the explanatory variables including the effective contribution levels of Person A (contributing zero is the baseline), the dummy variable "Take", which equals to 1 for the Take treatment, and 0 otherwise, and the dummy variable "Male", which equals to 1 if subjects were male and 0 otherwise. We also control for slope differences by including as independent variables interaction terms between the dummy variable "Take" and each scenario separately. Regression coefficients (see equation 1 in Table B.3 in Appendix B) reveal that the dummy variable "Take" is statistically significant at the 5% level, implying that subjects are significantly more condemning in the Give than in the Take treatment, since the coefficient is positive and mean evaluations are in the negative domain of the vertical axis. In aggregate, subjects evaluate free riding as being an immoral act, since the coefficient of the constant is negative and statistically significant. The moral evaluation function is negatively sloped, *ceteris paribus*, since the coefficients of the four scenario dummy variables are statistically different from zero and from each other (from F-test, p-value = 0.000), suggesting the corroboration of the "increasing condemnation hypothesis".

This implies that the same act of free riding (contributing nothing or withdrawing everything) is considered increasingly morally unacceptable the higher is the other player's effective contribution.

To understand better the observed patterns of moral judgments, we divided subjects into three groups: (1) subjects with a negatively sloped moral evaluation function, (2) subjects with a flat moral evaluation function, and (3) "others", including non-monotonic subjects and subjects with a positively sloped moral evaluation function.[10] The moral evaluation function for each of these three groups is shown in Figure 2.2. Along with the 95% confidence intervals for the mean moral evaluation in each of the five possible scenarios, the percentages of subjects falling under these categories for each framing manipulation are also reported.

[10] Non-monotonic subjects refer to those whose moral evaluation function is strictly negatively sloped in one range and strictly positively sloped in another.

Figure 2.2. Give vs. Take in Simultaneous treatments – The moral evaluation function for each group



From these three panels, we observe that the substantial majority of subjects have a flat and a nearly flat (that is, "others") moral evaluation function. This suggests that when there is no information about what the other has done, most of the subjects perceive free riding as equally reprehensible across all five scenarios. However, there exists a significant minority of subjects who display increasing condemnation of free rider and, in fact, drive the negative slope of the moral evaluation function observed in the aggregate level (see Figure 2.2). As noted above, the existence of such a group in Simultaneous treatments is somewhat surprising, from the perspective of the reason-based model, as they in effect condemn Person B on the basis of something he did not know. Focusing on the source of the framing difference, our subjects' classification provides us with useful insights (econometric evidence is given in Table B.3 in Appendix B). In particular, it turns out that subjects with a flat moral evaluation function and subjects classified as "others" are significantly more condemning in the Give than in the Take treatment.[11] In relation to subjects with a negatively sloped moral evaluation function, we find that there is no statistically significant difference across frames. In sum, the existence of a framing effect recorded by our data can be attributed to those subjects who condemn free riding equally across scenarios, regardless of how much the other

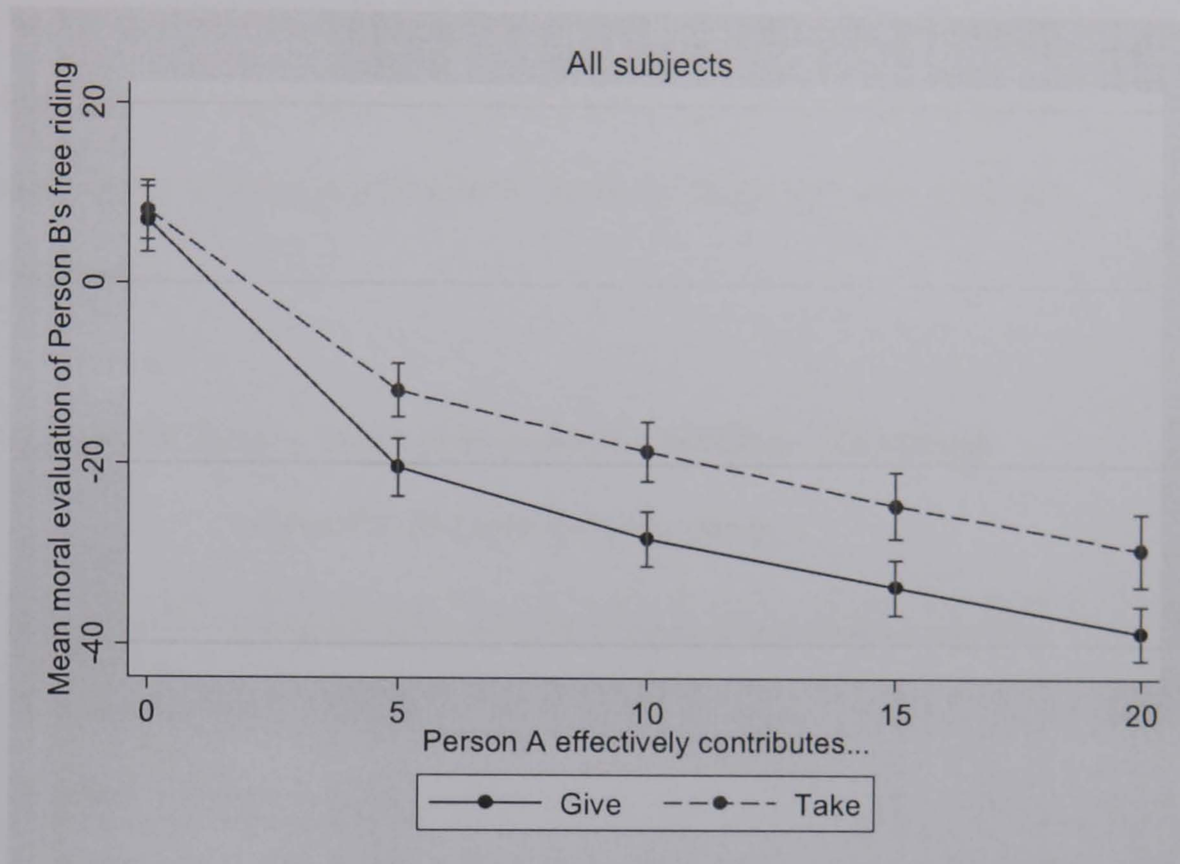
[11] Looking at the moral evaluation functions of subjects classified as "others" across frames, we observe that they are quite stable and flat. This is confirmed from the regression coefficients of this group (see equation 4 in Table B.3), since they are not statistically significant from each other. Pooling subjects with a flat moral evaluation function with "others", we find that the variable dummy "Take" has a positive sign and is statistically significant at the 1% level.

person has contributed, implying that subjects' views are difficult to be reconciled with judgments driven by consequentialist theories.

2.3.3 Do sequential moves make a difference?

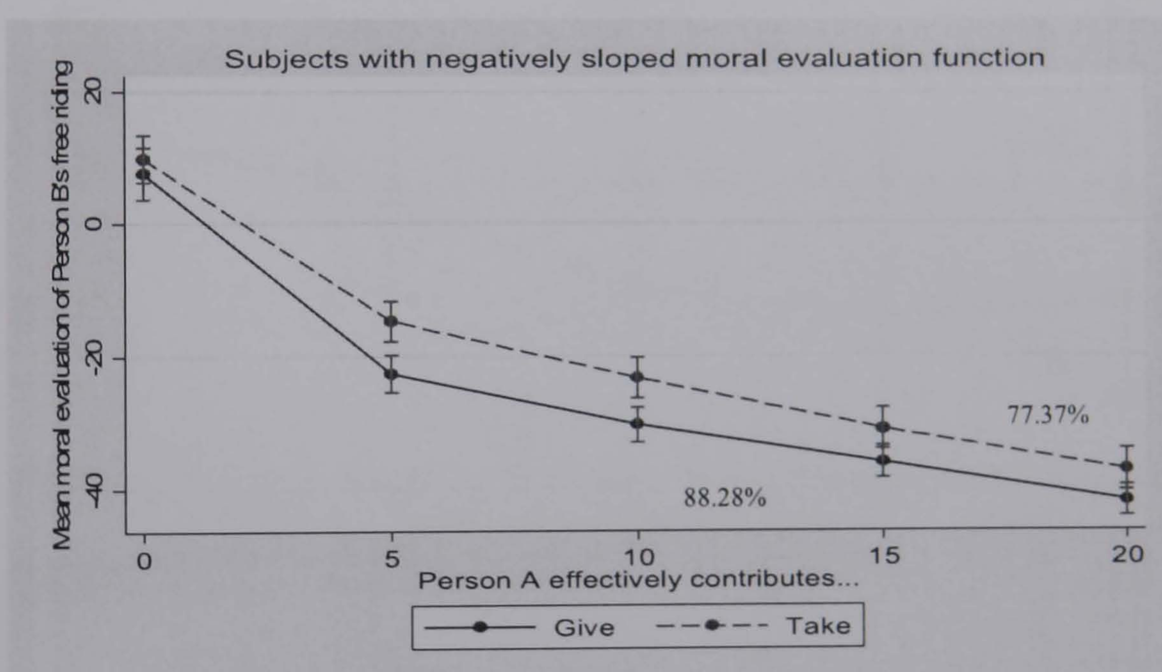
We now turn our attention to the investigation of how subjects' moral evaluations were affected in the Sequential treatments (that is, when Person A decides first and Person B observes Person A's choice before making his own decision). Figure 2.3 shows the mean moral evaluation function for all subjects. Similar to the Simultaneous treatments, we also identify a framing effect to the same direction as displayed in Figure 2.1: subjects tend to be more condemning in the Give than in the Take frame. A noteworthy aspect of subjects' moral evaluations is that, under both frames, free riding is exonerated and perceived to be a morally commendable action if Person B is aware that Person A is a free rider as well. However, as Person A's effective contribution increases, moving from scenario 1 to scenario 2, there is a large drop in moral evaluations, rendering free riding a morally impermissible action. Comparing the slope of the moral evaluation function in the Simultaneous treatments to its slope in the Sequential treatments for the same move, we observe that it is much flatter in the former case (see Table B.4 in Appendix B for econometric evidence).

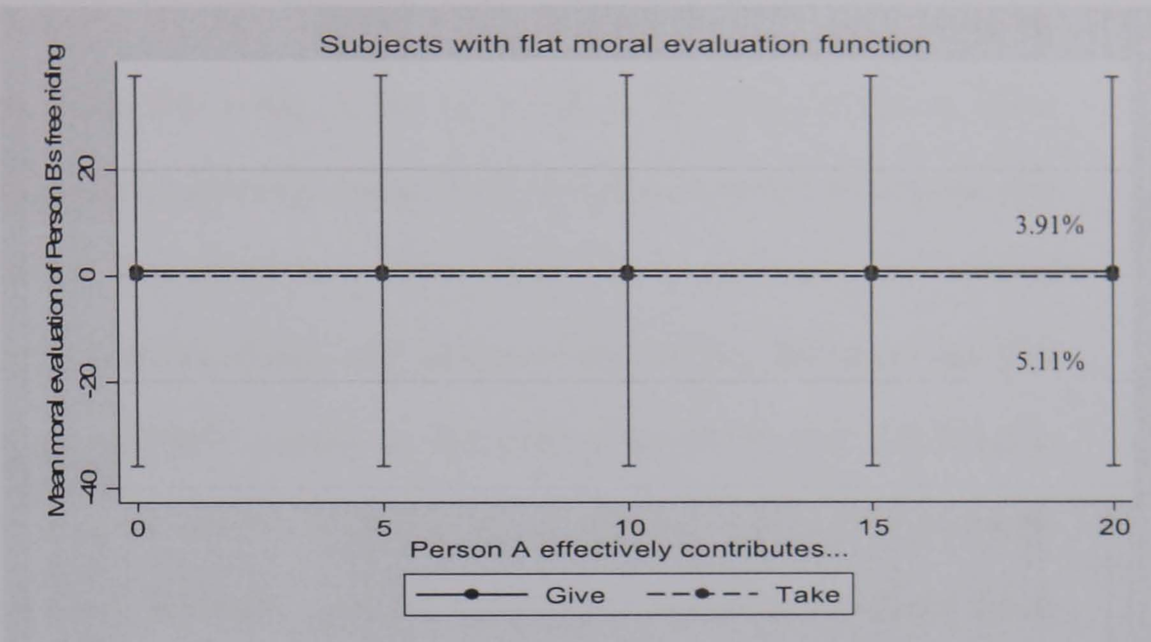
Figure 2.3. Give vs. Take in Sequential treatments – The moral evaluation function

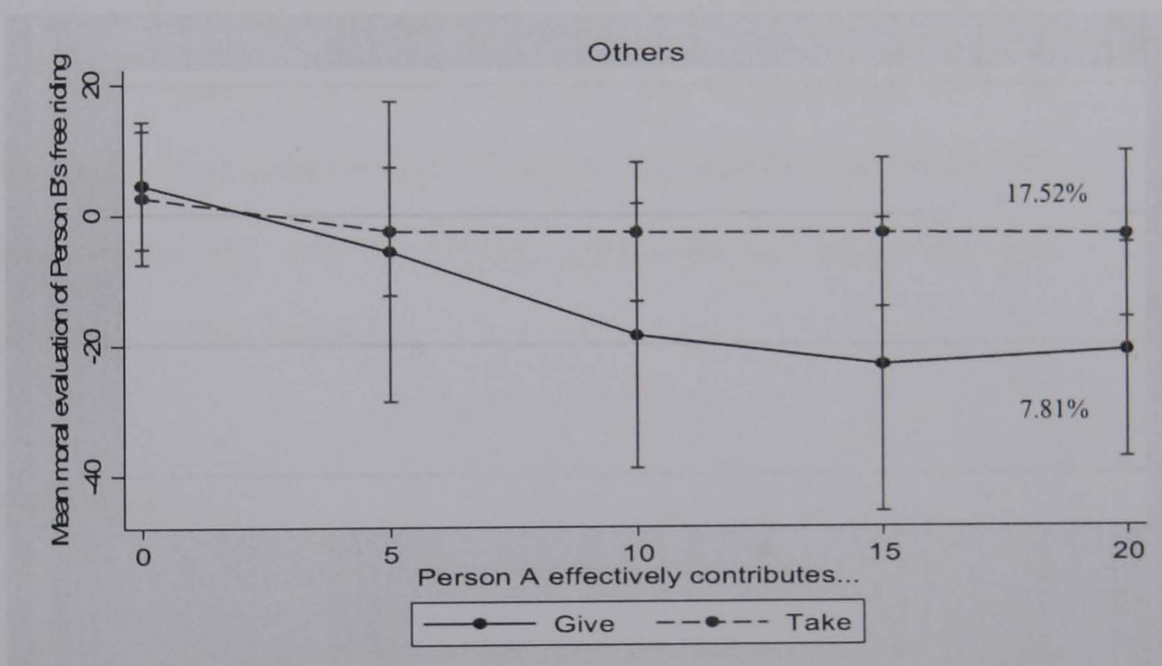


As for the effects of framing in Sequential treatments, formal regression analysis (see Table B.5 in Appendix B) suggests that the regression coefficients for each scenario separately is significantly different from zero and from each other (from F-test, p-value = 0.000). The constant of this equation has a positive sign and is also statistically significant from zero. We also find that there is no level difference with regards to the framing difference, but a slope difference. As it is clear from Figure 2.3, this is due to the substantial drop in moral judgments from the scenario where both players effectively contribute 0 to the scenario where Person A effectively contributes 5 and Person B effectively contributes 0. To understand better the patterns of moral judgments observed in the Sequential treatments, we again classify subjects into three categories as outlined above. The moral evaluation function for each group is shown in Figure 2.4. Each panel includes the 95% confidence interval of mean evaluation of each possible scenario, along with the percentages of subjects falling into each group for the two frames.

Figure 2.4. Give vs. Take in Sequential treatments – The moral evaluation function for each group







From these panels, we clearly observe that the composition of the groups changes dramatically, compared to the simultaneous treatments. More specifically, the increasing condemnation hypothesis characterise the overwhelming majority of subjects in the Sequential treatments, whereas, only a small percentage of subjects fall under the other two groups. Recall that in the Simultaneous treatments, this pattern is completely reversed.

Interestingly, the fact that free riding is perceived a morally acceptable action, when free riding occurs on a known free rider, is due to those subjects with a negatively sloped moral evaluation function. In the other two categories, free riding on a known free rider is perceived as a morally neutral action. Examining the source of the framing difference, we find again an opposite pattern to the one observed in the Simultaneous treatments: the significant majority of the negatively sloped subjects appear to drive this difference, whereas, those with a flat and a nearly flat moral evaluation function are frame insensitive. Under the Sequential treatments our findings suggest that moral judgments can be reconciled with the reason-based moral accounts only if subjects endorse ethical principles that license reciprocation and only if their judgments are driven by non-consequentialist ethical motives.

4. Conclusions

This chapter investigates experimentally people's moral impartial evaluations in the context of an experimental game that has played a key role in the literature: the public goods game. Our main findings can be summarized in the following three points.

First we establish that social dilemmas are perceived as having a moral dimension. Specifically, we find that free riding is a morally reprehensible action except for one case, in which it is perceived as a morally acceptable action: the free rider knows that the other group members have also free ridden. Second we show that contextual cues that would be irrelevant from a consequentialist perspective are a major determinant of moral evaluations. Subjects tend to condemn a failure to contribute more strongly than a withdrawal of support with the same net consequences. These findings are explained either by an emotions-based model or a reason-based model only if subjects endorse non-consequentialist moral principles. Third, we find evidence in support of the increasing condemnation hypothesis: the free rider is morally increasingly condemned, as the effective contribution of the other player increases. This pattern is more prevalent in the sequential treatments and can be attributed to subjects' moral views consistent with either an emotions-based model or a reason-based model assuming that subjects endorse moral principles that licence reciprocation. However, the presence of increasing condemnation by a significant minority of subjects in the simultaneous treatments cannot easily be reconciled with reason-based ethical view, but it is more readily explicable with an emotions-based model in which gut reactions and intuitions generate the negative slope of the moral evaluation function, since such reactions may respond to features of the scenarios not known to the free rider.

Appendix A. Technical Appendix

Appendix A1: Instructions

Questionnaire for the Give frame and the "Simultaneous" treatment

Imagine a group that consists of two group members, Person A and Person B. Each group member receives an endowment of 20 tokens and has to decide how many tokens to keep for himself and how many to contribute to a group project. Each token he keeps for himself has a value of one pound for him. Each token contributed to the group project has a value of 1.50 pounds to the project. The total value of the project is divided equally between the two group members. So, each token contributed to the project earns both group members 0.75 pounds each. The total income of a group member is the sum earned from tokens kept for himself and his share of the earnings of the group project. Each group member decides simultaneously, that is, without knowing what the other one has done.

A) Assume that Person A contributes 0 tokens to the group project and Person B contributes 0 tokens to the group project. Therefore, the value of the group project is 0 pounds and, thus, as a result of their contributions, Person A's total income is 20 pounds and Person B's total income is 20 pounds. How do you rate **Person B's** morality? (Please select –50:'extremely bad' to 50:'extremely good' by moving the slider; Clicking on a slider activates it):

B) Assume that Person A contributes 5 tokens to the group project and Person B contributes 0 tokens to the group project. Therefore, the value of the group project is 7.5 pounds and, thus, as a result of their contributions, Person A's total income is 18.75 pounds and Person B's total income is 23.75 pounds. How do you rate **Person B's** morality? (Please select –50:'extremely bad' to 50:'extremely good' by moving the slider; Clicking on a slider activates it):

C) Assume that Person A contributes 10 tokens to the group project and Person B contributes 0 tokens to the group project. Therefore, the value of the group project is 15 pounds and, thus, as a result of their contributions, Person A's total income is 17.5 pounds and Person B's total income is 27.5 pounds. How do you rate **Person B's** morality? (Please select –50:'extremely bad' to 50:'extremely good' by moving the slider; Clicking on a slider activates it):

D) Assume that Person A contributes 15 tokens to the group project and Person B contributes 0 tokens to the group project. Therefore, the value of the group project is 22.5 pounds and, thus, as a result of their contributions, Person A's total income is 16.25 pounds and Person B's total income is 31.25 pounds. How do you rate **Person B's** morality? (Please select –50:'extremely bad' to 50:'extremely good' by moving the slider; Clicking on a slider activates it):

E) Assume that Person A contributes 20 tokens to the group project and Person B contributes 0 tokens to the group project. Therefore, the value of the group project is 30 pounds and, thus, as a result of their contributions, Person A's total income is 15 pounds and Person B's total income is 35 pounds. How do you rate **Person B's** morality? (Please select –50:'extremely bad' to 50:'extremely good' by moving the slider; Clicking on a slider activates it):

Is there a difference among the five cases described above? Please explain why, or why not, in just a few sentences.

Questionnaire for the Take frame and the "Simultaneous" treatment

Imagine a group that consists of two group members, Person A and Person B. There are 40 tokens in a group project. Each group member has to decide how many, up to a maximum of 20, of these tokens to withdraw for himself and how many to leave in the group project. Each token he withdraws for himself has a value of one pound for him. Each token left in the group project has a value of 1.50 pounds to the project. The total value of the project is divided equally between the two group members. So, each token left in the project earns both group members 0.75 pounds each. The total income of a group member is the sum earned from tokens withdrawn by himself and his share of the earnings of the group project. Each group member decides simultaneously, that is, without knowing what the other one has done.

A) Assume that Person A withdraws 20 tokens from the group project and Person B withdraws 20 tokens from the group project. Therefore, the value of the group project is 0 pounds and, thus, as a result of their withdrawals, Person A's total income is 20 pounds and Person B's total income is 20 pounds. How do you rate **Person B's** morality? (Please select –50:'extremely bad' to 50:'extremely good' by moving the slider; Clicking on a slider activates it):

B) Assume that Person A withdraws 15 tokens from the group project and Person B withdraws 20 tokens from the group project. Therefore, the value of the group project is 7.5 pounds and, thus as a result of their withdrawals, Person A's total income is 18.75 pounds and Person B's total income is 23.75 pounds. How do you rate **Person B's** morality? (Please select –50:'extremely bad' to 50:'extremely good' by moving the slider; Clicking on a slider activates it):

C) Assume that Person A withdraws 10 tokens from the group project and Person B withdraws 20 tokens from the group project. Therefore, the value of the group project is 15 pounds and, thus as a result of their withdrawals, Person A's total income is 17.5 pounds and Person B's total income is 27.5 pounds. How do you rate **Person B's** morality? (Please select –50:'extremely bad' to 50:'extremely good' by moving the slider; Clicking on a slider activates it):

D) Assume that Person A withdraws 5 tokens from the group project and Person B withdraws 20 tokens from the group project. Therefore, the value of the group project is 22.5 pounds and, thus as a result of their withdrawals, Person A's total income is 16.25 pounds and Person B's total income is 31.25 pounds. How do you rate **Person B's** morality? (Please select –50:'extremely bad' to 50:'extremely good' by moving the slider; Clicking on a slider activates it):

E) Assume that Person A withdraws 0 tokens from the group project and Person B withdraws 20 tokens from the group project. Therefore, the value of the project is 30 pounds and, thus as a result of their contributions, Person A's total income is 15 pounds and Person B's total income is 35 pounds. How do you rate **Person B's** morality? (Please select –50:'extremely bad' to 50:'extremely good' by moving the slider; Clicking on a slider activates it):

Is there a difference among the five cases described above? Please explain why, or why not, in just a few sentences.

Questionnaire for the Give frame and the "Sequential" treatment

Imagine a group that consists of two group members, Person A and Person B. Each group member receives an endowment of 20 tokens and has to decide how many tokens to keep for himself and how many to contribute to a group project. Each token he keeps for himself has a value of one pound for him. Each token contributed to the group project has a value of 1.50 pounds to the project. The total value of the project is divided equally between the two group members. So, each token contributed to the project earns both group members 0.75 pounds each. The total income of a group member is the sum earned from tokens kept for himself and his share of the earnings of the group project. Assume that Person A decides first and Person B observes Person A's choice before making his own decision.

A) Assume that Person A contributes 0 tokens to the group project and, then, Person B contributes 0 tokens to the group project. Therefore, the value of the group project is 0 pounds and, thus, as a result of their contributions, Person A's total income is 20 pounds and Person B's total income is 20 pounds. How do you rate **Person B's** morality? (Please select –50:'extremely bad' to 50:'extremely good' by moving the slider; Clicking on a slider activates it):

B) Assume that Person A contributes 5 tokens to the group project and, then, Person B contributes 0 tokens to the group project. Therefore, the value of the group project is 7.5 pounds and, thus, as a result of their contributions, Person A's total income is 18.75 pounds and Person B's total income is 23.75 pounds. How do you rate **Person B's** morality? (Please select –50:'extremely bad' to 50:'extremely good' by moving the slider; Clicking on a slider activates it):

C) Assume that Person A contributes 10 tokens to the group project and, then, Person B contributes 0 tokens to the group project. Therefore, the value of the group project is 15 pounds and, thus, as a result of their contributions, Person A's total income is 17.5 pounds and Person B's total income is 27.5 pounds. How do you rate **Person B's** morality? (Please select –50:'extremely bad' to 50:'extremely good' by moving the slider; Clicking on a slider activates it):

D) Assume that Person A contributes 15 tokens to the group project and, then, Person B contributes 0 tokens to the group project. Therefore, the value of the group project is 22.5 pounds and, thus, as a result of their contributions, Person A's total income is 16.25 pounds and Person B's total income is 31.25 pounds. How do you rate **Person B's** morality? (Please select –50:'extremely bad' to 50:'extremely good' by moving the slider; Clicking on a slider activates it):

E) Assume that Person A contributes 20 tokens to the group project and, then, Person B contributes 0 tokens to the group project. Therefore, the value of the group project is 30 pounds and, thus, as a result of their contributions, Person A's total income is 15 pounds and Person B's total income is 35 pounds. How do you rate **Person B's** morality? (Please select –50:'extremely bad' to 50:'extremely good' by moving the slider; Clicking on a slider activates it):

Is there a difference among the five cases described above? Please explain why, or why not, in just a few sentences.

Questionnaire for the Take frame and the "Sequential" treatment

Imagine a group that consists of two group members, Person A and Person B. There are 40 tokens in a group project. Each group member has to decide how many, up to a maximum of 20, of these tokens to withdraw for himself and how many to leave in the group project. Each token he withdraws for himself has a value of one pound for him. Each token left in the group project has a value of 1.50 pounds to the project. The total value of the project is divided equally between the two group members. So, each token left in the project earns both group members 0.75 pounds each. The total income of a group member is the sum earned from tokens withdrawn by himself and his share of the earnings of the group project. Assume that Person A decides first and Person B observes Person A's choice before making his own decision.

A) Assume that Person A withdraws 20 tokens from the group project and, then, Person B withdraws 20 tokens from the group project. Therefore, the value of the group project is 0 pounds and, thus, as a result of their withdrawals, Person A's total income is 20 pounds and Person B's total income is 20 pounds. How do you rate **Person B's** morality? (Please select –50:'extremely bad' to 50:'extremely good' by moving the slider; Clicking on a slider activates it):

B) Assume that Person A withdraws 15 tokens from the group project and, then, Person B withdraws 20 tokens from the group project. Therefore, the value of the group project is 7.5 pounds and, thus as a result of their withdrawals, Person A's total income is 18.75 pounds and Person B's total income is 23.75 pounds. How do you rate **Person B's** morality? (Please select –50:'extremely bad' to 50:'extremely good' by moving the slider; Clicking on a slider activates it):

C) Assume that Person A withdraws 10 tokens from the group project and, then, Person B withdraws 20 tokens from the group project. Therefore, the value of the group project is 15 pounds and, thus as a result of their withdrawals, Person A's total income is 17.5 pounds and Person B's total income is 27.5 pounds. How do you rate **Person B's** morality? (Please select –50:'extremely bad' to 50:'extremely good' by moving the slider; Clicking on a slider activates it):

D) Assume that Person A withdraws 5 tokens from the group project and, then, Person B withdraws 20 tokens from the group project. Therefore, the value of the group project is 22.5 pounds and, thus as a result of their withdrawals, Person A's total income is 16.25 pounds and Person B's total income is 31.25 pounds. How do you rate **Person B's** morality? (Please select –50:'extremely bad' to 50:'extremely good' by moving the slider; Clicking on a slider activates it):

E) Assume that Person A withdraws 0 tokens from the group project and, then, Person B withdraws 20 tokens from the group project. Therefore, the value of the project is 30 pounds and, thus as a result of their contributions, Person A's total income is 15 pounds and Person B's total income is 35 pounds. How do you rate **Person B's** morality? (Please select –50:'extremely bad' to 50:'extremely good' by moving the slider; Clicking on a slider activates it):

Is there a difference among the five cases described above? Please explain why, or why not, in just a few sentences.

Appendix A.2: Advantages and Disadvantages of Internet Experiments

The advancement of the World Wide Web has been proved to be a useful tool of conducting experimental research in the domain of psychology. During the last decade, an increasing number of researchers use the Internet as a medium to collect data and their experience of how to conduct such experiments is of great importance for the economics science as well. A growing number of Web psychological experiments have been conducted and advertised via the Internet, with some informative links for the reader being as follows (although this list is not considered to be conclusive):

- http://psych.fullerton.edu/mbirnbaum/exp.htm (M. Birnbaum's judgment and decision making experiments)
- http://www.psychologie.unizh.ch/sowi/Ulf/Lab/WebExpPsyLab.html (Ulf-Dietrich Reips' Web experimental psychology lab)
- http://psych.hanover.edu/research/exponnet.html (American Psychological Society)
- http://vacognition.wjh.harvard.edu/ (Online experiments from the Harvard University)
- http://www.surf.to/experiments (Portal for Psychological Experiments on Language from the Universities of Edinburgh, Glasgow and Saarlandes)

The increasing conduct of Web experiments during the last years has resulted into recognising a number of advantages and disadvantages of this medium of experimental research.[12] The main advantage of the Web experiments is related to the generalisability of their findings in two main respects. First, the number of people using the Internet is growing rapidly and those who have access to online experiments increase over time. The population that can now take part in a Web experiment is more diverse and thus, more representative, compared to the usual student subject pool, used in traditional laboratory experiments. Participants with completely different demographic/socio-economic and cultural characteristics can be reached easily, facilitating the conduct of research, whose primary purpose is to examine these variables. Second, the sample diversity of the Internet users adds also to the external validity of the Internet experiments. Experiments in the lab set aside the external validity of the experiment at the benefit of gaining more internal validity (i.e. satisfying the requirements for drawing a causal relationship). However, with online experiments, external validity and generalisability of the results can be potentially increased, given the wide variety of the participant population.

Experiments on the Web give also the opportunity to subjects to participate at any time and day they are available. The experiment now goes to those interested, who can take part at their own convenience, and not at the experimenter's convenience. The number of participants increases with impressive speed since there are no scheduling difficulties or time constraints with regard to use of lab. The lack of scheduling

[12] For a survey and more detailed discussion about the pros and cons of Web psychological experiments, see Reips (2000; 2002) and Birnbaum (2004).

inconveniencies, along with the savings on the lab space and its equipment, make Internet experiments an economic medium of obtaining valid data.

The plethora of data that can be collected via online experiments has also positive implications in the statistical analysis, since one way to increase the statistical power of the data analysis is to increase the sample size. The participation rate is much higher in Internet experiments than in lab ones, and therefore, the data have more statistical power.

In most psychological lab experiments, subjects take part in the experiment, because they will receive extra credit for their course as a counter-reward for their coming. However, this simply means that not all subjects voluntarily participate in the experiment. The so-called "volunteer bias" in the psychological lab experiments is not present in the Web experiments, where subjects voluntarily participate in the experiment they wish to at their own time. This increasing degree of voluntary participation is coming at some cost: subjects in the Web experiments may show lower commitment in completing their task. They can leave the experiment and drop out at any time they want. This drawback can be turned into an advantage in a between-subjects design, since the drop out rate can help us detect motivational confounding. Put simply, a higher drop out rate indicates that the task is more difficult or less attractive.

A criticism that lab experiments have accepted is the interaction between experimenters and subjects. The presence of the experimenter in the lab when subjects make their decisions, known also as the "experimenter bias", can influence and contaminate experimental data. This bias is kept at a minimum level with the conduct of Internet experiments, given that subjects do not have any contact with the experimenter at all. The lack of the "experimenter bias" is an additional factor that increases the external validity of this type of experiments.

Yet, the advantages do not come without some cost. To begin with, the lack of experimental control can be mentioned as a first disadvantage, which is due to the high external validity of the Web experimental data, compared to those obtained in the lab. One aspect of this kind of lack refers to multiple submissions, especially in experiments, which are widely advertised online. A number of solutions have been put forward to detect multiple submissions, including, *inter alia*, that subjects could be asked at the beginning of the experiment to participate only once, experimenters could check for similarities in the IP addresses of the participants' computers, contact a random sample of participants after the experiment and checking for similarities in their answers (*sub-sampling technique*), or allow participation by password only (*password technique*). However, we need to emphasise that evidence from psychological mainly experiments suggests multiple submissions is not considered as a particularly important problem, since it is quite rare for subjects to participate more than once in the same experiment.

Another drawback which is present in online experiments is the so-called "self-selection" bias, since participation is voluntary for all subjects. A possible solution to this is the ***multiple site entry technique***. This technique uses multiple entry Web pages to participate in the experiment.

Comparing the data of participants from these different sites, it is possible to detect whether self-selection is an important factor, affecting the results, or not. The high drop out rates because of the voluntary participation in Web experiments is another potential disadvantage. To mitigate this downside of Web experiments, suggested solutions are to create an attractive website, put emphasis on the trustworthiness of the site, explain to subjects the importance of the study and how valuable the data are for the science, provide information with the current status of the participants' answers and the total length of the experiment.

Finally, conducting experiments on the Internet is limited to those cases, where no interaction between the subjects and the experimenter is required. For instance, in cases where such interaction is needed (for instance, in fMRI scans), then Web-based experiments are not the appropriate tool to use.

Appendix B. Data Analysis

Table B.1. Does paying a random participation fee affect response rates?

Independent Variables	**Dependent Variable: Participation = 1; No-participation = 0**
Payment	0.054***
	(0.015)
Male	-0.013
	(0.015)
Obs.	2,718

*Notes: Marginal effects listed. Robust standard errors are presented in parentheses. The variable "Payment" is a dummy variable equal to 1 for those subjects who participated in the "Payment" condition and 0 otherwise. The variable "Male" is a dummy variable equal to 1 for male subjects and 0 otherwise. ** denotes significance at the 5-percent level, and *** at the 1-percent level.*

Table B.2. Does paying a random participation fee affect moral evaluations?

Independent Variables	**Dependent Variable: Moral evaluations of the free rider**
Person A contributes 5 tokens	-15.571***
	(0.923)
Person A contributes 10 tokens	-20.314***
	(1.046)
Person A contributes 15 tokens	-23.942***
	(1.133)
Person A contributes 20 tokens	-27.507***
	(1.224)
Payment	1.529
	(1.534)
Male	-0.857
	(1.531)
Constant	0.108
	(1.526)
Obs.	2,690

*Notes: OLS estimates. Robust standard errors are presented in parentheses and clustered on individuals. ** denotes significance at the 5-percent level, and *** at the 1-percent level.*

Table B.3. Give vs. Take in Simultaneous treatments – Regression results

	Dependent variable: Moral evaluation of the free rider			
	All subjects	Subjects with negatively sloped function	Subjects with flat function	"Others"
Person A contributes 5 tokens	-9.481*** (1.374)	-19.327*** (2.483)	–	-13.75*** (4.273)
Person A contributes 10 tokens	-11.422*** (1.533)	-24.788*** (2.478)	–	-12.65** (4.852)
Person A contributes 15 tokens	-12.837*** (1.601)	-29.423*** (2.445)	–	-10.15** (3.914)
Person A contributes 20 tokens	-15.541*** (1.863)	-35.442*** (2.611)	–	-12.75** (4.855)
Take	5.834** (2.437)	4.367 (4.055)	7.066** (3.187)	10.883* (6.162)
Male	-2.342 (2.229)	-3.025 (3.644)	-2.183 (3.245)	-4.961 (5.161)
Person A contributes 5 tokens × Take	3.126 (1.923)	3.089 (3.856)	–	5.272 (6.538)
Person A contributes 10 tokens × Take	2.400 (2.200)	1.527 (4.124)	–	0.998 (7.171)
Person A contributes 15 tokens × Take	1.902 (2.319)	1.256 (4.151)	–	-4.024 (6.356)
Person A contributes 20 tokens × Take	2.874 (2.606)	0.561 (4.357)	–	0.446 (6.510)
Constant	-8.006*** (1.557)	-0.069 (2.121)	-15.721*** (2.361)	-2.769 (3.478)
Obs.	1,365	470	680	215

*Notes: (1) OLS estimates. Robust standard errors are presented in parentheses and clustered on individuals. ** denotes significance at the 5-percent level, and *** at the 1-percent level. (2) Scenario dummies were excluded for the subjects whose moral evaluation function was flat.*

Table B.4. Simultaneous vs. Sequential treatments – Regression results

	Dependent variable: Moral evaluation of the free rider
	All subjects
Person A contributes 5 tokens	-23.472*** (1.436)
Person A contributes 10 tokens	-30.725*** (1.557)
Person A contributes 15 tokens	-36.374*** (1.649)
Person A contributes 20 tokens	-41.332*** (1.719)
Simultaneous	-13.642*** (1.729)
Male	-1.100 (1.510)
Person A contributes 5 tokens × Simultaneous	15.571*** (1.729)
Person A contributes 10 tokens × Simultaneous	20.516*** (1.906)
Person A contributes 15 tokens × Simultaneous	24.498*** (2.015)
Person A contributes 20 tokens × Simultaneous	27.244*** (2.157)
Constant	8.008*** (1.452)
Obs.	2,690

*Notes: (1) OLS estimates. Robust standard errors are presented in parentheses and clustered on individuals. ** denotes significance at the 5-percent level, and *** at the 1-percent level. (2) Scenario dummies were excluded for the subjects whose moral evaluation function was flat.*

Table B.5. Give vs. Take in Sequential treatments – Regression results

	Dependent variable: Moral evaluation of the free rider			
	All subjects	Subjects with negatively sloped function	Subjects with flat function	"Others"
Person A contributes 5 tokens	-27.398*** (2.102)	-30.142*** (2.147)	–	-10.1 (8.046)
Person A contributes 10 tokens	-35.25*** (2.155)	-37.885*** (2.209)	–	-23.1*** (7.592)
Person A contributes 15 tokens	-40.586*** (2.148)	-43.531*** (2.103)	–	-27.6*** (9.437)
Person A contributes 20 tokens	-45.602*** (2.159)	-49.407*** (2.093)	–	-25.4*** (6.789)
Take	1.054 (2.515)	2.379 (2.705)	3.846 (10.989)	-2.695 (6.633)
Male	-0.731 (1.967)	-1.454 (1.750)	-9.423 (8.060)	10.935 (7.164)
Person A contributes 5 tokens × Take	7.596*** (2.848)	5.708** (2.883)	–	4.975 (9.886)
Person A contributes 10 tokens × Take	8.754*** (3.072)	4.828 (3.064)	–	17.85* (9.655)
Person A contributes 15 tokens × Take	8.148** (3.253)	2.814 (3.092)	–	22.267* (11.401)
Person A contributes 20 tokens × Take	8.258** (3.387)	2.426 (3.068)	–	19.733* (9.865)
Constant	7.309*** (1.968)	8.080*** (2.076)	2.885 (12.864)	1.220 (3.693)
Obs.	1,325	1,095	60	170

*Notes: (1) OLS estimates. Robust standard errors are presented in parentheses and clustered on individuals. ** denotes significance at the 5-percent level, and *** at the 1-percent level. (2) Scenario dummies were excluded for the subjects whose moral evaluation function was flat.*

Chapter 3

Framing and Free Riding: Emotional Responses and Punishment in Social Dilemma Games[13]

3.1 Introduction

The experimental literature has used a number of simple games to measure aspects of social preferences (Camerer and Fehr, 2004). Its findings have helped to inspire the development of new models of other-regarding preferences, such as those surveyed in Camerer (2003, ch. 2),

[13] The present chapter is a substantially expanded version of a joint paper with Robin Cubitt and Simon Gächter (for more details on this paper see Cubitt, R., Drouvelis, M., and Gächter, S., (2008), 'Framing and Free Riding: Emotional Responses and Punishment in Social Dilemmas', *CeDEx Working Paper 2008-02*).

which in turn have begun to be applied in other areas of economics (Gintis, Bowles, Boyd and Fehr, 2005). However, recent studies have cast doubt on the robustness of elicited social preferences to framing and contextual changes, especially in dictator games (e.g. Dana, Weber and Kuang, 2005; Haley and Fessler, 2005; Bardsley, 2007; List, 2007). Evidence from these studies suggests that behaviour is sensitive to seemingly irrelevant context-specific cues. If such a finding held more generally for the games used to measure social preferences in the laboratory, this would have serious implications for the external validity of the measures obtained, for the theories motivated by them, and for the explanation of subjects' behaviour.

In this chapter, we examine the issue of frame-sensitivity in relation to the darker side of social preferences by studying a different experimental game that has played a major part in the literature: the public goods game with punishment. The framing manipulation we consider is of the contributions stage of the game and has a Give versus Take form, previously studied in the social psychology literature and similar to that introduced to economics by Andreoni (1995).[14] In one frame, subjects decide how much to contribute to a public good; in the other frame, they decide how much to withdraw from it. The setup is such that the two decision problems are objectively equivalent, in terms of the feasible set of allocations available to a subject. However, studies from social psychology found related framing manipulations to have an effect on contributions (e.g. Brewer and Kramer, 1986; McDaniel and Sistrunk, 1991; McCusker and Carnevale, 1995; Sell

[14] For a conceptual discussion and classification of framing effects, see Levin, Schneider and Gaeth, 1998.

and Son, 1997; van Dijk and Wilke, 2000). Experimental economists have also identified framing effects, typically in a repeated game context (Andreoni, 1995; Sonnemans, et al., 1998; Willinger and Ziegelmeyer, 1999; Cookson, 2000 and Park, 2000).

Our study is distinctive in two main respects. First, we add a stage to the game in which subjects can punish one another and a further phase in which their self-reported emotional responses are elicited. Second, unlike most of the economics literature, we study a one-shot game.

Rather than considering only the impact of framing on contributions to the public good, we focus mainly on whether it affects emotions and punishment behaviour. The fact that emotions and punishment have played a central role in the social preference literature provides a general motivation for this aspect of our study (Fehr and Gächter, 2000; Bosman and van Winden, 2002; Andreoni, et al., 2003; Masclet, et al., 2003).

A more specific motivation stems from the findings reported in Chapter 2. Recall that in this experiment subjects, playing the role of impartial observers, pass moral judgments on agents in hypothetical scenarios. The scenarios concerned a simple public goods game. Some were described using a Give frame and others using a Take frame; and subjects' judgments of the agents proved to be sensitive to this difference. On average, subjects tended to condemn a failure to contribute more strongly than a withdrawal of support with the same net consequences. However, as there may be a difference between judgments made in the role of impartial observer and responses to agents with whom one is interacting, it is important to investigate whether a parallel finding holds for subjects who are materially affected by each other's decisions. For this case, self-reported emotions are analogous to moral judgments to the extent of being non-behavioural indicators of attitudes; whereas punishment provides a behavioural indicator of willingness to act on one's attitudes.

We study a one-shot game for two reasons. First, as we explain in Section 3.2, one-shot games provide cleaner tests for framing effects than repeated games.[15] Second, and for similar reasons, one-shot games are of particular interest to the study of social preferences, as they prevent other-regarding considerations from being confounded with strategic ones arising from repetition of the game.

