

An Exploration of Third and Second Party Punishment in Ten Simple Games

Andreas Leibbrandt¹ and Raúl López-Pérez²

18 June 2009

Abstract: This paper identifies the motives behind punishment from unaffected third parties and affected second parties using a within-subject design in ten simple games. We apply a classification analysis and find that a parsimonious model assuming that subjects are either inequity averse or selfish best explains the pattern of punishment from both third and second parties. Despite their unaffected position, we do not find that third parties punish in a more impartial or normative manner.

Keywords: Fairness, inequity aversion, norms, punishment, reciprocity, third parties.

JEL Classification: C70, C91, D63, D74, Z13.

¹ (Corresponding author) Institute for Empirical Research in Economics, University of Zurich, Bluemlisalpstr. 10, CH-8006 Zurich, Tel. + 41 446345269, E-Mail: leibbrandt@gmail.com.

² Department of Economic Theory, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain. E-mail: raul.lopez@uam.es

1. Introduction

Third parties play a crucial role in many institutions: They serve in courts, as referees or arbitrators.³ The US legal system, for instance, relies on their judgment in juries when it comes to the application of sanctions.⁴ Third parties are also important with regard to informal sanctions (Homans, 1961) and, in fact, their interventions seem to be essential in the explanation of norm enforcement, as they are often more numerous than affected second parties (Bendor and Swistak, 2001) or the only parties present and hence their sanctions are potentially more damaging than those from second parties.

Despite their importance, little is known about how third parties sanction others. In particular, it is unclear whether third parties sanction in a different manner than second parties. In principle, third parties might sanction in a more impartial, "normative", and controlled manner, and less egocentrically (Fehr and Fischbacher, 2004). Adam Smith apparently had this idea in mind when he introduced the concept of the "impartial spectator" in his *Theory of Moral Sentiments*, a party who is not personally affected, making decisions from beyond the limitations of egocentric biases. In fact, the prevalence of institutions (like juries) that rely on third parties is in accordance with the idea that third parties make more appropriate decisions. However, it also seems plausible that even third parties cannot completely eliminate egocentric biases (Ross et al., 1976; Babcock et al., 1995). The concerns about the selection of jury members in many law cases suggest that third parties can make very inappropriate decisions in the context of sanctioning (e.g. Kennedy, 1997).

Recent models of other-regarding preferences propose competing explanations for third and/or second party punishment. Thus, theories of inequity-aversion like Fehr and Schmidt (1999) predict punishment of richer co-players if that reduces the payoff distance, while reciprocity theories (Rabin 1993, Dufwenberg and Kirchsteiger 2004; Cox et al., 2007) are based on the idea that people harm those who harmed them. Further, Bolton and Ockenfels (2000) predict punishment of any co-player if that brings the aggressor's relative payoff closer to the average relative payoff, Levine (1998) posits the existence of spiteful types who punish indiscriminately and type-reciprocal agents who punish selfish or spiteful co-players, Falk and Fischbacher (2006) combine ideas from inequity-aversion and reciprocity models, and López-Pérez (2008) predicts punishment of norm deviators.

³ We say that a player C is a third party with respect to a player A if her material payoff does *not* depend on the decisions of A (note however that it is possible that the material payoff of A depends on the decisions of C; for instance, it could be the case that C sanctions A and thus reduces her material payoff). We also say that a player B is a second party with respect to A if her material payoff depends on A's decisions.

⁴ On the use of juries in other countries, see *The Economist*, February 14th 2009.

This paper applies a within-subject experimental analysis and the classification method by El-Gamal and Grether (1995), with two key objectives: (i) to provide a test of recent models of other-regarding preferences and determine which one best accounts for punishment in a large range of situations, and (ii) to study and compare the motives behind third and second party punishment.⁵ Our paper suggests several insights into the key motives behind punishment. First, we find in our games that both third and second party punishment is predominantly targeted towards richer co-players. Second, the classification analysis shows that a model assuming that subjects are either inequity-averse or selfish captures the *occurrence* of third *and* second party punishment across our ten games better than any other equally parsimonious alternative. While models that also include small fractions of spiteful (for third and second parties) and reciprocal (for second parties) types are slightly more accurate, they come at the cost of increased complexity. Third, we observe that the *strength* of punishment depends heavily on the size of the payoff disadvantage (in 3P and 2P). Overall, these results suggest that second and third party punishment need not be qualitatively different, and that inequity-aversion is crucial to explain the occurrence and strength of second and third party punishment, while reciprocity and spite play a *relatively* minor role.

The rest of the paper proceeds as follows. The next section compares our study with the related literature, while section three presents the experimental design and procedure. In section four, we report the results from the classification analysis and study which factors affect the occurrence and strength of second and third party punishment. The fifth section concludes.

2. Related Literature

A large body of experimental research shows that subjects are often willing to spend money to reduce another player's payoff – i.e. to punish her – even if no future benefits can follow from this behavior. In the ultimatum game, responders frequently punish proposers for making unfair offers (Güth et al. 1982, Camerer and Thaler 1995, Roth 1995), while non-contributors are often punished in public goods games with a punishment stage (Ostrom et al. 1992, Fehr and Gächter 2000). Further, Nikiforakis (2008) show that punishment in public goods games is frequently retaliated. However, this literature has a completely different focus than our study because it is restricted to second party punishment and because the analyzed games are not well suited for discriminating the motives behind punishment. In the ultimatum

⁵ The reader interested in similar approaches to understand behavioral decision rules may consult Engle-Warnick (2003).

game, responders might reject offers due to inequity-aversion, reciprocity, spite, or to punish a violation of an equity norm, while punishment in the public goods game can be explained in terms of inequity-aversion, reciprocity, spite, or as a reaction to a transgression of a cooperation norm.