An important issue in studying the effect of the framing of the contribution stage on subsequent punishment behaviour and emotional responses is to recognise that the framing might affect punishment and emotions *either* because it affects the levels of contributions *or* because it affects how subjects see a given level of contributions. We are most interested in the latter mechanism, since the former is already well-understood. Our results suggest that, on average, neither punishment nor emotion varies with the framing of the contribution stage, when one controls for the level of contributions. Thus, the social preferences we observe are robust to our framing manipulation. Although our main research questions are whether and if so, how punishment and emotions, as two separate

[15] Conditional on the existence of a framing effect, it may be of interest to study the persistence or otherwise of the effect in repeat play. But, persistence is a separate issue from that of the initial existence of the effect. It is for study of the latter issue that we suggest one-shot games are superior.

indicators of negative reciprocity, are context dependent, our design allows us to investigate the relationship between punishment and emotions. We maintain that it is worth exploring this possibility and present our findings from this analysis in an appendix.

The remainder of the chapter is structured as follows. Section 3.2 provides a literature review of the relevant framed social dilemma experiments in this topic through the lens of social psychologists and experimental economists. Section 3.3 describes the framework, the design and the hypotheses of our experiment. Section 3.4 discusses the results and Section 3.5 concludes.

3.2 Framing and Free Riding in Social Dilemma Games: A Literature Review

3.2.1 Psychological experiments

Social psychologists have investigated extensively the role of framing in social dilemma games. Despite the fact that the methodology of psychological experiments exhibits significant differences to that employed in economics experiments,[16] evidence from the former type of experiments can provide us with useful insights, since the frames we study are equivalent

---

[16] Some of the methodological disparities between psychological and economics experiments lie in the fact that, in the former case, subjects' decisions are not incentivised according to their performance, false feedback is given to subjects, decision-making is based on hypothetical scenarios and description of the game is often modelled on real-life examples, making wording more loaded (see also for some discussion Hertwig and Ortmann, 2008).

in terms of final consequences, but may not be equivalent from a psychological or a moral (see Chapter 2) viewpoint. For our purposes, we focus our attention on those psychological experiments which are closest to our framing manipulation. Recall that we are specifically interested in Give versus Take game structures; that is, environments where subjects can either give to a common resource or take from it. Findings from psychological experiments are mixed and equivocal with respect to how framing affects cooperation.

In a classic study by Brewer and Kramer (1986) subjects' cooperative behaviour between give and take dilemmas is explored. For the give treatment, each subject is endowed with 25 points in their personal accounts and the total number of points in their personal account would determine their earnings. On every trial the size of the common pool temporarily dropped by either 200 points (in the case where the group size was equal to 8) or 800 (in the case where the group size was equal to 32). Subjects had to decide how many of these points to return/give to the common pool. For the next trial the size of the common pool was determined by the total contributions and the random replenishment rate ranging from 1% to 10%. In total, there were 20 trials, in which trial the common pool was completely depleted by the computer. For the take frame, each subject had now to decide how many points to withdraw (up to 25) from the common resource. These points were then added to their personal accounts.

To explain choice behaviour in these two dilemmas, Brewer and Kramer rely on the risk ingredient of Kahneman and Tversky's (1984) prospect theory. Specifically, under the give frame, subjects' contribution involves giving up something subjects already posses (certain loss) in order to obtain an uncertain benefit (replenishment rate and total contributions were unknown from trial to trial). In this treatment, subjects will be risk-seeking and thus, will contribute less to the public good. However, under the take frame, whatever subjects withdraw constitutes a certain gain and this makes individuals more risk averse. It is, therefore, expected that subjects will prefer to take a small amount from the common pool (definite gain) and leave more in the public good. On this basis, Brewer and Kramer hypothesise that a take frame evokes more cooperation than a give frame. Their findings provide support for this hypothesis.

In another series of experiment, Sell and Son (1997) analysed choice behaviour in give versus take dilemma under both static and dynamic contexts.[17] Their results provide, to some extent, support to Brewer and Kramer's effect. Specifically, in both static and dynamic contexts, the take dilemma produces higher cooperation than the give one on the first trial. However, this difference diminishes over time, in dynamic contexts and when group interaction (that is, knowledge of how much others contribute) is added in the static context. Following a similar experimental design as Brewer and Kramer (1986), McCusker and Carnevale (1995) explain

---

[17] In the Give static context, subjects were endowed with 25 tokens and have to decide whether to contribute all or a portion of their tokens to the common pool. In the Take static context, subjects had to decide whether to take up to 25 tokens and transfer it to their private fund. In the Give dynamic context, the group account was endowed with 600 tokens. Subjects were also endowed with 25 tokens per period and have to decide how many of these 25 tokens to keep for themselves and how many to invest to the group account. Note that the size of the group account was subject to a replenishment rate ranging from 1% to 10%. In the Take dynamic context, the decisions involved keeping tokens in the group account or withdrawing tokens from the group account. In this condition, the group account was endowed by 700 tokens, 100 of which could be withdrawn by group members.

behaviour in give and take dilemmas using the loss aversion hypothesis. Since losses loom larger than gains, they argue that people are motivated to avoid a loss more that they are motivated to obtain a gain, anticipating thus higher levels of cooperation in take than in give games. Their findings replicate Brewer and Kramer's (1986) framing effect. Interestingly, in the give frame, subjects do not match their fellow group members' contribution behaviour, while, the opposite effect is observed under the take frame. This was due to the fact that when others were cooperative, the Give dilemma evoked lower levels of cooperation than the Take dilemma. A similar behavioural pattern is corroborated by Fleishman (1988): subjects conformed to others' behaviour under the take frame, but not under the give frame, where the average group contribution was greater when others had contributed little.[18] When choice behaviour between give and take dilemmas is compared, Fleishman does not find any significant difference. The absence of a difference in cooperation rates between frames has been also documented by Rutte, Wilke and Messick (1987). Contrary to the evidence reported above, findings by McDaniel and Sistrunk (1991) suggest that the give context produces higher cooperation rates than the corresponding take context. Additionally, as the expected cooperation of others increased, levels of cooperation for both game structures increased, but the increase in the take game was not higher than the corresponding increase in the give game. However, it is important to note that this finding is based on a within-subjects analysis.

[18] Note that the behaviour of the others was pre-programmed by the experimenter but remained constant across frames.

In sum, evidence from psychological experiments suggests that different game contexts are an important factor that affects levels of cooperation, making the Give and Take frame psychologically different, albeit economically equivalent. However, much of the inconsistent behaviour observed in these experiments appears to be produced by different methodologies and environments in which subjects interact. For instance, Brewer and Kramer (1986) consider a dynamic context, where the size of the common pool replenishes at a random rate from trial to trial: an environment different to that used in Fleishman (1988). Further, McDaniel and Sistrunk (1991) measure cooperative choices in a context where subjects act as hypothetical CEOs of an agricultural firm. This context is quite different from the one used in other studies, such as in Sell and Son (1997) or in Brewer and Kramer (1986). What is more, in Fleishman's (1988) and McCusker and Carnevale's (1995) study, decisions are made on repeated trials; whereas in Rutte, et al., (1987) experiment decisions are made in one-shot contexts.

The next sub-section reviews the relevant social dilemma experiments from the economics literature, the framework of which is very similar to ours. As it will become clear below, in the economics literature, the effects are more pronounced to the direction of being more cooperative under the give than under the take context.

3.2.2 Economics experiments

A seminal paper in this area is due to Andreoni (1995), who analysed the effects of positive versus negative framing on cooperation. The type of framing manipulation implemented by Andreoni emphasises the difference between the positive externality generated to other group members from investing to the public account and the negative externality generated to other group members from investing to one's own account. That is, in the positive frame case, subjects' initial endowment was placed in their private account; whereas, in the negative-frame case, their endowment was placed in the public account. Although presentation of the choice problems differs, they are objectively equivalent, in terms that subjects' economic incentives are identical in both conditions.

In Andreoni's experiment, each subject was given an endowment of 60 tokens and each group consisted of 5 subjects. The payoff functions for both conditions are given in equations (1) and (2), respectively:

Positive frame: $\pi_i = x_i + \frac{1}{2} \cdot g_i + \frac{1}{2} \cdot \sum_{j \neq i} g_j$ (1)

Negative frame: $\pi_i = x_i + \frac{1}{2} \cdot g_i - \frac{1}{2} \cdot \sum_{j \neq i} x_j + 120$ (2)

where x_i and g_i denote contributions to the private and public good, respectively. The positive-frame payoff function is equivalent to the negative-frame one by simply substituting the budget constraint of the other group members, $x_j + g_j = 60$, from equation (1) into (2). This allows economic incentives to remain the same, with the dominant-strategy Nash equilibrium being complete free riding under both manipulations. In a ten-period game, Andreoni's findings suggest that subjects are far more cooperative in the positive-frame than in the negative frame condition. In particular, mean contribution levels, in the positive condition, reach 34% of players' total endowment, whereas, corresponding levels in the negative condition were only about 16%. This behavioural asymmetry is attributed to the fact that the "warm-glow" of doing something good exceeds the "cold-prickle" of doing something bad. That is, people feel much better if they contribute to the public good, since they consider their action as a positive one in which many others have cooperated as well; while the same level of cooperation in the negative frame may be considered as a negative action, as long as investment in the private good is made at the expense of the other group members.

Building on Andreoni's type of framing, Willinger and Ziegelmeyer (1999) and Park (2000) tested the robustness of his results in different experimental settings. Specifically, Willinger and Ziegelmeyer designed an experiment where the unique dominant strategy of the game was to partially contribute to the public good; whereas, Park examined the connection between subjects' value orientation and contribution levels in different frames. In the latter experiment, subjects were classified to types according to the value orientation method (see Offerman, et al., 1996) before they made their contribution decisions.[19] Both investigations provide support to Andreoni's framing effect: average levels of contributions are significantly higher in the positive than in the negative frame. Looking at the specific

[19] The value orientation method revealed that subjects are mainly classified as individualistic (65.5%) and cooperative (32%).

type classifications in Park's experiment, it is reported that both individualists and co-operators contributed more in the positive compared to the negative frame, although the difference is not significant for co-operators.

The role of framing effects on cooperative choices has been also investigated by Cookson (2000). In particular, Cookson examined three kinds of presentational variations. The first one was related to the investigation of whether the description of a $4 \cdot N$ stage game as four games with N stages creates a positive restart effect and thus, this effect documented in earlier public goods games (see Andreoni, 1988) can be interpreted as a framing effect. In the second variation, the payoff function is decomposed in two different, but strategically equivalent, ways. In one frame, a standard public goods game was implemented, where subjects have to decide how many tokens to keep in their private account and how many to contribute to a public account that benefits everyone; whereas, in another frame, the payoff function is decomposed in such a way that subjects have to decide how many tokens to keep in their private account and how many to donate to the other group members. In this second frame, donation benefits only the other group members and not the donor.[20] The third framing variation was associated with the use of written instructions: in the first one ("I-task"), the competitive features of the game are stressed (the

[20] In the first frame, the material payoff functions is $\pi_i = 400 - g_i + 0.5 \cdot (g_i + G_{-i})$, where g_i is player i's contribution to the public good; while in the second frame, the payoff function is $\pi_i = 400 - d_i + D_{-i}$, where d_i is player i's donation to the other group members. Both frames are strategically equivalent using, for each player i, the relationships $d_i = 0.5 \cdot g_i$ and $D_{-i} = 0.5 \cdot G_{-i}$.

more a player contributes individually, the less he earns); while, in the second one ("We-task"), the cooperative features of the game are stressed (the more tokens a player contributes, the more income someone else earns). Cookson finds a positive and significant restart effect under the standard payoff description and also that the effects of decomposing the payoff are positive, in the sense that players make significantly more cooperative choices when presented with this frame. Regarding the comprehension tasks, subjects who were given the "We-task" were more cooperative than those who were given the "I-task".

The literature discussed above records the presence of a rather strong framing effect when a linear public goods game is finitely repeated. However, it seems that the existing literature does not provide an explanation for the documentation of such framing effects. Our previous chapter sheds some light by investigating subjects' moral judgments. Our current investigation extends the previous literature (and Chapter 2) by assessing the reasons for the observed framing difference and examining punishment and self-reported emotions, as means of understanding how subjects perceive the Give versus Take manipulation. In particular, we implemented a design where the same group of subjects interacted only once, and also added a punishment stage to the game and a phase in which subjects' self-reported emotional responses are elicited. Such an experimental design presented in more detail in the following section will allow us to shed light on our research questions which are concerned with whether framing affects the way subjects treat, and respond emotionally to, a given deviation from contributions.

3.3 Our laboratory experiment with punishment stage

3.3.1 Framework

The basic building block of our study is a framing manipulation of a voluntary contributions game played by the members of a group. Our framings follow a similar formulation of the payoff function as in Dufwenberg, et al. (2006). In the Give frame, each player is endowed with 20 tokens and has to decide how many of them he keeps for himself and how many he contributes to the public good (described as a "project" to subjects). Each token kept for himself increases his own monetary payoff by one Guilder (our experimental currency). Each token contributed to the public good increases the payoff of every group member by 0.5 Guilders. The payoff function is given by equation (3).

$$\pi_i^1 = 20 - g_i + 0.5 \cdot \sum_{j=1}^{n} g_j \qquad (3)$$

where π_i^1 denotes group member i's payoff, g_i the number of tokens contributed to the public good by group member i, and n the number of group members.[21]

In the Take frame, there are initially $20 \cdot n$ tokens in the public good (described again as a "project" to subjects) for each group. Each group

[21] The reason for the superscript in group member i's payoff emerges below.

member has control of 20 tokens and has to decide how many of them he withdraws from the project and how many he leaves in it. The payoff function for this framing treatment is now given by equation (4).

$$\pi_i^1 = t_i + 0.5 \cdot (20 \cdot n - \sum_{j=1}^{n} t_j) \qquad (4)$$

where t_i indicates the number of tokens withdrawn from the public good by group member i.

In each frame, subjects have the same opportunities, regarding the final allocation of tokens. Equations (3) and (4) are equivalent using, for each player i, the relationship $g_i = 20 - t_i$. Thus, from a consequentialist perspective, the framing manipulation has no strategic significance.

3.3.2 Procedures and Design

Our experimental design centres on a game with two stages. The first stage is a standard linear public goods game presented in two different framings, as described in Section 3.3.1; and the second is a punishment stage. We refer to the two resulting treatments as Give-P (Give frame for contributions, with punishment opportunities) and Take-P (Take frame for contributions, with punishment opportunities). In the second stage of the voluntary contributions game, players are allowed to punish each other. Each subject can assign up to five punishment points to each of the other group members. Punishment is costly both for the punishing and the punished parties. Adapting Fehr and Gächter (2002), we choose a punishment technology in which each punishment point assigned costs the punished player two Guilders and the punishing player one Guilder.

The material payoff function from the whole experiment for a given subject i is given by equation (5).

$$\pi_i = \pi_i^1 - \sum_{j=1}^{n} p_{ij} - 2 \cdot \sum_{j=1}^{n} p_{ji} \tag{5}$$

where π_i^1 denotes group member i's payoff from the contribution stage, and p_{ij} the punishment points group member i assigns to group member j. Conditional on each subject i being motivated to maximise equation (5), the unique subgame perfect equilibrium requires that subjects free ride completely in the first stage and refrain completely from punishing in the second stage.

We asked subjects to state their beliefs about contributions of the other group members after they had made their own contribution decision (but before the punishment stage). We also elicited beliefs about how much punishment a subject expected to receive. We elicited these beliefs after subjects had had the opportunity to punish. Elicitation of beliefs was non-incentivised in order to exclude potential income effects in the punishment treatments. We also wanted to avoid punishment being motivated by disappointment about low payoffs resulting from inaccurate beliefs.

In each treatment, subjects were asked at the end of the game to indicate the intensity of emotions they felt about the actual contribution behaviour of each member of their group. Note that at the time of the emotions' elicitation, subjects were aware of the amount of tokens contributed to the public good by each group member separately (in both no-punishment and punishment treatments). However, they were not informed at that point about the punishment points assigned to them by the other group members (for the corresponding treatments).

The procedure we used to elicit self-reports on perceived emotions is due to Bosman and van Winden (2002). In particular, subjects were given a list of thirteen emotions, and were then asked to indicate the intensity with which they felt each emotion when they saw the contribution of each other group member. The intensity for each emotion was recorded on a 7-point scale (1 = "not at all", ..., 7 = "very much"). Appendix A.1 provides a screenshot of the interface we used for eliciting self-reports on emotions. The list of the thirteen emotions with the order presented to subjects is as follows: warmth, anger, fear, envy, sadness, happiness, shame, irritation, contempt, guilt, joy, jealousy and surprise.

Our design is completed by two treatments without punishment opportunities. In these treatments, the second stage consisted only of the elicitation of beliefs and emotions. We refer to the no-punishment treatments as Give-N (Give frame for contributions, no punishment opportunities) and Take-N (Take frame for contributions, no punishment opportunities). The reason for including the no-punishment treatments is to check for two possibilities: (i) if it turns out that there are differences in emotions across frames, we want to be able to check whether emotions are responses to contributions themselves or *ex post* rationalisations of punishment behaviour; and (ii) if it turns out that there is no difference in emotions across frames, we want to be able to check whether this is because emotional response to contributions is the same or because the act of punishment expunges (or arouses) emotional response.

Each subject was assigned at random to a group of three members and played a one-shot voluntary contributions game under one of the treatments just described. (Sessions were allocated to treatments at random.) The advantage of a one-shot game is that it eliminates confounding effects that might come from repeated interaction, allowing us to focus on pure framing effects. In contrast, in a repeated game, there is always the possibility that subjects think that other subjects' future behaviour may be influenced by their own current behaviour. This could confound the investigation of framing effects in two ways. First, sufficiently strong repeated game effects could swamp framing effects that would otherwise be present. Alternatively, if subjects' views of the dependence of other subjects' behaviour on their own are frame-sensitive, repetition could create a framing effect that would not otherwise be present.

In total, 42 subjects took part in the Give-N treatment; 45 in the Take-N treatment; 42 in the Give-P treatment; and 39 in the Take-P treatment. All subjects were recruited at the University of Nottingham. The vast majority were undergraduate students from different academic fields including, but not confined to, economics. The experiment was conducted in the Centre for Decision Research and Experimental Economics (CeDEx) lab. All treatments were computerised and programmed with the software z-Tree (Fischbacher, 2007). The no-punishment treatments lasted about 50 minutes and the punishment treatments about 70 minutes. At the end of each session, guilders were converted to UK pounds at the pre-announced exchange rate of £0.40 per guilder and subjects were paid in cash. On average, subjects earned about £9 in the no-punishment and £12.30 in the punishment treatments. Before subjects played the game, they received the instructions reproduced in Appendix A.2. As we wanted to ensure that subjects understood the decision situation and the mechanics of payoff calculations, all participants answered several computerised test questions, concerning what the payoffs would be for various hypothetical configurations of behaviour. The experiment did not proceed until every subject had answered these questions correctly.

3.3.3 Hypotheses

We now state our hypotheses concerning the effect of framing on contributions, and, more importantly, on punishment behaviour, and reported emotions. To make them comparable between the Give and Take frame, we express everything in terms of the amount of tokens left in the public good. That is, the notion of "contribution" used in the remainder of this chapter matches the concept of "effective contribution" from Chapter 2.

For contributions, the null hypothesis is that framing has no effect on their level, at least in no-punishment treatments. An alternative hypothesis for these treatments, suggested by previous evidence from repeated public goods games and the findings on moral judgments described in Chapter 2, is that contributions will be higher in the Give frame than in the Take frame.

Previous evidence suggests that punishment opportunities will tend to raise contributions (Fehr and Gächter, 2000; 2002), but whether this should be expected to induce a difference in contributions between Give and Take frames depends on the first of the two questions to which we now turn.

These questions which are our main concern are: whether the propensity to punish differs between frames; and whether emotional responses to contributions do likewise. We investigate them using the "punishment function" and the "emotions' function", respectively, as tools.

The "punishment function" gives the punishment points assigned by the punisher, as a function of the recipient's deviation from the punisher's contribution. Consistent with previous evidence (e.g., Fehr and Gächter, 2000), our expectation is that the punishment function will be downward sloping for the negative part of the deviation (horizontal) axis, implying that a subject punishes his co-player more, the more the co-player negatively deviates from the punisher's contribution.

Having defined the punishment function, we can now state our derived hypotheses in relation to framing. The null hypothesis predicts that the punishment function does not depend on framing, ceteris paribus. This hypothesis is implied by any consequentialist theory of behaviour, even by those which allow subjects to contribute and punish to some extent. For instance, although the inequity aversion model of Fehr and Schmidt (1999) or the ERC model of Bolton and Ockenfels (2000) can explain contribution and punishment, these theories cannot account for any framing effect.22

[22] Consequentialist theories suppose that subjects' actions are determined only by the final consequences of the actions. Almost all economic theories are consequentialist in this

Even if framing leads to different contribution levels, punishment may not be different for a given deviation by the recipient from the punisher's contribution. This prediction parallels the finding of Fehr and Gächter (2000) that, even if subjects contribute more under a fixed matching protocol than under a random matching one, these treatment manipulations do not affect the level of punishment for given deviations from the average group contribution. An effect of different treatments on contribution levels need not imply any corresponding effect on punishment, for a given deviation from the punisher's contribution.

In principle, the punishment function could be tilted either upwards or downwards by the Take frame relative to the Give frame.23 However, the evidence from Chapter 2, in which subjects morally rate free riding, suggests that failing to contribute is perceived as morally worse than withdrawing from the public good. Motivated by these findings, the alternative hypothesis states that the "punishment function" would be steeper in the Give, compared to the Take frame. Similar hypotheses can be derived using j's deviation from the punisher's beliefs about j's contributions in place of j's deviation from i's contribution.

The second tool we use to analyse subjects' perception of free riding is the "emotions' function", which gives aggregate emotions in a given category (i.e. positive or negative) as a function of deviations from

---

sense. However, psychological game theory (Geanakoplos, et al., 1989) includes beliefs in the payoff function and can therefore explain framing effects (see Dufwenberg, et al., 2006).

[23] The reason for expecting tilts is that we assume there to be no punishment if the deviation from the punisher's contribution is zero.

individual's own contribution.24 We expect the "emotions' function" to be negatively sloped in the negative deviation interval for the negative emotions, and positively sloped in the negative deviation interval for the positive emotions. Although our main concern is with the negative deviation interval, since we are mostly interested in how subjects treat lower contributions by their counterparts, we can also state hypotheses for the emotions' function for the positive deviation interval. We hypothesise that, within the range of this interval, for the negative emotions the function will be negatively sloped; whereas, for the positive emotions, it will be positively slope. As with the "punishment function", we can also state hypotheses for the slope and level of the "emotions' function". Since punishment and (negative) emotions are likely to be closely related, we expect that any differences observed in the "punishment function" will be reflected in the "emotions' function" as well (and in the same direction).

3.4 Results

3.4.1 Contributions

Table 3.1 shows the average absolute level of contribution across all subjects in each treatment. In absolute levels, the average contribution is largest in the Give-P treatment (i.e. 7.21 tokens) and smallest in the Take-N treatment (i.e. 4.47 tokens), as previous literature on framed public goods experiments would suggest.

[24] Positive emotions comprise warmth, happiness and joy; whereas, negative emotions comprise anger, fear, envy, sadness, shame, irritation, contempt, guilt and jealousy. We consider surprised as a "neutral" emotion as it can either be positive or/and negative.

Table 3.1 Average absolute levels of contribution

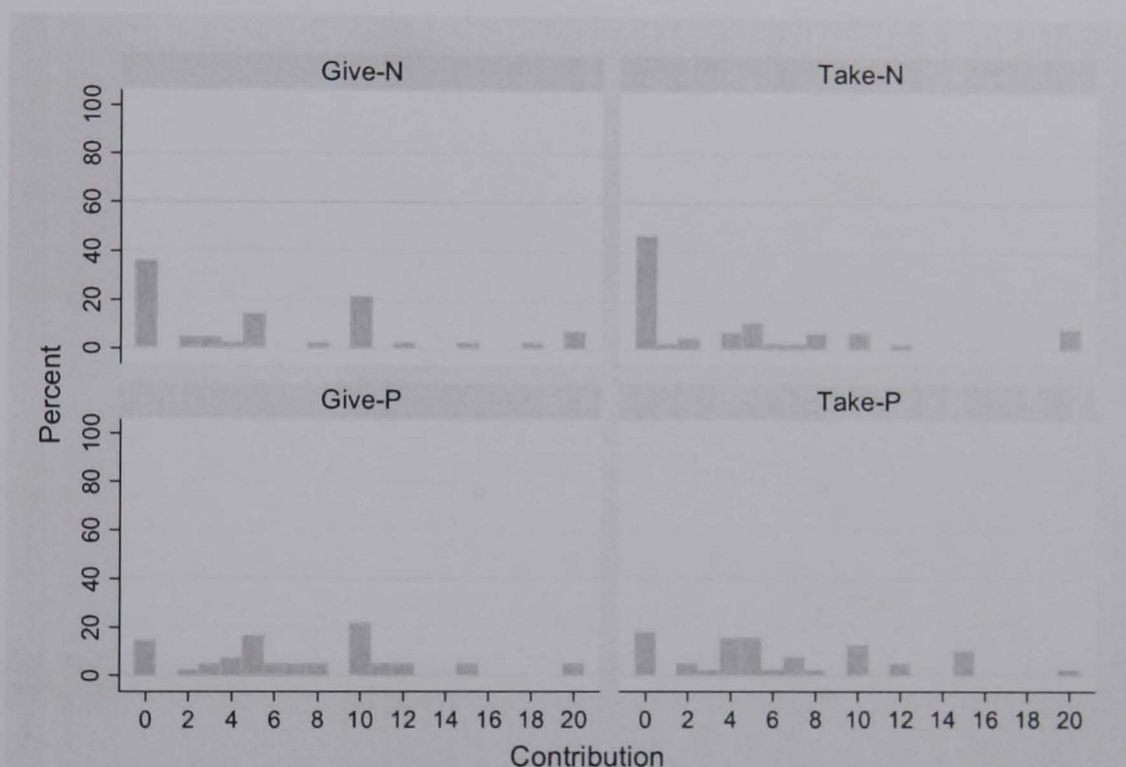
	Average contribution levels
Give-N	5.88
	(6.24)
Take-N	4.47
	(6.04)
Give-P	7.21
	(5.03)
Take-P	6.41
	(5.12)

Note: Standard deviations in parentheses.

A non-parametric Kruskal-Wallis test suggests significant differences among treatments ($\chi^2(3) = 10.089$, $p = 0.0178$). Using a Wilcoxon rank-sum test shows that the only significant differences at the 5% level are between the Give-P and Take-N treatments ($p = 0.002$) and between the Take-N and Take-P treatments ($p = 0.025$).

The distribution of contributions across all frames is shown in Figure 3.1, which suggests three interesting observations. First, the introduction of punishment reduces the zero contributions both in the Give and the Take treatments. Second, with respect to the Give frame, there exists a spike for contributions equal to ten, while this tendency does not seem to exist for the Take frame, implying that there may be a salience for giving half but not for taking half. Third, in all treatments, those contribution levels not located at either zero or ten are fairly widely spread. Performing a Kolmogorov-Smirnov test for the equality of distributions, we find significant differences only for the Give-P and Take-N treatments (p = 0.006), and the Take-N and Take-P treatments (p = 0.037)).

Figure 3.1 Distribution of contributions across frames



We now examine subjects' beliefs about contributions across frames. Table 3.2 shows the average beliefs about contributions across all four treatments. A Kruskal-Wallis test suggests weakly significant differences between treatments ($\chi^2(3) = 6.626$, p = 0.085). Using a Wilcoxon rank-sum test shows that the only significant differences at the 5% level are between the Give-P and Take-N treatments (p = 0.032) and between the Give-P and Take-P treatments (p = 0.023).

Table 3.2 Average absolute levels of beliefs about contributions

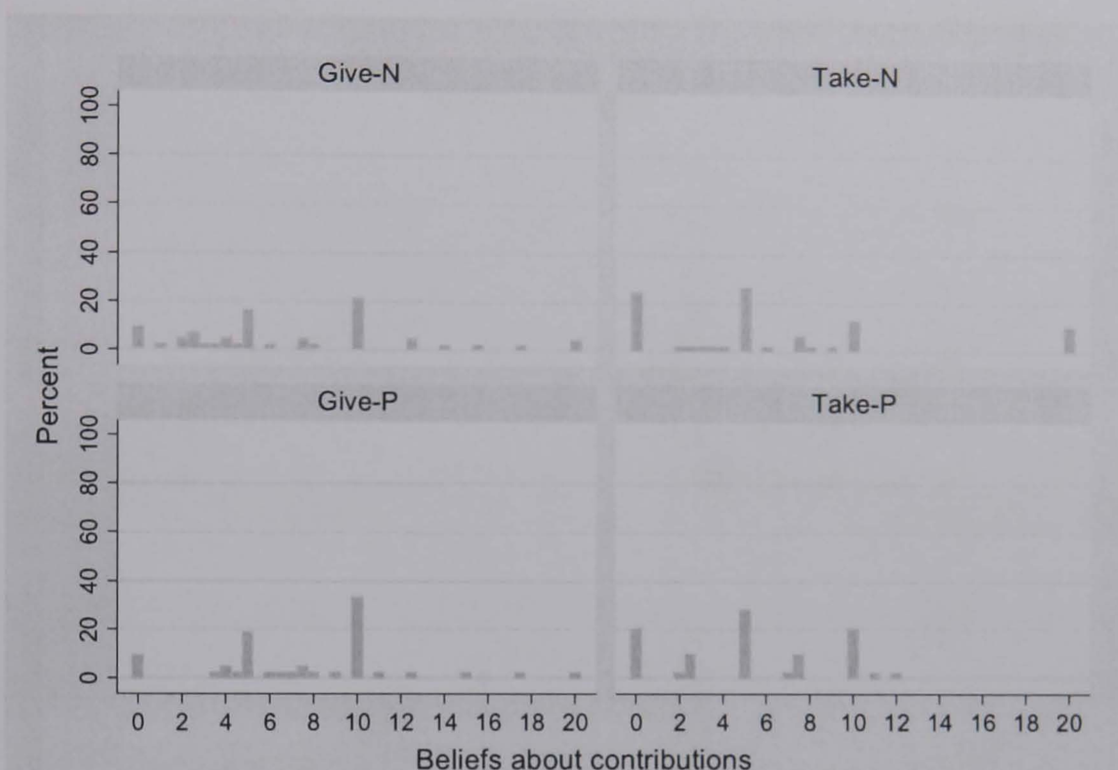
	Average beliefs about contributions
Give-N	7.08 (5.26)
Take-N	6.23 (5.93)
Give-P	7.70 (4.32)
Take-P	5.31 (3.78)

Note: Standard deviations in parentheses.

From a visual examination of Figure 3.2, we observe that in the Take treatments there is a spike of beliefs about zero contributions, which is less pronounced in the Give treatments. In addition, the salience observed in actual contributions in relation to giving half but not taking half seems to be reflected in subjects' beliefs about contributions.

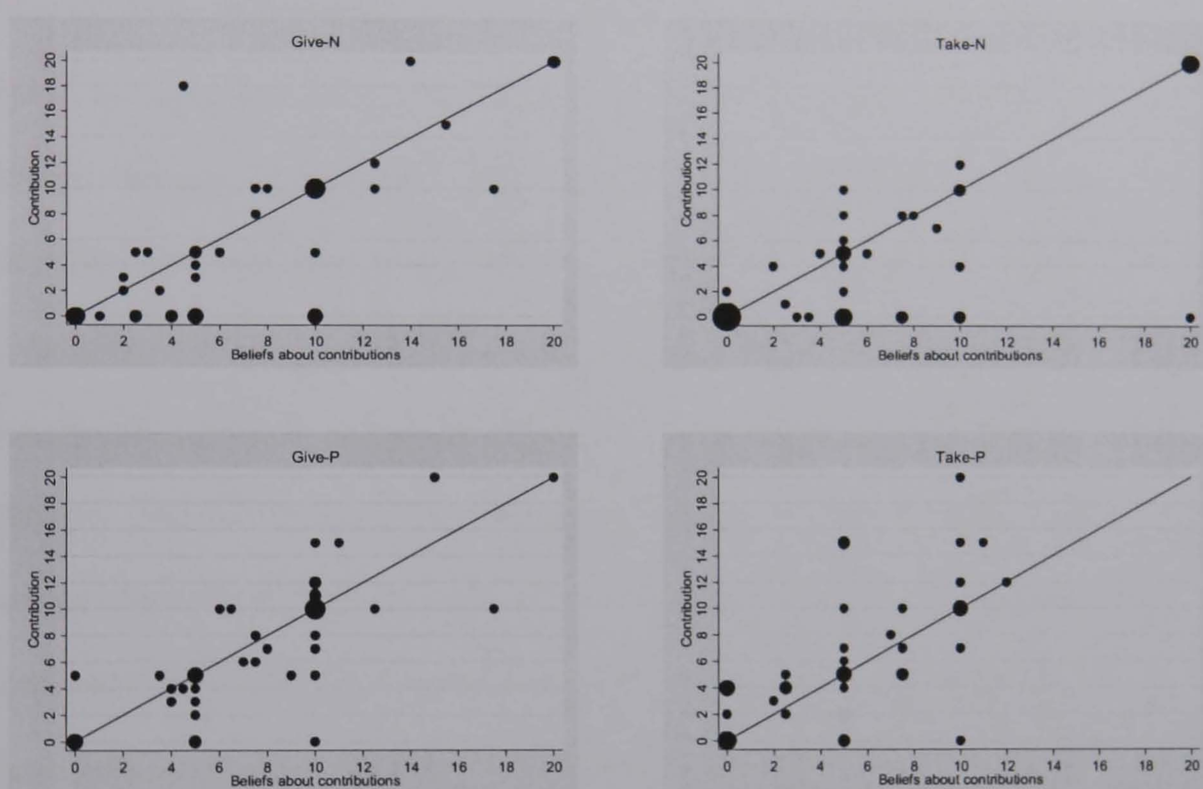
Testing for differences in the distribution of beliefs about contributions across frames, a Kolmogorov-Smirnov test suggests no significant differences for any pair-wise treatment comparison. A visual representation of how beliefs about contributions are distributed for each treatment separately is provided in Figure 3.2.

Figure 3.2 Distribution of beliefs about contributions across frames



Thus far, we observe that framing affects contributions and beliefs about contributions in a similar way. Next we investigate how contributions and beliefs are related as, in our analysis of punishment and emotions, we are interested in deviations from contributions and beliefs about contributions. Note also that we derive similar hypotheses for both definitions of deviations and it is thus of interest to examine the interdependence between contributions and beliefs. Figure 3.3 provides a graphical illustration of contributions as a function of beliefs about contributions for each frame separately. Circles represent combinations of contributions and beliefs per treatment, with the size of the circles being proportional to the number of observations. Circles on the horizontal axis correspond to zero contributions, whereas, circles on the diagonal indicate contributions that exactly match beliefs.

Figure 3.3 Contributions as a function of beliefs about contributions per treatment



As Figure 3.3 shows, there is a positive relationship between contributions and beliefs for all treatments. The Spearman correlation coefficients are as follows: Give-N: $\rho = 0.693$; Take-N: $\rho = 0.593$; Give-P: $\rho = 0.796$; and Take-P: $\rho = 0.708$. For each treatment, the corresponding p-value is equal to zero, implying that contributions and beliefs about contributions are positively and significantly correlated. This observation suggests that at least in our one-shot context subjects' contribution decisions are closely matched with their expectations about how much the other group members contribute, and we thus expect that the findings reported in the following sections to be similar across the two definitions of deviations.

However, in this chapter, our main concern is with punishment behaviour and emotions, as these are our vehicles to answer the question of whether, and if so, how framing influences the way subjects treat a given deviation from contributions or from beliefs about contributions. Before turning our attention to the investigation of both indicators, we look at whether standard deviations about contributions and beliefs about contributions of each group were different between treatments (see Tables B.1-B.4, Appendix B). Our analysis suggests that performing a Wilcoxon rank-sum test yields no significant differences for any pair-wise comparison. However, despite the similarity of standard deviations across frames, our findings reported in Chapter 2 provide us with motivation to expect that punishment behaviour and emotions might be different, to the extent that moral judgments were different, controlling for the level of standard deviations. In the next two sections, we explore whether subjects treat free-riding differently by punishing differently or/and displaying different feelings across framing manipulations, using the "punishment function" and the "emotions' function" as our tools.

3.4.2 The punishment function

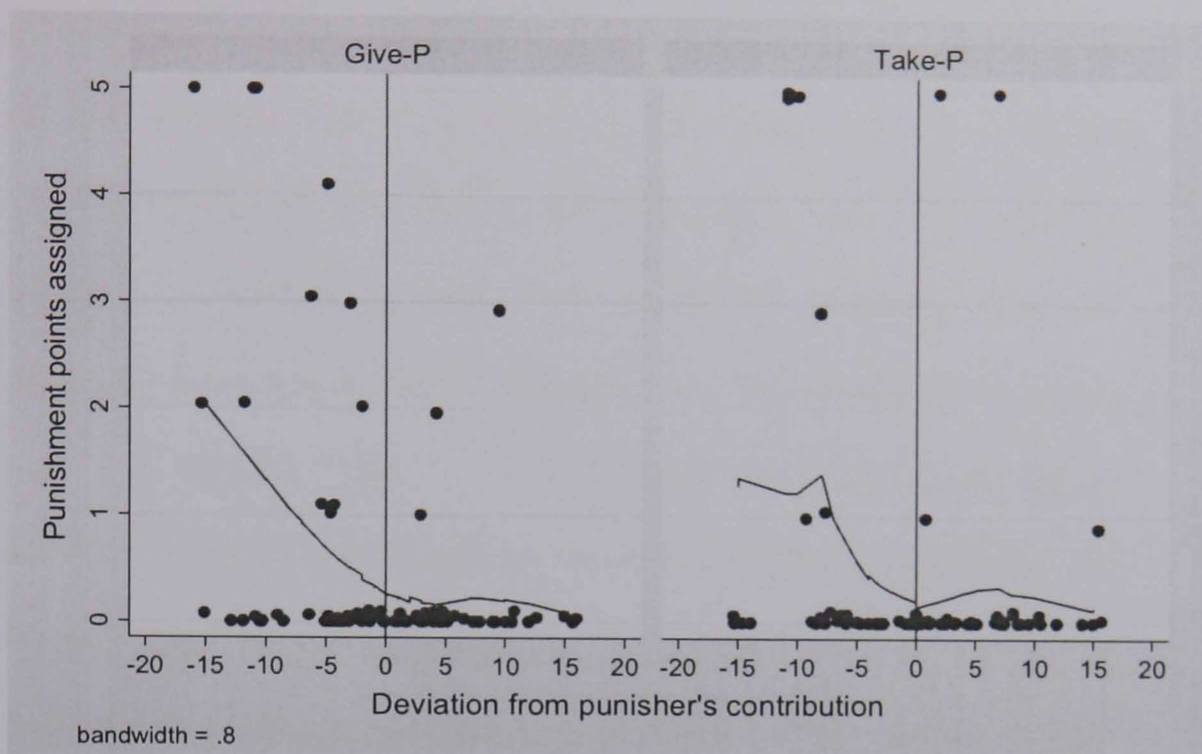
We start our analysis by examining subjects' punishment behaviour for each framing context. Figure 3.4 below shows the punishment points assigned by the punisher to another player j as a function of j's deviation from the punisher's contribution.[25] The punishment function is given by the solid line, which indicates the fitted line of the locally weighted regression of punishment assigned on the deviation from the punisher's contribution.

RESULT 1: The punishment function does not depend on framing, ceteris paribus.

Support. In Figure 3.4, each dot represents a single observation. The punishment functions appear quite similar across frames, having the anticipated negative slope.

[25] We refer to the punisher as player i, the recipient of punishment as player j, and the third group member as player k.

Figure 3.4 Punishment as a function of deviation from punisher's contribution



However, the graphs in Figure 3.4 are simply visual representations of the punishment function and do not control for any factors other than j's deviation from i's contribution. To test econometrically whether the slope of the punishment function differs across frames, we estimated a Tobit regression model. In this regression, the dependent variable is the "punishment assigned by player i to player j" and the independent variables comprise "Player k's contribution deviation", "Player j's absolute negative (contribution) deviation", "Player j's positive (contribution) deviation", and the dummy variable "Take". We also included two interaction terms, which indicate whether the slope of the punishment function differs with respect to negative and positive deviations across frames. Note that all deviations are calculated with respect to the punisher's contribution. We include "absolute negative deviation" and "positive deviation" as separate regressors, since Figure 3.4 suggests that these two different sorts of deviation elicit different punishment responses. The variable "absolute negative deviation" is the absolute value of the actual deviation of subject j 's contribution from the punisher's contribution, when subject j 's contribution is below the punisher's contribution; and zero otherwise. The variable "positive deviation" is constructed in an analogous way. The variable "Player k 's contribution deviation" is the actual deviation of player k 's contribution from the punisher's contribution. The reason for including such a variable is that player i 's attitude to player j may differ according to the behaviour of player k. The dummy variable "Take" equals 0 for the Give frame, and 1 for the Take frame. The regression results are given in Table 3.3.

Table 3.3 The punishment function – Regression results

	Dependent variable: Punishment assigned by player i to player j
Player j's absolute negative deviation from punisher's contribution	0.790** (0.331)
Player j's positive deviation from punisher's contribution	-0.102 (0.337)
Player k's contribution deviation from punisher's contribution	0.083 (0.096)
Take	-1.699 (2.739)
Take × Player j's absolute negative deviation from punisher's contribution	-0.007 (0.497)
Take × Player j's positive deviation from punisher's contribution	0.378 (0.396)
Constant	-9.025*** (3.425)
Observations	162

*Notes: Tobit estimates. Standard errors are presented in parentheses (clustered on groups). Results are corrected for heteroskedasticity. ** denotes significance at the 5-percent level, and *** at the 1-percent level.*

The results from Table 3.3 indicate that the vertical intercept of the punishment function does not differ across treatments, since the coefficient of the dummy variable "Take" is not statistically significant. The coefficients of the interaction terms are also insignificant, implying that the slope of the punishment function with regard both to negative and positive deviations is the same, irrespective of framing. These findings fail to reject the null hypothesis that subjects punish given (negative and positive) deviations from their own contribution equally in different frames. According to our findings, the only variable that has a significant effect on the assignment of punishment is the absolute negative deviation from the punisher's contributions: a subject punishes a co-player more, the less the co-player contributes relative to the punisher.

We next examine whether punishment as a function of beliefs about contributions, varied across frames. As we have already seen in Section 3.4.1 contributions closely match beliefs about contributions, being positively and highly correlated, and as a result, our expectation is that it is very likely to observe similar patterns of punishment behaviour as a function of beliefs about contributions compared to the case of deviations from punisher's contributions. It is therefore of interest to look at whether our conjecture actually holds by investigating the punishment function with respect to deviations from punisher's beliefs about contributions. Our findings again yield no evidence that the "punishment function" differs between frames. Result 2 records our finding.

RESULT 2: Punishment as a function of deviation from punisher's beliefs about contributions does not depend on framing, ceteris paribus.

Support. Figure 3.5 provides a graphical illustration of the punishment assigned as a function of deviations from the punisher's beliefs about contributions. It turns out that the punishment function with respect to these deviations is very similar in both framing contexts.

Figure 3.5 Punishment as a function of deviation from punisher's beliefs about contributions

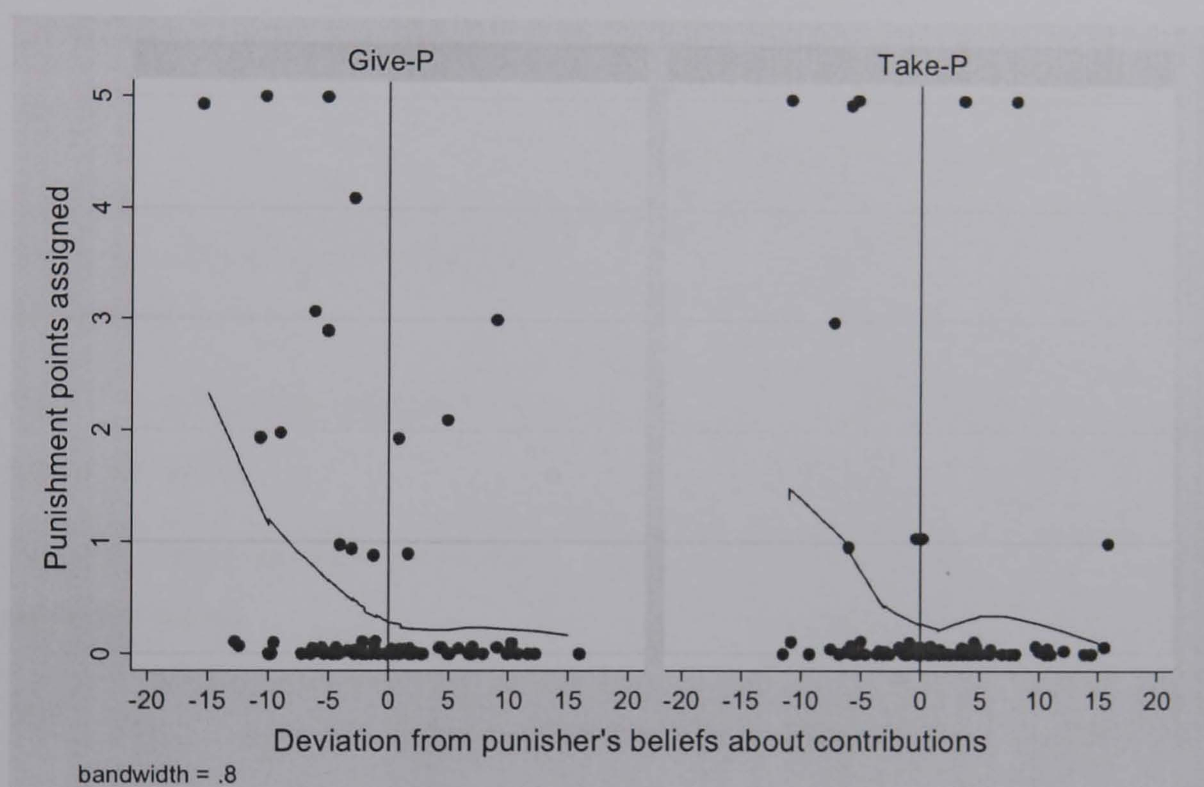


Table 3.4 provides econometric support for our Result 2. We have estimated a Tobit model, including the same independent variables as in Table 3.3, with the only difference that we now express deviations with respect to the punisher's beliefs about contributions. We find evidence that punishment again does not differ as a function of these deviations under both framing contexts, since the coefficients of the dummy variable 'Take'

and the interaction terms are statistically insignificant. Analogously with the case of deviations from the punisher's actual contribution, the only significant variable that explains punishment is player j's absolute negative deviation from punisher's beliefs about contributions.

Table 3.4 Punishment as a function of punisher's beliefs about contributions – Regression results

	Dependent variable: Punishment assigned by player i to player j
Player j's absolute negative deviation from punisher's beliefs	0.762** (0.371)
Player j's positive deviation from punisher's beliefs	-0.093 (0.261)
Player k's contribution deviation from punisher's beliefs	0.117 (0.101)
Take	-1.445 (2.945)
Take × absolute negative deviation from punisher's beliefs	0.189 (0.659)
Take × positive deviation from punisher's beliefs	0.251 (0.357)
Constant	-8.866** (3.483)
Observations	162

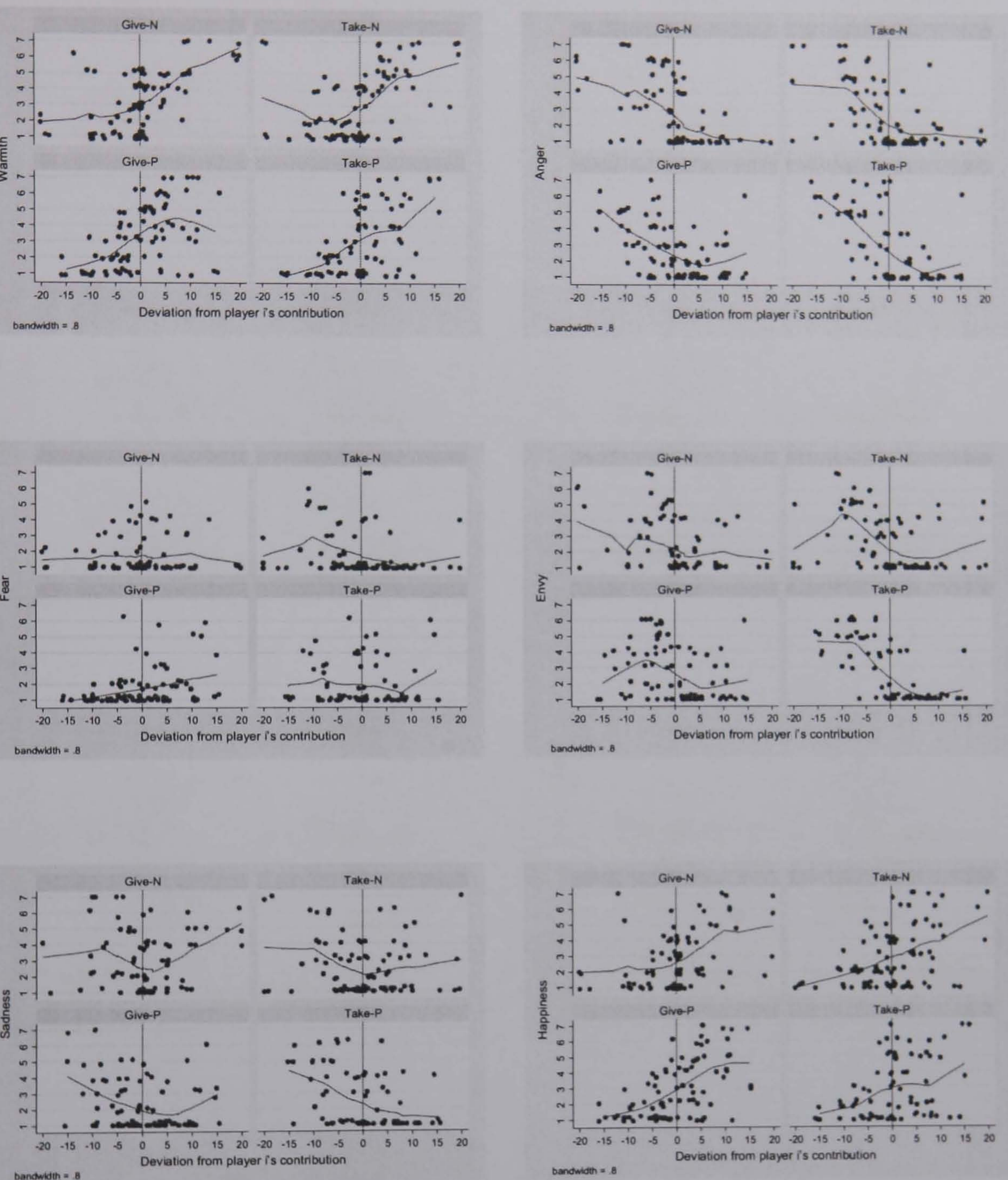
*Notes: Tobit estimates. Standard errors are presented in parentheses (clustered on groups). Results are corrected for heteroskedasticity. ** at the 5-percent level, and *** at the 1-percent level.*

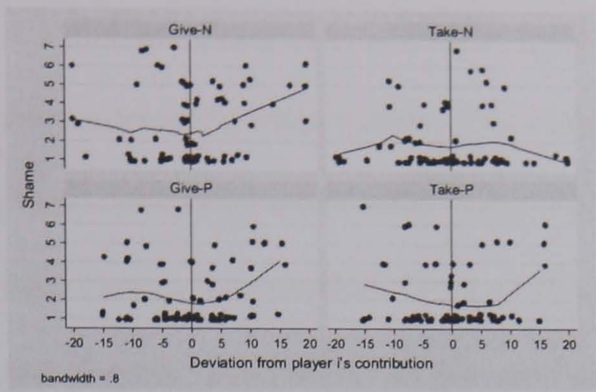
Thus far, we have found that subjects do not treat deviations from either own contribution or beliefs about others' contributions differently across frames, in terms of the punishment they mete out. However, this observation could be attributed either to subjects perceiving deviations differently across frames, but not being prepared to act differently on the basis of this perception; or to subjects not perceiving deviations differently across frames at all. In the next section, we use the elicited emotions to disentangle these two possibilities.

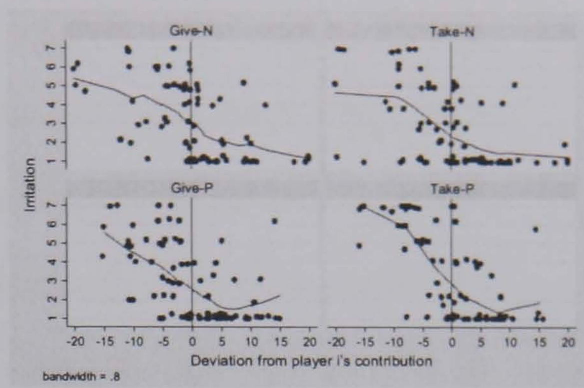
3.4.3 The emotions' function

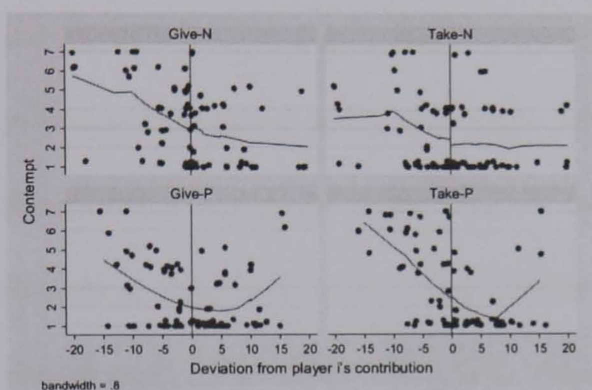
In this section, we explore whether the self-reported emotions differ across frames. Figure 3.6 displays each emotion separately for each treatment. The horizontal axis indicates the deviation of an individual's contribution from player i's contribution, while the vertical axis indicates for each emotion the intensity (ranging from 1 = "not at all" to 7 = "very much") with which a group member felt each emotion when they saw the contribution of the other group member.

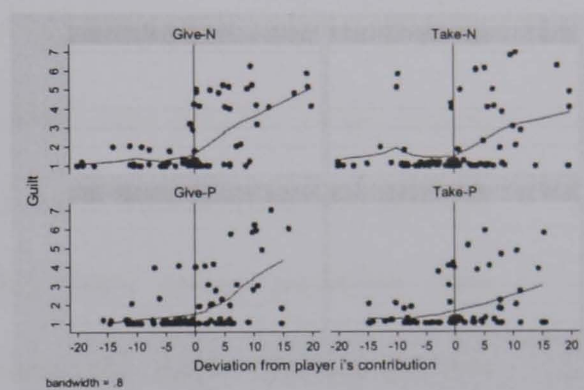
Figure 3.6 Emotions as a function of deviations from player i's contributions

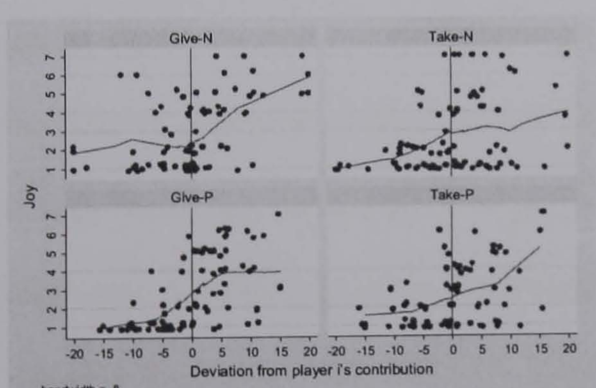


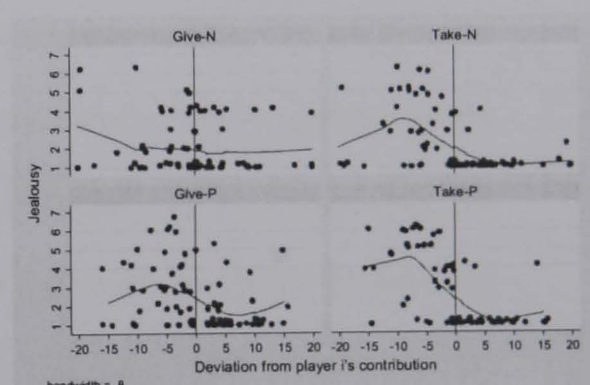

Give-N
Take-N
Give-P
Take-P
Shame
Deviation from player i's contribution
bandwidth = .8

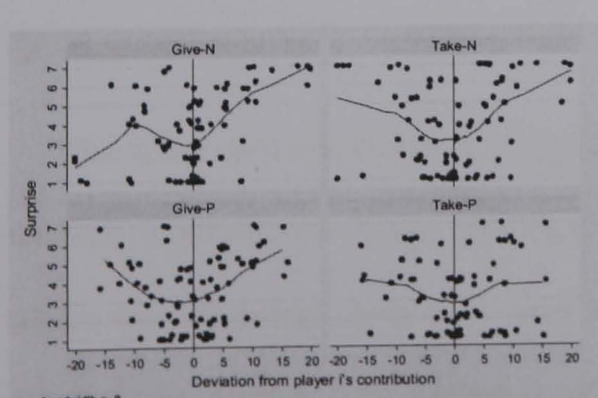

Give-N
Take-N
Give-P
Take-P
Irritation
Deviation from player i's contribution
bandwidth = .8


Give-N
Take-N
Give-P
Take-P
Contempt
Deviation from player i's contribution
bandwidth = .8


Give-N
Take-N
Give-P
Take-P
Guilt
Deviation from player i's contribution
bandwidth = .8


Give-N
Take-N
Give-P
Take-P
Joy
Deviation from player i's contribution
bandwidth = .8


Give-N
Take-N
Give-P
Take-P
Jealousy
Deviation from player i's contribution
bandwidth = .8


Give-N
Take-N
Give-P
Take-P
Surprise
Deviation from player i's contribution
bandwidth = .8

From the examination of each panel separately, it is worth mentioning three interesting observations. First, subjects are feeling sadder about their own contribution when other group members positively deviate from their contribution (see positive slope of the function for sadness in the Give-N and Give-P treatments in the positive deviation interval). This is an indication that perhaps if they have been given a second chance to make a contribution, they would have contributed more. Second, there are some subjects displaying more contempt for others (see positive slope of the function for contempt in the Give-P and Take-P treatments in the positive deviation interval) when they realize that other group members have contributed more than they did, suggesting that these subjects did not appreciate their counterparts' good action of contributing even in a one-shot game. Third, the slope of the function with respect to shame is almost flat in the Take-N treatment, indicating that subjects did not feel any pressure even if they were low contributors under this treatment. In all other cases, the emotions' functions have the anticipated slope with respect to positive and negative deviations. However, we have to bear in mind that in some cases this tendency is driven by a few observations and need to be careful when drawing conclusions.

Our main concern is with framing and for a formal statistical analysis of whether emotions are frame sensitive, we estimate ordered probit models where the dependent variable is each emotion separately. In these models, all data are pooled. The inclusion of the independent variables follows similar reasoning as for the case of the punishment function. We also include a dummy variable called "No-Punishment", which takes on the value '1' for the no-punishment treatments and '0' for the punishment treatments, and its interaction terms with respect to absolute negative deviation and positive deviation to test for any difference in the elicited emotions between punishment and no-punishment treatments, for the reasons explained in Section 3.3.2. The results are presented in Table 3.5.

Table 3.5 Emotions as a function of deviations from player i's contributions – Regression results

	Dependent Variables												
	Warmth	Anger	Fear	Envy	Sadness	Happiness	Shame	Irritation	Contempt	Guilt	Joy	Jealousy	Surprise
Player j's absolute negative deviation from player i's contribution	-0.177*** (0.038)	0.109*** (0.023)	-0.038 (0.038)	0.072** (0.032)	0.093** (0.039)	-0.090** (0.035)	0.069** (0.031)	0.166*** (0.033)	0.139*** (0.035)	-0.040 (0.036)	-0.100*** (0.038)	0.085** (0.034)	0.040 (0.029)
Player j's positive deviation from player i's contribution	0.035 (0.034)	-0.060*** (0.037)	0.026 (0.032)	-0.086* (0.049)	0.012** (0.039)	0.072** (0.034)	0.102*** (0.029)	-0.064 (0.041)	0.031 (0.028)	0.121*** (0.035)	0.105*** (0.034)	-0.060 (0.043)	0.102*** (0.028)
Player k's contribution deviation from player i's contribution	-0.009 (0.009)	0.004 (0.010)	0.029*** (0.011)	0.020* (0.011)	-0.011 (0.010)	0.004 (0.008)	-0.004 (0.011)	0.015* (0.009)	0.005 (0.009)	0.007 (0.012)	0.001 (0.009)	0.029*** (0.011)	0.002 (0.009)
Take	-0.252 (0.208)	-0.249 (0.224)	-0.060 (0.243)	-0.284 (0.244)	-0.082 (0.237)	0.125 (0.193)	-0.149 (0.206)	-0.143 (0.187)	-0.098 (0.193)	-0.113 (0.280)	0.125 (0.205)	-0.140 (0.263)	-0.219 (0.182)
Take × Player j's absolute negative deviation from player i's contribution	0.030 (0.046)	0.051 (0.041)	0.058 (0.040)	0.065 (0.047)	0.012 (0.044)	-0.024 (0.036)	-0.022 (0.035)	0.049 (0.042)	0.012 (0.038)	0.032 (0.042)	-0.018 (0.040)	0.063 (0.047)	0.053 (0.038)
Take × Player j's positive deviation from player i's contribution	0.032 (0.030)	0.032 (0.040)	-0.023 (0.037)	0.007 (0.039)	-0.042 (0.031)	-0.033 (0.033)	-0.059* (0.032)	0.001 (0.042)	-0.010 (0.031)	-0.023 (0.034)	-0.052* (0.027)	-0.056 (0.048)	-0.018 (0.033)
No-Punishment	-0.609*** (0.201)	-0.022 (0.210)	-0.228 (0.250)	-0.235 (0.238)	0.303 (0.245)	-0.240 (0.199)	0.440** (0.213)	0.222 (0.186)	0.513** (0.201)	0.028 (0.294)	-0.081 (0.211)	-0.364 (0.254)	-0.003 (0.185)
No-Punishment × Player j's absolute negative deviation from player i's contribution	0.139*** (0.040)	-0.017 (0.035)	0.068 (0.042)	-0.005 (0.042)	-0.040 (0.043)	0.050 (0.038)	-0.045 (0.038)	-0.088** (0.039)	-0.075* (0.039)	0.013 (0.043)	0.064 (0.043)	-0.025 (0.044)	-0.003 (0.036)
No-Punishment × Player j's positive deviation from player i's contribution	0.086 (0.031)	-0.035 (0.037)	-0.046 (0.036)	0.061 (0.045)	0.059* (0.036)	0.031 (0.036)	-0.048 (0.030)	-0.027 (0.042)	-0.046 (0.032)	0.005 (0.038)	-0.003 (0.030)	-0.041** (0.047)	0.063 (0.035)
Observations	336	336	336	336	336	336	336	336	336	336	336	336	336

*Notes: Ordered probit estimates. Standard errors are presented in parentheses (clustered on groups). Results are corrected for heteroskedasticity. ** denotes significance at the 5-percent level, and *** at the 1-percent level. All data are pooled.*

Table 3.5 indicates that the variable "Take" and its associated interaction terms are not statistically significant for any of the elicited emotions.[26] The only independent variables which typically have a significant impact on emotions (depending on whether they are positive or negative) are the absolute negative and the positive deviation from player i's contribution.

We therefore proceed by categorising emotions into positive and negative ones, and using as tools the mean positive and mean negative emotions' functions, which are aggregate measures of positive and negative emotions, respectively, expressed as functions of other variables. Put simply, we plot mean positive and mean negative emotions as a function of deviations from player i's contribution. Positive emotions comprise warmth, happiness and joy; whereas, negative emotions comprise anger, fear, envy, sadness, shame, irritation, contempt, guilt and jealousy. To control whether the presence of punishment opportunities has influenced emotions, we also examine emotions for the no-punishment treatments. Our findings are recorded in Result 3.

<u>RESULT 3.</u> *(a) The (mean positive and mean negative) emotions' function does not depend on framing, ceteris paribus. (b) The availability of punishment affects the elicited mean positive emotions, ceteris paribus;*

[26] The only exception is shame and joy which appear to be marginally significant with respect to the slope of the function in the positive deviation's interval.

whereas, the elicited mean negative emotions are not affected by the presence of punishment, ceteris paribus.

Support. Figures 3.7 and 3.8 provide a graphical illustration of the positive and negative emotions' function for each of the four treatments, respectively. In all figures, each dot represents a single observation and the solid line indicates the fitted line of the locally weighted regression of emotions expressed on the deviation from the punisher's contribution.

Figure 3.7 Mean positive emotions for each treatment

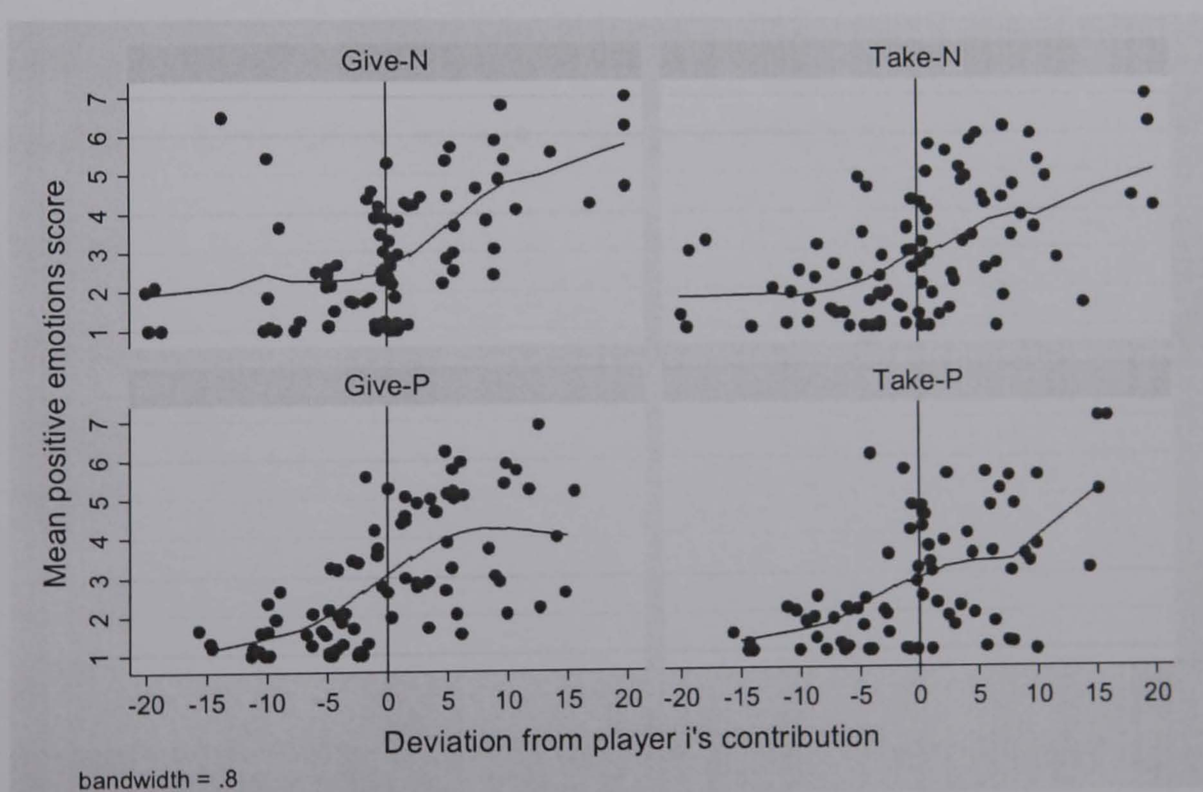


Figure 3.8 Mean negative emotions for each treatment

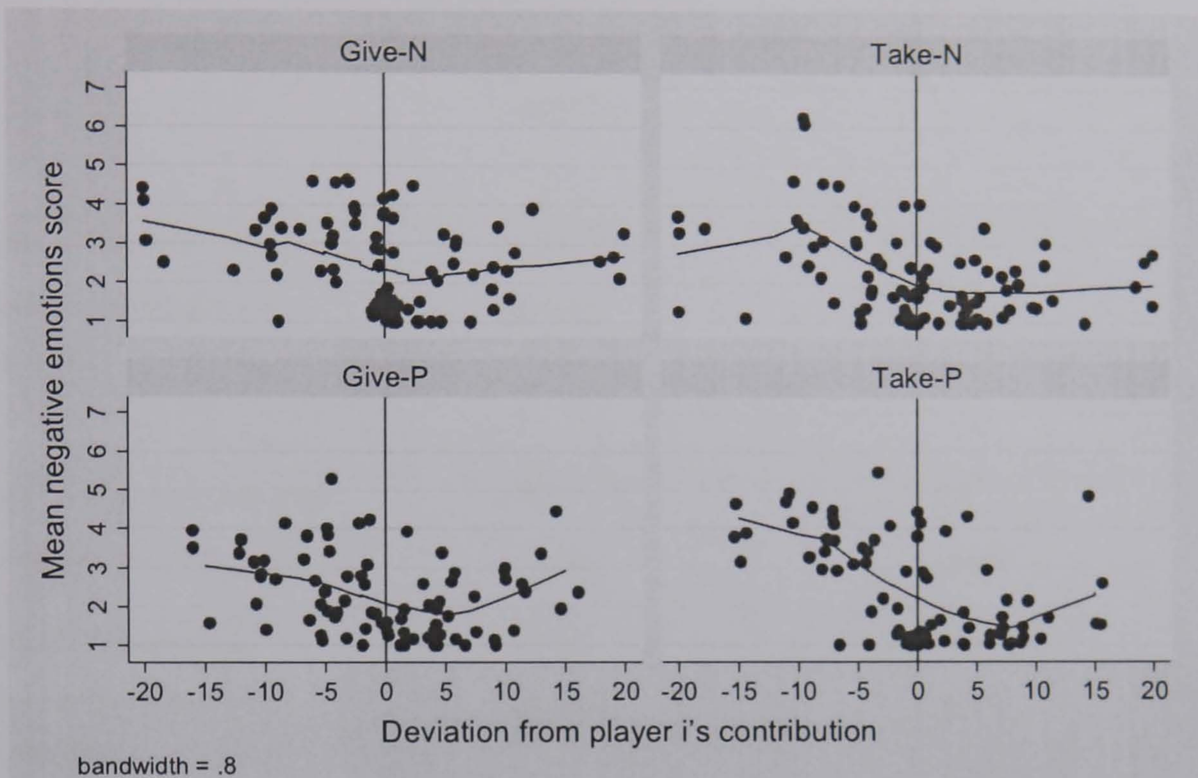


Table 3.6 provides econometric evidence for Result 3. The equations reported in this Table include the same covariates as in Table 3.5.

Table 3.6 The positive and negative emotions' function – Regression results

	Dependent variable: Mean positive emotions	Dependent variable: Mean negative emotions
Player j 's absolute negative deviation from player i 's contribution	-0.139*** (0.034)	0.131*** (0.030)
Player j 's positive deviation from player i 's contribution	0.087**** (0.030)	0.025 (0.027)
Player k 's contribution deviation from player i 's contribution	-0.001 (0.009)	0.011 (0.010)
Take	-0.023 (0.194)	-0.155 (0.217)
Take × Player j 's absolute negative deviation from player i 's contribution	0.003 (0.038)	0.035 (0.041)
Take × Player j 's positive deviation from player i 's contribution	-0.023 (0.028)	-0.022 (0.026)
No-Punishment	-0.354* (0.199)	0.182 (0.1216)
No-Punishment× Player j 's absolute negative deviation from player i 's contribution	0.091** (0.037)	-0.057 (0.038)
No-Punishment× Player j 's positive deviation from player i 's contribution	0.045 (0.029)	-0.005 (0.026)
Observations	336	336

*Notes: Ordered probit estimates. Standard errors are presented in parentheses (clustered on groups). Results are corrected for heteroskedasticity. ** denotes significance at the 5-percent level, and *** at the 1-percent level. All data are pooled.*

Examining the coefficients related to framing, it is clear that they are not statistically different across frames, implying that the aggregate "emotions' function" is not frame-sensitive, for either positive or negative emotions. This strengthens our conclusion that subjects do not consider negative deviations from their own contribution differently across frames.

Looking at the negative emotions, we observe that the coefficient on the variable "No-Punishment" and its interaction terms are not statistically significant in the equation reported in Table 3.6. This finding provides no support for the idea that punishment opportunities expunge negative emotional responses. However, for the positive emotions, we find that there is a significant level difference at the 10% level and a significant slope difference for the negative deviation interval at the 5% level.

We investigate further whether elicited emotions vary with the presence or absence of a punishment option by estimating four ordered probit models: two relating to the no-punishment case and two to the punishment case (in each case, taking mean positive and mean negative emotions, separately).[27] The rationale and interpretation of the independent variables in these models is parallel to that described above. Our regression results indicate that the coefficients for the dummy variable 'Take' and its interaction terms are not statistically significant. Elicited emotions do not differ between the Give

[27] It would be potentially interesting to explore the correlation between punishment and emotions, investigating whether one's own emotions affect one's own punishment assigned. Since this is beyond the scope of our main research questions, we provide such an analysis in Appendix B.

and Take frames for either the no-punishment or the punishment treatment, taken separately. The results are presented in Table 3.7.

Table 3.7 The positive and negative emotions' function in the absence and presence of punishment opportunities – Regression results

	Dependent variable: Mean positive emotions		Dependent variable: Mean negative emotions	
	No-Punishment	Punishment	No-Punishment	Punishment
Player j's absolute negative deviation from player i's contribution	-0.039 (0.041)	-0.162*** (0.043)	0.092*** (0.029)	0.104*** (0.033)
Player j's positive deviation from player i's contribution	0.141*** (0.029)	0.066 (0.038)	0.017 (0.021)	0.018 (0.038)
Player k's contribution deviation from player i's contribution	-0.006 (0.013)	0.007 (0.009)	0.012 (0.014)	0.013 (0.014)
Take	0.144 (0.273)	-0.315 (0.284)	-0.242 (0.300)	-0.204 (0.327)
Take × Player j's absolute negative deviation from player i's contribution	-0.024 (0.049)	0.059 (0.057)	-0.000 (0.053)	0.097 (0.054)
Take × Player j's positive deviation from player i's contribution	-0.039 (0.034)	0.014 (0.045)	-0.018 (0.030)	-0.013 (0.045)
Observations	174	162	174	162

*Notes: Ordered probit estimates. Standard errors are presented in parentheses (clustered on groups). Results are corrected for heteroskedasticity. ** denotes significance at the 5-percent level, and *** at the 1-percent level. All data are pooled.*

Table 3.7 suggests that, for positive emotions, in the absence of punishment opportunities, the only significant variable is the positive deviation from player i's contribution when we examine positive emotions (the more player j positively deviates from player i, the more positive emotions player i is displaying); whereas, in the presence of punishment opportunities, we find that the only significant variable is the absolute negative deviation from player i's contribution (the more player j negatively deviates from player i, the less positive emotions player i is displaying). This can be interpreted that if player j deviates positively from i, then player i feels positive emotions towards player j, but *not* if player i is anticipating that player j might have punished him. If player j negatively deviates from player i, then player i is not inclined to feel positive emotions towards j, and when player i can punish, he feels positive (negative) emotions with less (more) intensity.

When we examine negative emotions, it turns out that the main determinant of both positive and negative emotions is the negative deviation from player i's contribution either in the presence or in the absence of punishment (note that, as expected, the sign of this variable is negative for positive emotions and positive for negative emotions). In sum, our results imply that for negative emotions the presence of punishment makes very little difference; whereas, for positive emotions, it makes a significant difference. When j has deviated positively from i, i seems positively disposed to j when no punishment opportunities are available. However, when j has deviated negatively from i, he tends not to be positively disposed to j and punishment strengthen this effect.

Just as with the punishment function, we examine whether deviations from beliefs about contributions evoke differences among treatments. Appendix B (see Figure B.1 and Table B.5) provides both graphical and econometric analysis for each emotion separately expressed as a function of deviations from player i's beliefs about contributions. From this analysis, we find that there is little difference across frames for each emotion and we thus restrict our attention to mean negative and mean positive emotions. Result 4 records our finding.

RESULT 4. *(a) Mean positive and mean negative emotions as a function of deviations from player i's beliefs about contributions do not depend on framing, ceteris paribus. (b) The availability of punishment affects the elicited mean positive emotions, ceteris paribus; whereas, the elicited mean negative emotions are affected less by the presence of punishment, ceteris paribus.*

Support. A graphical representation of the emotions' function in relation to deviations from player i's beliefs about contributions is provided in Figures 3.9 and 3.10 below.

Figure 3.9 Mean positive emotions as a function of deviations from player i's beliefs about contributions for each treatment

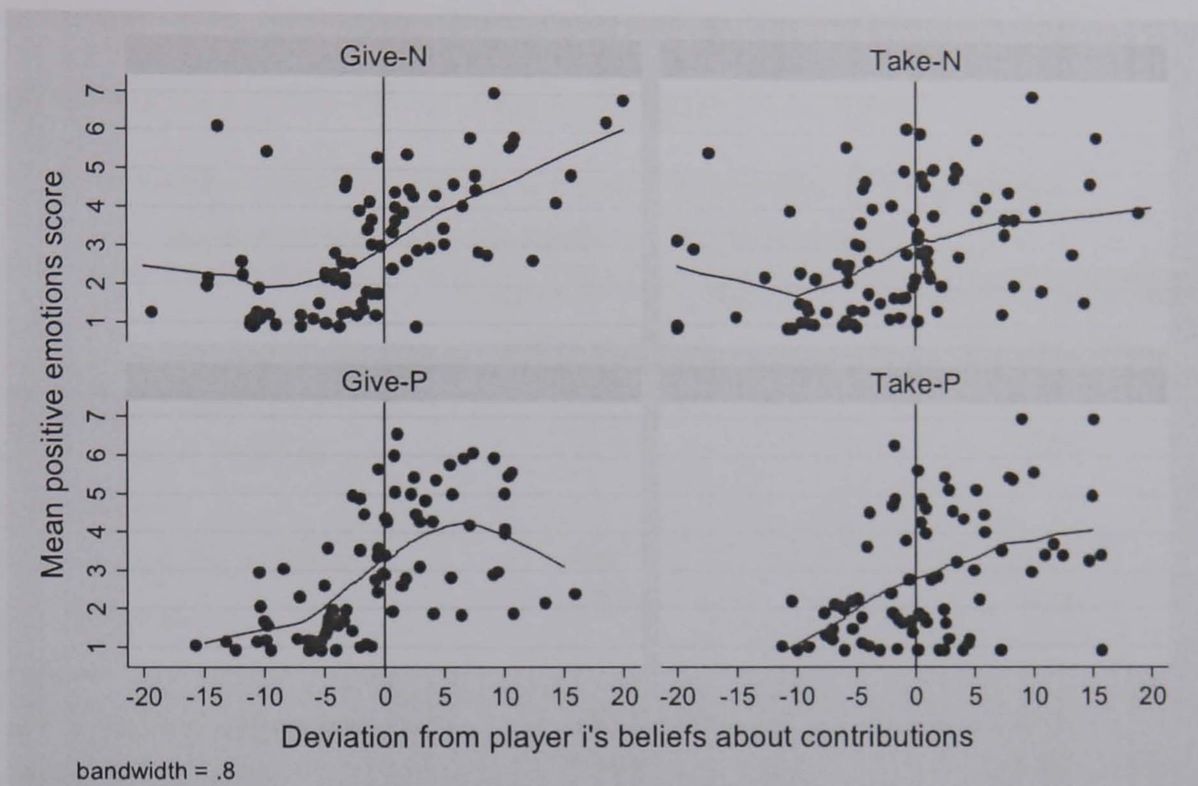
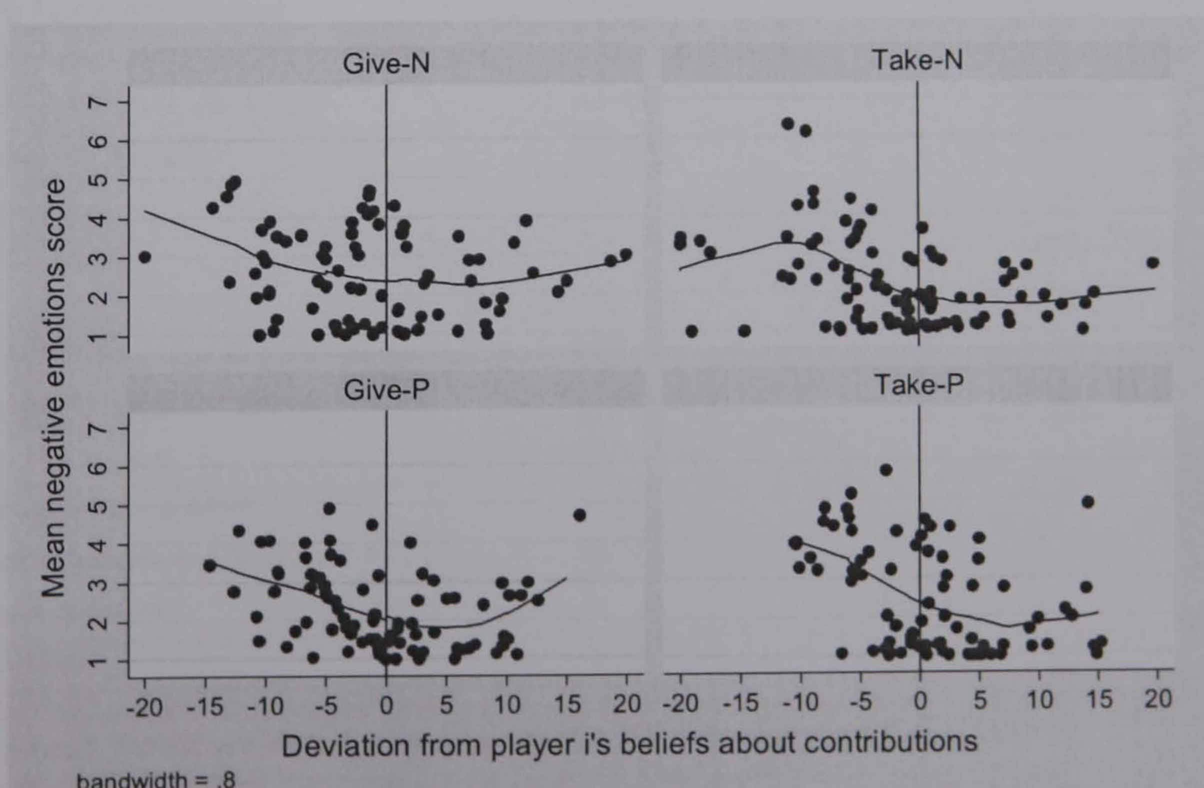


Figure 3.10 Mean negative emotions as a function of deviations from player i's beliefs about contributions for each treatment



Econometric models reported in Table 3.8 test for any emotions' differences across frames. It turns out that replacing deviations from contributions with deviations from beliefs about contributions in the emotions functions makes little qualitative difference to the results.

Table 3.8 Mean positive and negative emotions as a function of deviations from player i's beliefs about contributions – Regression results

	Dependent variable: Mean positive emotions	Dependent variable: Mean negative emotions
Player j 's absolute negative deviation from player i 's beliefs about contributions	-0.207*** (0.043)	0.134*** (0.030)
Player j 's positive deviation from player i 's beliefs about contributions	0.061** (0.031)	0.022 (0.028)
Player k 's contribution deviation from player i's beliefs about contributions	-0.009 (0.007)	0.006 (0.011)
Take	-0.145 (0.219)	0.014 (0.217)
Take × absolute negative deviation from player i 's beliefs about contributions	0.038 (0.054)	-0.008 (0.043)
Take × positive deviation from player i 's beliefs about contributions	-0.032 (0.032)	-0.040 (0.028)
No-Punishment	-0.217 (0.215)	-0.039 (0.214)
No-Punishment × absolute negative deviation from player i 's beliefs about contributions	0.101** (0.046)	-0.064* (0.038)
No-Punishment × positive deviation from player i 's beliefs about contributions	0.044 (0.031)	0.006 (0.027)
Observations	336	336

*Notes: Ordered probit estimates. Standard errors are presented in parentheses (clustered on groups). Results are corrected for heteroskedasticity. * denotes significance at the 10-percent level, ** denotes significance at the 5-percent level, and *** at the 1-percent level. All data are pooled.*

Econometric analysis finds no framing differences for either positive or negative emotions, whereas the only significant variables is the absolute negative (for both positive and negative emotions) and the positive deviation (for positive emotions) from player i's beliefs about contributions. With respect to the presence of punishment, our findings suggest that for positive and for negative emotions, there does not appear to be a level difference, so much as an interaction.[28] In particular, in both equations the interaction variable "No-Punishment × absolute negative deviation from player i's beliefs about contributions" is significant (especially for the mean positive emotions), implying that the slope of the corresponding emotions' functions when player j negatively deviates from player i's beliefs about contributions is affected by the availability of punishment opportunities.