The studies by Falk et al. (2005) and Dawes et al. (2007) have provided progress in our understanding of *second-party* punishment, although considerable uncertainties remain regarding the main motives. First of all, the two studies come to different conclusions. Falk et al. (2005) find that "retaliation seems to be the most important motive behind fairness-driven informal sanctions" (*ibid*, p. 2017) whereas Dawes et al. (2007) emphasize the importance of the "egalitarian motives". Moreover, these studies do not allow for perfect discrimination between some motives. For instance, the results from Falk et al. (2005) are not only consistent with reciprocity (i.e., retaliation) but also with a model predicting punishment of players who deviate from a norm of cooperation/efficiency (see López-Pérez, 2008). In turn, reciprocity might explain part of the results in Dawes et al. (2007). In this study, subjects were placed in small groups and allocated a randomly determined sum of money which could be used to reduce the income of their co-players'. Since all players could engage in damaging, they could damage conditional on their expectations about the damaging from the other subjects, i.e. retaliate (see Zizzo, 2003).

In addition, the studies by Falk et al. (2005) and Dawes et al. (2007) pay little attention to players' heterogeneity because they consider very few games and use a between-subject design. This seems to be a limitation because punishment is very likely caused by multiple motivational forces. On the one hand, different players might have different reasons to punish in the same game –the evidence from Falk et al. (2005) is indeed consistent with this. On the other hand, the same player might punish for one reason (say, inequity-aversion) in one game, and for another reason (say, reciprocity) in a different game. Consequently, it seems important to investigate which forces are *relatively* more powerful for each player across a large range of games, thus classifying each subject as (predominantly) inequity-averse, reciprocal, etc.

Few studies address *third party* punishment (e.g. Zizzo 2003; Carpenter and Matthews, 2005; Charness et al., 2008) and Fehr and Fischbacher (2004) is the only study which compares it to second party punishment. The authors report that third parties punish unfair allocation choices in a dictator game and defectors in a prisoner's dilemma game, although less strongly than second parties do. However, it remains unclear why third parties punish in these games (it could be because they punish violations from norms of

cooperation/equity, but also because of inequity-aversion, spite, or because they are type-reciprocal á la Levine, 1998) and why they punish less than second parties (this might be an artifact of their experimental design, as the payoff disadvantage was larger between first and second parties than between first and third parties). To the best of our knowledge, our study is the first to address these issues in a comprehensive design.

3. Experimental Design and Procedures

There are two treatments in our experimental design: A second party punishment treatment (2P) and a third party punishment treatment (3P). Participants in 2P play ten two-player games, while participants in 3P play ten three-player games. All these games have a two-stage structure. In the first stage of both treatments, one player (the *first party*) chooses between a left-hand and a right-hand allocation of payoffs between herself and another player (the *second party*). Table 1 shows the two allocations available in each game (as we argue later, this selection of games renders it possible to discriminate between a large number of theories). They are identical in 2P and 3P and presented in points (10 points = 1 Swiss Franc).

TABLE 1—THE ALLOCATIONS IN THE 10 GAMES

		Game									
		1	2	3	4	5	6	7	8	9	10
Allocation	Left	(150,150)	(100,100)	(560,60)	(150,90)	(220,260)	(280,240)	(250,80)	(100,100)	(250,150)	(250,150)
	Right	(590,60)	(50,530)	(120,140)	(50,630)	(220,400)	(390,240)	(80,250)	(50,150)	(110,290)	(330,70)

The second stage differs in the two treatments. In any game of 2P, the second party can spend points out of her allocation share to reduce the first party’s payoff –i.e., to punish her. In any game of 3P, a third player (the *third party*) can punish the first *or/and* the second party, while the second party in 3P makes no decision, i.e. she is a “bystander”.⁶ The third party is endowed with 200 points in each allocation of each game meaning the first party’s choice never affects her payoff in the first stage. The punishment technology is the same in 2P and 3P: Up to 50 points can be used to punish and each point spent reduces the payoff of the punished player by three points. Hence, if the first party chooses the allocation (x_{FP}, x_{SP}) in a game in 2P and the second party punishes her with $0 \leq p \leq 50$ points, the first party’s payoff

⁶ We allowed the third party to sanction the passive second party because the available theories make very different predictions in this regard (e.g., reciprocity predicts no sanctioning, while inequity aversion is consistent with sanctioning of richer second parties), so that we test them further.

in that game is $x_{FP} - 3p$ and the second party's payoff is $x_{SP} - p$. In 3P, if the first party chooses allocation (x_{FP}, x_{SP}) in a game and the third party punishes her with p_1 points and the second party (the bystander) with p_2 points ($p_1 + p_2 \leq 50$), the payoffs in this game are $x_{FP} - 3p_1$ for the first party, $x_{SP} - 3p_2$ for the second party, and $200 - p_1 - p_2$ for the third party.

We ran eight sessions where we observed a total of 3100 punishment decisions. Each session proceeded as follows. Subjects were randomly assigned to be a first or second party (or third party in 3P) and anonymously matched in groups of two (in 2P) or three (in 3P). Each subject received written instruction sheets (dependent on role and treatment) which explained the extensive form of the games (without giving information about the payoff constellations of the ten games). Subjects had to fill out control questions to make sure that they understood the rules. We used neutral language and avoided terms such as "punishment". Every subject always played the ten games in the same role on the computer and no subject participated in both treatments. The ten games were presented one at a time, and the order in which they were played was randomly predefined for each group. At the end of each game, we additionally asked first parties in 2P and 3P (and second parties in 3P) about their expectations of punishment in each allocation. Subjects were never told about their counterparts' previous choices to prevent repeated game effects. After the subjects played the ten games, only one game was randomly selected for payment in order to prevent income effects.⁷ The last two points imply that each game can be treated as a one-shot game –even though subjects were always matched with the same subject(s).

In the eight sessions, we employed the strategy method to elicit the punishment behavior in the second stage, i.e. the subjects had to indicate for both allocations in each of the ten games the number of points (0–50) they wanted to assign to the other subject(s). In principle, the strategy method might induce a different behavior than the specific response method, where subjects face given, known choices for one allocation or the other.⁸ However, Falk et al. (2005) investigate this issue and find no differences in subjects' punishment patterns, although the strength of punishment is somewhat lower overall with the strategy method. Thus, the existing evidence suggests that the strategy method does not affect the

⁷ It could be argued that this dilutes monetary incentives because subjects make more decisions for the same amount of money. However, a meta-study by Camerer and Hogarth (1999) suggests that this is not the case.

⁸ For other decisions than punishment, there is evidence of no systematic differences in behavior between the strategy and specific response method (Cason and Mui, 1998; Brandts and Charness, 2000).

pattern of punishment, but might possibly lead to an under-representation of actual punishment.