3.5 Conclusions

Motivated by previous findings about altruistic behaviour in dictator games and our evidence from moral judgments reported in Chapter 2, the present chapter reports an investigation of the frame-sensitivity of two indicators of negative reciprocity, an equally-important, if darker, source of social preferences. Specifically, we study punishment and emotional responses in a one-shot public goods game. Our findings suggest that neither is sensitive to the Give versus Take framing manipulation we

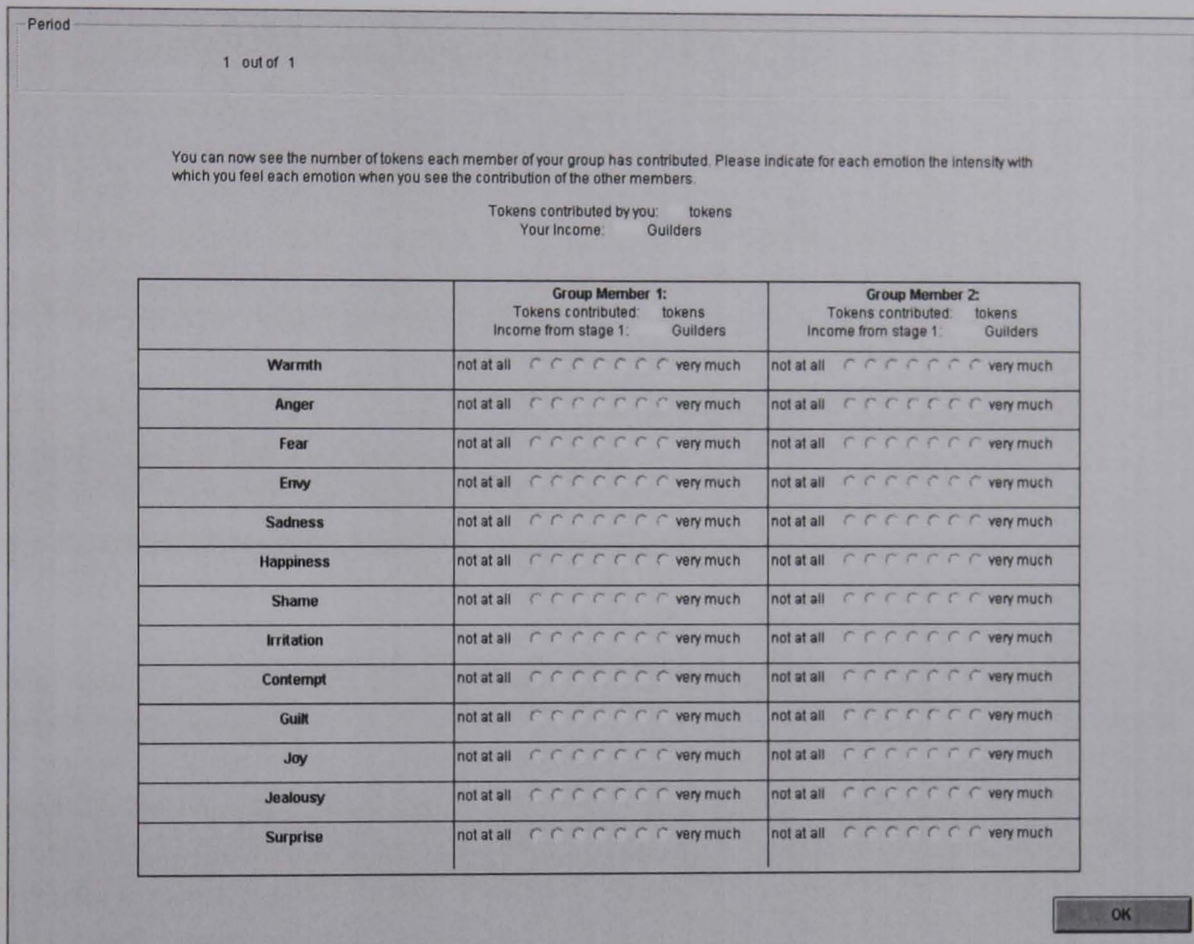
[28] As in the case of deviations from player i's contributions, a similar sort of analysis for elicited emotions has been performed with respect to deviation from player i's beliefs about contributions for either the no-punishment or the punishment treatment, taken separately. These results are reported in Table B.6 in Appendix B.

consider, controlling for the level of contributions. We conclude that the main determinant of punishment is the difference between the contributions of the punisher and the punished; and the same difference is also the main source of emotional responses. We essentially reach the same conclusion when we examine deviations from beliefs about contributions. In contrast to the frame-sensitivity of moral judgments in public goods games or altruism in dictator games, our framing has little effect on our indicators of negative reciprocity. Interestingly, the availability of punishment affects elicited emotions. More specifically, the presence of punishment opportunities affects the mean positive emotions (but not the mean negative emotions) when player j negatively deviates from player i's contributions; whereas, the mean positive emotions and, to a lesser extent, the mean negative emotions are affected by the presence of punishment looking at negative deviations from player i's beliefs about contributions.

Appendix A. Technical Appendix

Appendix A.1: Screenshot for eliciting emotions

[Note: The screenshot for eliciting self-reports on emotions is presented below. The order of emotions was exactly the same in all four treatments.]


Period

1 out of 1

You can now see the number of tokens each member of your group has contributed. Please indicate for each emotion the intensity with which you feel each emotion when you see the contribution of the other members.

Tokens contributed by you: tokens
Your Income: Guilders

	Group Member 1: Tokens contributed: tokens Income from stage 1: Guilders	Group Member 2: Tokens contributed: tokens Income from stage 1: Guilders
Warmth	not at all very much	not at all very much
Anger	not at all very much	not at all very much
Fear	not at all very much	not at all very much
Envy	not at all very much	not at all very much
Sadness	not at all very much	not at all very much
Happiness	not at all very much	not at all very much
Shame	not at all very much	not at all very much
Irritation	not at all very much	not at all very much
Contempt	not at all very much	not at all very much
Guilt	not at all very much	not at all very much
Joy	not at all very much	not at all very much
Jealousy	not at all very much	not at all very much
Surprise	not at all very much	not at all very much

OK

Appendix A.2: Instructions

[Note: Instructions used for the Give and Punishment treatments. Amendments to the Stage 1 instructions for the No Punishment treatments are given in curly brackets, respectively.]

Instructions

You are now taking part in an economic experiment financed by the University of Nottingham. You can earn a considerable amount of money depending on the decisions made by you and other participants. It is therefore very important that you read these instructions with care.

These instructions are solely for your private use. **It is prohibited to communicate with other participants during the experiment.** If you have any questions, please ask me. If you violate this rule, you will be dismissed from the experiment and forfeit all payments.

During the experiment we will not speak in terms of Pounds, but in Guilders. During the experiment your entire earnings will be calculated in Guilders. At the end of the experiment the total amount of Guilders you have earned will be converted to Pounds at the following rate:

1 Guilder = 0.40 Pounds

At the end of the experiment your entire earnings from the experiment will be paid to you **in cash**.

During the experiment, you will be asked to fill in a few questionnaires. The answers you provide in these questionnaires are completely anonymous. They will not be revealed to anyone either during the experiment or after it. Furthermore, your answers to these questionnaires will not affect your earnings during the experiment.

At the beginning of the experiment, all participants will be randomly divided into groups of three. Apart from you, there will be two more members in your group. **You will not learn who the other people in your group are at any point.**

The experiment consists of two stages {one stage}. In the following pages we describe the experiment in detail. At the end of this introductory information we ask you to do several computerised control exercises which are designed to check that you have understood the decision situation.

Detailed Information on the Experiment

Stage 1

Each participant receives an endowment of **20 tokens**. At stage 1, you have to decide how many of these 20 tokens you contribute to a group project and how many you keep for yourself. The other two members of your group have to make the same decision. They can also either contribute tokens to a project or keep tokens for themselves. You and the other members of the group can each choose any amount between 0 and 20 tokens to contribute.

Every token that you do not contribute to the project automatically belongs to you and earns you one Guilder.

For the tokens contributed to the project the following happens: **the project's value will be multiplied by 1.5 and this amount will be divided equally among all three members of the group.** For example, if 1 token is contributed to the project, the project's value increases to 1.5 Guilders. This amount is divided equally among all three members of the group. Thus every group member receives 0.5 Guilders.

Your income from the project rises by 0.5 Guilders if you contribute one token more to the project. At the same time, the income of the other two members of the group also rises by 0.5 tokens, because they receive the same income from the project as you do. Therefore, if you contribute one token more to the project, the income from the project received by the whole group together increases by 1.5 Guilders. It is also true that your income rises by 0.5 Guilders if another group member contributes one token more to the project.

After all three members of the group have made their decisions about the amounts of tokens they contribute to the project the total income achieved by each participant is determined.

How is your income calculated from your decision?

The income of every member of the group is calculated in the same way. As you can see, your **income** consists of two parts:

(1) The tokens which you have kept for yourself ('income from tokens kept') whereby **1 token = 1 Guilder**.
(2) The 'income from the project' calculated as follows: **Your income from the project = 0.5 *times* sum of all tokens contributed to the project by members of your group.**

Your **total income in Guilders at stage 1 of the experiment** is therefore:

(20 – tokens contributed to the project by you) + 0.5*(sum of all tokens contributed to the project by members of your group)

If you do not contribute anything to the project the income from tokens kept is 20. If you contribute for instance 7 tokens to the project your income from tokens kept is 13. At the same time, the total sum of tokens contributed to the project increases and so does your 'income from the project'.

In order to explain the income calculation we give some examples. Please read them carefully:

Example 1:
If each of the three members of the group contributes 0 tokens to the project, all three will receive an 'income from tokens kept' of 20. Nobody receives anything from the project, because no one contributed anything. Therefore the total income of every member of the group is 20 tokens.
*Calculation of the income from stage 1 for every participant: (20-0) + 0.5 * (0) = 20*

Example 2:
If each of the three members of the group contributes 20 tokens, there will be a total of 60 tokens contributed to the project. The 'income of tokens kept' is 0 for everyone, but each member receives an income from the project of 0.5 * 60 = 30 tokens.
*Calculation of the income from stage 1 for every participant: (20-20) + 0.5 * (60) = 30*

Example 3:
If you contribute 20 tokens, the second member 10 tokens and the third 0 tokens, the following incomes are calculated.

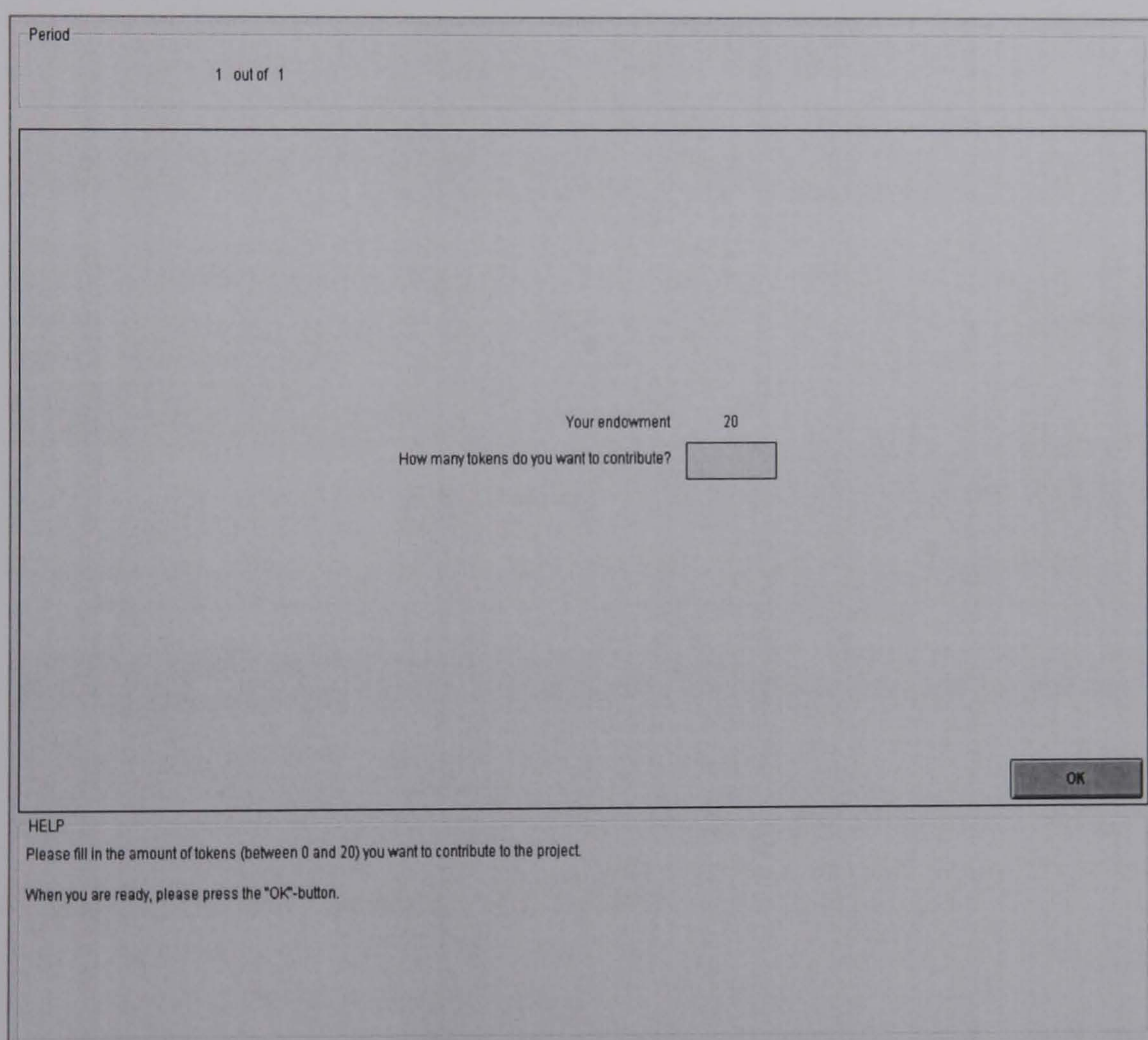
- Because you and the second member of the group have together contributed 30 tokens, everyone will receive 0.5 * 30 = 15 Guilders from the project.
- You contributed all your 20 tokens to the project. You will therefore receive 15 Guilders in total from the project.
- The second member of the group also receives 15 Guilders from the project. In addition, he receives 10 Guilders as the 'income from tokens kept', because he contributed 10 tokens to the project. Thus, he receives 15 + 10 = 25 Guilders altogether.
- The third member of the group, who did not contribute anything, also receives the 15 Guilders from the project and additionally the 20 Guilders from the 'income from tokens kept', which means 20 + 15 = 35.

*Calculation of your income from stage 1: (20-20) + 0.5 * (30) = 15*

*Calculation of the income from stage 1 for the 2nd group member: (20-10) + 0.5 * (30) = 25*
*Calculation of the income from stage 1 for the 3rd group member: (20-0) + 0.5 * (30) = 35*

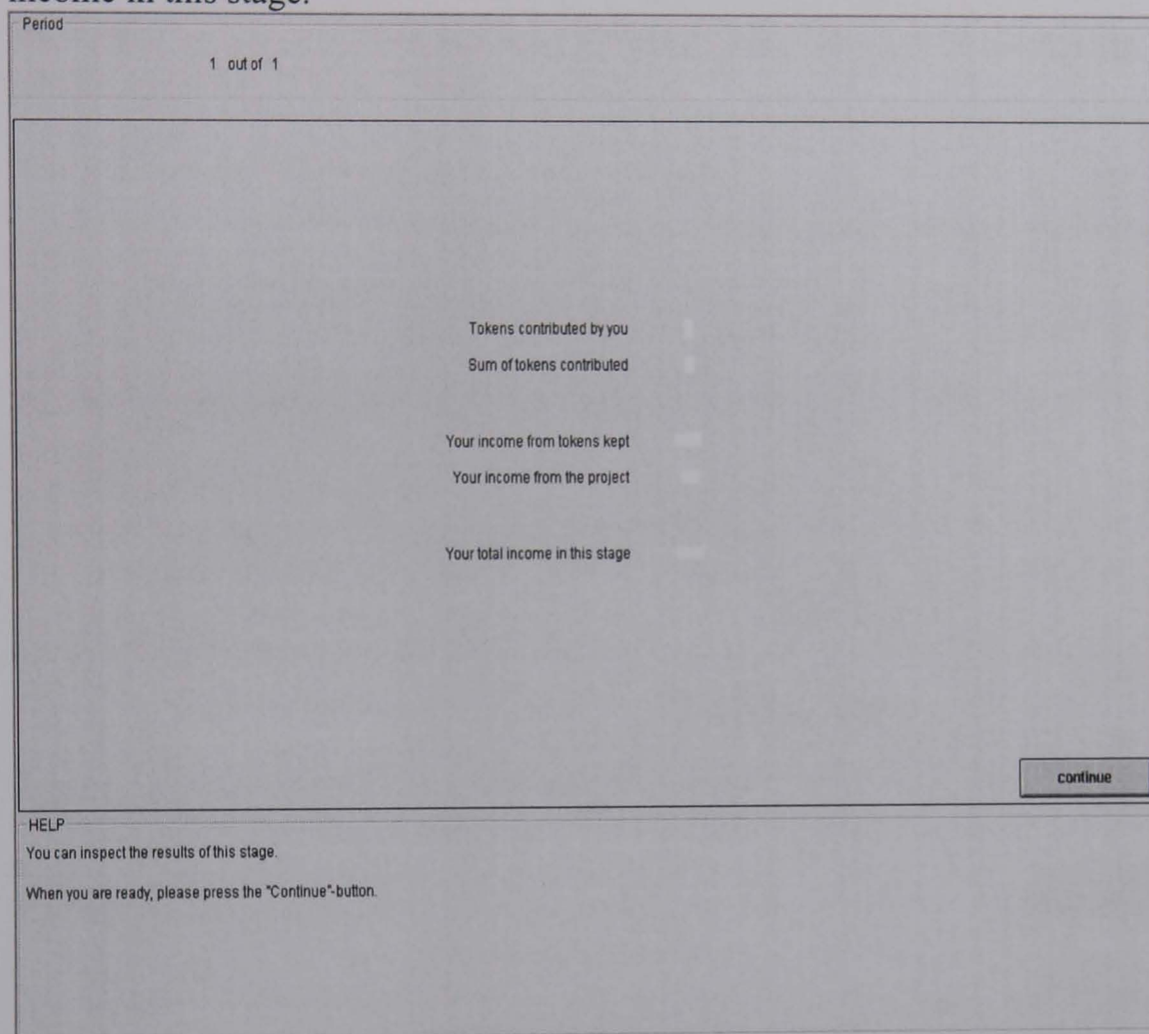
Example 4:
The other two members of your group contribute 20 tokens each to the project. You do not contribute anything. In this case the income will be calculated as follows:
*Calculation of your income from stage 1: (20-0) + 0.5 * (40) = 40*
*Calculation of the income from stage 1 for the 2nd group member: (20-20) + 0.5 * (40) = 20*
*Calculation of the income from stage 1 for the 3rd group member: (20-20) + 0.5 * (40) = 20*

When making your decision, the following input-screen will appear:



As mentioned above, your **endowment in the experiment is 20 tokens**. You have to decide how many tokens you contribute to the project by typing a number between 0 and 20 in the input field. This field can be reached by clicking it with the mouse. By deciding how many tokens to contribute to the project, you automatically decide how many tokens you keep for yourself. After entering the amount of tokens you contribute you must press the O.K. button using the mouse. Once you have done this, your decision can no longer be revised.

After that, you will be informed about the amount of tokens you contributed to the project, the sum of tokens contributed to the project and your total income in this stage.

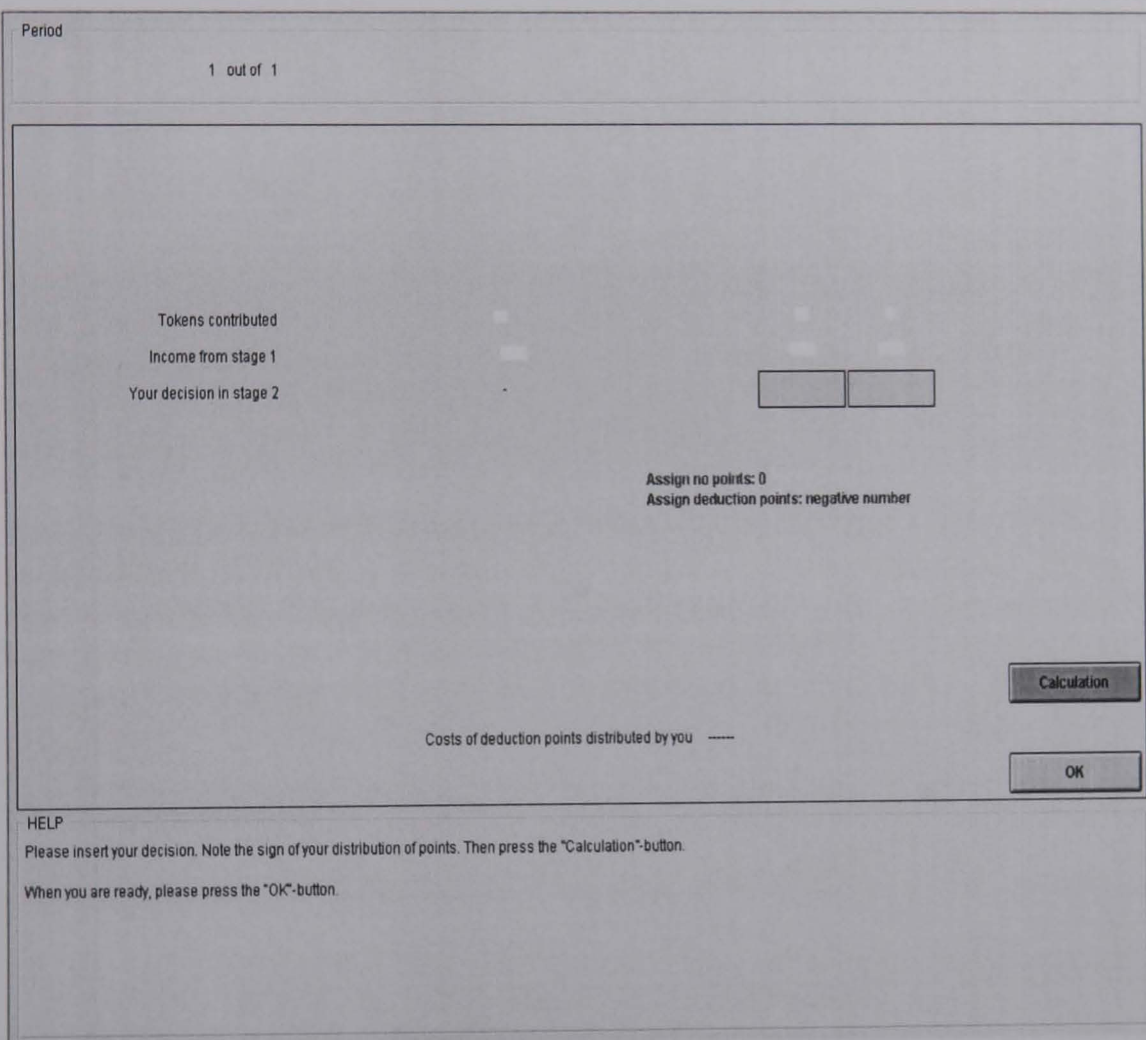


Stage 1 is now over and stage 2 commences.

Stage 2

At this stage, you will see how many tokens each of the other two group members has contributed to the project and his or her corresponding income from stage 1. Moreover, you can either **decrease** or **leave unchanged** the income of each other group member by assigning **deduction points** to them. The other group members can also decrease **your** income, by assigning deduction points to you, if they wish to do so.

You will see the following input screen at stage 2:



You must now decide how many deduction points to assign to each of the other two group members. In the first column you can see your contribution and your income from stage 1. In the other two columns, you can see the same information for each of the other two members of the group.

If you do not wish to change the income of a specific group member then enter 0 in the large box for that group member. If you do wish to reduce a group member's income, enter instead the number of deduction points that you wish to assign to them, preceded by minus sign (without spaces between them). For example, to assign 2 deduction points to a group member, type -2 in the relevant box. You can move from one input field to the other by pressing the tab -key (→|) or by using the mouse. You must enter a response in each large box.

You can assign **between 0 and 5 deduction points to each other group member**.

For each deduction point that you assign, there is a cost to you of one Guilder. Thus, the total cost to you in Guilders of assigning deduction points to other group members is given by the total number of deduction points that you assign. You can check the total cost on the computer, by pressing the '**Calculation**' button after you have assigned deduction points. Until you press the **OK-button**, you can still change your decision. To recalculate the costs after making a change, simply press the cost calculation button again.

The effects of assigning deduction points to other group members are as follows: If you give 0 points to a particular group member, you will not have any effect on his or her income. However, **for each deduction point that you assign to a particular group member, you will decrease their income by 2 Guilders** (unless their income is already exhausted). For example, if you give a group member 2 deduction points (i.e., enter –2), you will decrease their income by 4 Guilders. And so on.

Your own income will be reduced by 2 Guilders for each deduction point that is assigned to you by the other two group members, except that, if all of your income from the first stage is exhausted as a result of deduction points, your income cannot be reduced any further by other group members. Therefore, your total income from the two stages is calculated as follows:

Total income (in Guilders) after stage 2

= income from stage 1 (1)
– 2*(sum of deduction points assigned to you) (2)
– costs of deduction points assigned by you

if (1) + (2) is greater than or equal to 0;

= **0** – costs of deduction points assigned by you

if (1) + (2) is less than 0

Please note that your income in Guilders after stage 2 can be negative, if the cost of deduction points assigned by you exceeds your income from stage 1 less any reduction in your income caused by deduction points assigned to you by other group members.

However, at the end of the experiment and in addition to the calculation just given, you and the other members of your group will each receive a lump sum payment of **10 Guilders.** This payment is enough to cover any losses that you could incur.

Do you have any questions?

Control Questionnaire

1. Each group member has an endowment of 20 tokens. Suppose that nobody (including yourself) contributes any token to the project. What is:

Your income ?...........
The income of the other group members?...........

2. Each group member has an endowment of 20 tokens. Suppose that you contribute 20 tokens to the project. All other group members each contribute 20 tokens to the project. What is:

Your income?...........
The income of the other group members?...........

3. Each group member has an endowment of 20 tokens. Suppose that the other two group members contribute together a total of 30 tokens to the project.

a) What is your income if you contribute 0 tokens to the project?...........
b) What is your income if you contribute 4 tokens to the project?...........

4. Each group member has an endowment of 20 tokens. Suppose that you contribute 8 tokens to the project.

a) What is your income if the other group members together contribute a total of 14 tokens to the project?...........
b) What is your income if the other group members together contribute a total of 22 tokens to the project?...........

5. At the second stage you distribute the following deduction points to the other two other group members: -2,-4. What are the total costs of your distributed deduction points?...........

6. What are your costs if you distribute a total of 0 points?...........

7. By how many Guilders will your income from the first stage be changed if you receive a total of 0 deduction points from the other group members?...........

8. By how many Guilders will your income from the first stage be changed if you receive a total of 4 deduction points from the other group members?...........

[Note: Instructions used for the Take and Punishment treatments. Amendments to the Stage 1 instructions for the No Punishment treatments are given in curly brackets, respectively.]

Instructions

You are now taking part in an economic experiment financed by the University of Nottingham. You can earn a considerable amount of money depending on the decisions made by you and other participants. It is therefore very important that you read these instructions with care.

These instructions are solely for your private use. **It is prohibited to communicate with other participants during the experiment.** If you have any questions, please ask me. If you violate this rule, you will be dismissed from the experiment and forfeit all payments.

During the experiment we will not speak in terms of Pounds, but in Guilders. During the experiment your entire earnings will be calculated in Guilders. At the end of the experiment the total amount of Guilders you have earned will be converted to Pounds at the following rate:

1 Guilder = 0.40 Pounds

At the end of the experiment your entire earnings from the experiment will be paid to you **in cash**.

During the experiment, you will be asked to fill in a few questionnaires. The answers you provide in these questionnaires are completely anonymous. They will not be revealed to anyone either during the experiment or thereafter. Furthermore, your answers to these questionnaires will not affect your earnings during the experiment.

At the beginning of the experiment, all participants will be randomly divided into groups of three. Apart from you, there will be two more members in your group. **You will not learn who the other people in your group are at any point.**

The experiment consists of two stages {one stage}. In the following pages we describe the experiment in detail. At the end of this introductory information we ask you to do several computerised control exercises which are designed to check that you have understood the decision situation.

Detailed Information on the Experiment

Stage 1

There are **60 tokens** in a project for your group. You have to decide how many of these 60 tokens you withdraw from the project for yourself and how many of them you leave in the project. Each participant can withdraw up to 20 tokens from the project. The other two members of your group have to make the same decision. They can also either withdraw tokens from the project for themselves or leave tokens in the project. You and the other members of the group can each choose any amount between 0 and 20 tokens to withdraw.

Every token that you withdraw from the project for yourself automatically belongs to you and earns you one Guilder.

For the tokens that are not withdrawn from the project the following happens: **the project's value will be multiplied by 1.5 and this amount will be divided equally among all three members of the group.** For example, if 1 token is not withdrawn from the project, the project's value increases to 1.5 Guilders. This amount is divided equally among all three members of the group. Thus every group member receives 0.5 Guilders.

Your income from the project rises by 0.5 Guilders if you withdraw one token less from the project. At the same time, the income of the other two members of the group also rises by 0.5 tokens, because they receive the same income from the project as you do. Therefore, if you withdraw one token less from the project, the income from the project received by the whole group together increases by 1.5 Guilders. It is also true that your income rises by 0.5 Guilders if another group member withdraws one token less from the project.

After all three members of the group have made their decisions about the amounts they withdraw from the project the total income achieved by each participant is determined.

How is your income calculated from your decision?

The income of every member of the group is calculated in the same way. As you can see, your **income** consists of two parts:

(1) The tokens which you have withdrawn for yourself ('income from tokens withdrawn') whereby **1 token = 1 Guilder**.
(2) The 'income from the project' calculated as follows: **Your income from the project = 0.5 *times* (60 – sum of all tokens withdrawn from the project by members of your group)**

Your **total income in Guilders at stage 1 of the experiment** is therefore: **(Tokens withdrawn from the project by you) + 0.5*(60 – sum of all tokens withdrawn from the project by members of your group)**

If you withdraw all 20 tokens from the project the income from tokens withdrawn is 20. If you leave for instance 7 tokens in the project your income from tokens withdrawn is 13. At the same time, the total sum of tokens left in the project increases and so does your 'income from the project'.

In order to explain the income calculation we give some examples. Please read them carefully:

Example 1:
If each of the three members of the group withdraws 20 tokens from the project, all three will receive an 'income from tokens withdrawn' of 20. Nobody receives anything from the project, because no one left anything. Therefore the total income of every member of the group is 20 tokens.
*Calculation of the income from stage 1 for every participant: (20) + 0.5 * (60-60) = 20*

Example 2:
If each of the three members of the group withdraws 0 tokens, there will be a total of 60 tokens left in the project. The 'income of tokens withdrawn' is 0 for everyone, but each member receives an income from the project of 0.5 * 60 = 30 tokens.
*Calculation of the income from stage 1 for every participant: (0) + 0.5 * (60-0) = 30*

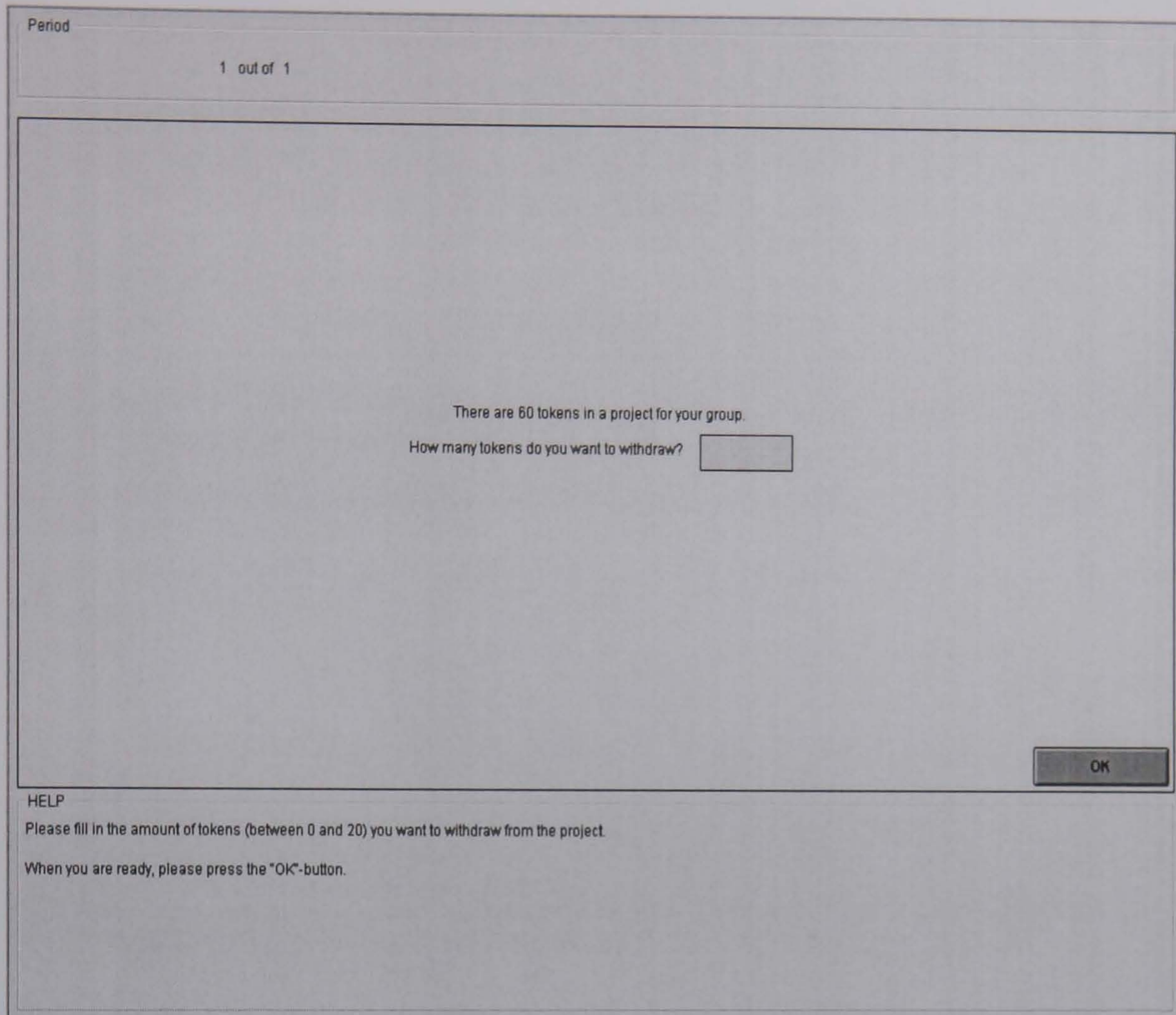
Example 3:
If you withdraw 0 tokens, the second member 10 tokens and the third 20 tokens, the following incomes are calculated.

- Because the second and the third member of the group have together withdrawn 30 tokens, everyone will receive 0.5 * 30 = 15 Guilders from the project.
- You withdrew 0 tokens from the project. You will therefore receive 15 Guilders in total from the project.
- The second member of the group also receives 15 Guilders from the project. In addition, he receives 10 Guilders as the 'income from tokens withdrawn', because he withdrew 10 tokens from the project. Thus, he receives 15 + 10 = 25 Guilders altogether.
- The third member of the group, who withdrew all tokens, also receives the 15 Guilders from the project and additionally the 20 Guilders from the 'income from tokens withdrawn', which means 20 + 15 = 35.

*Calculation of your income from stage 1: (0) + 0.5 * (60-30) = 15*
*Calculation of the income from stage 1 for the 2nd group member: (10) + 0.5 * (60-30) = 25*
*Calculation of the income from stage 1 for the 3rd group member: (20) + 0.5 * (60-30) = 35*

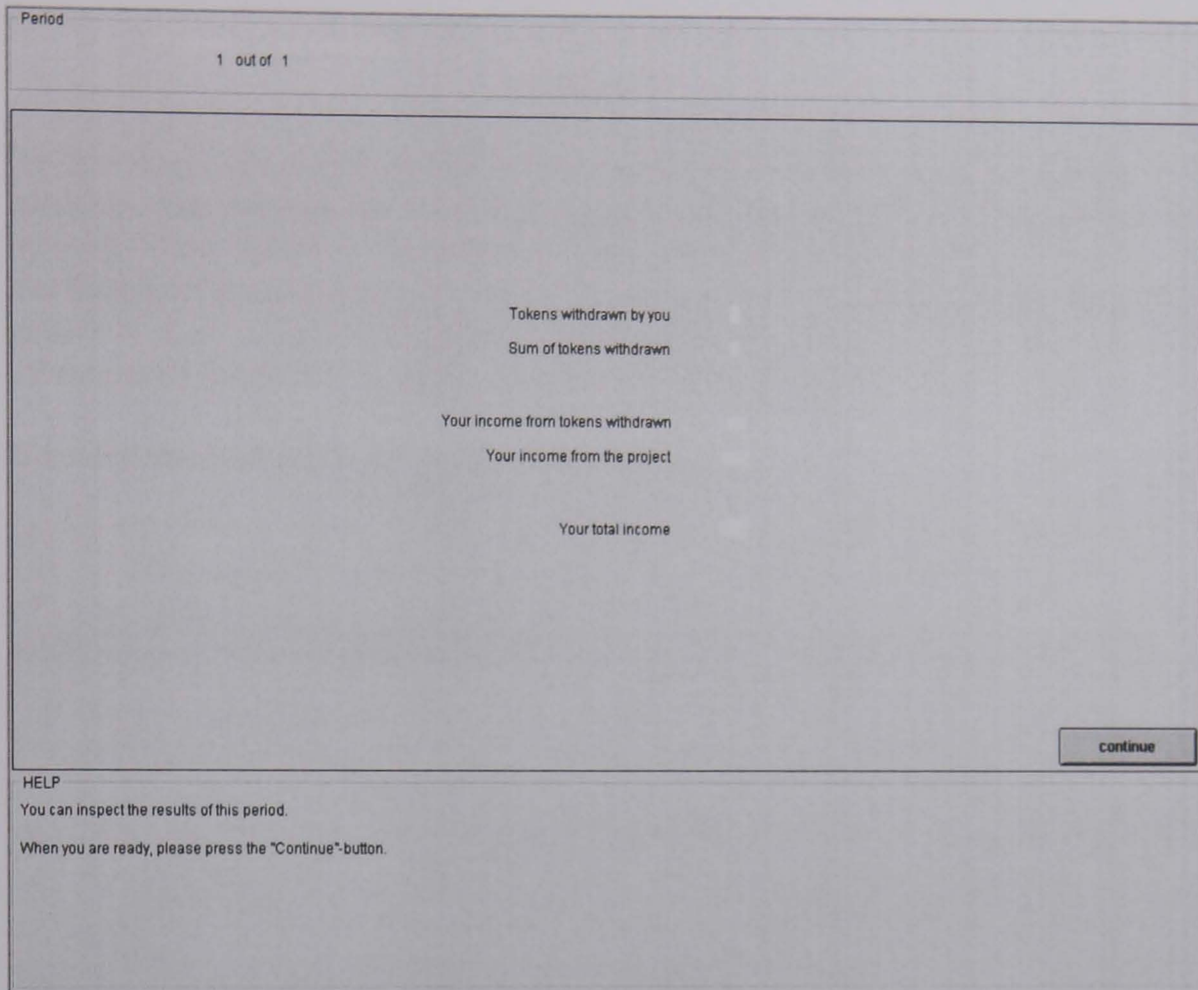
Example 4:
The other two members of your group withdraw 0 tokens each from the project. You withdraw all tokens. In this case the income will be calculated as follows:
*Calculation of your income from stage 1: (20) + 0.5 * (60-20) = 40*
*Calculation of the income from stage 1 for the 2nd group member: (0) + 0.5 * (60-20) = 20*
*Calculation of the income from stage 1 for the 3rd group member: (0) + 0.5 * (60-20) = 20*

When making your decision, the following input-screen will appear:



As mentioned above, there are **60 tokens** in a project for your group. You have to decide how many of these 60 tokens you withdraw from the project by typing a number between 0 and 20 in the input field. This field can be reached by clicking it with the mouse. By deciding how many tokens to withdraw from the project, you automatically decide how many tokens you leave in the project. After entering the amount of tokens you withdraw you must press the O.K. button using the mouse. Once you have done this, your decision can no longer be revised.

After all participants have made their decisions, your total income will be displayed on the following screen:

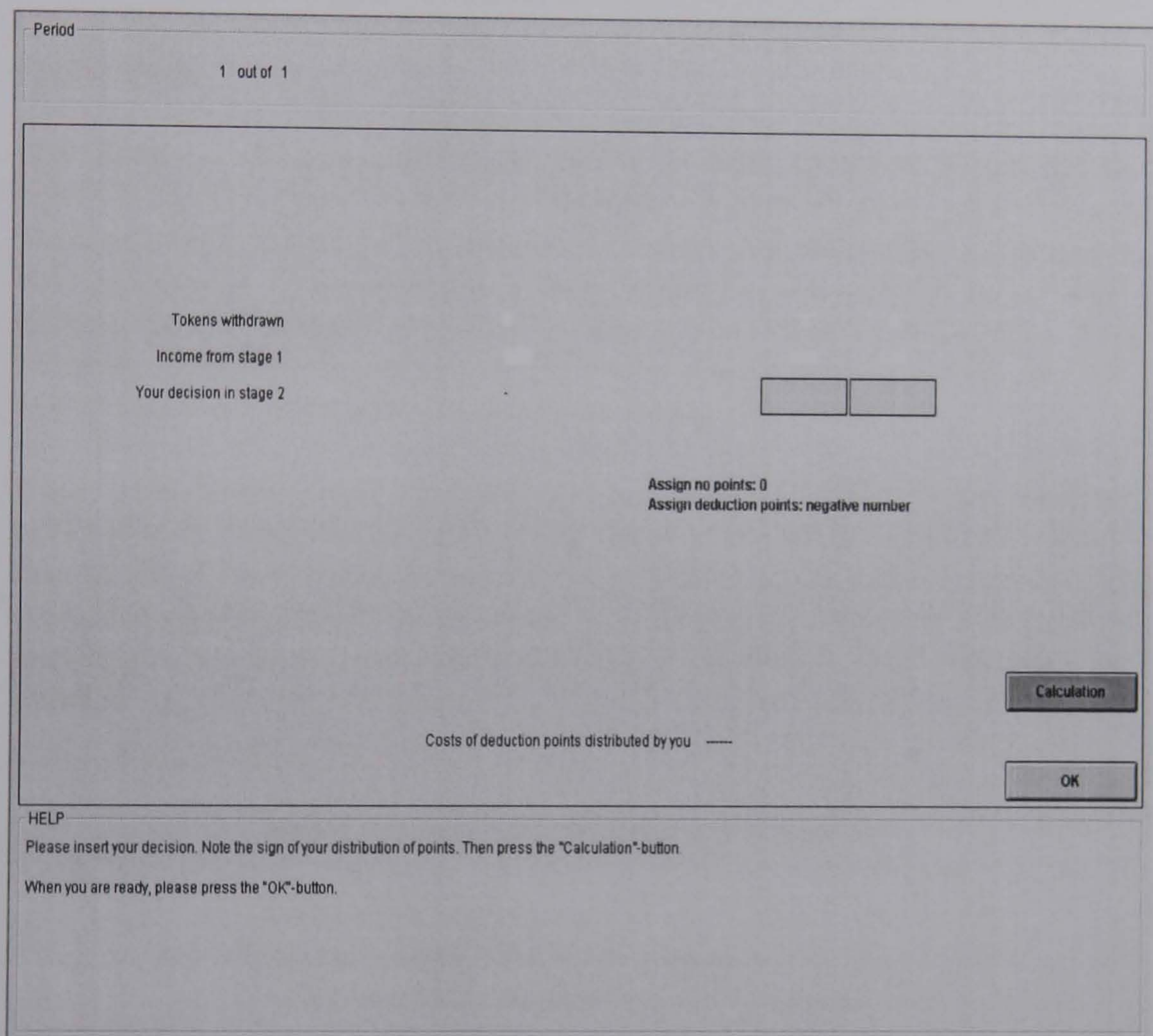


Stage 1 is now over and stage 2 commences.

Stage 2

At this stage, you will see how many tokens each of the other two group members has withdrawn from the project and his or her corresponding income from stage 1. Moreover, you can either **decrease** or **leave unchanged** the income of each other group member by assigning **deduction points** to them. The other group members can also decrease **your** income, by assigning deduction points to you, if they wish to do so.

You will see the following input screen at stage 2:



You must now decide how many deduction points to assign to each of the other two group members. In the first column you can see your withdrawal and your income from stage 1. In the other two columns, you can see the same information for each of the other two members of the group.

If you do not wish to change the income of a specific group member then enter 0 in the large box for that group member. If you do wish to reduce a group member's income, enter instead the number of deduction points that you wish to assign to them, preceded by minus sign (without spaces between them). For example, to assign 2 deduction points to a group member, type -2 in the relevant box. You can move from one input field to the other by pressing the tab -key (→|) or by using the mouse. You must enter a response in each large box.

You can assign **between 0 and 5 deduction points to each other group member**.

For each deduction point that you assign, there is a cost to you of one Guilder. Thus, the total cost to you in Guilders of assigning deduction points to other group members is given by the total number of deduction points that you assign. You can check the total cost on the computer, by pressing the '**Calculation**' button after you have assigned deduction points. Until you press the **OK-button**, you can still change your decision. To recalculate the costs after making a change, simply press the cost calculation button again.

The effects of assigning deduction points to other group members are as follows: If you give 0 points to a particular group member, you will not have any effect on his or her income. However, **for each deduction point that you assign to a particular group member, you will decrease their income by 2 Guilders** (unless their income is already exhausted). For example, if you give a group member 2 deduction points (i.e., enter –2), you will decrease their income by 4 Guilders. And so on.

Your own income will be reduced by 2 Guilders for each deduction point that is assigned to you by the other two group members, except that, if all of your income from the first stage is exhausted as a result of deduction points, your income cannot be reduced any further by other group members. Therefore, your total income from the two stages is calculated as follows:

Total income (in Guilders) after stage 2

= income from stage 1 (1)
– 2*(sum of deduction points assigned to you) (2)
– costs of deduction points assigned by you

if (1) + (2) is greater than or equal to 0;

= **0** – costs of deduction points assigned by you

if (1) + (2) is less than 0

Please note that your income in Guilders after stage 2 can be negative, if the cost of deduction points assigned by you exceeds your income from stage 1 less any reduction in your income caused by deduction points assigned to you by other group members.

However, at the end of the experiment and in addition to the calculation just given, you and the other members of your group will each receive a lump sum payment of **10 Guilders.** This payment is enough to cover any losses that you could incur.

Do you have any questions?

Control Questionnaire

1. There are 60 tokens in a project of your group. Suppose that everyone (including yourself) withdraws 20 tokens from the project. What is:

Your income from the first stage?...........
The income of the other group members?...........

2. There are 60 tokens in a project of your group. Suppose that you withdraw 0 tokens from the project. All other group members each withdraw 0 tokens to the project. What is:

Your income from the first stage?...........
The income of the other group members?...........

3. There are 60 tokens in a project of your group. Suppose that the other two group members withdraw together a total of 10 tokens from the project.

a) What is your income from the first stage if you withdraw 20 tokens from the project at the first stage?...........
b) What is your income from the first stage if you withdraw 16 tokens from the project at the first stage?...........

4. There are 60 tokens in a project of your group. Suppose that you withdraw 12 points from the project.

a) What is your income from the first stage if the other two group members together withdraw a total of 26 tokens from the project?...........
b) What is your income from the first stage if the other group members together withdraw a total of 18 tokens from the project?...........

5. At the second stage you distribute the following deduction points to the other two group members: -2,-4. What are the total costs of your distributed deduction points?...........

6. What are your costs if you distribute a total of 0 points?...........

7. By how many Guilders will your income from the first stage be changed if you receive a total of 0 deduction points from the other group members?...........

8. By how many Guilders will your income from the first stage be changed if you receive a total of 4 deduction points from the other two group members?...........

Appendix B. Data Analysis

Table B.1. Standard deviations for contributions and beliefs about contributions for the Give-N treatment

Groups	Standard deviation for contributions	Standard deviation for beliefs about contributions
1	10	10
2	5.774	6.292
3	1.155	2.255
4	11.547	6.658
5	1.155	1.732
6	2.887	8.520
7	0	5.204
8	2.887	3.464
9	2.887	2.309
10	0.577	1.732
11	4.619	5.204
12	2.887	2.5
13	6.807	6.538
14	9.019	3.329

Table B.2. Standard deviations for contributions and beliefs about contributions for the Take-N treatment

Groups	Standard deviation for contributions	Standard deviation for beliefs about contributions
1	10	10.408
2	0	3.775
3	1.155	2.887
4	4.041	5.196
5	4.041	2.291
6	2.309	5.774
7	3.055	0
8	5.774	1.443
9	6.110	6.614
10	10.408	7.211
11	2.887	0
12	11.015	8.660
13	4.583	4.481
14	4.619	2.887
15	2.887	2.646

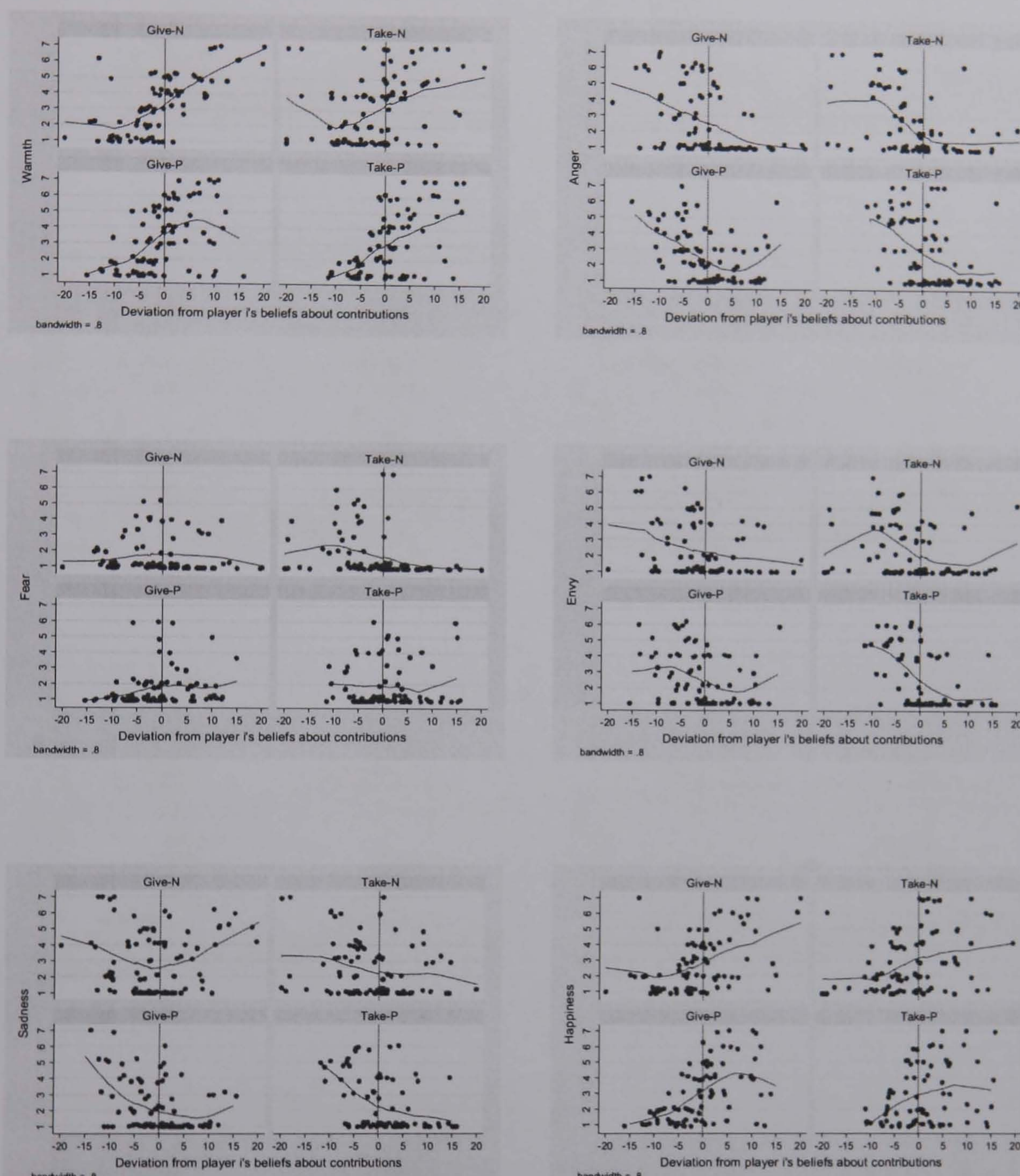
Table B.3. Standard deviations for contributions and beliefs about contributions for the Give-P treatment

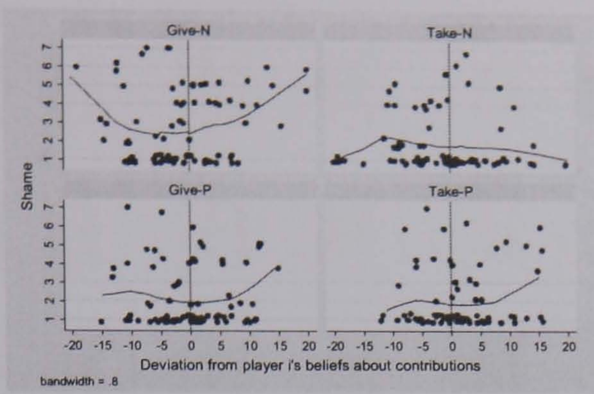
Groups	Standard deviation for contributions	Standard deviation for beliefs about contributions
1	5.774	5.074
2	3.512	2.517
3	6.429	1.443
4	5	3.464
5	3.512	3.014
6	2.646	0
7	2.887	7.006
8	7.767	3.215
9	2	0.287
10	7.638	7.638
11	1	1.155
12	4.041	5.774
13	5.033	3.215
14	7.937	3.969

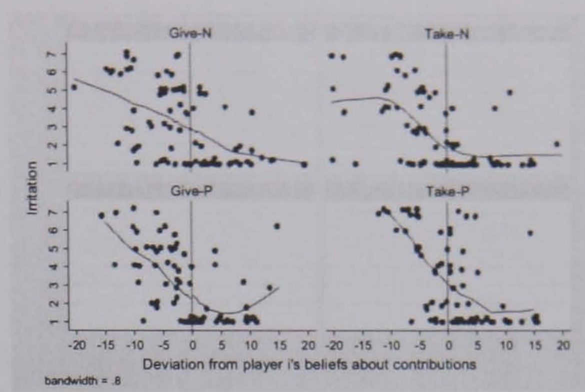
Table B.4. Standard deviations for contributions and beliefs about contributions for the Take-P treatment

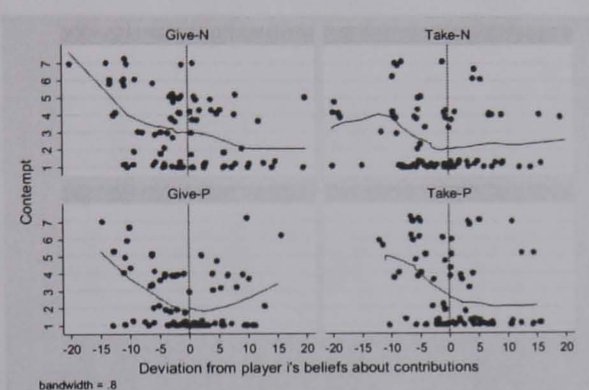
Groups	Standard deviation for contributions	Standard deviation for beliefs about contributions
1	0	2.5
2	2.887	3.819
3	4.359	5.204
4	2.309	2.5
5	7.638	2.887
6	2	1.443
7	5.292	5.774
8	7.506	1.443
9	6.351	5.204
10	3.512	2.517
11	4.359	1.732
12	8.660	6.351
13	4.041	3.547

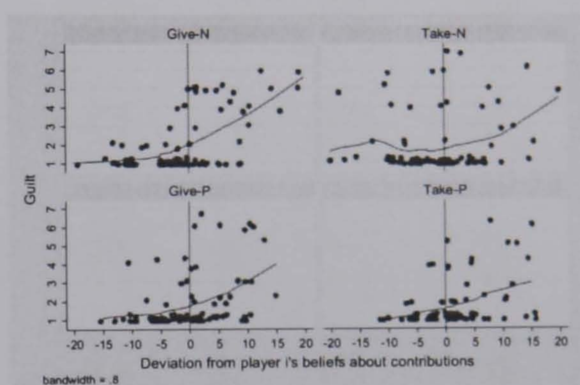
Figure B.1: Emotions as a function of deviations from player i's beliefs about contributions

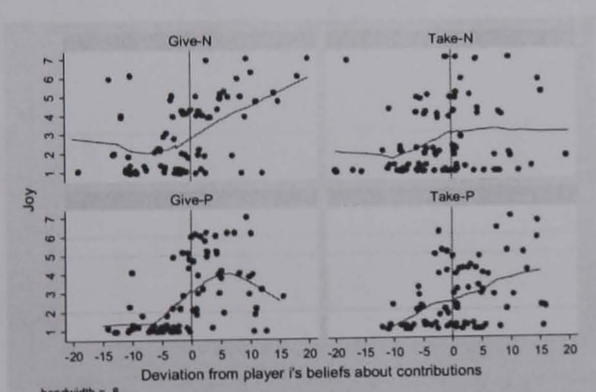


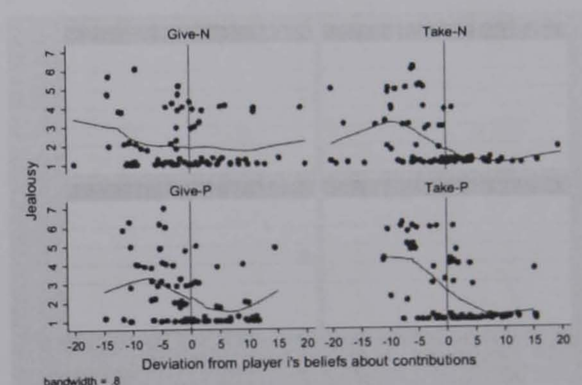

Give-N
Take-N
Give-P
Take-P
Shame
Deviation from player i's beliefs about contributions
bandwidth = .8


Give-N
Take-N
Give-P
Take-P
Irritation
Deviation from player i's beliefs about contributions
bandwidth = .8


Give-N
Take-N
Give-P
Take-P
Contempt
Deviation from player i's beliefs about contributions
bandwidth = .8


Give-N
Take-N
Give-P
Take-P
Guilt
Deviation from player i's beliefs about contributions
bandwidth = .8


Give-N
Take-N
Give-P
Take-P
Joy
Deviation from player i's beliefs about contributions
bandwidth = .8


Give-N
Take-N
Give-P
Take-P
Jealousy
Deviation from player i's beliefs about contributions
bandwidth = .8

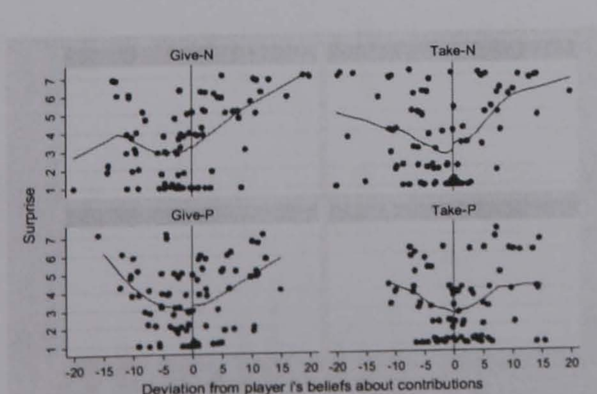

Give-N
Take-N
Give-P
Take-P
Surprise
Deviation from player i's beliefs about contributions
bandwidth = .8

Table B.5. Emotions as a function of deviations from player i's beliefs about contributions – Regression results

	Dependent Variables												
	Warmth	Anger	Fear	Envy	Sadness	Happiness	Shame	Irritation	Contempt	Guilt	Joy	Jealousy	Surprise
Player j's absolute negative deviation from player i's beliefs about contributions	-0.217*** (0.044)	0.120*** (0.026)	-0.079** (0.037)	0.107*** (0.034)	0.114*** (0.040)	-0.150*** (0.037)	0.048 (0.035)	0.220*** (0.034)	0.147*** (0.037)	-0.116** (0.055)	-0.176*** (0.048)	0.118*** (0.037)	0.074*** (0.028)
Player j's positive deviation from player i's beliefs about contributions	0.040 (0.040)	-0.044 (0.039)	-0.003 (0.027)	-0.041 (0.044)	0.013 (0.031)	0.050*** (0.031)	0.083*** (0.029)	-0.035 (0.039)	0.032 (0.027)	0.088*** (0.033)	0.062** (0.032)	-0.040 (0.038)	0.113*** (0.030)
Player k's contribution deviation from player i's beliefs about contribution	-0.016* (0.009)	0.007 (0.011)	0.022** (0.011)	0.017 (0.011)	-0.025** (0.011)	-0.005 (0.008)	-0.002 (0.010)	0.015 (0.009)	0.009 (0.008)	-0.007 (0.013)	-0.004 (0.008)	0.028** (0.011)	-0.004 (0.009)
Take	-0.315 (0.221)	-0.057 (0.224)	-0.217 (0.253)	-0.008 (0.232)	0.214 (0.247)	0.053 (0.212)	-0.122 (0.218)	0.063 (0.215)	0.063 (0.225)	-0.218 (0.301)	-0.047 (0.221)	0.034 (0.234)	-0.140 (0.201)
Take × Player j's absolute negative deviation from player i's beliefs about contributions	0.072 (0.055)	0.018 (0.041)	0.085* (0.042)	0.028 (0.052)	-0.030 (0.043)	-0.002 (0.042)	-0.044 (0.038)	0.014 (0.042)	-0.019 (0.043)	0.054 (0.042)	0.014 (0.047)	0.037 (0.045)	0.039 (0.043)
Take × Player j's positive deviation from player i's about contributions	0.011 (0.038)	0.017 (0.042)	-0.022 (0.039)	-0.036 (0.048)	-0.087*** (0.030)	-0.047 (0.034)	-0.052* (0.032)	-0.009 (0.041)	-0.018 (0.030)	-0.020 (0.032)	-0.043 (0.032)	-0.072 (0.046)	-0.023 (0.036)
No-Punishment	-0.300 (0.225)	-0.168 (0.221)	-0.201 (0.255)	-0.152 (0.224)	0.218 (0.247)	-0.154 (0.211)	0.190 (0.217)	0.094 (0.213)	0.301 (0.226)	-0.015 (0.304)	-0.147 (0.220)	-0.378 (0.238)	-0.054 (0.198)
No-Punishment × Player j's absolute negative deviation from player i's beliefs about contributions	0.105** (0.047)	-0.026 (0.034)	0.061 (0.044)	-0.045 (0.045)	-0.056 (0.043)	0.072* (0.042)	-0.002 (0.042)	-0.113*** (0.040)	-0.065 (0.042)	0.059 (0.054)	0.109** (0.050)	-0.055 (0.045)	-0.023 (0.038)
No-Punishment × Player j's positive deviation from player i's about contributions	0.060 (0.038)	-0.027 (0.042)	-0.070** (0.035)	-0.031 (0.046)	0.067 (0.032)	0.029 (0.034)	-0.023 (0.030)	-0.033 (0.041)	-0.026 (0.030)	0.014 (0.034)	0.022 (0.1032)	0.036 (0.044)	0.086 (0.036)
Observations	336	336	336	336	336	336	336	336	336	336	336	336	336

*Notes: Ordered probit estimates. Standard errors are presented in parentheses (clustered on groups). Results are corrected for heteroskedasticity. ** denotes significance at the 5-percent level, and *** at the 1-percent level. All data are pooled.*

Table B.6. Mean positive and negative emotions as a function of deviations from player i's beliefs about contributions in the absence and presence of punishment options – Regression results

	Dependent variable: Mean positive emotions		Dependent variable: Mean negative emotions	
	No-Punishment	Punishment	No-Punishment	Punishment
Player j's absolute negative deviation from player i's contribution	-0.103** (0.053)	-0.212*** (0.047)	0.072** (0.036)	0.120*** (0.032)
Player j's positive deviation from player i's contribution	0.128*** (0.029)	0.032 (0.036)	0.021 (0.025)	0.018 (0.040)
Player k's contribution deviation from player i's contribution	-0.016** (0.007)	0.003 (0.010)	0.004 (0.014)	0.010 (0.017)
Take	0.130 (0.287)	-0.485 (0.329)	-0.220 (0.314)	0.078 (0.313)
Take × Player j's absolute negative deviation from player i's contribution	0.024 (0.068)	0.060 (0.063)	0.004 (0.054)	0.060 (0.060)
Take × Player j's positive deviation from player i's contribution	-0.074** (0.036)	0.021 (0.050)	-0.030 (0.032)	-0.036 (0.047)
Observations	174	162	174	162

*Notes: Ordered probit estimates. Standard errors are presented in parentheses (clustered on groups). Results are corrected for heteroskedasticity. ** denotes significance at the 5-percent level, and *** at the 1-percent level. All data are pooled.*

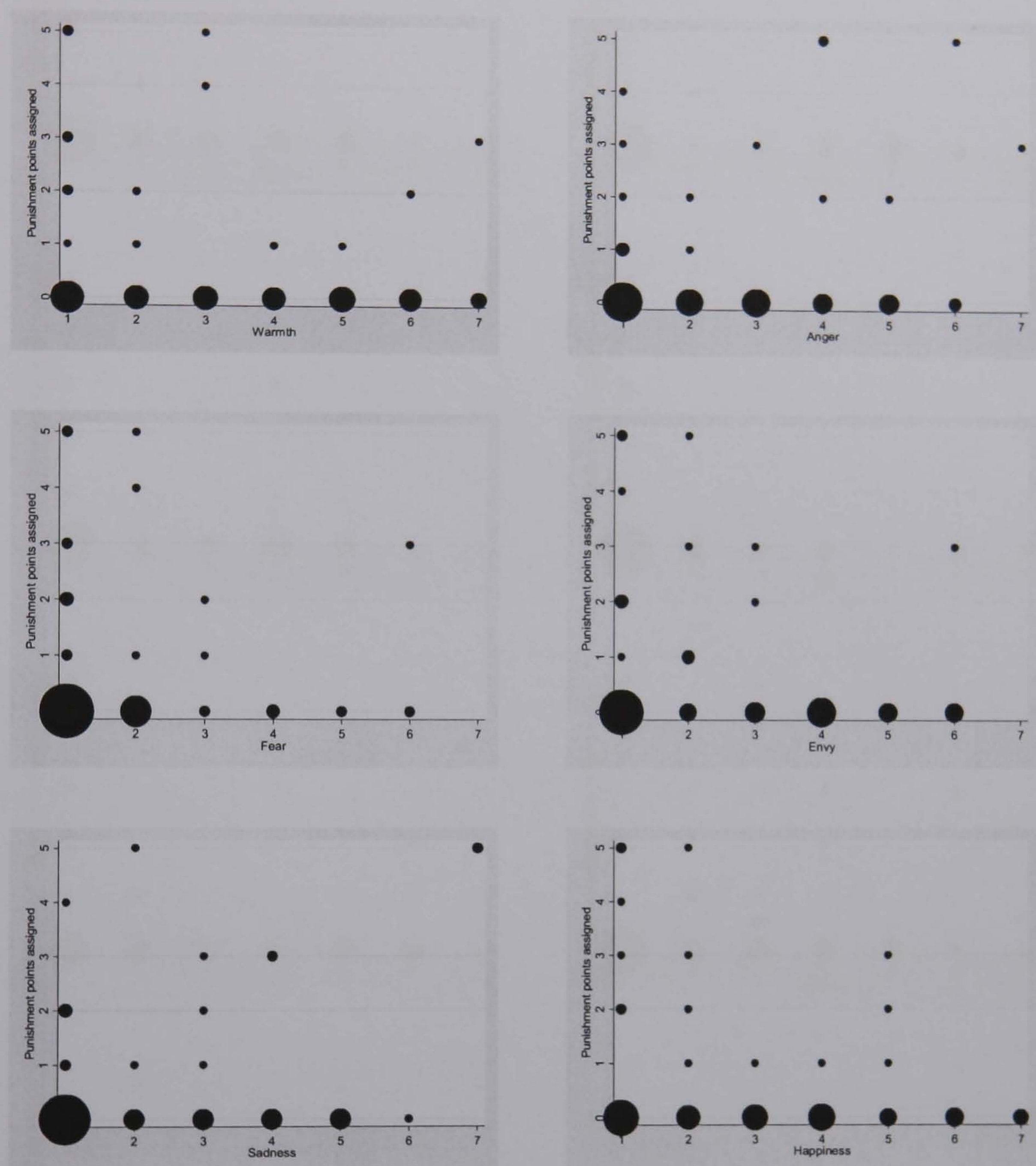
Correlation between punishment and emotions

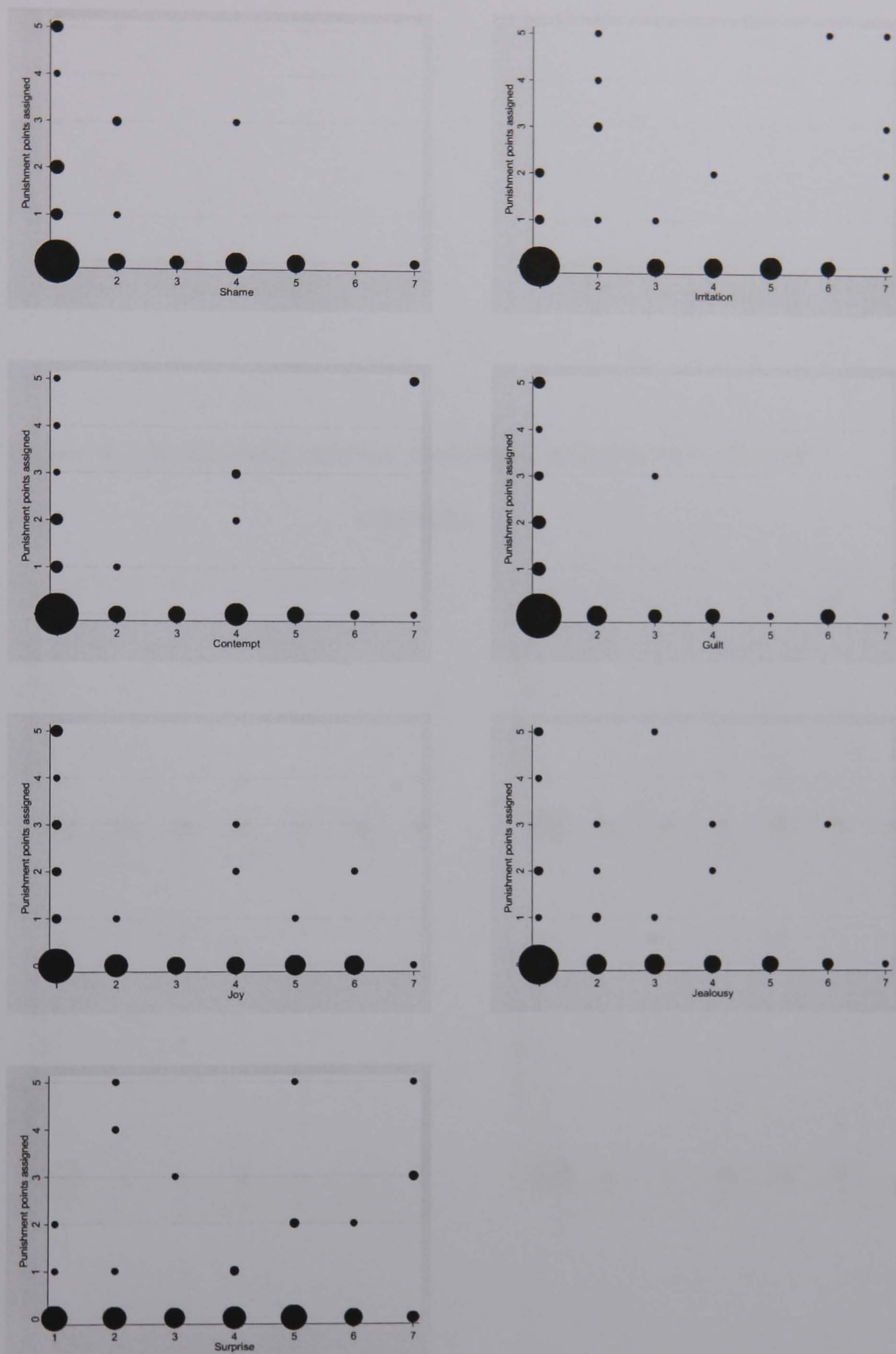
In the analysis provided earlier in the main text, both punishment and emotions are treated as separate indicators to measure the frame sensitivity of negative reciprocity. However, it is likely that these two notions are interdependent, in the sense that one's own emotions may affect one's own punishment assigned. The aim of this appendix is to investigate the correlation between punishment and emotions, as emotions could be an important determinant of how punishment is assigned. It is worth noting that investigating the correlation between punishment and emotions does of course not establish any causation (which is beyond the scope of this experiment and analysis).[29]

A graphical representation of the relationship between punishment points assigned and each emotion separately is given in Figures B.3.1a and B.3.1b for the Give-P and the Take-P treatments, respectively. In these figures, we depict punishment points assigned by player i to player j as a function of each emotion separately (mean positive and mean negative emotions defined earlier are also shown) for the punishment treatments. Circles represent combinations of punishment points and emotions, with the size of the circles being proportional to the number of observations.

[29] As earlier analysis suggests the causation could go the other way for the case of positive emotions.

Figure B.2. Relationship between punishment and emotions – Give-P treatment




Punishment points assigned
Shame
Irritation
Contempt
Guilt
Joy
Jealousy
Surprise

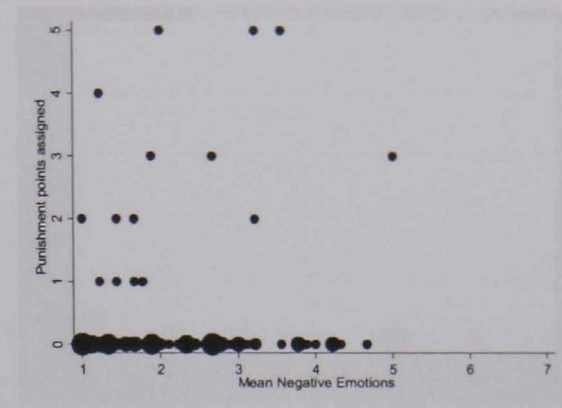
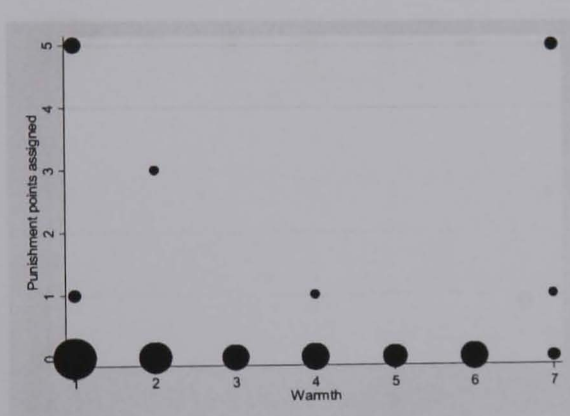
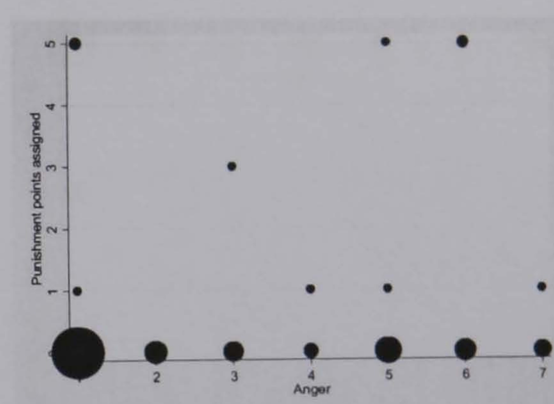
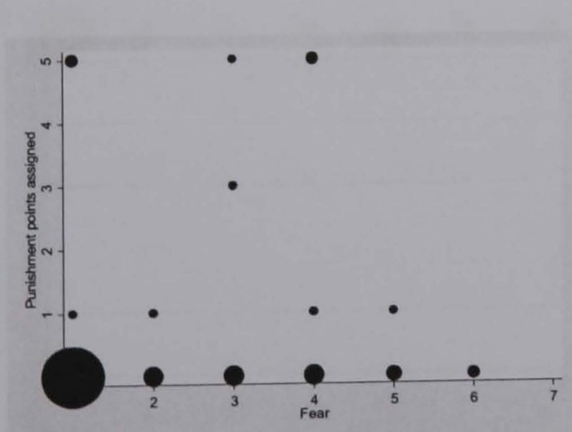
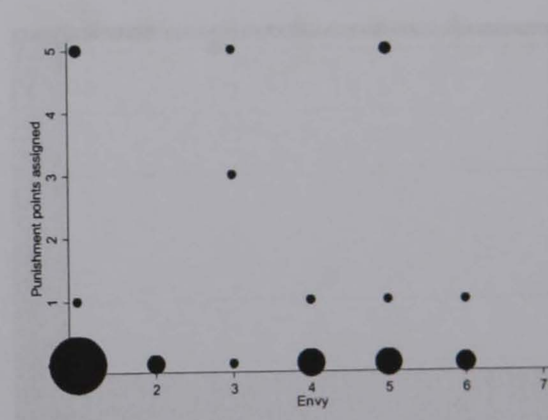


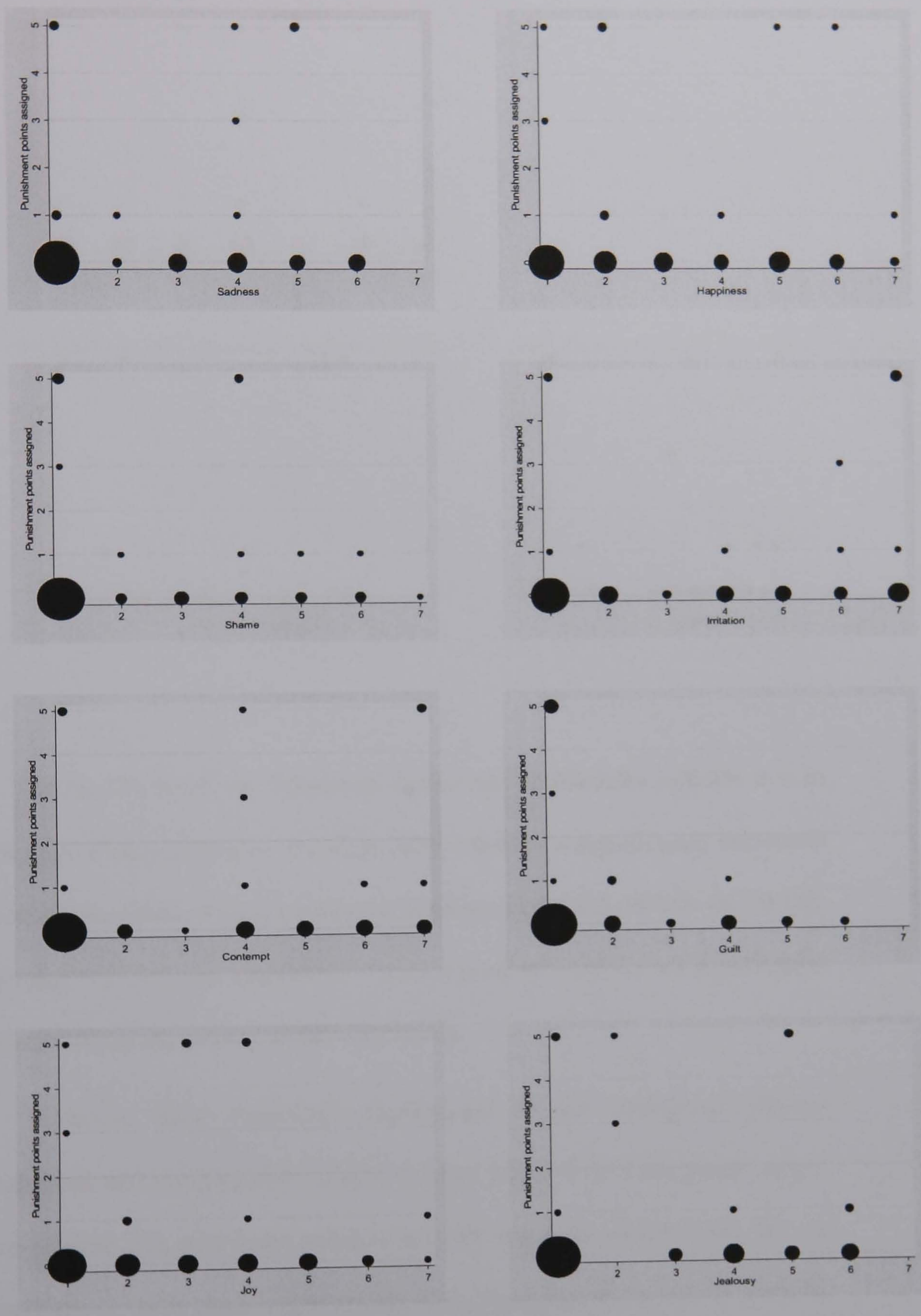
Figure B.3. Relationship between punishment and emotions – Take-P treatment

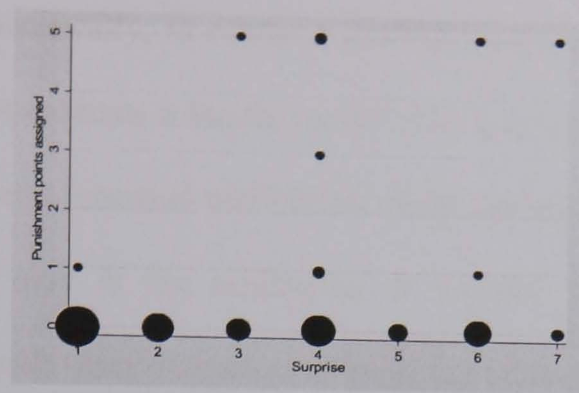


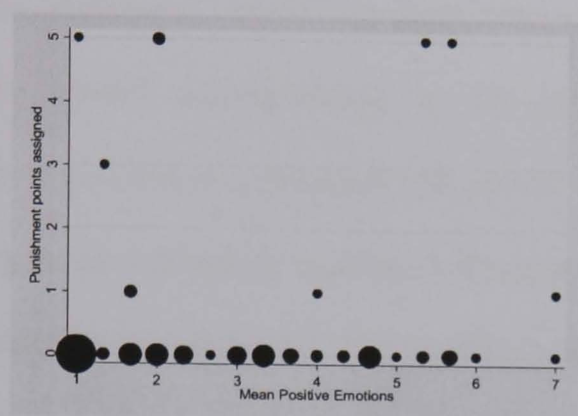


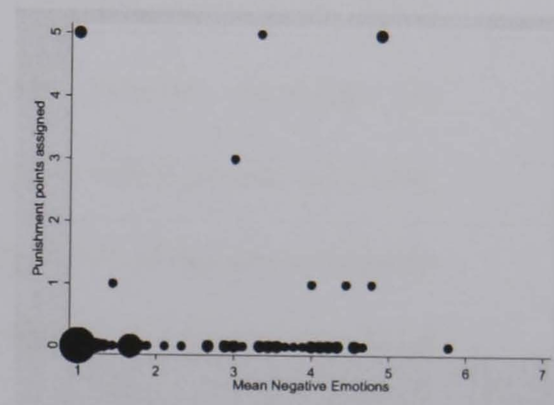





Punishment points assigned
Sadness
Happiness
Shame
Irritation
Contempt
Guilt
Joy
Jealousy







At the 5% level, the Spearman correlation coefficients indicate that in the Give-P only guilt ($\rho = -0.238$; p-value = 0.030) is significantly correlated with the punishment points assigned; whereas, in the Take-P fear ($\rho = 0.276$; p-value = 0.015) and surprise ($\rho = 0.232$; p-value = 0.041) are significantly correlated (at the same conventional level).

However, these correlation coefficients do not distinguish between punishers and non-punishers. From a visual inspection of the panels above, we observe that only some subjects actually decide to punish, with the vast majority of them choosing to refrain from punishment. Hence, it could be the case that the process determining whether to punish or not and, given that a subject punishes, the process determining how much to punish may be different. We therefore model these two processes differently. More specifically, to examine the correlation between punishment and emotions, we estimate a hurdle model. The idea underlying the hurdle formulations is that a binomial probability model governs the decision of whether to punish or not. If the realization is positive, the "hurdle is crossed", and the conditional likelihood of those who punish is governed by a truncated model (see McDowell, 2003). Our regression results are presented in Table B.3.1. The model corresponding to the punishment decision estimates the likelihood that an individual will punish by using a Probit model, where the dependent variable is equal to 1 if a given subject has assigned punishment points and 0 otherwise. The model corresponding to the punishment level estimates the conditional likelihood of an individual who actually punishes and examine whether *his* punishment decision is explained by *his* emotions using as an estimation method a truncated regression. The reason we use a truncated regression is that in this analysis the observations of those who decided to assign zero punishment points are not included. In both models (for the punishment decision and the punishment level), the only independent variable included is the emotion expressed, with each line representing a different regression.

Table B.7. Relationship between punishment assigned and emotions across frames – Regression Results

	Give-P		Take-P	
	Punishment decision	Punishment level	Punishment decision	Punishment level
Warmth	-0.030	-0.139**	0.003	0.023
	(0.018)	(0.057)	(0.029)	(0.317)
Anger	0.018	0.370*	0.027	-0.081
	(0.024)	(0.190)	(0.023)	(0.366)
Fear	0.012	-0.038	0.045*	-0.184
	(0.044)	(0.121)	(0.024)	(0.486)
Envy	-0.041*	-0.053	0.029	-0.272
	(0.027)	(0.259)	(0.027)	(0.395)
Sadness	0.033	0.452***	0.017	0.377
	(0.025)	(0.120)	(0.029)	(0.414)
Happiness	-0.031**	-0.449**	0.013	-0.106
	(0.014)	(0.176)	(0.028)	(0.372)
Shame	-0.065**	-0.000	0.035	-0.541**
	(0.031)	(0.244)	(0.025)	(0.261)
Irritation	0.019	0.268	0.029	0.009
	(0.024)	(0.165)	(0.020)	(0.328)
Contempt	0.007	0.389***	0.030	-0.076
	(0.028)	(0.119)	(0.019)	(0.350)
Guilt	-0.096**	0.179	-0.019	-1.294**
	(0.035)	(0.145)	(0.026)	(0.576)
Joy	-0.035	-0.295***	0.023	-0.215
	(0.022)	(0.107)	(0.024)	(0.340)
Jealousy	-0.005	-0.054	0.026	-0.314
	(0.024)	(0.307)	(0.026)	(0.327)
Surprise	0.023	0.175	0.038*	0.360
	(0.023)	(0.156)	(0.019)	(0.351)
Mean positive emotions	-0.041*	-0.336***	0.014	-0.072
	(0.022)	(0.119)	(0.031)	(0.388)
Mean negative emotions	-0.011	0.589*	0.049	-0.245
	(0.033)	(0.356)	(0.037)	(0.488)

*Notes: Hurdle model estimates. Punishment decision is estimated using a Probit specification and punishment level is estimated using a truncated regression. Standard errors are presented in parentheses (clustered on groups). Results are corrected for heteroskedasticity. * denotes significance at the 10-percent level, ** denotes significance at the 5-percent level, and *** at the 1-percent level.*

Two interesting observations emerge from Table B.3.1. First, the regression coefficients from the first model (punishment decision) suggest that in the Give-P treatment, it is less likely for a subject to punish the more shameful, guilty and happy he is. The likelihood to punish is marginally significant ($p = 0.096$) and negatively correlated with those subjects who feel envy. On the other hand, looking at the Take-P treatment, we find that the likelihood to punish or not is weakly affected by fear and surprise (which is classified as a neutral emotion). Second, Table B.3.1 indicates that, once a subject decides to punish, his decision is also correlated with different emotions between Give-P and Take-P. More specifically, regarding the Give-P treatment, the higher the intensity of the sanction, the less intense all positive emotions (namely, warmth, happiness and joy) are and the more intense the emotions of anger, sadness and contempt are. However, regarding the Take-P treatment, we observe that only shame and guilt are negatively correlated with the decision of how much to punish (the more shameful and guilty a subject is, the less punishment points he is willing to assign).