The key reason for using the strategy method was to prevent subjects from receiving any feedback about the first party's choices in any of the ten games, something that would lead to serious confounds: Punishers' mood could change depending on the first party's prior behavior, and this could generate order or history effects which would severely complicate the data analysis.⁹ Since we employ a within-subject analysis in a large range of games, the use of the strategy method seems unavoidable (unless the researcher has access to huge samples in order to control for order effects). Additionally, it maximizes the amount of statistical data gathered, and facilitates comparison with the data from Fehr and Fischbacher (2004), which was also obtained using the strategy method.

We chose our selection of games for two main reasons. First of all, it allows us to discriminate between many different models of other-regarding preferences (including the most prominent in the literature). We show this in section 4.2.1 and Table 4, where we report the predictions of the theories across our games. Note also that the large number of games makes inferences possible with respect to how consistently individuals follow each behavioral rule, thus providing a stress test of the models. Second, we aimed to have a balanced selection, thus taking care of potential confounds. In this respect, note that the games can be categorized according to four criteria (see also the last four columns of Table 4): (1) JPM, i.e. whether a unique joint-payoff maximizing allocation exists in the respective game, (2) PARETO, i.e. whether a Pareto-dominant allocation exists in the respective game, (3) STRICT, i.e. whether a strictly equal allocation exists in the respective game and, (4) EQUITY, which specifies whether the party who can punish is monetarily disadvantaged by the allocation that is less equal. We believe that these four criteria help to omit potential biases towards certain theories. It is possible, for instance, that second and/or third parties are less willing to punish the choice of unequal allocations if they are socially or Pareto efficient: Maybe some third and/or second parties are willing to accept a small disadvantage for a great advantage of the other player. This means that inequity-aversion might be less influential at any social or Pareto efficient allocation, which seems plausible given the evidence from several studies that suggest that deciders often choose socially efficient allocations even at

⁹ As an illustration, consider a second party who first plays against an "unkind" first party and gets angry as a result. This negative emotional state could affect her posterior behavior, even if the new opponent (players should be re-matched when using the specific response method in order to prevent repeated game effects) makes a "kind" choice. In this regard, Fehr and Fischbacher (2004) report spillover effects when using the specific response method in their two treatments where participants played two games with re-matching. To keep this spillover from contaminating their results, they had to restrict the analysis to the games that were played first.

their own material disadvantage (Charness and Rabin 2002, Engelmann and Strobel 2004; Fehr et al. 2006). In contrast, inequity-aversion might be more influential if the game has a strictly equal allocation –in line with this, Güth et al. (2001) find that in ultimatum games an unfair offer is more often rejected if the alternative is a strictly equal instead of a slightly unequal offer. This justifies the category STRICT. Finally, the category EQUITY complements the previous category.¹⁰

The experiment was conducted with the Z-tree software (Fischbacher, 2007) and the participants were recruited with the software “ORSEE” (Greiner, 2004). 255 subjects participated in our experiment, 90 in 2P and 165 in 3P, that is, we observed 45 second and 55 third parties. Most subjects were students from different disciplines of the University of Zurich or the Swiss Federal Institute of Technology in Zurich (9 percent of them came from the faculty of economics and management). They earned on average 30 Swiss Francs (around \$ 24) which included a show-up fee of 10 Swiss Francs (this fee could be accordingly reduced if one subject got a negative point score as a result of heavy punishment, although this never happened). The sessions lasted approximately 60 minutes. The instructions for the second (in 2P) and third party (in 3P) are in the appendix.

4. Experimental Results

This section starts with a very brief overview of third and second party punishment on an aggregate level. The major part of the section is devoted to the analysis of third and second party punishment on the *individual* level, where we present a classification procedure to thoroughly study the driving motivations behind third and second party punishment. We finish this section with an analysis and comparison of the strength of third and second party punishment.

4.1 Third and Second Party Punishment: Aggregate Overview

We observe frequent punishment in both treatments. In 3P, 54 percent of the third parties punish at least once. Furthermore, third parties spend on average 12.7 points per game to punish, more precisely, 8.6 and 4.1 points on the first and the second party, respectively.

¹⁰ Recall that the two main objectives of this paper are (i) testing recent theories of other-regarding preferences, and (ii) comparing the pattern of second and third party punishment. To avoid over-length, we decided against presenting in this paper the full analysis of issues like the punishment of socially efficient (or even Pareto efficient) choices, the sanctioning of by-standers, or the punishment in the presence of a strictly equal allocation. This analysis can be obtained from the authors and will be provided in a separate working paper.

Table 2 summarizes the frequency and strength of third party punishment in each allocation of each game, distinguishing between punishment for first and second parties.

Table 2— FREQUENCY AND STRENGTH OF PUNISHMENT
THIRD PARTIES

Game			First Party				Second p. (By-stander)			
			Left		Right		Left		Right	
1	(150,150)	vs. (590,60)	.06	(0.3)	.44	(14.7)	.09	(0.5)	.04	(0.3)
2	(100,100)	vs. (50,530)	.11	(2.9)	.06	(0.4)	.04	(0.9)	.26	(9.3)
3	(560,60)	vs. (120,140)	.45	(14.7)	.07	(0.8)	.06	(0.3)	.15	(1.5)
4	(150,90)	vs. (50,630)	.29	(3.8)	.07	(1.2)	.04	(0.7)	.26	(6.9)
5	(220,260)	vs. (220,400)	.24	(3.2)	.09	(0.9)	.13	(1.5)	.22	(5.5)
6	(280,240)	vs. (390,240)	.22	(3.6)	.33	(7.7)	.11	(0.9)	.13	(1.0)
7	(250,80)	vs. (80,250)	.29	(6.6)	.02	(0.1)	.02	(0.1)	.24	(4.1)
8	(100,100)	vs. (50,150)	.06	(0.4)	.04	(0.5)	.06	(0.4)	.18	(2.0)
9	(250,150)	vs. (110,290)	.26	(5.0)	.06	(1.3)	.02	(0.4)	.22	(4.5)
10	(250,150)	vs. (330,70)	.26	(5.0)	.44	(12.8)	.11	(0.5)	.04	(0.1)

Note: Average points spent for punishment by all participants in parentheses. The endowment of the third party is always 200 points. 55 observations in each allocation of each game.