Our main conclusions can be summarised as follows. First, the emotions that affect punishment seem to be different across frames. Second, for a given frame, the decision of whether to punish or not and the decision of how much to punish seem to be driven by different emotions.

Chapter 4

The behavioural consequences of unfair punishment

4.1 Introduction

A central theme in the experimental literature is the examination of the ability of punishment to sustain high cooperation rates and to regulate behaviour in social dilemma games. Thus far, the vast majority of this literature establishes the importance of punishment as a successful norm enforcement mechanism that fosters cooperation (e.g., Fehr and Gächter, 2000, 2002; Masclet et al., 2003; Noussair and Tucker, 2005; Page et al., 2005; Carpenter, 2007a). Recent experimental research demonstrates that the assignment of punishment mainly depends on three factors: (i) cost and effectiveness of punishment (Anderson and Putterman, 2006; Carpenter, 2007b; Egas and Riedl, 2008; Nikiforakis and Normann, 2008), (ii) second-round punishment opportunities (Cinyabuguma et al., 2006; Denant-Boemont et al., 2007; Nikiforakis, 2008), and (iii) antisocial punishment (Gächter and Herrmann, 2007; Herrmann et al., 2008).[30]

Our present investigation is integrated into this literature by exploring the impact that the unfair assignment of punishment has on people's willingness to punish. But why are we interested in the unfair assignment of punishment? We are motivated by the observation that evidence from public goods experiments with punishment suggests that one condition for punishment to work is that individuals assign it fairly by sanctioning non-cooperators only (e.g., Fehr and Gächter, 2000; Fehr and Gächter, 2002; Masclet et al., 2003; Bochet et al., 2006; Herrmann et al., 2008). These findings are interpreted as evidence that individuals punish non-cooperators because they violate a norm of, or a predisposition towards, reciprocity. We extend this line of investigation by looking at how an unfair environment that violates such norms impacts on subjects' willingness to assign punishment. Previous evidence from bargaining and public good games (see Henrich et al., 2004 and Herrmann et al., 2008, respectively) suggests that people's everyday experiences are reflected in their observed experimental behaviour. However, in these experiments, experiences have been shaped exogenously, outside of the lab. Our experiment explores how the experience of a corrosive environment experienced in the lab affects individuals' expectations of how punishment might work, adding thus to the

[30] For a more analytical discussion on these experiments, see Section 4.2.

existing literature that identifies possible candidates for the assignment of punishment.

To generate an environment where punishment is assigned unfairly, we propose a variant of the standard punishment game (Fehr and Gächter, 2000) to which we refer as the default punishment game. In this game, members of a group participate in a two-stage game. In the first stage, they are engaged in a standard linear public goods game, in which they have to decide how much of their initial endowment they are willing to contribute (see Ledyard, 1995). However, at the beginning of the second stage, we introduce a new element: all group members are exogenously sanctioned by having imposed on them an automatic penalty – the default punishment. This implies that all group members receive a decrease in their monetary income, irrespective of their first stage behaviour. Because the default punishment is unrelated to first stage behaviour, we assume that it is perceived as unfair. During the second stage, after contribution decisions have been anonymously revealed, subjects are given the opportunity to alleviate the exogenous default punishment of others at some cost to themselves.

A noteworthy aspect of the default punishment game is that it resembles the reward game previously studied in the experimental literature (see Sefton et al., 2007). Specifically, helping behaviour (i.e. reward via alleviation of the automatic penalty) in the default punishment game is tantamount to rewarding in the reward game. Therefore, the main substantive difference between the default punishment game and the reward game is the presence of the automatic penalty in the former but not in the latter game. This implies that behaviour from the default punishment game can be affected either by the existence of the default punishment which is allocated to each group member *or* by the opportunity given to group members to reciprocate positively by helping each other. In order to disentangle these effects, we also explore subjects' behaviour in the reward game where they can only reciprocate positively via rewards. In this reward game, there is no automatic penalty, but the reward dimension remains unchanged relative to the default punishment game. Thus, investigation of the default punishment game automatically extends to the investigation of the reward game as well.

Our two main research questions can be formulated as follows: First, is behaviour in the default punishment game different from behaviour in the standard punishment game and in the reward game? Secondly, how do subjects behave (a) when they have previously experienced the default punishment game and (b) when they have previously experienced the reward game?

Our findings suggest that contribution levels in the default punishment game are not significantly different from those in the standard punishment game. The same holds for the contribution levels generated by an environment where there are only rewards. Yet, we document significant time trends in the contribution behaviour of the default punishment game: at the beginning of the game we observe an increase in the levels of contributions, which after the second half of the game begin to decline. We also find that contribution levels in the standard punishment game remain unaffected after a history of the default punishment game and the reward game.

Regarding second stage behaviour, our findings suggest that assigned alleviation and reward are not significantly different. Furthermore, the assignment of punishment is unaffected by the previous experience of a corrosive environment. However, it turns out that subjects' reactions to alleviation, reward and punishment differ. Those subjects who contributed at least as much as the group average decreased their contributions per reward point received, but did not change their contributions when reward takes place via alleviation. However, those who contributed less than the group average decreased their contributions at least weakly significantly in the default punishment game but did not change their contributions in the reward game. Relative to reactions to punishment, we observe that those subjects who contributed less than the group average increased their contributions per punishment point received when they have already experienced the reward game (or when they do not have any previous experience at all), but not when they have experienced the default punishment game, in which case they do not change their contributions significantly.

The remainder of the chapter is organised as follows. Section 4.2 gives a brief literature review on public goods experiments related to our experiment. Section 4.3 presents the design and the procedures of the experiment. Section 4.4 reports the results and Section 4.5 concludes.

4.2 What affects the assignment of punishment: A Literature Review

The linear public goods game with punishment is one of the most frequently used experimental games for studying human cooperation and measuring social preferences. This game, as designed by Fehr and Gächter (2000), consists of two stages: in the first stage, players make their decisions about how much to contribute to the public good and after that, in the second stage, they decide whether, and if so, how much to punish other fellow group members. Subjects can assign up to ten punishment points and punishment is costly both for the punishing and the punished party. Fehr and Gächter compared subjects' behaviour in the presence and absence of punishment opportunities. They also investigated contribution and punishment behaviour in the cases where the composition of the group remained the same throughout the whole experiment (Partner matching protocol) and where it randomly changed from period to period (Stranger matching protocol).

The findings from their experiment are impressive. Contributions dramatically increase when a punishment option is added, contrary to the theoretical predictions, and thus, punishment is found to lower the return on self-interested behaviour and provide a credible and strong incentive to sustain high cooperation. This trend is observed when the experiment with punishment is played either in the first or in the second sequence. With respect to matching, contributions are significantly higher under a Partner than under a Strangers matching protocol.

Another important feature of their experiment is that subjects are prepared to expend their own resources and punish those who contribute less than the average group contribution. Their findings indicate that the average punishment points received by a subject increase with this subject's negative deviation from the other group members' average contribution. Their data suggest a consistent pattern with the norm for conditional cooperation, which is a norm that prescribes cooperation if the other group members cooperate as well.

In another study, Fehr and Gächter (2002) examine the role of punishment in explaining human cooperation when no subject ever met another subject more than once (Perfect Stranger matching protocol). Their findings suggest altruistic punishment as a key motive for the explanation of cooperation. We refer to altruistic punishment as the case where punishment is costly for an individual but does not yield a future material benefit. That is, a punished defector will refrain from free riding and this will benefit other individuals in future interactions.

These two seminal studies by Fehr and Gächter boosted the experimental literature on public good games with punishment, with subsequent experiments having replicated their findings (see, for instance, citations in Section 4.1). From these experiments, it is clear that punishment at least under some circumstances is an effective mechanism that fosters cooperation and remedies the free rider problem. However, the effectiveness of punishment is not immune in all treatment manipulations. Three main factors have been identified to affect its assignment and as a result, subjects' contribution behaviour: (i) cost and effectiveness of punishment, (ii) second-round punishment opportunities, and (iii) antisocial punishment. We comment on these factors in turn.

(i) Cost and effectiveness of punishment. A stylised fact from public good experiments with punishment demonstrates that subjects are willing to demand punishment and spend their own resources to "buy" punishment. The extent to which this demand varies with the price of punishment (i.e., the amount by which a punishment point reduces the punisher's payoff) is explored by Anderson and Putterman (2006) and Carpenter (2007b). Both studies converge to the conclusion that the demand for punishment does obey the law of demand, implying that the higher its price is, the lower is the use of punishment. In a different series of experiments, Egas and Riedl (2008) investigated how subjects' contribution behaviour and punishment attitudes change as the cost and the impact of punishment vary across treatments, contrary to Anderson and Putterman (2006) and Carpenter (2007b), where the price of punishment changes within treatments. It turns out that cooperation is only maintained when the cost for the punisher is low and the impact on the punished is high. Additionally, as the effectiveness of punishment decreases, holding constant its cost, the threshold up to which a deviation in contribution is not punished increases. In a similar study, Nikiforakis and Normann (2008) analyse the effectiveness of punishment (i.e., the amount by which a punishment point reduces the recipient's payoff) by systematically varying it across treatments. With regard to contributions, they find that the more effective punishment is, the higher average contributions are. Turning to punishment behaviour, their findings suggest a positive and concave relationship between punishment assigned and its effectiveness.

(ii) Second-round punishment opportunities. Thus far we have examined punishment in cases where there is only one opportunity to sanction in each period and no possibility to identify individual punishment behaviour. Nikiforakis (2008) added a third stage to the standard public good game with punishment of Fehr and Gächter (2000), in which subjects are given the opportunity to avenge the sanctions received. This second round of sanctions is termed counter-punishment. It turns out that the addition of such a stage affects subjects' punishment attitudes significantly and as a consequence, their contribution behaviour. More specifically, subjects avenge sanctions assigned to them in the previous punishment stage. As a result, punishment of high contributors leads to a decline of contributions over time.

Cinyabuguma et al. (2006) also give subjects the opportunity for a second round of punishment, but their design differs significantly from Nikiforakis' design with respect to the rules governing punishment. In particular, Cinyabuguma et al. allow for a second stage of punishment in every third period of the experiment, in which stage each individual was informed about how much punishment other group members had assigned to below average, above average and average contributors. Then, each individual was allowed to punish, on the basis of others' punishment patterns. Therefore, since individual punishment decisions were not linked to subject identifiers (contrary to Nikiforakis' design where subjects had the possibility to identify the punisher and as a result, to avenge sanctions), motives for counter-punishment were precluded. In their experiment, second order punishment can be motivated by those who failed to punish free riders in the first stage of punishment and/or by those who punish high contributors. These kinds of sanctions are termed sanction enforcement. Cinyabuguma et al. also find that punishment of high contributors in the first stage of punishment decreases significantly, but this was offset by second order punishment of punishers of low contributions. Contrary to Nikiforakis' findings, a second round of sanctions increases contributions and earnings; however, this increase is not significantly different from those treatments where only one stage of punishment is allowed. Based on these two studies and building on their design, Denant-Boemont et al. (2007) conduct an experiment in order to isolate the effects of counter-punishment and sanction enforcement, and measure their magnitude. To do so, they employ treatments in which (i) each subject is fully informed about the sanctioning decisions of other group members; (ii) each subject is informed about how much punishment each other group member assigned to him (this treatment allows for counter-punishment); (iii) each subject is informed about how much punishment was assigned to each group member except for himself (this treatment precludes counter-punishment). Their data confirm the negative effect on contributions from the possibility of counter-punishment as suggested by Nikiforakis (2008) and the positive effect of sanction enforcement on contributions as identified by Cinyabuguma et al. (2006). Comparing both effects, it turns out that the effect of sanction enforcement in increasing contributions is smaller than that of counter-punishment in decreasing contributions, suggesting that the overall effect of a second stage of punishment and a third stage of punishment that allows for unrestricted sanctions is a reduction in contributions.

(iii) Antisocial punishment. Gächter and Herrmann (2007), Herrmann, Thöni and Gächter (2008) and Gächter and Herrmann (2009) examine the extent to which the assignment of punishment differs across subject pools. All experiments document the widespread existence of antisocial punishment (that is, punishment of high contributors) in a number of subject pools around the world. In sum, their findings suggest that the cooperation-enhancing effect of punishment opportunities cannot be taken for granted. The reason for that is the existence of antisocial punishment: the higher antisocial punishment was in a participant pool, the lower was the rate of increase in cooperation when punishment was available relative to when it was not. The explanation of Herrmann et al. (2008) for the presence of antisocial punishment hinges on the intensity of norms of civic cooperation and the rule of law across societies. In particular, the stronger norms of civic cooperation are, the stronger should free riders be punished. In other words, subjects in countries with strong norms of civic cooperation would perceive free riding more unacceptable and thus, would not punish high contributors.[31] On the other hand, if such norms are weak, the restraint on using antisocial punishment is weak too. Similarly, the stronger a country's rule of law is, the higher the perceived confidence in this country's law enforcement institutions is and in consequence, this indicator mirrors the way social norms are enforced.[32] It turns out that both norms of civic cooperation and the rule of law are significantly and negatively correlated with antisocial punishment. Put differently, punishment of cooperators is harsher in subject pools with weak norms of civic cooperation and a weak rule of law.

From the hitherto reviewed literature, we can summarize that previous research suggests that punishment works well only if free riders get punished, the punished free riders accept that their behaviour gets punished, and as a result, free riders increase their contributions accordingly. Yet, possibilities for subjects to retaliate or the existence of antisocial punishment are significant factors that undermine the effectiveness of punishment. In our experiment, we extend this literature and provide a framework where the assignment of punishment is unfair from the beginning of the game. The next section describes how we implemented such a framework and presents our experimental design that addresses our research questions stated earlier.

---

[31] The variable "norms of civic cooperation" is derived from answers of representative residents of a country to questions such as how justified behaviours like tax evasion, benefit fraud and dodging public transport fares are. Norms of civic cooperation are stronger with the condemnation of these behaviours.

[32] The rule of law indicator based on a number of different variables measures the extent to which individuals have confidence in and abide by the rules of society. A strong rule of law implies high levels of confidence of how a law is enforced in a given society.

4.3 Experimental Design and Procedures

4.3.1 Experimental design

Our experimental design consists of three conditions: the "default punishment condition" (D-condition), the "standard punishment condition" (S-condition) and the "reward condition" (R-condition). In all three conditions, subjects were involved in a two-stage game. The first stage of the game was common to all three conditions. Yet, regarding the second stage, some of its features were common, while others varied.

To begin with, the first stage involves a voluntary contributions mechanism game with linear payoffs. For this stage, subjects, being randomly assigned to a four-person group, are privately endowed with 20 tokens and have to decide how many of these to keep for themselves and how many to contribute to a public good (described to subjects as "project"). For each token kept, each subject earns 1 Money Unit; whereas, for each token contributed the return is equal to 0.5 Money Units, resulting in a total of 2 Money Units for the whole group. Subjects make their decisions in private and, at the end of the first stage, they are informed about the sum of the contributions to the public good by the whole group and about their own first stage income.

After the first stage has finished, a second stage begins. In each of our three conditions, the common characteristics of the second stage are as follows. At the beginning of the second stage, subjects can see the profile of contributions of the other three group members. However, no subject could identify the particular contribution of any other subject, since the order of contributions shown in each screenshot randomly changed from period to period. Therefore, subject-specific reputations cannot develop across periods. After being informed about how much each member in their group has contributed in the first stage, each subject can assign adjustment points to other group members. Since subject-specific reputation cannot build up, the possibility that player i assigns adjustment points to player j in period t for contribution decisions made in a previous period from t is ruled out. Subjects could assign between 0 and 2 adjustment points to each other group member. Assignment of adjustment points is always costly, with each adjustment point having a cost of one Money Unit per token. In addition, assigning an adjustment point has an impact on the payoff of the assignee whose absolute magnitude is equal to three. How these adjustment points can be used depends on each condition, but in any case, subjects are given a message/suggestion about how they might use their adjustment points. Finally, note that, at the end of the second stage, subjects were informed about their own cost of assigning adjustment points, the total number of adjustment points assigned to them, and their earnings. No information about the number of adjustment points received by each group member was available.

Depending on the condition, there are three features that vary: (i) the presence or absence of an automatic penalty; (ii) the sign of the impact of an adjustment point on the assignee; and (iii) the message regarding the use of the adjustment points.

In particular, under the D-condition, all subjects incurred an automatic penalty irrespective of their first stage contributions. We refer to this penalty as default punishment and assume it to be unfair since it was unrelated to subjects' past behaviour and was assigned to all group members. In this stage, the message subjects were given was that they can reward other group members by alleviating their automatic penalty, which was costly for the alleviator, but beneficial for the person receiving the adjustment points. Note that if a subject received more alleviation points than the automatic penalty, their income did not increase by this extra amount. In our experiment, the automatic penalty was set equal to 10 Money Units. We did so for two reasons. First, complete alleviation of the automatic penalty was possible only if the majority of the group members decided to assign adjustment points. Recall that each group member can assign up to 2 adjustment points, with each point decreasing the automatic penalty by 3 Money Units. This essentially implies that the automatic penalty is fully alleviated only if two or more group members assign the total amount of points they control. Second, we did not want to create a situation where subjects would be very likely to end up with substantial losses due to the automatic penalty at the end of the experiment. In the case that this could occur, subjects would receive a large lump sum payment to cover possible losses in the D-condition. However, since the lump sum payment has to be kept constant across conditions, an unnecessarily high level of it might affect behaviour, which we wanted to avoid for our purposes.

Contrary to the D-condition, the S-condition does not include any automatic penalty, but subjects are given the opportunity to decrease other group members' income. In other words, the message sent to subjects in this condition was that they can penalise each other group member. Assignment of adjustment points is costly both for the punisher and the recipient of the punishment.

In order to assess the extent to which behaviour differs in a situation where punishment is assigned unfairly and exogenously and a situation in which it is not assigned in such a way, we compare contribution levels between the D- and the S-condition. However, by comparing behaviour between these two conditions, we end up that such a difference could be due to two reasons: either the sign of the impact of an adjustment point on the assignee or the automatic penalty components (that is, the presence/absence of the automatic penalty and the message suggested to subjects).

It is therefore crucial to include a treatment which allows us to disentangle between these two effects. This can be done by the inclusion of the R-condition, in which subjects are given the opportunity to increase the earnings of each other group member. That is, under this condition, subjects were given the message that they can reward their counterparts. Contrary to the S-condition, assignment of adjustment points is costly for the donor, but benefits its recipient. The cost-to-impact ratio was identical to the one used in the D-condition. Clearly, the inclusion of the R-condition allows us to identify the source of a potential difference between the S-condition and the D-condition. If it turns out that the R-condition is the same as the S-condition, the difference between the S-condition and the D-condition will be due to the automatic penalty. However, if it turns out that the R-condition is the same as the D-condition, the difference between the S-condition and the D-condition will be due to the sign of adjustment points. In sum, the differences between our three conditions are given in Table 4.1 below.

Table 4.1 Differences between D-, S- and R-conditions

	S-condition	D-condition	R-condition
Automatic penalty	No	Yes	No
Sign of impact of adjustment points	– 3	+ 3	+ 3
Message	"You can penalise the other group members."	"You can reward the other group members. This can alleviate the automatic penalty of 10 Money Units."	"You can reward the other group members."

To answer our research questions, we implement a within-subjects design under which the same individual participates in two conditions, with each condition being played for 10 periods. In total, we have three experimental sequences: the DS sequence, in which the D-condition is followed by the S-condition; the SD sequence, in which the S-condition is followed by the D-condition; and the RS sequence, in which the R-condition is followed by the S-condition. Each of the three sequences described above was conducted twice, yielding a total of 6 sessions. For both conditions within a sequence, we implemented a Partners' matching protocol meaning that the group composition remained the same across all 20 periods. In the DS sequence 40 subjects participated, resulting in 10 independent observations; while in the SD and RS sequences 36 subjects participated separately, resulting in 9 independent observations per sequence. At the beginning of each sequence, subjects were informed that the session consists of two conditions in order to reduce the possibility for having wrong expectations about the nature of the experiment. However, they were not told what will happen in the second condition. The design information is summarised in Table 4.2.

Table 4.2 Design Overview

Sequence	Periods 1-10	Periods 11-20
DS sequence	D-condition	S-condition
SD sequence	S-condition	D-condition
RS sequence	R-condition	S-condition

First, our experimental design as presented in Table 4.2 enables us to explore whether behaviour in the D and R conditions differs from behaviour in the S-condition. In other words, our design allows us to investigate the pure behavioural effects of each of the three conditions in terms of contributions. For instance, comparing the D-condition of the DS sequence with the S-condition of the SD sequence allows us to assess whether there is any difference in contribution levels between these two environments as subjects experience them both for the first time. This is also the case for the comparison between the R-condition of the RS sequence with the S-condition of the SD sequence. Second, our design allows us also to investigate the robustness of both contribution and punishment behaviour after the experience of either the D-condition or the R-condition. Recall that we chose a cost-to-impact ratio which is equal to 1:3. We did so in the light of previous experimental findings on public goods experiments with punishment (see Nikiforakis and Normann, 2008; Egas and Riedl, 2008) which demonstrate that these ratios can induce high and stable contribution levels. Based on this observation, we can test whether the effectiveness of the S-condition is still maintained after the experience of the D and R conditions. Thus, in order to investigate the robustness of behaviour in the S-condition of the SD sequence, we compare it with the S-condition of the DS sequence and the S-condition of the RS sequence.

4.3.2 Procedures

All sessions took place in April and May 2008 in the Centre for Decision Research and Experimental Economics (CeDEx) lab. Recruitment was conducted via the software ORSEE (Greiner, 2004) at the University of Nottingham using subjects from a university-wide pool of registered students. All conditions were computerised and programmed with the software z-Tree (Fischbacher, 2007). At the beginning of each sequence subjects received instructions for the first condition and at the end of it for the second condition.[33] All participants answered several test questions, concerning the calculation of payoffs for various hypothetical configurations of behaviour. None of the treatments proceeded until every subject had answered these questions correctly. At the end of a sequence, subjects were privately paid according to their accumulated earnings from all 20 periods, using an exchange rate of 0.015p per Money Unit. Average earnings per sequence were as follows: £10.44 for the DS sequence; £9.48 for the SD sequence; £11.65 for the RS sequence. Sessions lasted, on average, 75 minutes, with no session taking longer than 90 minutes.

4.4 Results

4.4.1 First stage behaviour: Does the default punishment game differ from the standard punishment game and the reward game?

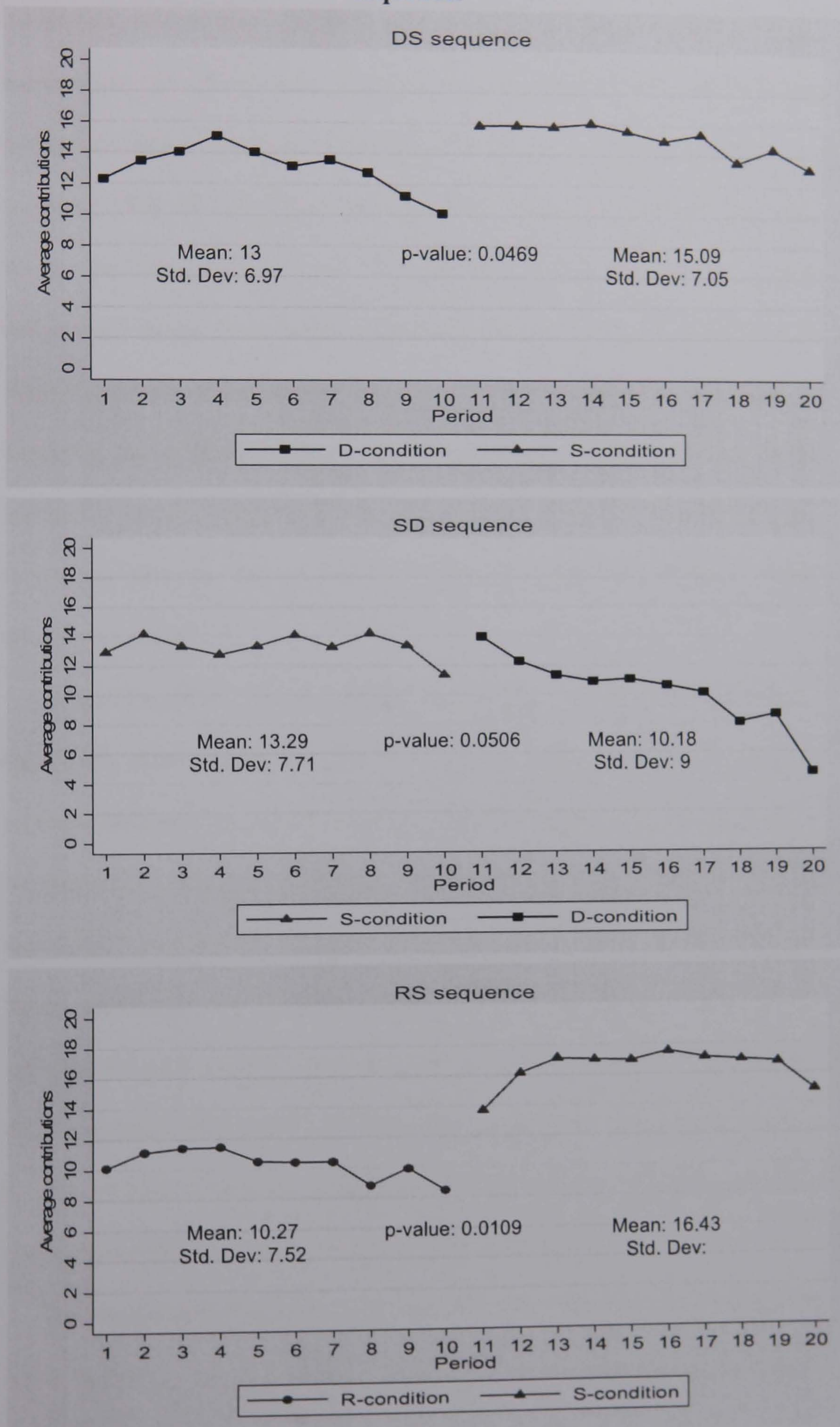
We begin our data analysis by looking at how contribution levels evolved in our three sequences. Data are presented as the amount of tokens

[33] A copy of the instructions is reproduced in Appendix A.

contributed to the group account. Starting with the DS sequence and averaging across all ten periods, we find that subjects' mean contributions were 13 and 15.09 tokens for the D and the S conditions, respectively. Regarding the SD sequence, average contributions across all ten periods were 13.29 tokens for the S-condition and 10.18 tokens for the D-condition; while regarding the RS sequence, the corresponding mean contribution levels were 10.27 tokens for the R-condition and 16.43 tokens for the S-condition.[34] The average contribution pattern as a function of periods is illustrated in Figure 4.1 for each sequence separately. In each panel, the mean contributions across periods for each condition, the corresponding standard deviation and the p-values for within sequence comparisons are also shown.

[34] The average contributions for each matching group in each treatment and condition is given in Appendix B, Table B.1.

Figure 4.1 Time series of average contributions for the DS, SD and RS sequences



To assess the extent to which contributions differ among our three conditions, we contrast the average contributions for the following three comparisons: (i) the S-condition of the SD sequence versus the D-condition of the DS sequence, (ii) the R-condition of the RS sequence versus the S-condition of the SD sequence, and (iii) the R-condition of the RS sequence versus the D-condition of the DS sequence. Recall that if we find a difference between the S and D conditions, then the R-condition will help us understand such a difference. That is, if the R-condition is like the S-condition, the difference between the D versus the S condition is due to the automatic penalty; whereas, if the R-condition is like the D-condition, the difference between the D and S conditions is due to the sign of the adjustment points.

A non-parametric Kruskal-Wallis test suggests significant differences among the three conditions ($\chi^2(3) = 35.763$, p-value = 0.0001). To identify whether there are any differences between the comparisons of our interest, we perform a Wilcoxon rank-sum test. We find that contributions are not statistically significantly different between any of the three comparisons. More specifically, we find that comparing the S-condition (SD sequence) with the D-condition (DS sequence) yields a p-value of 0.775; comparing the R-condition (RS sequence) with the S-condition (SD sequence) yields a p-value of 0.2332; and comparing the R-condition (RS sequence) with the D-condition (DS sequence) yields a p-value of 0.1651.

Our econometric analysis also suggests no difference in contribution levels relative to the three different examined conditions. We estimate three

OLS models, in which the dependent variable is the contribution to the group account and the independent variable comprises four dummies, where each one corresponds to a block of two periods and an additional dummy variable called "condition" that captures the possible difference between our condition comparisons. The dummy variable "condition" equals 1 for the first condition in each comparison. For instance, for the first comparison of Table 4.3, "condition" equals 1 for the S-condition of the SD sequence and 0 for the D-condition of the DS sequence. Our regression results are presented in Table 4.3. Robust standard errors are reported in parentheses.

Table 4.3 Condition differences in contribution levels – Regression Results

	Dependent variable: Contribution		
	S-condition (SD sequence) vs. D-condition (DS sequence) (1)	R-condition (RS sequence) vs. S-condition (SD sequence) (2)	R-condition (RS sequence) vs. D-condition (DS sequence) (3)
Periods 1&2	1.704	1.319	1.836
	(1.171)	(1.066)	(1.164)
Periods 3&4	2.362**	1.472*	3.112***
	(0.895)	(0.775)	(1.016)
Periods 5&6	2.178**	1.319*	2.138**
	(0.796)	(0.656)	(0.863)
Periods 7&8	1.954***	0.896*	1.514**
	(0.573)	(0.426)	(0.672)
Condition	0.289	-3.025	-2.736
	(2.378)	(2.496)	(2.015)
Constant	11.363***	12.290***	11.283***
	(1.529)	(2.141)	(1.547)
Obs.	760	720	760

*Note: OLS estimates with robust standard errors (clustered on independent matching groups) presented in parentheses. For the first comparison (1), the dummy variable "condition" equals 1 for the S-condition in the SD sequence and 0 otherwise. For the second comparison (2), the dummy variable "condition" equals 1 for the R-condition in the RS sequence and 0 otherwise. For the third comparison (3), the dummy variable "condition" equals 1 for the R-condition in the RS sequence and 0 otherwise. Periods 9 & 10 are the baseline. * denotes significance at the 10-percent level, ** denotes significance at the 5-percent level and *** denotes significance at the 1-percent level.*

Regression coefficients from Table 4.3 suggest that subjects do not contribute differently on average in any of the three conditions we examine, when subjects experience them for the first time. In other words, the default punishment game does not yield significantly different contribution levels from the standard punishment game and the reward game.

However, observing Figure 4.1, we notice that the contribution patterns over time seem to be different among the three conditions. We next examine this possibility by exploring whether there are any period effects that influence the observed contribution patterns illustrated in each panel above. To do so, we run three OLS regression models in which the dependent variable is the contribution to the group account. To control for period effects, we include four dummies, where each one corresponds to a block of two periods (as in Table 4.3). The regression results from this analysis are given in Table 4.4. Robust standard errors are also presented in parentheses.

Table 4.4 Period effects in the D-, S- and R-conditions

	Dependent variable: Contribution		
	D-condition (DS sequence) (1)	S-condition (SD sequence) (2)	R-condition (RS sequence) (3)
Periods 1&2	2.175	1.181	1.458
	(1.795)	(1.564)	(1.548)
Periods 3&4	3.875**	0.681	2.264
	(1.482)	(0.627)	(1.418)
Periods 5&6	2.913*	1.361	1.278
	(1.327)	(0.814)	(1.083)
Periods 7&8	2.488**	1.361***	0.431
	(1.026)	(0.406)	(0.746)
Constant	10.712***	12.375***	9.181***
	(1.821)	(2.336)	(1.863)
Obs.	400	360	360

*Note: OLS estimates with robust standard errors (clustered on independent matching groups) presented in parentheses. Periods 9 & 10 are the baseline. * denotes significance at the 10-percent level, ** denotes significance at the 5-percent level and *** denotes significance at the 1-percent level.*

Table 4.4 reveals that period effects are significant for the D-condition, as suggested by the first panel of Figure 4.1. Specifically, we observe that the contribution patterns for the D-condition follow a hump-shaped pattern. Contributions increase up to a certain level, but after that they decline till the end of the game. In contrast, for both other conditions (S- and R-conditions), contribution levels are not strongly affected by periods and show a rather stable pattern across time.

4.4.2 First stage behaviour: Does experience of the default punishment game and the reward game affect behaviour in the standard punishment game?

Having documented these period effects, it is of interest to investigate whether they can affect the effectiveness of the S-condition. We explore this possibility in this subsection. In particular, we assess whether the ability of the S-condition to sustain contributions can survive after subjects have experienced the D-condition and the R-condition. To answer this question, we compare the contribution levels in the S-condition when it is played first versus the S-condition when the D-condition preceded the S-condition. We also compare the contribution levels in the S-condition when it is played first versus the S-condition when the R-condition preceded the S-condition, as the R-condition isolates the reward element incorporated in the D-condition. Performing a non-parametric Wilcoxon rank sum test reveals no significant differences from either comparison.[35]

This finding is corroborated by our formal econometric analysis. Following a similar econometric methodology as previously, our regression results are presented in Table 4.5. Note that in this Table our dependent variable comprise contributions to the group account and the independent variables comprise the four block dummies and the dummy variable

[35] The corresponding p-values are 0.3691 and 0.2004, respectively.

"condition" which equals 1 for the first condition in each comparison as explained earlier.

Table 4.5 Robustness of contribution in the S-condition after a history of the D-condition and the R-condition – Regression Results

	Dependent variable: Contribution	
	S-condition (SD sequence) vs. S-condition (DS sequence) (1)	S-condition (SD sequence) vs. S-condition (RS sequence) (2)
Periods 1&2	1.645	0.139
	(1.052)	(1.187)
Periods 3&4	1.447	0.951
	(0.864)	(0.666)
Periods 5&6	1.408	1.375**
	(0.876)	(0.581)
Periods 7&8	1.007*	1.271***
	(0.558)	(0.322)
Condition	-1.801	-3.133
	(2.521)	(2.216)
Constant	13.991***	15.678***
	(1.908)	(1.291)
Obs.	760	720

*Note: OLS estimates with robust standard errors (clustered on independent matching groups) presented in parentheses. For the first comparison (1), the dummy variable "condition" equals 1 for the S-condition in the SD sequence and 0 otherwise. For the second comparison (2), the dummy variable "condition" equals 1 for the S-condition in the SD sequence and 0 otherwise. Periods 9 & 10 are the baseline. * denotes significance at the 10-percent level, ** denotes significance at the 5-percent level and *** denotes significance at the 1-percent level.*

Table 4.5 suggests that the downward trend on contributions observed after the second half of the default punishment game cannot affect the effectiveness of the punishment game to generate high contribution levels. Not surprisingly, a history of the reward game in which the contribution pattern is rather stable does not also affect the ability of the punishment game to sustain contributions.

However, although contributions are not significantly different across sequences, it is worth exploring whether one condition yields different contribution levels from those in another condition experienced by the same subjects. Our design allows us to make within-subject comparisons since in a given sequence subjects experience two conditions. This implies that we can identify whether there is any impact of a history of one condition on another. Specifically, we are able to explore how the default punishment game impacts on the standard punishment game and vice versa, as well as how the reward game impacts on the standard punishment game.

A Wilcoxon signed-rank test suggests that average contributions in the D-condition are significantly lower than in the S-condition for both the DS ($p = 0.0469$) and the SD sequence ($p = 0.0506$). It also suggests that regarding the RS sequence, average contribution levels are higher in the S-condition than in R-condition ($p = 0.0109$). This evidence is corroborated by our formal econometric analysis presented in Table 4.6. Our dependent variable comprise contributions to the group account and the independent variables comprise four block dummies and the dummy variable "condition" which equals 1 for the first condition in each comparison. For instance, regarding the DS sequence, the variable "condition" equals 1 for the D-condition and 0 for the S-condition.

Table 4.6 Contribution differences within a given sequence

	Dependent variable: Contribution		
	DS sequence	SD sequence	RS sequence
	(1)	(2)	(3)
Periods 1&2	2.119	3.826***	0.278
	(1.513)	(0.775)	(0.937)
Periods 3&4	3.006*	2.569***	1.743**
	(1.423)	(0.751)	(0.673)
Periods 5&6	2.181	2.764***	1.333**
	(1.325)	(0.602)	(0.483)
Periods 7&8	1.588*	1.903**	0.806
	(0.804)	(0.690)	(0.488)
Condition	-2.09**	3.117**	-6.158***
	(0.835)	(1.066)	(1.395)
Constant	13.314	7.963***	15.593***
	(2.168)	(2.151)	(1.185)
Obs.	800	720	720

*Note: OLS estimates with robust standard errors (clustered on independent matching groups) presented in parentheses. For the first comparison, the dummy variable "condition" equals 1 for the D-condition in the DS sequence and 0 otherwise. For the second comparison, the dummy variable "condition" equals 1 for the S-condition in the SD sequence and 0 otherwise. For the third comparison, the dummy variable "condition" equals 1 for the R-condition in the RS sequence and 0 otherwise. Periods 9 & 10 are the baseline. * denotes significance at the 10-percent level, ** denotes significance at the 5-percent level and *** denotes significance at the 1-percent level.*

Regression results from Table 4.6 suggest that contribution levels in the S-condition are always higher than those in either the D-condition or the R-condition when subjects experience both conditions. This implies that when subjects can compare two conditions the tool of punishment can discipline them more efficiently than the corresponding tools of rewards and rewards via alleviation. From Table 4.6, we also observe significant and strong period effects with respect to the SD sequence. It is worth mentioning that these period effects are due to the D-condition of this sequence. Specifically, running a similar regression as in Table 4.4 (see Appendix B, Table B.2), we notice that the coefficients of the block dummies are positive in decreasing order and statistically significant, indicating a clear downward trend in contribution levels for the D-condition. This observation implies that subjects realise the unfairness of the D-condition and react to it by decreasing their contributions faster, when they have already experienced the S-condition, which is arguably a fairer condition, and can compare it with the D-condition. However, when there is no previous experience of another condition and thus, no other means of comparison, subjects start lowering their cooperation rates after the second half of the game.

The results from the analysis of the first stage behaviour can be summarised as follows. First, contributions in the D-condition and the R-condition are not significantly different from those in the S-condition. Second, significant time trends are recorded for the contribution pattern in the D-condition; whereas, these trends are not significant for the contribution patterns in the S- and R-conditions. Third, the ability of the S-condition to sustain contributions survives a history of the D-condition and R-condition. Fourth, within subjects comparisons suggest that contributions in the S-condition are always significantly higher than contributions in any of the DS, SD and RS sequences.

4.4.3 Second stage behaviour: Does assignment of adjustment points in the default punishment game differ from that in the reward game?

In the following we investigate behaviour in the second stage of our conditions. We begin by comparing whether second stage behaviour differs between the D- and the R-condition. In both conditions, assignment of an adjustment point reduces that group member's earnings by 1 Money Unit, but increases the recipient's earnings by 3 Money Units. We can compare the assignment of alleviation and rewards, since the cost-to-impact ratio and the sign of the adjustment points is identical across both conditions and thus, rewarding is tantamount to alleviating. In the following, we investigate the extent to which the D-condition differs from the R-condition by examining subjects' willingness to alleviate and reward, respectively. This is a test for whether the automatic penalty affects helping behaviour (i.e. either alleviation or reward).

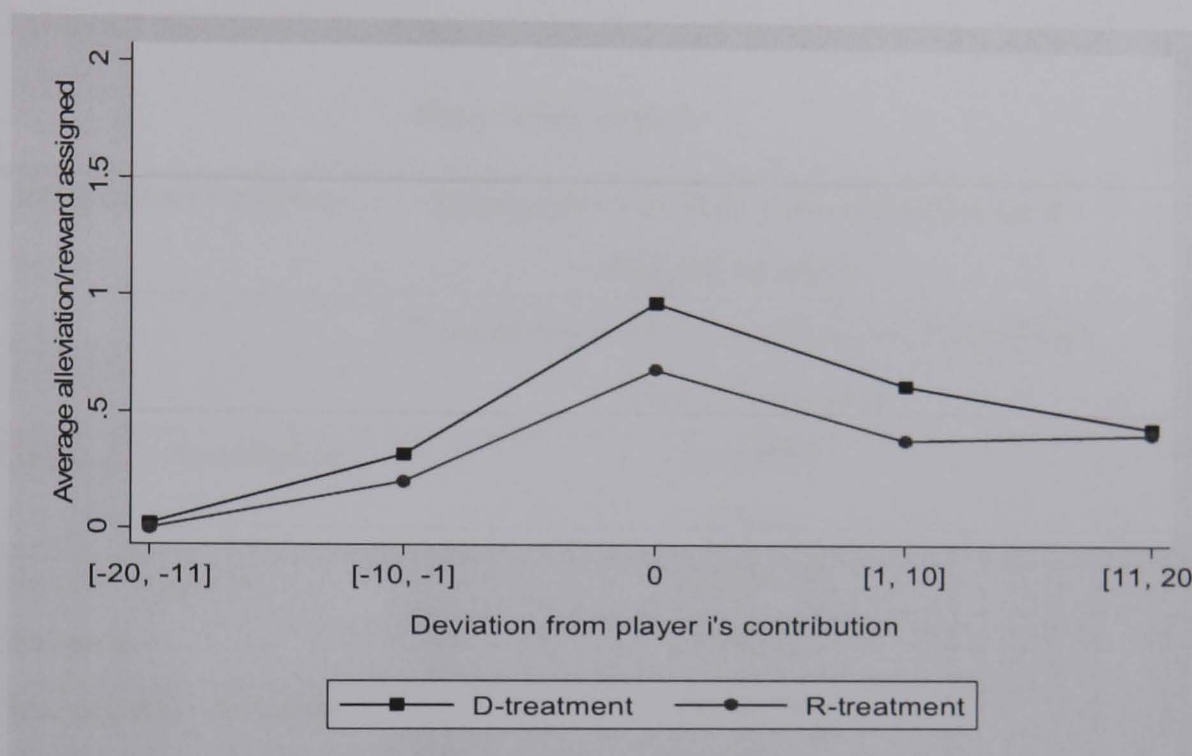
Figure 4.2 provides a graphical illustration of how subjects alleviated and rewarded as a function of the recipient's deviation from the alleviator's/donor's contribution. The vertical axis indicates the average alleviation and reward assigned to a group member. The horizontal axis indicates the deviation in discrete intervals of the recipient's contribution from the contribution of the alleviator/donor. We refer to the solid lines of Figure 4.2 as the "alleviation function" or the "reward function" depending on the condition. From a visual examination of this figure, we observe similar patterns both with respect to the level and the slope of the alleviation and the reward function. Specifically, both functions are positively sloped for negative deviations, suggesting that the less a group member contributes relative to the alleviator/donor, the less alleviation/reward is assigned to him. For non-negative deviations, the slope of the function is negatively sloped as well, indicating that higher contributions from the alleviator's/donor's contribution trigger less alleviation/reward. The intuition behind this negative slope of both functions is that in the positive deviation intervals subjects are low contributors and not willing to incur costs in order to reward high contributors. Consequently, as we move further down to the right of the horizontal axis, low contributors are less and less willing to give up some of their earnings for the sake of costly alleviation and reward of other group members.

Performing a Wilcoxon rank sum test, we test for differences between average assigned alleviation and reward for each of the five deviation intervals. All performed tests fail to reject the null hypothesis that the assigned alleviation is equal to the assigned reward across conditions (p-values > 0.142).[36] Note that adapting the deviation intervals that are mentioned in Fehr and Gächter (2000, pp. 991) yields similar conclusions. It turns out that alleviation and reward are not statistically different from each

[36] The corresponding p-values for each deviation interval, starting from the largest negative deviation interval and ending to the largest positive deviation interval, are: 0.2226, 0.1416, 0.6830, 0.1208 and 0.7322, respectively.

other in all deviation intervals, but for the one in which the recipient positively deviates from the donor's contribution between 8 (not inclusive) and 14 (inclusive) tokens.

Figure 4.2 Average alleviation/reward assigned



For a formal analysis, we estimate an ordered probit regression model, supporting the patterns observed in the above figure. As dependent variable, we include the expenditures on alleviation or reward by player i. The independent variables comprise the recipient's contribution, the absolute negative (nonnegative) deviation from the alleviator's/donor's contribution, and the dummy variable "condition" to capture level differences between the two conditions. The variable "absolute negative deviation" is the absolute value of the actual deviation of subject j's contribution from subject i's contribution, when subject j's contribution is below the subject

i's contribution; and zero otherwise. The variable "non-negative deviation" is constructed in an analogous way. The variable "condition" equals to 1 for the D-condition of the DS sequence and 0 otherwise. The regression results are reported in Table 4.7.

Table 4.7 Assigned alleviation for negative and nonnegative deviations – Regression Results

Independent variables	**Dependent Variable: Alleviation/Reward assigned by player i**
	D-condition in the DS sequence vs. R-condition in the RS sequence
Player j's contribution	0.128***
	(0.012)
Absolute negative deviation	-0.080***
	(0.024)
Non-negative deviation	-0.059***
	(0.016)
Condition	0.141
	(0.193)
Condition × Absolute negative deviation	0.023
	(0.031)
Condition × Non-negative deviation	-0.002
	(0.020)
Obs.	2,280

*Notes: Ordered probit estimates. Standard errors presented in parentheses (clustered on independent matching groups). * denotes significance at the 10-percent level, ** at the 5-percent level, and *** at the 1-percent level.*

The regression coefficients corroborate our findings from the non-parametric Wilcoxon test, failing to reject the null hypothesis that alleviation and reward as a function of both nonnegative and negative deviations are the same. Both level and slope differences are insignificant at conventional levels. As illustrated in Figure 4.2, we find econometric evidence that both alleviation and reward functions have a positive slope for the negative deviation interval (the sign of the "absolute negative deviation" is negative) and a negative slope for the non-negative deviation interval (the sign of the "non-negative deviation" is negative as well). Additionally, assigned alleviation and reward are both found to be higher the more the recipient contributes. In sum, the main finding from Table 4.7 implies that the assignment of alleviation is the same as rewarding.

However, from our first stage behaviour analysis, we have recorded significant period effects with respect to how subjects contributed when unfair punishment is present. In the following, we investigate whether experiencing the automatic penalty has an impact on the average contribution level as a function of received alleviation or reward. In particular, we examine how a subject reacted who got alleviated/rewarded for a contribution above (or equal to) the group average and for a contribution below the group average. In our econometric model, the dependent variable is the change in the recipient's contribution between period t and period t+1. The independent variables are the amount of alleviation/reward received from the other three group members in period t and variables measuring the time trend ("Period" and "Final period"). We estimate this model either for the cases in which a group member contributed less than the other group members or at least as much as the other group members. Table 4.8 and Table 4.9, respectively, report our regression results for each case separately.

Table 4.8 Reactions to alleviation/reward – Regression results (1)

	Change in contribution if a subject contributed *less* than the other group members	
	D-condition of the DS sequence	R-condition of the RS sequence
Received alleviation/reward	-1.395* (0.748)	-0.983 (0.705)
Period	-0.342 (0.280)	0.022 (0.170)
Final period	1.921 (2.500)	-0.581 (2.941)
Constant	5.785** (1.893)	2.291** (0.933)
Obs.	138	140

*Note: OLS estimates with robust standard errors (clustered on independent matching groups) presented in parentheses. * denotes significance at the 10-percent level, ** denotes significance at the 5-percent level and *** denotes significance at the 1-percent level.*

Table 4.8 suggests that subjects who have contributed less than the group do not change their contributions significantly with the received reward, whereas they lowered their contributions at least weakly significantly per adjustment point alleviated. Looking at those who have contributed more than (or equal to) the other three group members, we also observe different reactions across conditions. Table 4.9 below reports the results of OLS estimations for these cases.

Table 4.9 Reactions to alleviation/reward – Regression results (2)

	Change in contribution if a subject contributed *more* than the other group members or equal	
	D-condition of the DS sequence	R-condition of the RS sequence
Received alleviation/reward	0.484 (0.300)	-0.578*** (0.171)
Period	-0.304* (0.152)	-0.246 (0.203)
Final period	-0.339 (1.534)	-0.723 (1.396)
Constant	-1.591 (1.177)	-0.646 (1.012)
Obs.	222	184

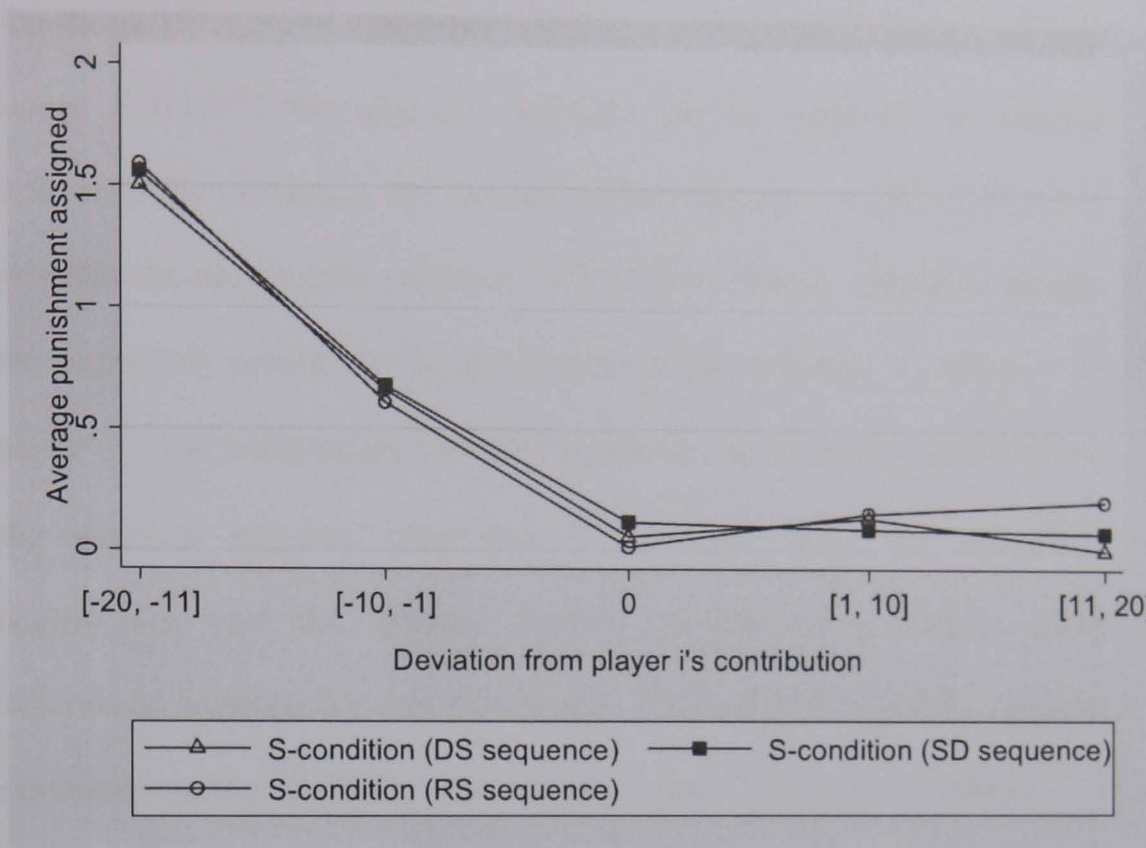
*Note: OLS estimates with robust standard errors (clustered on independent matching groups) presented in parentheses. * denotes significance at the 10-percent level, ** denotes significance at the 5-percent level and *** denotes significance at the 1-percent level.*

Two observations emerge from Table 4.9. First, for the D-condition of the DS sequence, those subjects who have contributed at least as much as the group average do not change significantly their contributions per adjustment point received. Second, and contrary to the D-condition, we find that for the R-condition of the RS sequence, those subjects who have contributed at least as much as the group average significantly decreased their contributions per adjustment point received.

4.4.4 Second stage behaviour: Does experience of the default punishment game and the reward game affect behaviour in the standard punishment game?

Since subjects indicate different reactions with respect to received alleviation and reward, it is of interest to investigate whether and if so, how the D-condition and the R-condition impact on the punishment assigned. In this subsection, we contrast punishment behaviour in the S-condition of the SD condition versus the S-condition of the DS treatment. Recall that helping behaviour in the D-condition is tantamount to rewarding behaviour in the D-condition. We thus isolate the effects of the rewarding element (in the D-condition) by examining subjects' willingness to punish after they experience the R-condition where there are only rewards. In sum, the comparisons that we are examining are the following: (i) the S-condition of the SD sequence versus the S-condition of the DS sequence; and (ii) the S-condition of the SD sequence versus the S-condition of the RS sequence. Figure 4.3 below shows the assignment of punishment in each of the three sequences.

Figure 4.3 Punishment assigned in each sequence



In the above figure, the horizontal axis indicates the deviation in discrete intervals of the recipient's contribution from the contribution of the punisher. The vertical axis shows the average punishment assigned. We refer to the solid lines of Figure 4.3 as the "punishment function", which gives the average punishment points assigned by the punisher as a function of the recipient's deviation from the punisher's contribution. As expected, looking at the negative deviation intervals, the punishment function is negatively sloped, indicating that the more an individual negatively deviates from the punisher's contribution the higher the punishment assigned to him or her. We also observe some punishment targeted at high contributors. In the following, we examine whether punishment assigned differs for comparisons (i) and (ii) mentioned earlier.

We find that a non-parametric Wilcoxon rank sum test yields no significant differences either for negative or non-negative deviations (p-values > 0.13)[37]. We also test formally for the existence of possible differences by estimating two ordered probit regression models. Table 4.10 provides the econometric evidence. In this table, for all regression models the dependent variable is the punishment points assigned by player i to player j. The independent variables comprise the recipient's contribution, the absolute negative (nonnegative) deviation from the punisher's contribution, and the dummy variable "condition" to capture level differences between the two conditions. The variable "absolute negative deviation" is the absolute value of the actual deviation of subject j's contribution from subject i's contribution, when subject j's contribution is below the subject i's contribution; and zero otherwise. The variable "non-negative deviation" is constructed in an analogous way. The variable "condition" equals to 1 for the first condition of each comparison and 0 otherwise.[38] To capture slope differences between conditions relative to negative and non-negative deviations, we additionally include two interaction terms for each deviation interval separately. The interaction variable "condition × absolute negative deviation" equals to the product of the dummy variable "condition" and the variable "absolute negative

[37] More specifically, for the comparison between the S-condition of the SD sequence versus S-condition of the DS sequence, the corresponding p-values were 0.9648 for the negative deviations and 0.1381 for the non-negative deviations. Regarding the comparison between the S-condition of the SD sequence versus S-condition of the RS sequence, the corresponding p-values were 0.8253 and 0.2004 for the negative deviations and for the non-negative deviations, respectively.

[38] For instance, for the comparison between the S-condition of the SD sequence and the S-condition of the DS sequence, the variable "condition" takes the value 1 for the former condition and 0 for the latter one.

deviation"; and the variable "condition × non-negative deviation" is constructed analogously. The regression results are shown in the table below.

Table 4.10 Differences in punishment assigned after a history of D- and R-condition – Regression Results

Independent variables	**Dependent Variable: Punishment assigned by player i**	
	S-condition in the SD sequence vs. S-condition in the DS sequence	S-condition in the SD sequence vs. S-condition in the RS sequence
Player j's contribution	-0.032	-0.032
	(0.012)	(0.021)
Absolute negative deviation	0.125***	0.125***
	(0.025)	(0.027)
Nonnegative deviation	-0.025	-0.025
	(0.021)	(0.021)
Condition	0.101	-0.184
	(0.233)	(0.273)
Condition × absolute negative deviation	0.006	0.051
	(0.026)	(0.033)
Condition × non-negative deviation	-0.004	0.061**
	(0.025)	(0.030)
Obs.	2,280	2,280

*Notes: Ordered probit estimates. Standard errors presented in parentheses (clustered on independent matching groups). For the comparison between the S-condition of the SD sequence and the S-condition of the DS sequence, the variable "condition" takes the value 1 for the former condition and 0 for the latter one. For the comparison between the S-condition of the SD sequence and the S-condition of the RS sequence, the variable "condition" takes the value 1 for the former condition and 0 for the latter one. * denotes significance at the 10-percent level, ** at the 5-percent level, and *** at the 1-percent level.*

Starting from the first comparison, we observe that the experience of the D-condition does not have any impact on the assigned punishment either with respect to negative or non-negative deviations. Both the variable "condition" and the corresponding interaction terms are insignificant indicating that subjects' willingness to punish is insensitive to the experience of an unfair environment (that is, the D-condition). Turning to the second comparison of Table 4.10 we also reach the same conclusion. A history of an environment where there are only rewards (and not automatic penalty) does not affect the assignment of punishment. Notice also that we observe some antisocial punishment in the S-condition of the RS sequence, as the interaction term "condition × non-negative deviation" turns out to be statistically significant, but it is not economically significant as identified in other subject pools (see e.g. Herrmann et al, 2008).

The message from Table 4.10 is that the assignment of punishment is unaffected from either a history of the D- or the R-condition. However, from our analysis in Section 4.4.1, we have recorded significant period effects with respect to how subjects contributed when unfair punishment is present. In the following analysis, we examine how a subject reacted who got punished for a contribution above (or equal to) the group average and for a contribution below the group average. In our econometric model, the dependent variable is the change in the recipient's contribution between period t and period t+1. The independent variables are the amount of punishment received in period t and variables measuring the time trend ("Period" and "Final period"). We estimate this model either for the cases in which a group member contributed less than the other group members or at least as much as the other group members. Table 4.11 and Table 4.12, respectively, report our regression results for each case separately.

Table 4.11 Reactions to punishment – Regression results (1)

	Change in contribution if a subject contributed *less* than the other group members		
	S-condition of the DS sequence	S-condition of the SD sequence	S-condition of the RS sequence
Received punishment	0.288 (0.359)	1.341** (0.615)	1.179* (0.619)
Period	0.124 (0.185)	0.198 (0.220)	-0.617** (0.239)
Final period	-3.010** (1.165)	-3.557 (3.331)	2.751 (2.241)
Constant	1.376 (1.231)	-1.438 (1.340)	2.765 (1.560)
Obs.	104	109	82

*Note: OLS estimates with robust standard errors (clustered on independent matching groups) presented in parentheses. * denotes significance at the 10-percent level, ** denotes significance at the 5-percent level and *** denotes significance at the 1-percent level.*

Table 4.11 suggests that a history of either the D-condition or the R-condition produces different reactions with respect to punishment when a subject contributes less than the other three group members. Specifically, when there is a history of the R-condition the estimated coefficient of

"Received punishment" is statistically positive, indicating that subjects who contributed less than the average contribution of the other three group members increased their contributions per punishment point received. This is exactly the case when there is no previous history of any environment. Yet, a history of the D-condition renders the relationship between change in contributions and punishment received insignificant, implying that subjects with such an experience did not change their contributions significantly in the S-condition that was followed. We next turn to a situation where a subject contributes more than the group average.

Table 4.12 Reactions to punishment – Regression results (2)

	Change in contribution if a subject contributed *more* than the other group members or equal		
	S-condition of the DS sequence	S-condition of the SD sequence	S-condition of the RS sequence
Received punishment	-0.055 (0.370)	0.680 (0.382)	-0.529 (0.849)
Period	-0.157 (0.212)	-0.040 (0.105)	-0.022 (0.067)
Final period	-0.085 (1.335)	-1.079 (1.019)	-2.404 (1.352)
Constant	-0.448 (0.967)	-1.456 (0.894)	0.270 (0.498)
Obs.	256	215	242

Note: OLS estimates with robust standard errors (clustered on independent matching groups) presented in parentheses. * denotes significance at the 10-percent level, ** denotes significance at the 5-percent level and *** denotes significance at the 1-percent level.

Findings from Table 4.12 suggest that similar reactions across conditions. In particular, it turns out that in any of the three comparisons subjects did not change their contributions per punishment point received when they contribute at least as much as the group average.

Our analysis regarding punishment behaviour suggests that the assignment of punishment is not affected by a previous experience of either the D- or the R-condition. However, the reactions to punishment depend on previous history. It turns out that those subjects, who contributed less than the group average, increase their contributions if they experience a game where there are only rewards or if they do not have any experience. Yet, after the experience of a game with unfair punishment the correlation between received punishment and change in contributions becomes insignificant.

Summarising our findings from this section we conclude that the assignment of alleviation and reward as a function of the deviation from the alleviator's/donor's contribution is not significantly different between the D- and the R-condition. Additionally, the assignment of punishment is unaffected after a history either of the D-condition or the R-condition. However, we observe significant differences with respect to how subjects react to alleviation, reward and punishment. In particular, those subjects who contributed at least as much as the group average decreased their contributions per adjustment point received in the R-condition, but did not change their contributions in the D-condition. Regarding those who contributed less than the group average, we find that they decreased their contributions in the D-condition, but not in the R-condition. Reactions to punishment also differ depending on previous history. Our findings suggest that those subjects who contributed less than the group average increased their contributions after a history of the R-condition (or after no history at all), but not after a history of the D-condition.

4.5 Conclusions

Previous research on public good games with punishment suggests that punishment is effective when subjects assign it fairly by sanctioning non-cooperators. In this chapter, we report an experiment in which punishment is assigned unfairly. Specifically, in our experiment, punishment is meted out exogenously to all members (default punishment), irrespective of their behaviour. We tested whether an unfair environment with default punishment generates a difference relative to the standard punishment game, both in terms of contribution behaviour and punishment attitudes. Notice that our default punishment game has also a reward element incorporated in its structure because subjects can alleviate the exogenously assigned punishment. As an auxiliary condition, we therefore included a condition in which group members are only given the opportunity to reward their fellow group members, without having been exogenously punished.

Our findings suggest that contributions do not differ significantly among the default punishment game, the standard punishment game and the reward game. Yet, it is worth noting that the contribution pattern in the default punishment game is characterised by strong period effects, which are not present neither in the standard punishment game nor in the reward game. Specifically, after the second half of the game contribution levels decline over time. In addition, a history of an unfair environment does not affect the ability of the standard punishment game to sustain high levels of contributions. Interestingly, a history of the standard punishment game causes contribution levels in the default punishment game to collapse from the very beginning of the game.

Turning to second stage behaviour, we find that assigned alleviation and reward are not significantly different. We also find that the assignment of punishment is unaffected by the previous experience of a corrosive environment. However, it turns out that reactions to alleviation, reward and punishment differ. More specifically, those subjects who contributed at least as much as the group average decreased their contributions per adjustment point received in the reward game, but did not change their contributions in the default punishment game. Regarding those who contributed less than the group average, we find that they decreased their contributions in the default punishment game but did not change their contributions in the reward game. Relative to reactions to punishment, we observe that those subjects who contributed less than the group average increase their contributions per punishment point received when they have already experienced the reward game (or when they do not have any previous experience at all), but not when they have experienced the default punishment game, in which case they do not change their contributions significantly.

Appendix A. Technical Appendix

Appendix A.1: Instructions

{Note: There are three experiments in total. We refer to each experiment as "D-condition", "S-condition" and "R-condition". Each experiment consists of two stages. Instructions for the first stage in all experiments are identical. Instructions for the second stage in each experiment are presented separately. The control questionnaire was identical for all experiments.}

Instructions

You are now about to take part in an experimental economics session financed by the University of Nottingham. If you read the following instructions carefully, you can, depending on the decisions that you and other participants make, earn a considerable amount of money. It is therefore very important that you read these instructions with care.

These instructions are solely for your private use. **It is prohibited to communicate with the other participants during the session.** Should you have any questions, please ask us.

During the session we will not speak in terms of Pounds, but of Money Units. Your entire earnings will, thus, be calculated in Money Units. At the end of the session the total amount of Money Units you have earned will be converted to Pounds at the following rate:

1 Money Unit = 0.015 Pounds

At the end of the session your entire earnings will be paid to you **in cash**.

During this session, you will take part in two experiments. You will now undertake the first experiment. You will learn about the second experiment at the beginning of that experiment, where you will receive new instructions.

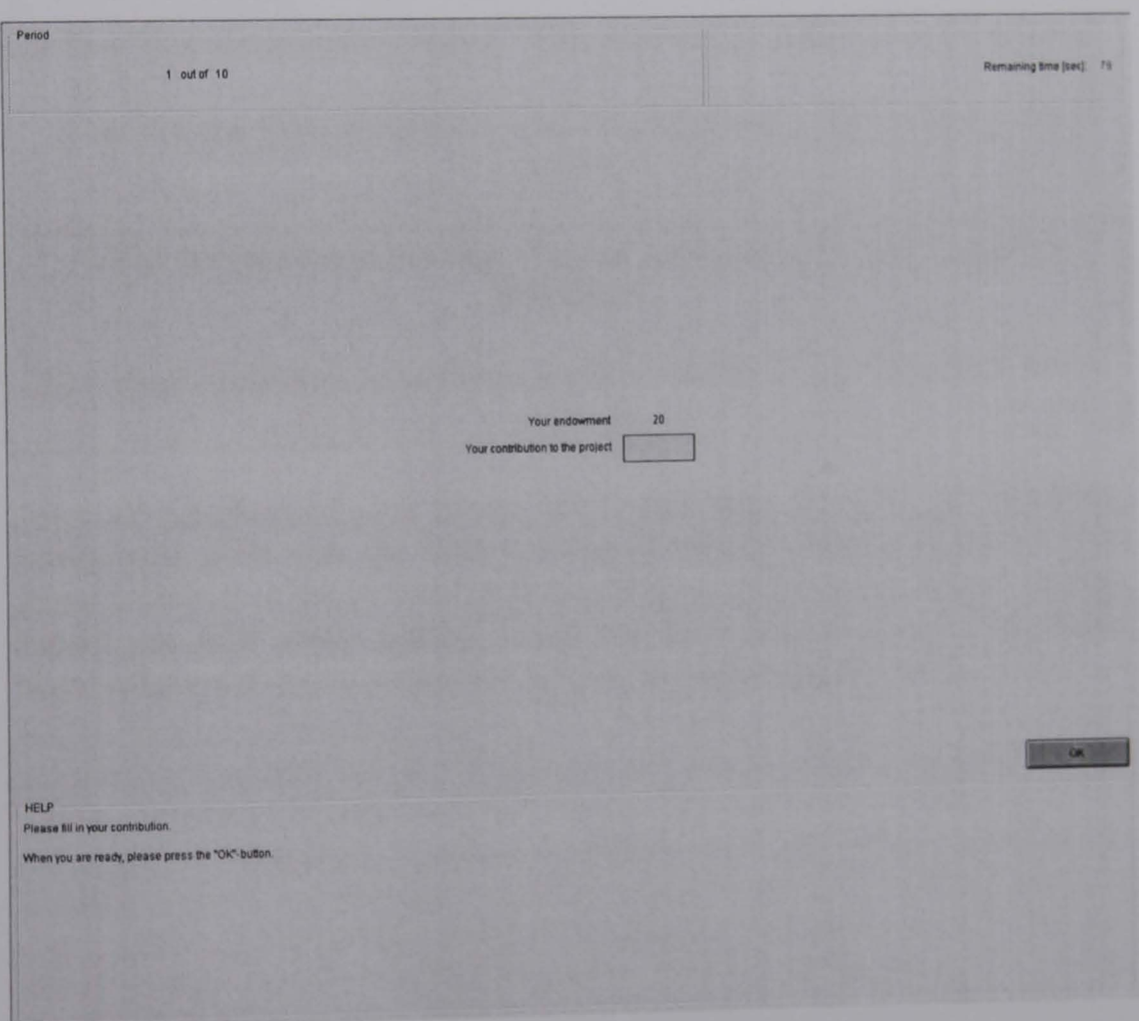
FIRST EXPERIMENT

This experiment has ten periods. In each period the participants are divided into groups of four. You will therefore be in a group with 3 other participants. **The composition of the groups will remain the same throughout the experiment.** Each period has two stages, which are described below.

The first stage

At the beginning of each period each participant receives **20 tokens**. We call this his or her **endowment**. Your task is to decide how to use your endowment. You have to decide how many of the 20 tokens you want to contribute to a **project** and how many of them to keep for yourself. The consequences of your decision are explained in detail below.

At the beginning of each period the following input-screen for the first stage will appear:



The period identifier appears in the top left corner of the screen. In the top right corner you can see how many more seconds remain for you to decide on the distribution of your tokens. You will have 90 seconds to decide in the first period and 30 seconds in the remaining periods. Your decision must be made before the time displayed is 0 seconds.

Your endowment in each period is 20 tokens. You have to decide how many tokens you want to contribute to the project by typing a number between 0 and 20 in the input field. This field can be reached by clicking it with the mouse. By deciding how many tokens to contribute to the project, you automatically decide how many tokens you keep for yourself: This is **(20 – your contribution to the project)** tokens. After entering your contribution you must press the O.K. button. Once you have done this, your decision can no longer be revised.

Your **income** consists of two parts:

(1) The tokens which you have kept for yourself ("Income from retained tokens") whereby

1 token = 1 Money Unit.

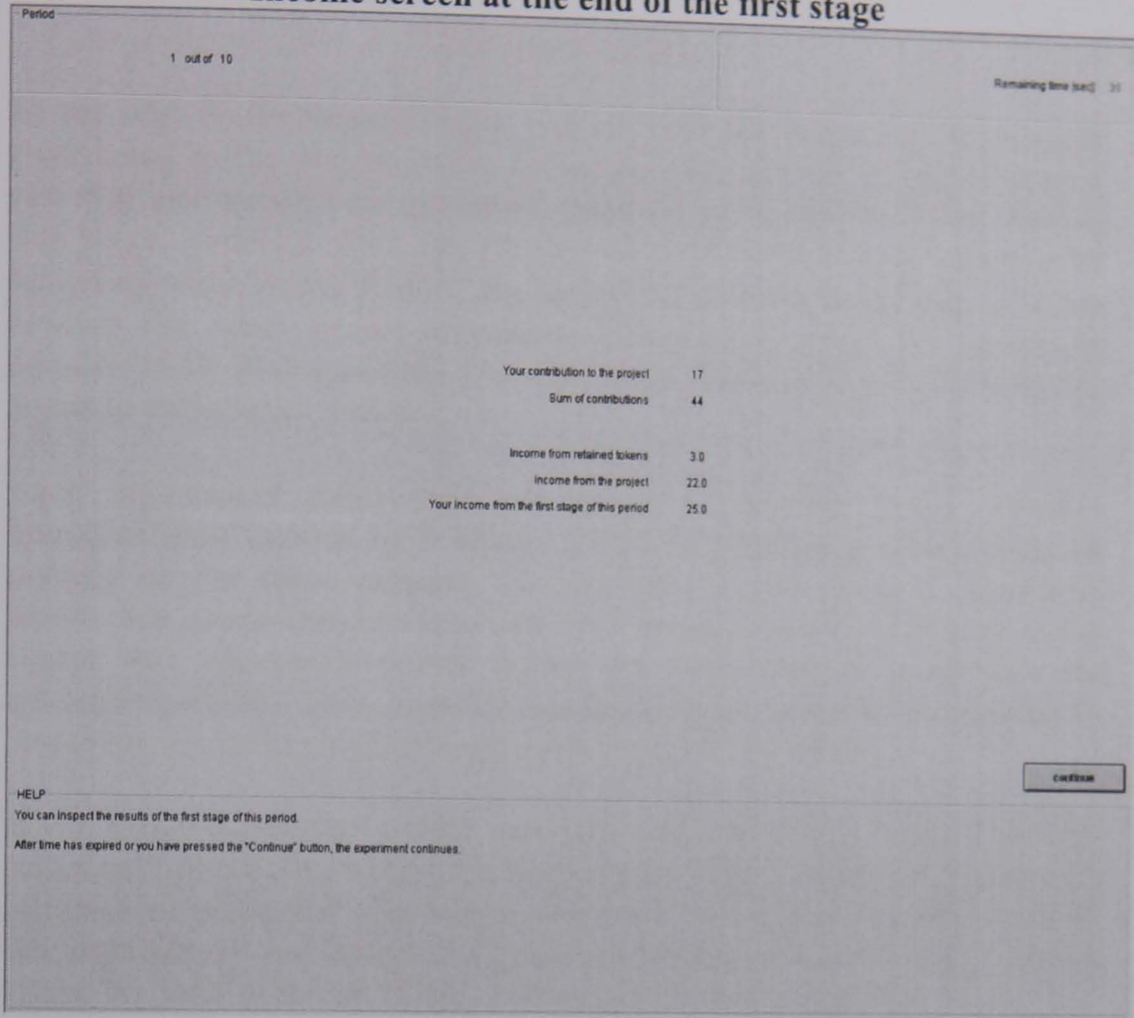
(2) The "Income from the project". This income is calculated as follows:

Your income from the project = 0.5 *times* the total contributions to the project.

Your income from the first stage of a period in Money Units is therefore: **(20 – your contribution to the project) + 0.5*(total contributions to the project)**

After all members of your group have made their decision the following screen will show you the total amount of tokens contributed by all four group members to the project (including your contribution). This screen also shows you how many Money Units you have earned at the first stage. Numbers shown in this screenshot are for example purposes only.

Income screen at the end of the first stage



The income of each group member from the project is calculated in the same way, i.e., each group member receives the same income from the project. Assume, for example, that the sum of the contributions of all group members is 44 tokens. In this case each member of the group receives an income from the project of: 0.5*44 = 22 Money Units.

For each token which you keep for yourself you earn an income of 1 Money Unit. Supposing you contributed this token to the project instead, then the total contributions to the project would rise by one token. Your income from the project would rise by 0.5*1=0.5 Money Units. However the income of the other group members would also rise by 0.5 Money Units each, so that the total income of the group from the project would rise by 2 Money Units. Your contribution to the project therefore also raises the income of the other group members. On the other hand you earn an income for each token contributed by the other members to the project. For each token contributed by any member you earn 0.5*1=0.5 Money Units.

To view the income screen at the end of the first stage, you have 45 seconds in the first period and 20 seconds in the remaining periods. If you are finished with it before the time is up, please press the continue button. As long as you have inspected the results of the first stage of a period, the second stage starts.

{Note: Second stage for the D-condition}

The second stage

At the start of the second stage, you see how much each group member contributed to the project in the first stage. **Regardless of contributions, you will also receive an automatic penalty of 10 Money Units. During this stage, you can alter the income of each other group member by assigning adjustment points. By assigning adjustment points, you can reward the other group members. This can alleviate the automatic penalty of 10 Money Units.** You can assign between 0 and 2 adjustment points to each group member.

Each adjustment point that you assign to another group member increases their income by 3 Money Units, so alleviating their automatic penalty by the same amount. For example, if you assign 2 adjustment points, this group member's income will be increased by 6 Money Units, except that adjustment points cannot do more than fully alleviate the automatic penalty. Thus, a group member's income cannot be increased by more than ten, through adjustment points assigned by others.

If you assign adjustment points, you have costs in Money Units. The more adjustment points you assign, the higher your costs. Specifically, **for each adjustment point that you assign, there is a cost to you of 1 Money Unit.** For example, if you assign 2 adjustment points, this costs you 2 Money Units. We refer to this as "Cost of adjustment points assigned by you".

Just as you can alleviate other players automatic penalty by assigning adjustment points to them, so they can also alleviate your automatic penalty by the same method. We refer to this as "Number of adjustment points assigned to you".

Your total income from the two stages is therefore calculated as follows:

Total income (in Money Units) at the end of the second stage = Period income =

= Income from the first stage
– Automatic penalty
– Cost of adjustment points assigned by you
+ 3*(Number of adjustment points assigned to you)

if the impact of the adjustment points assigned to you is less than the automatic penalty;

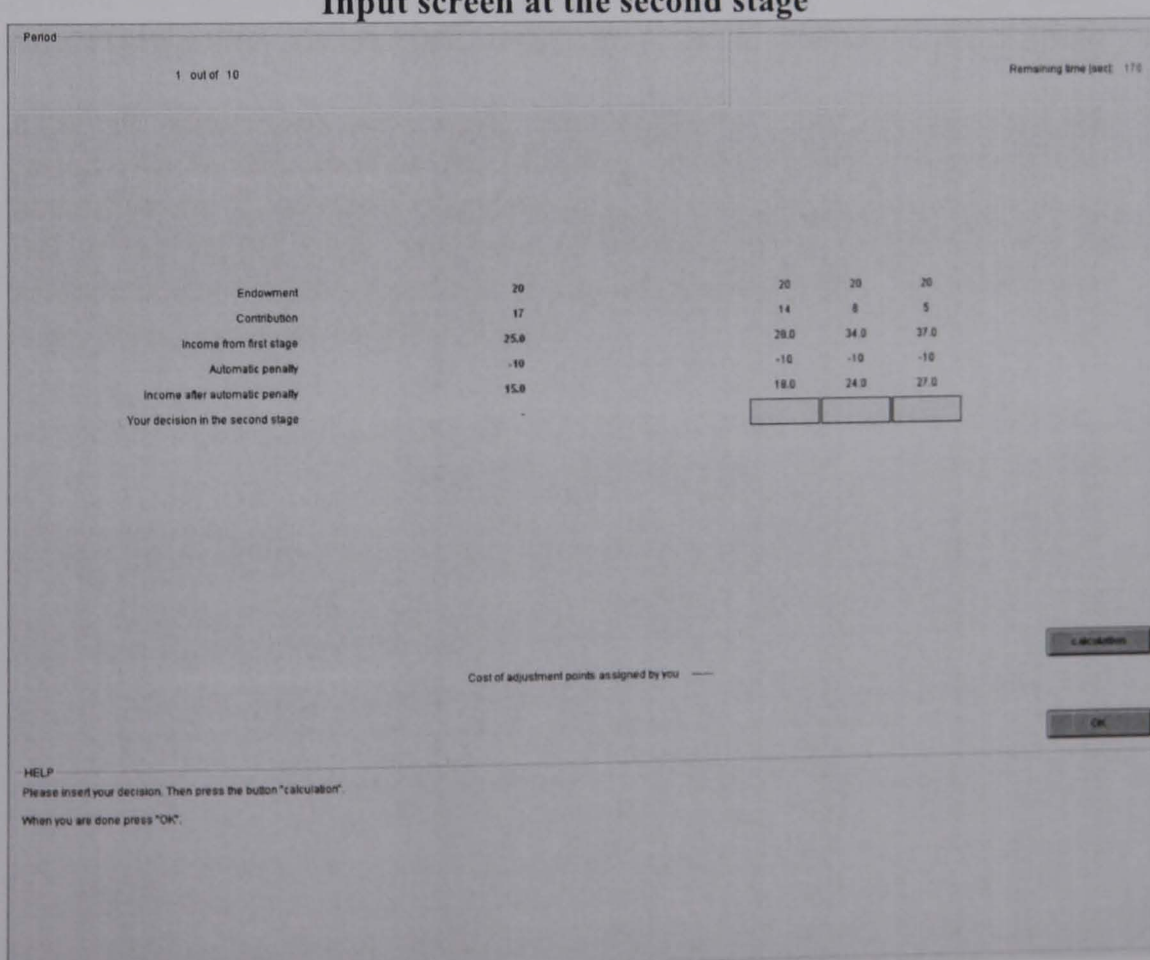
OR

= Income from the first stage
– Cost of adjustment points assigned by you

if the impact of the adjustment points assigned to you is greater than the automatic penalty.

The way with which you can assign adjustment points is apparent from the input screen at the second stage. Numbers shown in this screenshot are for example purposes only.

Input screen at the second stage



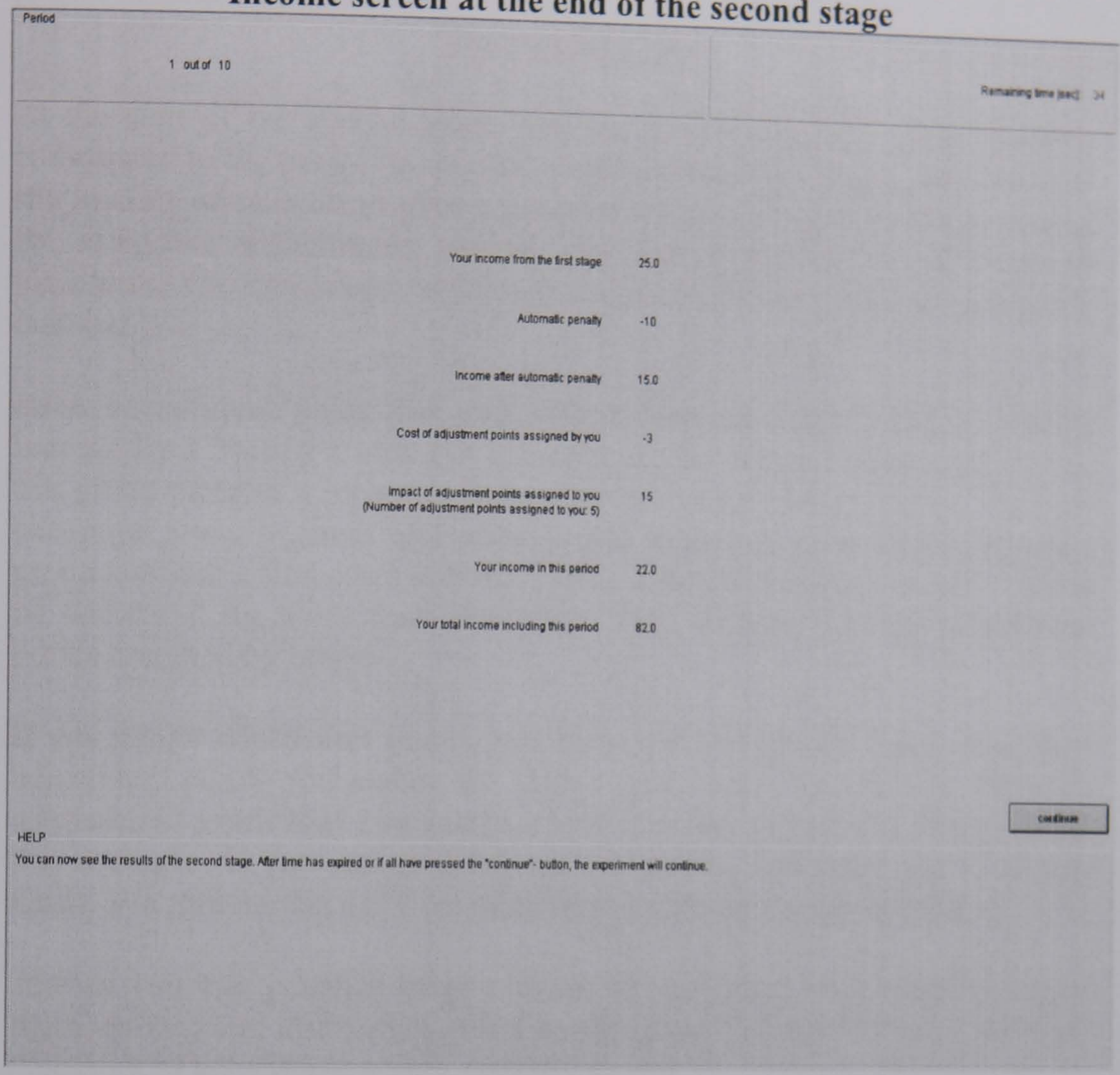
Besides the period and time display, you now see how much each group member contributed to the project in the first stage. **Your contribution is displayed in blue in the first column,** while the contributions of the other group members of this period are shown in the remaining three columns. Please note that the order in which contributions are displayed changes in a random order in each period. The contribution in the second column, for example, generally represents a different group member each time. The same holds for the contributions in the other columns. That way you are informed about the contributions but not about the identities of the other group members. Besides the absolute contributions, the income from the first stage and the income after the automatic penalty are also displayed.

You must decide whether and if so how many adjustment points to assign. In any case you must enter a number in the large blue box for each group member. For this decision, you have 180 seconds in the first period and 60 seconds in the remaining periods. You can move from one input field to the other by pressing the tab-key (→ |) or by using the mouse.

You can determine the total costs you incur on the computer. To perform the calculation you have to press the button "**Calculation**" (see the input screen at the second stage). You can do this after you have made an input. On the screen you will see the total costs of the points you assigned. As long as you have not yet pressed the **OK-button**, you can still change your decision (within the remaining time). To recalculate the cost after a change of the adjustment points you assigned, simply press the "Calculation" button again.

After all participants have made their decisions, your income from the period will be displayed on the following screen. Numbers shown in this screenshot are for example purposes only. To view the income screen at the end of the second stage, you have 45 seconds in the first period and 30 seconds in the remaining periods. If you are finished with it before the time is up, please press the continue button.

Income screen at the end of the second stage



Depending on the decisions that you and others take, it is possible for the net effect on your income of the adjustment points assigned in the second stage to be negative. However, taking all periods together, any such losses will always be outweighed by the income from the first stage and a lump sum payment of 60 Money Units that you receive at the beginning of the experiment.

Do you have any questions?

{Note: Second stage for the S-condition}

The second stage

At the start of the second stage, you see how much each group member contributed to the project in the first stage. **During this stage, you can alter the income of each other group member by assigning adjustment points. By assigning adjustment points, you can penalise the other group members.** You can assign between 0 and 2 adjustment points to each group member.

Each adjustment point that you assign decreases this group member's income by 3 Money Units. For example, if you assign 2 adjustment points, this group member's income will be decreased by 6 Money Units. The only exception arises because adjustment points cannot do more than eliminate a group member's first stage income. Thus, a group member's income cannot be decreased by more than their first stage income, through adjustment points assigned by others.

If you assign adjustment points, you have costs in Money Units. The more adjustment points you assign, the higher your costs. Specifically, **for each adjustment point that you assign, there is a cost to you of 1 Money Unit.** For example, if you assign 2 adjustment points, this costs you 2 Money Units. We refer to this as "Cost of adjustment points assigned by you".

Just as you can penalise other players by assigning adjustment points to them, so they can also penalise you by the same method. We refer to this as "Number of adjustment points assigned to you".