Table 3— FREQUENCY AND STRENGTH OF PUNISHMENT
SECOND PARTIES

Game			Left		Right	
1	(150,150)	vs. (590,60)	.02	(0.2)	.42	(14.7)
2	(100,100)	vs. (50,530)	.18	(4.1)	.11	(2.3)
3	(560,60)	vs. (120,140)	.31	(10.3)	.13	(2.9)
4	(150,90)	vs. (50,630)	.40	(9.6)	.16	(2.7)
5	(220,260)	vs. (220,400)	.40	(10.6)	.16	(4.0)
6	(280,240)	vs. (390,240)	.31	(8.2)	.36	(12.7)
7	(250,80)	vs. (80,250)	.38	(9.1)	.16	(2.9)
8	(100,100)	vs. (50,150)	.07	(1.7)	.16	(2.6)
9	(250,150)	vs. (110,290)	.40	(13.6)	.13	(4.0)
10	(250,150)	vs. (330,70)	.31	(8.1)	.47	(14.3)

Note: Average points spent for punishment by all participants in parentheses. 45 observations in each allocation of each game.

In 2P, 60 percent of the second parties punish at least once. Second parties spend on average 13.8 points per game to punish. Table 3 illustrates the frequency and strength of second party punishment in each allocation of each game. We find that the pattern of actual punishment is very well anticipated in 3P and 2P (recall that we elicited the punishment

expectations of first parties in 2P and first and second parties in 3P). For instance, there are eight games in which both third and second parties significantly punish the first party more strongly in one allocation ($p < 0.05$; Wilcoxon-Signed Rank-Test), and this is anticipated by the first parties in each of these eight games ($p < 0.01$; Wilcoxon-Signed Rank-Test) (see also Figures A and B in the appendix). The behavior of the first parties in 3P and 2P, which is not the focus of our study, can be seen in Table A in the appendix.

4.2 The Occurrence of Third and Second Party Punishment: *Individual Analysis*

We now turn to a precise analysis on an individual level and provide answers to important questions like: Do third and second parties follow any consistent behavioral patterns? Can we classify the punishers into different types? Which *parsimonious* theory fits our data best? For this, we use the classification procedure from El-Gamal and Grether (1995). This procedure has the crucial advantage that it circumvents the multicollinearity problems that would appear in a classical regression analysis if the decision theories entered as independent variables. Moreover, it allows appropriate inferences even when testing all possible theories –no matter how similar their predictions are– at the same time.¹¹

The procedure posits that third and second parties follow deterministic decision rules which may differ from subject to subject, but also that they tremble with probability $\varepsilon > 0$, in which case their behavior is random. By selecting the decision rule that best fits each subject's behavior, we can classify subjects in types. Further, we can also find the best single decision rule in 2P and 3P, or the combination of two, three, etc. decision rules that best account for the behavior in all ten games. Given this, we can then apply the Akaike information criterion to infer the number of decision rules necessary to provide a parsimonious explanation of punishment in our games.

4.2.1 Decision Rules in 3P and 2P

We specify here the examined decision rules for 3P and 2P, many of which correspond to recent models of other-regarding preferences. Although one could describe the rules for both treatments at the same time, we do it separately for expositional convenience; starting with the formal description of the 2P rules. Since our main focus here is on the occurrence of punishment, we restrict our analysis to *binary* decision rules, that is, rules indicating only whether the subject punishes and not the strength of punishment –i.e., the precise number of

¹¹ Multicollinearity problems may occur as soon as theories share predictions in some allocations (a common thing in our games, given the large number of theories that we test). For instance, this is the case for the theories that predict no punishment of the second party in 3P and hence share predictions in 20 out of 40 allocations.

punishment points assigned. Since second parties in 2P make a total of 20 decisions (one for each of the two allocations in each of the ten games), a decision rule in 2P consists of a vector of 20 ones and zeros: It takes value one if the rule predicts punishment at the corresponding allocation and zero if it predicts no punishment. Thus, there are in principle 2^{20} possible binary decision rules in 2P. For simplicity, however, we focus our attention on nine binary rules that correspond to the different theories which provide a rationale for costly punishment.¹² Letting $[x(L)_{FP}, x(L)_{SP}]$ refer to the left-hand and $[x(R)_{FP}, x(R)_{SP}]$ to the right-hand allocation at any game (with FP denoting first party and SP denoting second party), they are defined as follows:

- The “selfish” rule consists of a vector of 20 zeros and predicts never punishment. A standard, self-interested player would follow this rule.
- The “inequity-aversion” (IA) rule indicates that punishment should only occur at allocation $k = (L, R)$ if $x(k)_{FP} > x(k)_{SP}$. A second party with a utility function as in Fehr and Schmidt (1999) would punish according to this rule if her coefficient α of aversion to disadvantageous inequity was larger than 0.5.¹³ A second party with a utility function as in Falk and Fischbacher (2006) would also follow this rule –a proof of this can be requested to the authors.
- The “reciprocity” (RE) rule indicates that punishment should only occur at allocation $k = (L, R)$ if $x(k)_{SP} < x(a)_{SP}$ (a denotes the alternative allocation in the corresponding game). Second parties with a utility function as in Dufwenberg and Kirchsteiger (2004) would punish according to this rule under convenient parameterizations.¹⁴ Something similar can be said with respect to Cox et al. (2007).¹⁵ Intuitively, second parties punish the first party if the latter chooses the allocation of the game giving the second party the strictly lowest payoff –i.e. if the first party harmed the second party.

¹² We do not report the complete analysis here. For instance, we also tested a large number of decision rules that are composed of more than one motivation, e.g. an “inequity-aversion and reciprocity” rule predicting punishment when either inequity aversion or reciprocity models predict it. Including such more complex rules did not significantly improve the model. The results are available upon request.

¹³ In contrast, a player with a smaller α would never punish and hence would follow the selfish rule (something analogous occurs with the rest of the rules that we consider here, that is, they are conditional on the parameters values). Note however that we do not need to be precise a priori on the distribution of parameters in the population: Indeed this is part of what we aim to clarify with the classification analysis.