Your total income from the two stages is therefore calculated as follows:

Total income (in Money Units) at the end of the second stage = Period income =

= Income from the first stage
– Cost of adjustment points assigned by you
– 3*(Number of adjustment points assigned to you)

if the impact of the adjustment points assigned to you is less than the income from the first stage;

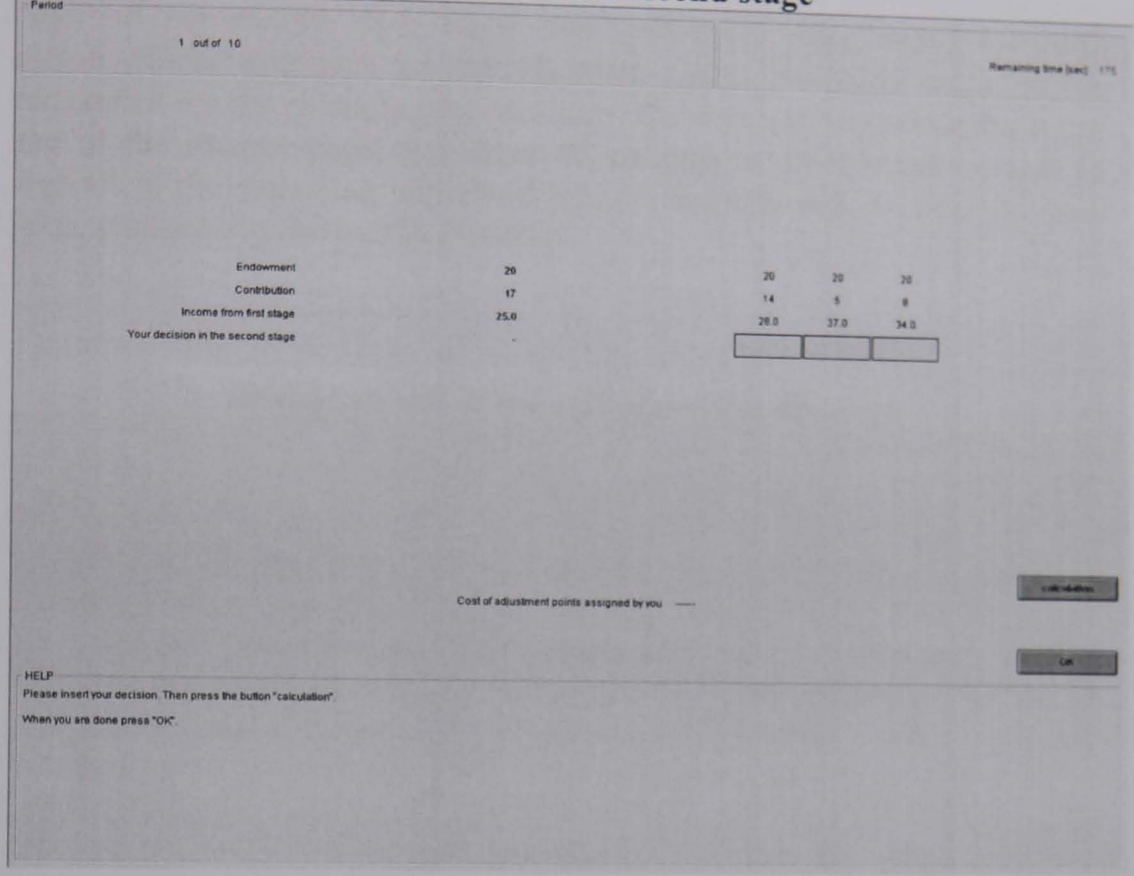
OR

= 0 – Cost of adjustment points assigned by you

if the impact of the adjustment points assigned to you is greater than the income from the first stage.

The way with which you can assign adjustment points is apparent from the input screen at the second stage. Numbers shown in this screenshot are for example purposes only.

Input screen at the second stage



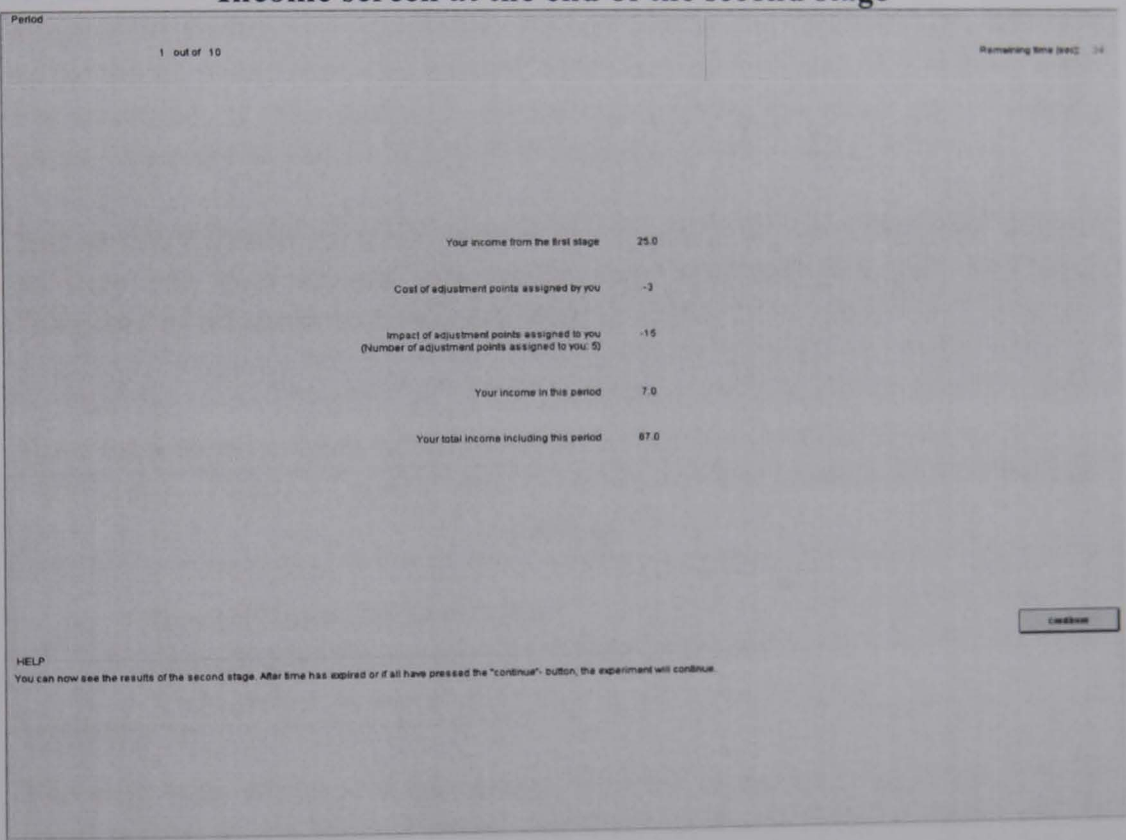
Besides the period and time display, you now see how much each group member contributed to the project in the first stage. **Your contribution is displayed in blue in the first column,** while the contributions of the other group members of this period are shown in the remaining three columns. Please note that the order in which contributions are displayed changes in a random order in each period. The contribution in the second column, for example, generally represents a different group member each time. The same holds for the contributions in the other columns. That way you are informed about the contributions but not about the identities of the other group members. Besides the absolute contributions, the income from the first stage is also displayed.

You must decide whether and if so how many adjustment points to assign. In any case you must enter a number in the large blue box for each group member. For this decision, you have 180 seconds in the first period and 60 seconds in the remaining periods. You can move from one input field to the other by pressing the tab-key (→|) or by using the mouse.

You can determine the total costs you incur on the computer. To perform the calculation you have to press the button "**Calculation**" (see the input screen at the second stage). You can do this after you have made an input. On the screen you will see the total costs of the points you assigned. As long as you have not yet pressed the **OK-button**, you can still change your decision (within the remaining time). To recalculate the cost after a change of the adjustment points you assigned, simply press the "Calculation" button again.

After all participants have made their decisions, your income from the period will be displayed on the following screen. Numbers shown in this screenshot are for example purposes only. To view the income screen at the end of the second stage, you have 45 seconds in the first period and 30 seconds in the remaining periods. If you are finished with it before the time is up, please press the continue button.

Income screen at the end of the second stage



Depending on the decisions that you and others take, it is possible for the net effect on your income of the adjustment points assigned in the second stage to be negative. However, taking all periods together, any such losses will always be outweighed by the income from the first stage and a lump sum payment of 60 Money Units that you receive at the beginning of the experiment.

Do you have any questions?

{Note: Second stage for the R-condition}

The second stage

At the start of the second stage, you see how much each group member contributed to the project in the first stage. **During this stage, you can alter the income of each other group member by assigning adjustment points. By assigning adjustment points, you can reward the other group members.** You can assign between 0 and 2 adjustment points to each group member.

Each adjustment point that you assign to another group member increases their income by 3 Money Units. For example, if you assign 2 adjustment points, this group member's income will be increased by 6 Money Units.

If you assign adjustment points, you have costs in Money Units. The more adjustment points you assign, the higher your costs. Specifically, **for each adjustment point that you assign, there is a cost to you of 1 Money Unit.** For example, if you assign 2 adjustment points, this costs you 2 Money Units. We refer to this as "Cost of adjustment points assigned by you".

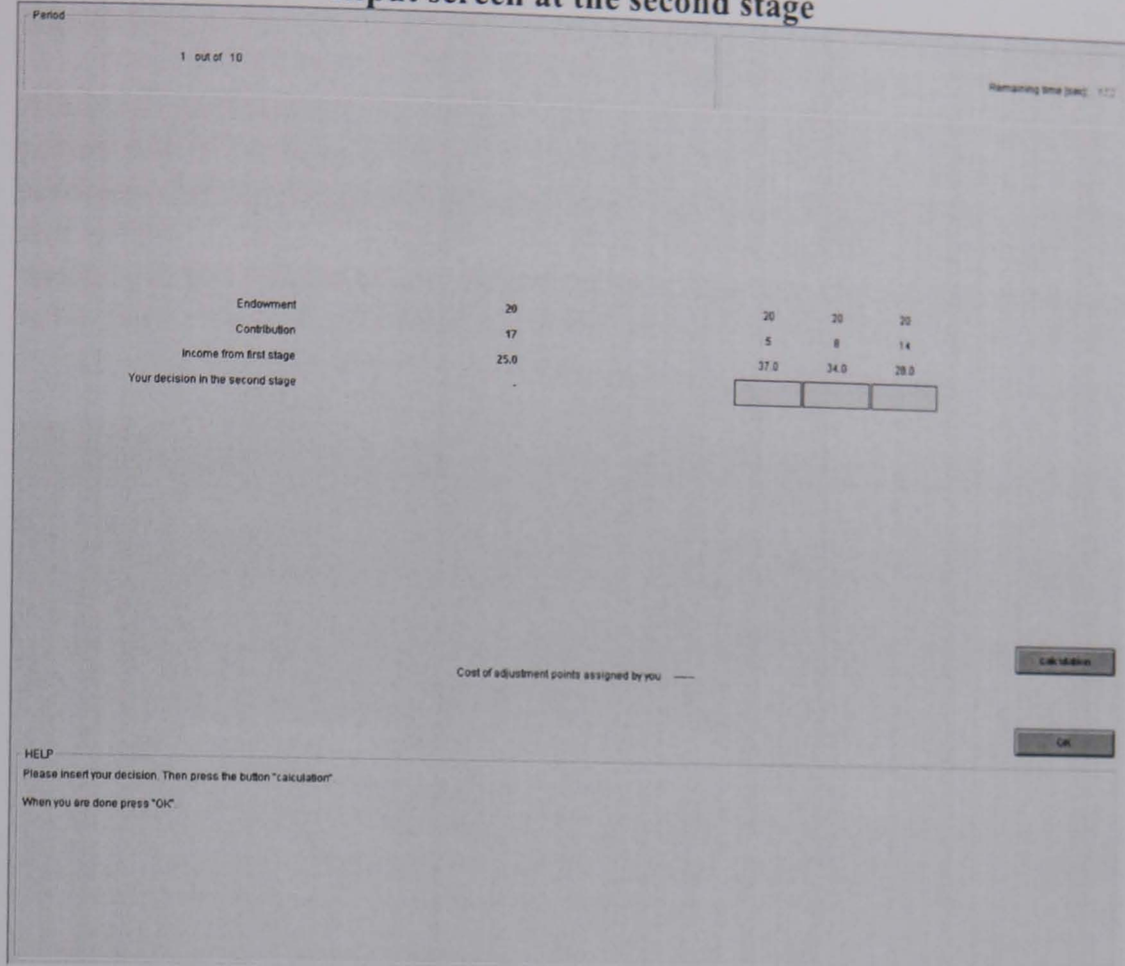
Just as you can reward other players by assigning adjustment points to them, so they can also reward you by the same method. We refer to this as "Number of adjustment points assigned to you".

Your total income from the two stages is therefore calculated as follows:

Total income (in Money Units) at the end of the second stage = Period income =
= Income from the first stage – Cost of adjustment points assigned by you + 3*(Number of adjustment points assigned to you)

The way with which you can assign adjustment points is apparent from the input screen at the second stage. Numbers shown in this screenshot are for example purposes only.

Input screen at the second stage



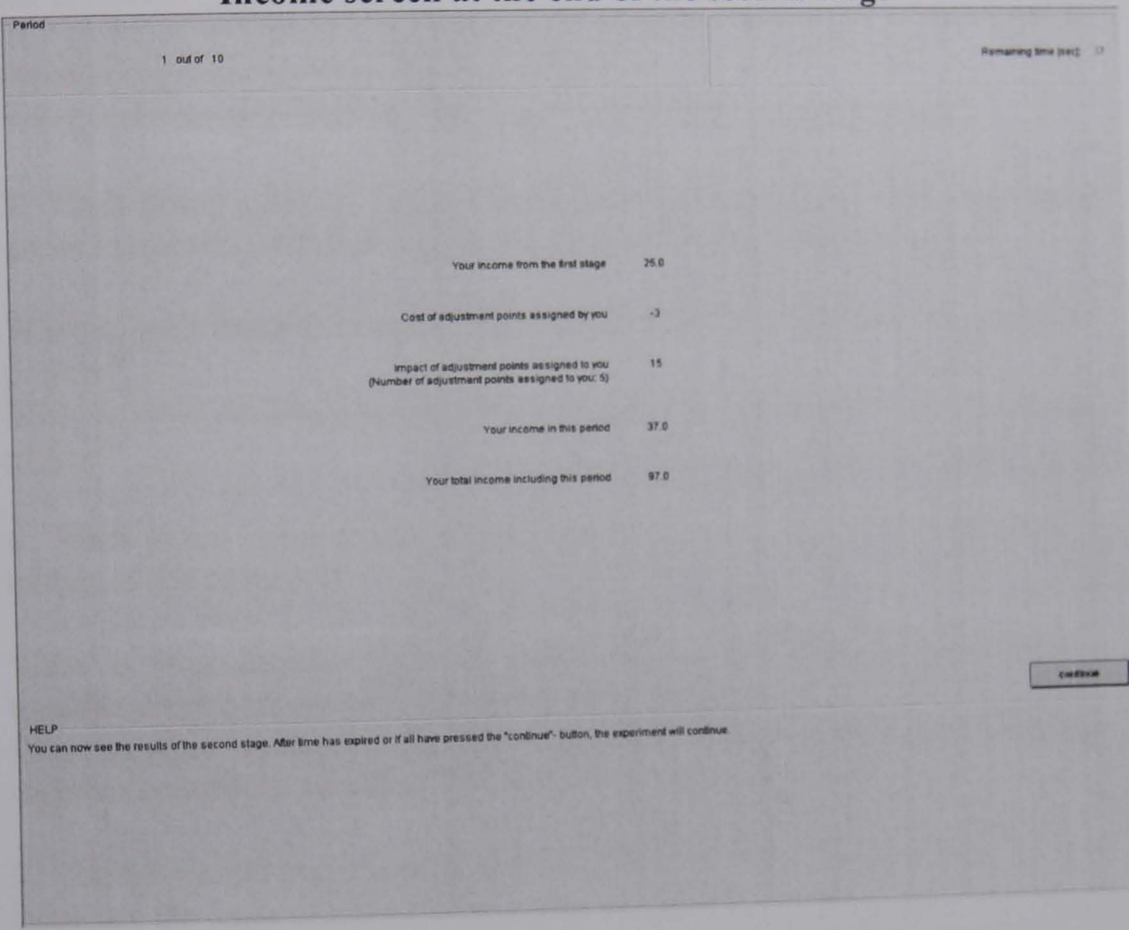
Besides the period and time display, you now see how much each group member contributed to the project in the first stage. **Your contribution is displayed in blue in the first column,** while the contributions of the other group members of this period are shown in the remaining three columns. Please note that the order in which contributions are displayed changes in a random order in each period. The contribution in the second column, for example, generally represents a different group member each time. The same holds for the contributions in the other columns. That way you are informed about the contributions but not about the identities of the other group members. Besides the absolute contributions, the income from the first stage is also displayed.

You must decide whether and if so how many adjustment points to assign. In any case you must enter a number in the large blue box for each group member. For this decision, you have 180 seconds in the first period and 60 seconds in the remaining periods. You can move from one input field to the other by pressing the tab-key (→|) or by using the mouse.

You can determine the total costs you incur on the computer. To perform the calculation you have to press the button "**Calculation**" (see the input screen at the second stage). You can do this after you have made an input. On the screen you will see the total costs of the points you assigned. As long as you have not yet pressed the **OK-button**, you can still change your decision (within the remaining time). To recalculate the cost after a change of the adjustment points you assigned, simply press the "Calculation" button again.

After all participants have made their decisions, your income from the period will be displayed on the following screen. Numbers shown in this screenshot are for example purposes only. To view the income screen at the end of the second stage, you have 45 seconds in the first period and 30 seconds in the remaining periods. If you are finished with it before the time is up, please press the continue button.

Income screen at the end of the second stage



Depending on the decisions that you and others take, it is possible for the net effect on your income of the adjustment points assigned in the second stage to be negative. However, taking all periods together, any such losses will always be outweighed by the income from the first stage and a lump sum payment of 60 Money Units that you receive at the beginning of the experiment.

Do you have any questions?

Control Questionnaire

1. Each group member has an endowment of 20 tokens. Nobody (including yourself) contributes any tokens to the project.

What is your income from the first stage? ....
What is the income from the first stage of the other group members? ....

2. Each group member has an endowment of 20 tokens. You contribute 20 tokens to the project. All other group members each contribute 20 tokens to the project.

What is your income from the first stage? ....
What is the income from the first stage of the other group members? ....

3. Each group member has an endowment of 20 tokens. The other three group members contribute together a total of 30 tokens to the project.

What is your income from the first stage if you contribute 0 tokens to the project? ....
What is your income from the first stage if you contribute 15 tokens to the project? ....

4. Each group member has an endowment of 20 tokens. You contribute 8 tokens to the project.

What is your income from the first stage if the other group members together contribute a total of 7 tokens to the project? ....
What is your income from the first stage if the other group members together contribute a total of 22 tokens to the project? ....

5. At the second stage you assign the following adjustment points: 1, 1, 0. What are the costs of adjustment points assigned by you? ....

6. What are your costs if you assign a total of 0 adjustment points? ....

7. By how many Money Units will your income from the first stage be changed by the adjustment points assigned to you by other group members if the other group members assign a total of 0 adjustment points to you? ....

8. By how many Money Units will your income from the first stage be changed by the adjustment points assigned to you by other group members if the other group members assign a total of 3 adjustment points to you? ....

Appendix B. Data Analysis

Table B.1 Average contributions for each matching group for each condition per sequence

	DS sequence		SD sequence		RS sequence	
	D-condition	S-condition	S-condition	D-condition	R-condition	S-condition
Matching group 1	17.95	20	9.5	5.125	9.65	13.425
Matching group 2	14.675	18.1	17.95	18	4.3	17.2
Matching group 3	11.75	10.45	7.475	2.675	10.35	18.5
Matching group 4	18.5	19.75	19.125	18	1.65	13.175
Matching group 5	13.175	19.275	10.4	4.375	12.275	12.075
Matching group 6	10.75	15	17.4	9.875	12.35	19.525
Matching group 7	8.25	13.425	16.725	17.75	15.475	19.75
Matching group 8	5.15	3.675	18.925	13.35	10.6	14.775
Matching group 9	17.075	16.9	2.125	2.425	15.75	19.4
Matching group 10	12.75	14.35	--	--	--	--

Table B.2 Period effects for the D-condition of the SD sequence

	Dependent variable: Contribution
	D-condition of the DS sequence
Periods 1&2	6.472***
	(1.535)
Periods 3&4	4.458**
	(1.611)
Periods 5&6	4.167***
	(1.006)
Periods 7&8	2.444*
	(1.300)
Constant	6.667**
	(1.987)
Obs.	360

*Note: OLS estimates with robust standard errors (clustered on independent matching groups) presented in parentheses. Periods 9 & 10 are the baseline. * denotes significance at the 10-percent level, ** denotes significance at the 5-percent level and *** denotes significance at the 1-percent level.*

Chapter 5

Conclusions

This thesis reports three experimental studies related to the provision of public goods when economic agents face free rider incentives. Its unifying theme is the investigation of different measures for how subjects perceive free riding under a number of treatment manipulations. In this Chapter, we conclude our findings and suggest further research avenues motivated by our experimental investigations presented earlier.

Our first study begins with the examination of whether a social dilemma game is perceived as having a moral dimension. We were particularly interested in how subjects perceive the action of free riding from a moral perspective, as free riding may be seen to constitute a violation of a basic moral principle (that is, causing harm to another person) and, thus, it is of moral significance. To do so, we elicit subjects' moral evaluations towards a free rider when they are detached from the decision situation and act as impartial observers. Our results are summarised as follows. First, we establish that free riding is indeed usually perceived as a morally inappropriate action. However, there is a notable exception on this. When the free rider moves second and observes that the other player has also free ridden, then subjects perceive free riding as a morally acceptable (even praiseworthy) action. We also find that framing of the decision situation and the order of moves are major determinants of moral evaluations. Specifically, withdrawing resources from the public good is condemned less strongly than failing to contribute, and sequential moves trigger stronger moral condemnation of free riding than simultaneous moves if the other player contributes. A noteworthy aspect of all these treatment manipulations is that as the contribution of the non-judged player rises, the free rider is condemned increasingly strongly, especially when he moves second. Interestingly, when moves are made simultaneously, the increasing condemnation hypothesis holds for a substantial minority of subjects; whereas, when the order of moves is sequential, an overwhelming majority of subjects take this view.

Having identified that the public good game generates a dilemma of moral significance, our study provides further motivation to explore how moral norms are formed when individuals are involved in the decision situation, acting as players materially affected by their and other players' choices. The relationship between morality and actual behaviour has so far been investigated in the context of bargaining games (Gächter and Riedl, 2005, 2006). However, there is a gap in the literature with regards to how subjects' normative views are shaped when they have to make actual choices that affect their monetary payoff in social dilemma situations. Specifically, further research on the interaction of moral norms and actual behaviour in social dilemma games is warranted since moral attitudes may be a candidate explanation for sanctioning behaviour, rendering morality as an additional non-behavioural indicator to measure social preferences. Recent evidence (e.g. Bosman and van Winden, 2002) indicates that emotions are identified as a proximate mechanism that generates economic behaviour and negative emotions drive the dark side of human behaviour. Yet, it is also possible that another mechanism, namely, moral judgments, is at work when basic moral principles are violated and individuals need to negatively reciprocate. Questions such as whether the moral legitimacy and purpose of sanctions are a direct function of the legitimacy of the action performed, and, to the extent that impartial moral evaluations are sensitive to framing, whether they are still context dependent when moral judgments are elicited from subjects involved in the decision situation are of great importance for the economic theories motivated by social preferences and for the explanation of subjects' behaviour, but still remain unanswered.

Motivated by our findings suggesting that subjects' moral evaluations of free riding are sensitive to the re-description of the social dilemma as a common resource problem, but also by previous experimental studies about altruistic behaviour in dictator games, Chapter 3 investigates whether frame sensitivity holds with respect to two measures of social preferences which have played a crucial role in the related public goods literature. Our two measures comprise of self-reported emotions, as a non-behavioural indicator of disapproval, and punishment, as a behavioural indicator of disapproval. The framing manipulation takes a Give vs. Take form and is identical to the one implemented in the first experiment. Our findings are that, for a given pattern of contributions, neither punishment nor emotional responses depends on the Give versus Take framing that we manipulate, suggesting that the social preferences we observe, using punishment and emotions as indicators, are robust to framing effects. The main determinant of both punishment and emotions is the difference between the contributions of the punisher and the punished group member. A possible reason why punishment and emotions are not frame sensitive, albeit moral judgments are, may be that being involved in the situation makes individuals see through the framing and/or concentrate more on consequences of a given situation.

Yet, the fact that punishment and emotional responses are context immune to the Give vs. Take manipulation does not preclude the possibility that these indicators are sensitive to other (perhaps, more suggestive) framing manipulations, such as label framing. In particular, evidence from public goods experiments that use a "Community" framing has recorded strong framing effects even in one-shot game contexts (see e.g., Rege and Telle, 2004 and Dufwenberg, et al., 2006). Investigation of whether framing is context dependent by examining whether, for a given level of contributions, punishment attitudes and self-reported emotional responses differ across this particular framing manipulation is on the agenda for future research. In addition, the finding that punishment as a measure of negative reciprocity cannot explain the existence of framing effects raises the interesting question of whether measures of positive reciprocity are able to give an explanation for these effects. Such research questions are still open and invite further investigation.

Our thesis is completed in Chapter 4 with the conduct of a third experiment which explores how subjects' assignment of punishment depends on the experience of an environment where punishment has already been assigned unfairly. To achieve such an environment, we implement a variant of the standard punishment game by Fehr and Gächter (2000), in which all members of a group have been punished exogenously, unrelated to their prior behaviour. Subjects are then given the opportunity to alleviate the automatic penalty imposed on them by incurring a cost. Since our default punishment game has also a reward element incorporated in its structure because subjects can alleviate the exogenously assigned punishment, we also examined as an auxiliary condition, a game in which group members are only given the opportunity to reward their fellow group members, without having been exogenously punished.

Our main findings suggest that, on average, contributions do not differ significantly among the three games investigated (i.e. the default punishment game, the standard punishment game and the reward game). Interestingly, we find that the time profile of contributions is different. Specifically, for the default punishment game, we document strong period effects, which are not present either in the standard punishment game or in the reward game. After the second half of the game contribution levels decline over time. In addition, a history of an unfair environment does not affect the ability of the standard punishment game to sustain high levels of contributions. However, a history of the standard punishment game causes contribution levels in the default punishment game to collapse from the very beginning of the game. Regarding behaviour in the second stage, we find that assigned alleviation and reward are not significantly different and that the assignment of punishment is unaffected by the previous experience of a corrosive environment. However, it turns out that reactions to alleviation, reward and punishment differ. More specifically, those subjects who contributed at least as much as the group average decreased their contributions per adjustment point received in the reward game, but did not change their contributions in the default punishment game. Regarding those who contributed less than the group average, we find that they decreased their contributions in the default punishment game but did not change their contributions in the reward game. Relative to reactions to punishment, we observe that those subjects who contributed less than the group average increase their contributions per punishment point received when they have already experienced the reward game (or when they do not have any previous experience at all), but not when they have experienced the default punishment game, in which case they do not change their contributions significantly.

An obvious research possibility from the findings of this experiment is whether and if so, how the severity of the exogenous punishment (which can alternatively be seen as the cost for the enforcement of a social norm) and its subsequent automatic losses have an impact on whether punishment attitudes in the standard punishment treatment alter.

To sum up, the present thesis has provided answers in questions pertaining to the broader topic of social preferences by making use of both non-traditional (that is, moral judgments and emotions) and traditional (that is, actual behaviour) data. Our view is that in the future, the science of experimental economics will benefit from the integration of non-behavioural measures into its tool kit. Furthermore, from our main findings described in this thesis we conclude that perceptions of free riding are indeed affected by various treatment manipulations, but this depends on the special circumstances in which economic agents interact. As a whole, we believe that the current thesis provides answers to pervasive questions in public good experiments and contributes to the related experimental literature in social preferences and economic moral psychology. At the same time, it opens new avenues for future research options that would give new insights and help us draw more reliable conclusions regarding issues of human cooperation.

References

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

Andreoni, J., (1988), "Why free ride? Strategies and learning in public goods experiments", *Journal of Public Economics*, vol. 37, pp. 291-304.

Andreoni, J., (1995), 'Warm-Glow versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments', *Quarterly Journal of Economics*, vol. 110, pp. 1-21.

Andreoni, J., Harbaugh, W., and Vesterlund, L., (2003), 'The Carrot or the Stick: Rewards, Punishments, and Cooperation', *American Economic Review*, vol. 93, pp. 893-902.

Bardsley, N., (2007, forthcoming), 'Dictator game giving: altruism or artefact?', *Experimental Economics.*

Baron, J. and Ritov, I., (1994), 'Reference points and omission bias', *Organizational Behavior and Human Decision Processes,* vol. 59, pp. 475-498.

Batina, R. and Ihori, T., (2005), 'Public Goods: Theories and Evidence', Springer.

Birnbaum, M. H., (2004), 'Human Research and Data Collection via the Internet', *Annual Review of Psychology*, vol. 55, pp. 803-832.

Bochet, O., Page, T., and Putterman, L., (2006), 'Communication and punishment in voluntary contribution experiments', *Journal of Economic Behavior and Organization*, vol. 60, pp. 11-26.

Bolton, G., and Ockenfels, A, (2000), 'ERC: A theory of Equity, Reciprocity, and Competition', *American Economic Review*, vol. 90, pp. 166-193.

Bosman, R., and van Winden, F., (2002), 'Emotional Hazard in a Power-to-Take Experiment', *Economic Journal*, vol. 112, pp. 147-169.

Bowles, S., Carpenter, J., and Gintis, H., (2001). Mutual Monitoring in Teams: Theory and Evidence on the Importance of Residual Claimancy and Reciprocity. Mimeo, University of Massachusetts, Amherst.

Brewer, M. B. and Kramer, R. M., (1986), 'Choice behavior in social dilemmas: Effects of social identity, group size, and decision framing', *Journal of Personality and Social Psychology*, vol. 50, pp. 543-549.

Camerer, C., (2003), *Behavioural game theory: Experiments on strategic interaction*, Princeton: Princeton University Press.

Camerer, C., and Fehr, E., (2004), 'Measuring Social Norms and Social Preferences using Experimental Games: A Guide for Social Scientists' in *Foundations of Human Sociality – Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*, edited by Joseph Henrich, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr and Herbert Gintis, pp. 55-95.

Carpenter, J., Matthews, H.P., and Ong'ong'a, O., (2004), 'Why Punish? Social Reciprocity and the Enforcement of Prosocial Norms', *Journal of Evolutionary Economics*, vol., 14, pp. 407-430.

Carpenter, J., (2007a), 'Punishing Free-Riders: How Group Size Affects Mutual Monitoring and the Provision of Public Goods', *Games and Economic Behavior*, vol. 60, pp. 31-51.

Carpenter, J., (2007b), 'The demand for Punishment', *Journal of Economic Behavior and Organization*, vol. 62, pp. 522-542.

Cinyabuguma, M., Page, T., and Putterman, L., (2006), 'Can second-order punishment deter perverse punishment?', *Experimental* Economics, vol. 9, pp. 265-279.

Cookson, R., (2000), 'Framing effects in Public Goods Experiments', *Experimental Economics*, vol. 3, pp. 55-79.

Cornes, R. and Sandler, T., (1996), 'The theory of externalities, public goods, and club goods', *Cambridge: Cambridge University Press.*

Croson, R., Konow, J., (2009), Social preferences and moral biases. *Journal of Economic Behavior & Organization*, vol. 69, pp. 201-212.

Cubitt, R., Drouvelis, M., and Gächter, S., (2008), 'Framing and Free Riding: Emotional Responses and Punishment in Social Dilemmas', *CeDEx Working Paper 2008-02.*

Cushman, F., Young, L., and Hauser, M., (2006), 'The Role of Conscious Reasoning and Intuition in Moral Judgment: Testing Three Principles of Harm', *Psychological Review*, vol. 17, pp. 1082-1089.

Dana, J., Weber, R., and Kuang, J. X., (2005), 'Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness', *Working Paper, Center for Behavioral Decision Research, Carnegie-Mellon University.*

Dawes, R.M., McTavish, J., and Shaklee, H., (1977), 'Behavior Communication and Assumptions about Other People's Behavior in a Commons Dilemma Situation', *Journal of Personality and Social Psychology*, vol. 35, pp. 1-11.

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

Dufwenberg, M., Gächter, S., and Hennig-Schmidt, H., (2006), 'The Framing of Games and the Psychology of Strategic Choice', *CeDEx Discussion Paper No. 2006-20.*

Egas, M., and Riedl, A., (2008), 'The economics of altruistic punishment and the maintenance of cooperation', *Proceedings of the Royal Society B.*

Elster, J., (1998), 'Emotions and Economic Theory', *Journal of Economic Theory*, vol. XXXVI, pp. 47-74.

Falk, A., Fehr, E., and Fischbacher, U., (2005), ,Driving forces behind informal sanctions', IZA Discussion Paper No. 1635.

Fehr, E., and Gächter, S., (2000), 'Cooperation and Punishment in Public Goods Experiments', *American Economic Review*, vol. 90, pp. 980-994.

Fehr, E., and Gächter, S., (2002), 'Altruistic Punishment in Humans', *Nature*, vol. 415, pp. 137-140.

Fehr, E., and Schmidt, K., (1999), 'A Theory of Fairness, Competition and Cooperation, *Quarterly Journal of Economics*, vol. 114, pp. 817-868.

Fischbacher, U., *(2007) 'z-Tree: Zurich Toolbox for Ready-made Economic Experiments'* Experimental Economics, vol. 10, pp.171-178.

Fleishman, J. A., (1988), 'The Effects of Decision Framing and Others' Behaviour on Cooperation in a Social Dilemma', *Journal of Conflict Resolution*, vol. 32, pp. 162-180.

Gächter, S., and Herrmann, B., (2007), 'The Limits of Self-Governance when Cooperators Get Punished: Experimental Evidence from Urban and Rural Russia', *CeDEx Working Paper 2007-11.*

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

Gächter, S., and Riedl, A., (2005), 'Moral Property Rights in Bargaining with Infeasible Claims', *Management Science*, vol. 51, pp. 249-263.

Gächter, S., and Riedl, A., (2006), 'Dividing Justly in Bargaining Problems with Claims: Normative Judgments and Actual Negotiations', *Social Choice and Welfare*, vol. 27, pp. 571-594.

Geanakoplos, J., Pearce, D., and Stacchetti, E., (1989), 'Psychological Games and Sequential Rationality', *Games and Economic Behavior*, vol. 1, pp. 60-79.

Gino, F., Moore, D. A., and Bazerman, N. H., (2008), 'No Harm, no foul: The outcome bias in ethical judgments', *Working Paper 08-080*.

Gintis, H., Bowles, S., Boyd, R., and Fehr, E., (2005), *Moral Sentiments and Material Interests – The Foundation of Cooperation in Economic Life*, MIT Press, Cambridge Massachusetts.

Greiner, B., (2004), 'An Online Recruitment System for Economic Experiments', In: Kurt Kremer and Volker Macho (Eds.): Forschung und wissenschaftliches Rechnen 2003, GWDG Bericht 63. Göttingen: Gesellschaft für Wissenschaftliche Datenverarbeitung, pp. 79-93.

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

Haidt, J., (2001), 'The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment', *Psychological Review*, vol. 108, pp. 814-834.

Haidt, J., (2007), 'The new synthesis in moral psychology', *Science*, vol. 316, pp. 998-1002.

Haidt, J., and Baron, J., (1996), 'Social roles and the moral judgement of acts and omissions', *European Journal of Social Psychology*, vol. 26, pp. 201-218.

Haley, K. J., and Fessler, D. M. T., (2005), 'Nobody's watching? Subtle cues affect generosity in an anonymous economic game', *Evolution and Human Behaviour*, vol. 26, pp. 245-256.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., and Gintis, H., (2004), Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies, *Oxford University Press.*

Herrmann, B., Thöni, C., and Gächter, S., (2008), 'Antisocial Punishment Across Societies', *Science*, vol. 319, pp. 1362-1367.

Hertwig, R. and Ortmann, A., (2008), 'Deception in Psychological Experiments: Two Misconceptions and a Research Agenda', *Social Psychology Quarterly*, vol. 71, pp. 222-227.

Hume, D., (1739), 'A treatise of human nature'.

Isaac, R.M., Walker, J.M., and Thomas, S.H., (1984), 'Divergent evidence on free riding: an experimental examination of possible explanations', *Public Choice*, vol. 43, pp. 113–49.

Isaac, R.M and Walker, J.M., (1988a), 'Communication and free-riding behavior: the voluntary contribution mechanism', *Economic Inquiry*, vol. 26, pp. 585–608.

Isaac, M.R. and Walker, J.M. (1988b), 'Group size effects in public goods provision: the voluntary contributions mechanism', *Quarterly Journal of Economics*, vol. 53, pp. 179–200.

Isaac, M. R., Walker, J. M., and Williams, A. W., (1994), 'Group size and the voluntary provision of public goods', *Journal of Public Economics*, vol. 54, pp. 1-36.

Kahneman, D., and Tversky, A., (1979), 'Prospect Theory: An Analysis of Decision Under Risk', *Econometrica*, vol. 47 (2), pp. 263-292.

Kahneman, D. and Tversky, A., (1984), 'Choices, values and frames', *American Psychologist*, vol. 46, pp. 1044-1057.

Kahneman, D., Knetsch, J. L., and Thaler, R., (1986), 'Fairness as a Constraint on Profit Seeking: Entitlements in the Market', *American Economic Review*, vol. 76, pp. 728-41.

Kelly, D., Stich, S., Haley, K. J., Eng, S. J., and Fessler, D. M. T., (2007), 'Harm, Affect and the Moral/Conventional Distinction', *Mind and Language*, vol. 22, pp. 117-131.

Kirchsteiger, G., Rigotti, L., and Rustichini, A., (2006), 'Your morals might be your moods', *Journal of Economic Behaviour and Organization*, vol. 59, pp. 155-172.

Kohlberg. L., (1969), 'Stage and sequence: The cognitive-developmental approach to socialization', In D. A. Goslin (Ed.), Handbook of socialisation theory and research (pp. 347-480), Chicago: Rand McNally.

Ledyard, J., (1995), 'Public Goods: A Survey of Experimental Research', in J. H. Kagel and A. E. Roth, eds., *Handbook of Experimental Economics,* Princeton University Press, pp. 111-194.

Levin, I., Schneider, S., and Gaeth G., (1998), 'All Frames are Not Created Equal: A Typology and Critical Analysis of Framing Effects', *Organizational Behavior and Human Decision Processes*, vol. 76, pp. 149-188.

List, J. A., (2007), 'On the interpretation of giving in dictator games', *Journal of Political Economy*, vol. 115, pp. 482-493.

Loewenstein, G., (2000), 'Emotions in economic theory and economic behavior', *American Economic Review*, vol. 90, pp. 426–432.

Marwell, G. and Ames, R., (1979), 'Experiments on provision of public goods I: resources, interest, group size, and the free riding problem', *American Journal of sociology*, vol. 84, pp. 1335-60.

Marwell, G. and Ames, R., (1980), "Experiments on provision of public goods II: provision point, stake, experience, and the free riding problem", *American Journal of sociology*, vol. 85, pp. 926-37.

Marwell, G. and Ames, R., (1981), 'Economists free ride, does anyone else?', *Journal of Public Economics*, vol. 15, pp. 295-310.

Masclet, D., Noussair, C., Tucker, S., and Villeval, M-C., (2003), 'Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism', *American Economic Review*, vol. 93, pp. 366-380.

McCusker, C., and Carnevale, P. J., (1995), 'Framing in Resource Dilemmas: Loss Aversion and the Moderating Effects on Sanctions', *Organizational Behavior and Human Decision Processes*, vol. 61, pp. 190-201.

McDaniel, W. C., and Sistrunk, F., (1991), 'Management dilemmas and decisions: Impact of framing and anticipated responses', *Journal of Conflict Resolution*, vol. 35, pp. 21-42.

McDowell, A., (2003), 'From the help desk: hurdle models', *Stata Journal*, vol. 3, pp. 178-184.

Niedermayer, L.Y., and Chapman, G.B., (2001), 'Action, inaction, and factors influencing perceived decision making', *Journal of Behavioral Decision* Making, vol. 14, pp. 295-308.

Nikiforakis, N., (2008), 'Punishment and Counter-Punishment in Public Goods Game: Can we still govern ourselves?', *Journal of Public Economics*, vol. 92, pp. 91-112.

Nikiforakis, N., and Normann, H.-T., (2008), 'A Comparative Static Analysis of Punishment in Public-Good Experiment', *Experimental Economics*, vol. 11, pp. 358-369.

Noussair, C., and Tucker, S., (2005), 'Combining Monetary and Social Sanctions to Promote Cooperation', *Economic Inquiry*, vol. 43, pp. 649-660.

Page, T., Putterman, L., and Unel, B., (2005), "Voluntary association in public goods experiments: reciprocity, mimicry, and efficiency", *Economic Journal*, 115, 1032-53.

Palfrey, T. R., and Prisbey, J. E., (1997), 'Anomalous Behavior in Public Goods Experiments: How Much and Why?', *American Economic Review*, vol. 87, pp. 829-846.

Park, E. S., (2000), 'Warm-Glow versus Cold Prickle: a further experimental study of framing effects on free-riding', *Journal of Economic Behavior and Organization*, vol. 43, pp. 405-421.

Piaget, J., (1965), 'The moral judgement of the child' (M. Gabain, Trans.). New York: Free Press. (Original work published 1932)

Prinz, J., (2006), 'The Emotional Basis of Moral Judgments', Philosophical *Explorations*, vol. 9, pp. 29-43.

Rege, M., and Telle, K., (2004), 'The impact of social approval and framing on cooperation in public good situations', *Journal of Public Economics*, vol. 88, pp. 1625-1644.

Reips, U-D., (2000), *Psychological Experiments on the Internet*, edited by Michael H. Birnbaum, Academic University Press.

Reips, U-D., (2002), 'Standards for Internet-Based Experimenting', *Experimental Psychology*, vol. 49, pp. 243-256.

Rutte, R. G., Wilke H. A. M., and Messick, D. M., (1987), 'The effects of framing social dilemmas as give-some or take-some games', *British Journal of Social Psychology*, vol. 26, pp. 103-108.

Samuelson, P. A., (1954), The Pure Theory of Public Expenditure, *Review of Economics and Statistics*, vol. 36, pp. 387-389.

Sefton, M., Shupp, R., and Walker, J. M., (2007), 'The effects of rewards and sanctions in provision of public goods', *Economic Inquiry*, vol. 45, pp. 679-690.

Sell, J., and Son, Y., (1997), 'Comparing public goods and Common Pool Resources: Three Experiments', *Social Psychology Quarterly*, vol. 60, pp. 118-137.

Smith, A., (1759), 'The Theory of Moral Sentiments'.

Sonnemans, J., and Frijda, N., (1994), 'The Structure Of Subjective Emotional Intensity', *Cognition & Emotion*, vol. 8, pp. 329-350.

Sonnemans, J., and Frijda, N., (1995), 'The Determinants Of Subjective Emotional Intensity', *Cognition Emotion*, vol. 9, pp. 483-506.

Sonnemans, J., Schram, A., and Offerman, T., (1998), 'Public Good Provision and Public Bad Prevention: The Effect of Framing', *Journal of Economic Behaviour and Organization*, vol. 34, pp. 143-161.

Spranca, M., Minsk, E., and Baron, J., (1991). 'Omission and commission in judgment and choice', *Journal of Experimental Social Psychology*, vol. 27, pp. 76-105.

Turiel, E., (1983), 'The development of social knowledge: Morality and Convention', Cambridge, England: Cambridge University Press.

Tversky, A., and Kahneman, D., (1981), 'The Framing of Decisions and the Psychology of Choice', *Science*, vol. 211, pp. 453-463.

Tversky, A., and Kahneman, D., (1986), 'Rational Choice and the Framing of Decisions', *Journal of Business*, vol. 59 (4), pp. 251-278

Van Dijk, E., and Wilke, H., (2000), 'Decision-Induced Focusing in Social Dilemmas: Give-Some, Keep-Some, Take-Some, and Leave-Some

Dilemmas', *Journal of Personality and Social Psychology*, vol. 78, pp. 92-104.

Wheatley, T., Haidt, J., (2005), 'Hypnotic Disgust Makes Moral Judgements More Severe', Psychological Science, vol. 18, pp. 780-84.

Willinger, M., and Ziegelmeyer, A., (1999), 'Framing and cooperation in public good games: an experiment with an interior solution', *Economics Letters*, vol. 65 (3), pp. 323-328.