¹⁴ A formal proof of this can be requested to the authors. Intuitively, this happens because in our games the second party always gets a smaller payoff if the first party chooses the allocation k such that $x(k)_{SP} < x(a)_{SP}$ and not the alternative one, *whatever* the second party is expected to do at these two allocations. The only exception to this is game 6, where we assume that second parties have second order beliefs such that no punishment is expected.

¹⁵ This model allows for some unconditional malevolence (measured by a parameter θ_0). A second party with a very negative θ_0 would follow the spite rule (see next), not the reciprocity rule.

- The “spite” rule consists of a vector of 20 ones. A sufficiently spiteful second party with a utility function as in Kirchsteiger (1994) and Levine (1998) should punish the co-player at all allocations.¹⁶
- The “anti-greed” (AG) rule predicts punishment at allocation $k = (L, R)$ if $x(k)_{FP} > x(a)_{FP}$. This is inspired by Levine (1998): A second party following this rule punishes the first party if the latter chose the allocation maximizing her own money payoff, the intuition being that second parties punish selfish or greedy first parties.
- The “efficiency” (EF) rule predicts punishment at allocation $[x(k)_{FP}, x(k)_{SP}]$ only if $x(k)_{FP} + x(k)_{SP} < x(a)_{FP} + x(a)_{SP}$. That is, a second party of this type punishes the first party if the latter chose the least efficient allocation of the game (i.e., the allocation with the smallest sum of payoffs); the intuition being that second parties punish deviations from a norm of social efficiency (López-Pérez, 2008).
- The “equity” (EQ) rule predicts punishment at allocation $[x(k)_{FP}, x(k)_{SP}]$ only if $|x(k)_{FP} - x(k)_{SP}| < |x(a)_{FP} - x(a)_{SP}|$. A second party of this type punishes the first party if the latter chose the least equitable allocation of the game (i.e., the allocation with the largest distance between players’ payoffs); the intuition being that second parties punish deviations from a norm of equity (Elster, 1989 and López-Pérez, 2008).
- The “maximin” (MA) rule predicts punishment in allocation $[x(k)_{FP}, x(k)_{SP}]$ only if $\min\{x(k)_{FP}, x(k)_{SP}\} < \min\{x(a)_{FP}, x(a)_{SP}\}$. In other words, this rule predicts punishment for the first party if she does not choose the maximin allocation, maybe because that constitutes a deviation from a “maximin norm”.
- The ‘competitiveness’ (C) rule, inspired by Levine (1998), predicts punishment at those allocations where $x(k)_{FP} \leq x(k)_{SP}$. This is the opposite of inequity-aversion, that is, second parties punish in order to increase an already positive income difference.¹⁷

In 3P, third parties make two different punishment decisions in each of the 20 allocations (they can punish the first and/or the second party). Therefore, decision rules in 3P consist of vectors of 40 ones and zeros. We focus on thirteen decision rules, which partly correspond to existing theoretical approaches. The first eight of them are based on the 2P

¹⁶ This might seem a very stringent prediction, but recall that our experimental design was such that only one allocation was chosen for payment in both treatments. A spiteful type would, therefore, punish in all allocations.

¹⁷ Although Levine allows for the existence of spiteful types that should punish indiscriminately, and estimates that around 20 percent of the population correspond to this type, he suggests later that “one explanation of spite is that it is really “competitiveness,” that is, the desire to outdo opponents” (Levine 1998, p. 614).

rules mentioned above: (1) The “selfish” rule consists of a vector of 40 zeros, (2) the “inequity-aversion” rule is a logical extension of the inequity-aversion rule in 2P predicting punishment of the first *and/or* second party when they have a larger payoff than the third party¹⁸, (3) the “spite” rule is a vector of 40 ones, the (4) “anti-greed”, (5) “efficiency”, (6) “equity” and (7) “maximin” rules are defined like in 2P (note that they never predict punishment of the second party in 3P), and (8) the “competitiveness” rule predicts punishment of the first *and/or* second party when they have a smaller or equal payoff than the third party.¹⁹ Further, (9) the “ERC” rule (Bolton and Ockenfels, 2000) predicts punishment of the first *and/or* the second party in allocation $[x(k)_{FP}, x(k)_{SP}, 200]$ if $400 < x(k)_{FP} + x(k)_{SP} < 600$, i.e., if the third party can use punishment to bring her relative payoff closer to 1/3, the equitable relative payoff in three-player games.²⁰ Finally, we include some rules for third parties that are based on ideas related to inequity-aversion and reciprocity and that allow us to investigate if third parties have different motivations than second parties. They are (10) the “indirect reciprocity” rule (inspired by Nowak and Sigmund, 2005 and Seinen and Schram, 2006) predicting punishment of the first party if $x(k)_{SP} < x(a)_{SP}$ and no punishment of the second party, (11) an “envy-active” rule predicting punishment of the first party in the same conditions as the inequity-aversion rule, but no punishment of the second party, i.e. people who follow this rule punish richer players only if they are responsible for the outcome, (12) an “egalitarian” rule (Dawes et al., 2007) that predicts punishment (in our games) of those co-players getting a payoff larger than the *average* one,²¹ and finally, (13) the “envy-

¹⁸ A third party with a utility function as in Fehr and Schmidt (1999) would punish according to this rule if her coefficient α of aversion to disadvantageous inequity was larger than 2. This is more demanding than in the case of the second party, as α is weighted in the model by $1/(n-1)$, where n denotes the number of players. If both co-players are richer, moreover, punishing one party increases the distance to the other party, so that a larger α is required.

¹⁹ We do not include a reciprocity rule. To start, Dufwenberg and Kirchsteiger (2004) predict multiple equilibria depending on second order beliefs. However, the third party never punishes in equilibrium if her second order beliefs are such that no punishment ever occurs. This prediction, therefore, coincides with the selfish rule. Finally, Cox et al. (2007) only applies to two-player games. The authors propose a extension to n -player games, however, which predicts no punishment in the 3P treatment unless the player’s unconditional term θ_0 is very negative –such a player would follow the spite rule.

²⁰ To see this, let $\sigma = x(k)_{FP} + x(k)_{SP} + 200$ denote the sum of players’ payoffs and hence $200/\sigma$ denote the third party’s relative payoff, and note that (i) this relative payoff is smaller than 1/3 if $400 < x(k)_{FP} + x(k)_{SP}$, and (ii) a unit of punishment increases the relative payoff if $\frac{200-1}{\sigma-3-1} > \frac{200}{\sigma} \Leftrightarrow \sigma < 800 \Leftrightarrow x(k)_{FP} + x(k)_{SP} < 600$. Note that we do not include an ERC rule in our analysis of second party punishment because it shares predictions with the inequity aversion rule.

²¹ More generally, this motive predicts punishment if that reduces the standard deviation from the group mean. This coincides with inequity aversion in two-player games and in many three-player games. In our 3P games, inequity aversion and the egalitarian motive can be discriminated in game 8. While inequity-averse third parties should not damage the bystander in the allocation (50/150), third parties following the egalitarian rule should do

perspective” rule predicting punishment of the first party if $x(k)_{FP} < x(k)_{SP}$, and no punishment of the second party (third parties who follow this rule put themselves in the shoes of an inequity-averse second party).

Table 4 presents the predictions of the non-trivial rules in each of our ten games for both third and second party punishment. The first two columns show the allocations available in each game. The following two columns correspond to the 2P treatment and indicate the rules that predict punishment in each game, distinguishing between the left-hand and right-hand allocation. As an illustration, take game 8 (100/100 vs. 50/150). We can observe that a reciprocal second party would punish in the left-hand allocation because that choice harms her, while a competitive individual would punish at every allocation in order to increase a positive distance. Note also that an inequity-averse second party never punishes the first party because, whatever the first party’s choice, the second party always gets a larger or equal payoff. The next four columns illustrate the punishment predictions for 3P; we first show the punishment predictions for the first party and then the punishment for the bystander.²² Finally, the last four columns indicate the characteristics of the games referred in section 3.

INSERT TABLE 4 ABOUT HERE

so. We find some support for the egalitarian rule since 18 percent damage the bystander then. However, in the classification analysis for third parties the inequity aversion rule outperforms the egalitarian rule.

²² If a theory predicts that the third party is indifferent between punishing the first party or the by-stander at one allocation, we take as compatible with the theory the punishment of any of those two parties.

4.2.2 Estimation of the Error Rate

The classification procedure posits that each subject follows one of the above mentioned decision rules but allows for mistakes. More precisely, subjects may tremble in each allocation with probability $\varepsilon > 0$, in which case it is assumed that they randomize with equal probability between punishing or not punishing.²³ This means that the probability that a subject s deviates from her rule at any allocation is $\frac{\varepsilon}{2}$. Consequently, if X_s denotes the actual number of times that s has acted in accordance with the rule out of her d choices in the experiment (20 in 2P, 40 in 3P), the probability that such behavior has been generated by her rule is:²⁴

$$\left(1 - \frac{\varepsilon}{2}\right)^{X_s} \times \left(\frac{\varepsilon}{2}\right)^{d - X_s} .$$

To find the maximum likelihood estimate $\hat{\varepsilon}$ of the error rate, consider first the simplest case: All subjects follow the same decision rule. In that case, $\hat{\varepsilon}$ maximizes the overall likelihood across all n players

$$\max \prod_{s=1}^n \left(1 - \frac{\varepsilon}{2}\right)^{X_s} \times \left(\frac{\varepsilon}{2}\right)^{d - X_s} . \quad (1)$$

One can then prove by applying standard optimization techniques (consult the appendix) that $\hat{\varepsilon}$ coincides with twice the proportion of overall deviations, that is,

$$\hat{\varepsilon} = \frac{2 \cdot (d \times n - \sum_s X_s)}{d \times n} . \quad (2)$$

By computing $\hat{\varepsilon}$ for every possible rule of each treatment, we can then find the optimal decision rule in the maximum likelihood sense, i.e. that maximizing function (1) given the data. This procedure can be extended to the case where different agents use different rules. If we assume that there are two types of players, for instance, we can find the optimal pair of rules by applying the following three-step algorithm to any pair of possible rules A and B: (a) We assign each individual s to the rule that minimizes the number of actual deviations $d - X_s$ (in case of a tie, we assign "half" of an individual to each rule), (b) we use

²³ To simplify the analysis, we assume that all subjects tremble with the same probability in any allocation. This is probably a realistic assumption in view that the punishers' decision problem is, from a strategic point of view, undemanding, so that no change of ε through time (due to learning effects) should be expected.

²⁴ In computing this, we posit that choices across allocations and games are independent –i.e., the probability of following the rule at any allocation does not depend on what the subjects did before. This seems reasonable in our experiment because (1) subjects are given no feedback and hence there appears to be no reason for changes in mood, and (2) since the punisher's decision problem is arguably easy, we do not expect any learning effects.

expression (2) and the experimental data to find $\hat{\varepsilon}$, and (c) we compute the probability that our data has been generated by the partition of the players generated in step (a), that is,

$$\prod_{i \in A} \left(1 - \frac{\varepsilon}{2}\right)^{X_i} \times \left(\frac{\varepsilon}{2}\right)^{d - X_i} \cdot \prod_{j \in B} \left(1 - \frac{\varepsilon}{2}\right)^{X_j} \times \left(\frac{\varepsilon}{2}\right)^{d - X_j} \quad (3)$$

The optimal pair of rules maximizes equation (3). Finally, if we assume that our subject pool follows three or more rules, the procedure applies analogously.

4.2.3 Results of the Classification Procedure

Tables 5 and 6 summarize the results of the classification procedure in 2P and 3P. The second column in each table indicates the best single rule in that treatment, the best pair of rules, and so on. The third column indicates the percentage of second and third parties that follow each rule. The fourth column reports the estimated error $\hat{\varepsilon}$ - recall that the probability that a subject deviates from her rule at any allocation is equal to $\frac{\hat{\varepsilon}}{2}$. Note in this regard that the success of our model (measured by how small $\hat{\varepsilon}$ is) increases as the number of rules k increases. This is intuitive as the overall likelihood (3) increases as k increases.²⁵ However, our model also becomes more complex as k increases and hence it would be desirable to introduce a penalty for allowing "too many" decision rules. To provide an indication of the optimal number of rules in each treatment, the fifth column of each table reports the log-likelihood - for the best two rules, for instance, this is the log value of (3) - less the number of parameters $(d + n) \cdot k$.²⁶ According to the Akaike information criterion (AIC), the optimal model should maximize this number. Finally, the sixth column of each table reports the results from a likelihood ratio test of goodness of fit, to be described later.

²⁵ The same logic applies here as in a linear regression model, where the coefficient of determination R^2 increases with the number of independent variables.

²⁶ In a model with k rules, we must first estimate each rule, which consists of d zeros and ones (hence the number $d \cdot k$) and moreover we have to find the rule each subject follows or those he or she does not follow (hence the number $n \cdot k$).

TABLE 5— RESULTS OF CLASSIFICATION PROCEDURE IN 3P (THIRD PARTIES)

Number of rules	Rule(s) chosen	Percentage of third parties (N = 55)	ϵ	AIC	Chi-squared (p-Value)
1	selfish	100%	0.309	-1042.2	
2	selfish, inequity-aversion	78%, 22%	0.182	-862.5	549.31 (0)
3	selfish, inequity-aversion, envy-perspective	66%, 20%, 14%	0.162	-905.6	653.08 (0)
4	selfish, inequity-aversion, envy-perspective, spite	66%, 16%, 14%, 4%	0.151	-971	712.24 (0)

RESULT 1: *A combination of inequity-averse and selfish types can sufficiently capture third parties' punishment patterns. If we allow for more complexity, a combination of two different inequity-averse types, selfish and spiteful types best explains the third parties' punishment pattern.*

Evidence for Result 1: As Table 5 shows, the classification procedure detects the following behavioral patterns for third parties: (1) if we force the algorithm to choose only one rule, the selfish rule is picked. A large number of subjects *never* punish and hence the selfish rule fits their behavior perfectly, so that the error rate is already considerably small (0.309 in 3P). The error rates of all other rules are at least twice as high (e.g. inequity-aversion rule: 0.769, envy-perspective rule: 0.667, spite rule: 1). (2) If we force the algorithm to choose the best pair of rules, it selects the selfish rule together with the inequity-aversion rule. Then 22 percent of the third parties are classified as inequity-averse and the error rate drops to 18.2 percent.²⁷ (3) Adding a third rule is suboptimal according to the AIC, which suggests that the assumption that there are just selfish and inequity-averse types can sufficiently capture the punishment pattern in 3P. (4) If we nevertheless add a third rule, the algorithm picks the envy-perspective rule. (5) If we add a fourth rule, spite is chosen. ▀

Although the AIC recommends not introducing a third rule if parsimony is our main goal, the comparatively good performance of the envy-perspective rule is an illustration of how this classification procedure can be used to provide new intuitions on punishment.²⁸ From our knowledge, no experimental paper has provided evidence on this rule before. We

²⁷ In comparison, El-Gamal and Grether (1995) study decisions under uncertainty and find an error rate of 0.312 when looking for the best pair of decision rules.

²⁸ Subjects following this rule punish as an inequity-averse second party would do in 2P. For this reason, one might be tempted to think that they just misunderstood the experimental instructions and thought that they were second parties. This is very unlikely, though, as their screens always indicated that they were third parties and they had to indicate their punishment for the first *and* the second party at each allocation.

speculate that the third parties who follow this rule might be motivated to alleviate the distress of the poorest, weakest party in case that party cannot defend herself (i.e., if she is passive). More experimental evidence, in any case, is required for a better understanding of this kind of behavior.

RESULT 2: *A combination of inequity-averse and selfish types can sufficiently capture second parties’ punishment patterns. If we allow for more complexity, a combination of inequity-averse, selfish, spiteful and reciprocal types best explains second parties’ punishment patterns.*

Evidence for Result 2: Table 6 indicates the following behavioral patterns for second parties: (1) If we force the algorithm to choose only one rule, the selfish rule is picked. This happens because a large number of subjects *never* punish and hence the selfish rule fits their behavior perfectly. The error rate of 0.502 is therefore quite small compared to that of other rules. The second lowest error rate comes from the inequity-aversion rule which is 0.731, the error rate of the reciprocity rule is 0.798, and the error rate of any other rule is 1. (2) If we force the algorithm to choose the best pair of rules, it selects the selfish rule together with the inequity-aversion rule. We can also see that a considerable fraction of 42 percent is then best classified as inequity-averse. Moreover, we observe that when using these two rules, the error rate is rather low (29 percent). (3) Adding a third rule is suboptimal according to the Akaike information criterion, which suggests that the punishment pattern of second parties can be sufficiently captured by the assumption that there are just selfish and inequity-averse types. (4) However, if we add a third rule, the algorithm picks the spite rule, and 29 percent of the second parties are now classified as inequity-averse and 13 percent as spiteful. (5) If we add a fourth rule, reciprocity is chosen. ▪

TABLE 6— RESULTS OF CLASSIFICATION PROCEDURE IN 2P (SECOND PARTIES)

Number of rules	Rule(s) chosen	Percentage of second parties (N = 45)	ϵ	AIC	Chi-squared (p-Value)
1	selfish	100%	0.502	-572.2	
2	selfish, inequity-aversion	58%, 42%	0.290	-503.4	267.52 (0)
3	selfish, inequity-aversion , spite	58%, 29%, 13%	0.220	-506.9	390.66 (0)
4	selfish, inequity-aversion , spite, reciprocity	53%, 22%, 13%, 11%	0.193	-545.9	442.58 (0)

We note that the results of our classification analysis are in line with Charness and Rabin (2002, p. 838) who suggest that, considering distributional preferences alone (i.e., no reciprocity) and when no self-interest is at stake, approximately 20 percent of their observed behavior can be attributed to difference (i.e., inequity) aversion and 10 percent to spite.²⁹ Further, and although Dawes et al. (2007) and Falk et al. (2005) do not perform any classification analysis, they cite some behaviors that seem to be explainable only by spite and which have a frequency in line with our previous results. Thus, around 15 percent of the participants in one treatment of Falk et al. (2005) defected *and* punished other players in a Prisoner's Dilemma with a punishment stage (the authors also show that this kind of punishment is very sensitive to its cost; the result that we have just cited corresponds to a punishment technology that is similar to the one available in our experiment).

Observe that the Akaike criterion suggests in both treatments that a model with two, three or four rules is better than one with just one single rule. To further clarify this point, we performed a likelihood ratio test to contrast the null hypothesis that a restricted model with only one rule fits the data similarly well as an unrestricted model with 2, 3, and 4 rules. From the table, we see that we always very strongly reject the null hypothesis.³⁰

To sum up, our classification analysis shows that a model assuming two types of players (selfish and inequity-averse) best explains the occurrence of punishment in our two treatments, while alternative and *equally parsimonious* models perform worse. This does not mean, of course, that inequity-aversion can account for the occurrence of all punishment in our games: As our classification analysis shows, reciprocity plays also a role in 2P, and other variables like spite affect third and second party punishment. Further, the fact that the error rate ε is never zero indicates that many punishers do not follow strictly a simple decision rule, but take several factors into account when deciding whether to punish.

4.3 The Strength of Third and Second Party Punishment

The disadvantage of the classification procedure is that, due to complexity, it makes more sense to investigate the occurrence of punishment only and abstract from its strength.

²⁹ Note: Charness and Rabin (2002, p. 823) use the term 'competitive preferences' to refer to what we call 'spite'.

³⁰ Since negative twice the log-likelihood ratio is asymptotically distributed as chi-squared with degrees of freedom equal to the number of restrictions, large values of the chi-squared statistic reject the null hypothesis. Note that the number of restrictions is d , $2d$, and $3d$ as we restrict 1, 2, and 3 rules, respectively, to coincide with another rule.

While this is not a problem when testing most theoretical models, we may lose information concerning some models of inequity-aversion and reciprocity which respectively forecast a positive relation between the strength of punishment and the difference in payoffs and the size of the harm, -we say that the second party was harmed if the difference $x(k)_{SP} - x(a)_{SP}$ (see the definition of the reciprocity rule in section 4.2.1) is negative; the size of the harm is equal to the absolute value of that term, that is, the net payoff loss of the second party.

We first briefly consider third parties. In this respect, an OLS analysis shows that their average punishment significantly ($p < 0.001$) increases by 6.75 (3.55) points when the difference in payoffs between the first (second) and the third party increases by 100 points (recall that each point spent reduces the payoff of the punished party by 3 points). Therefore, the bigger the difference in payoffs, the more the third party punishes the first and second parties.

We now consider second parties. Table 7 shows the results of five OLS regressions that show whether the difference in payoffs and the size or existence of harm predict the strength of second party punishment. Column (1) and (2) report that, considered in isolation, the difference in payoffs and the size of harm both predict the strength of punishment as suggested, but also that the coefficient for the difference in payoffs is more robust and twice as large as the coefficient for the size of harm. The coefficient of 0.0271 for difference in payoffs means that second parties spend on average 0.0271 points for a payoff disadvantage of one point, i.e. 2.71 points for a payoff disadvantage of 100 points. In column (4), we use the difference in payoffs and the size of harm at the same time in one regression. We can see that when controlling for the difference in payoffs, the size of the harm becomes insignificant. The coefficient for the difference in payoffs remains substantial; the amount of points spent by second parties to punish first parties increases by an average of 2.56 points when the payoff disadvantage increases by 100 points. We also investigate the effect of the sole *existence of harm* by itself. In column (3), we see that a dummy for the existence of harm (which equals one if there was harm, zero otherwise) is a highly significant predictor for the size of punishment when considered in isolation. Second parties are willing to spend 8.62 additional points to punish if they have been harmed. Further, column (5) indicates that the existence of harm alone is also important when we control for the difference in payoffs: The existence of harm then increases punishment by 6.15 points. In summary, subjects punish

more if they have been harmed, but apparently they do not increase the punishment the more they have been harmed. This leads us to our next result.

TABLE 7—DETERMINANTS OF SECOND PARTY PUNISHMENT (*OLS*)

Dependent Variable Model	Strength of Punishment for the First Party				
	(1)	(2)	(3)	(4)	(5)
Difference in payoffs	0.0271*** (0.0055)			0.0256*** (0.0054)	0.0158*** (0.0043)
Size of Harm		0.0133* (0.0055)		0.0043 (0.0055)	
Existence of Harm			8.6169*** (2.3935)		6.1489** (2.454)

Notes: Observations: 540. Data comes from all 27 second parties that punish at least once. Data is clustered on individual level. Robust standard errors in parentheses. Notes: *** 99-percent significance, ** 98-percent significance; * 95-percent significance.

RESULT 3: *The strength of both second and third party punishment positively depends on the difference in payoffs. Further, second party punishment is more intense if the second party was harmed, but the strength does not depend on the size of the harm. This implies that models that combine inequity-aversion with reciprocal motives, like Falk and Fischbacher (2006), perform well in predicting the strength of second and third party punishment.*

Evidence for Result 3: The predictions in column (5) are very much in line with Falk and Fischbacher (2006), who predict a relatively more intense punishment of a "richer" first party if she has *also* harmed the second party (independently on the amount of harm inflicted). We observe further support for their theory when comparing games 9 (250/150 vs. 110/290) and 10 (250/150 vs. 330/70). In both games, the first party can choose the allocation (250/150) and this choice leaves the second party in a disadvantageous position (hence some punishment is predicted). In addition, the choice for (250/150) "harms" the second party in game 9, where the alternative allocation is (110/290) but not in game 10 where the alternative is (330/70). As a result, Falk and Fischbacher predict less punishment by second parties of the choice (250/150) in game 10, a prediction which is supported by our data (Wilcoxon-Signed Rank Test, $z = 2.168$, $p=0.030$).³¹ ■

³¹ We make two remarks in this regard. First, this characteristic of the model is immaterial in the 3P treatment because third parties are never harmed by any other party. Second, a slightly different version of the model (appendix A of Falk and Fischbacher, 2006) predicts a relatively more intense punishment of a richer first party who harmed the second party *only if* the first party is richer than the second party in the *alternative* allocation.

4.4 Comparing the Strength of Third and Second Party Punishment

RESULT 4: *Third party punishment is not generally weaker than second party punishment. Yet, third parties appear to punish more selectively and are especially likely to spend money when their opponents are richer.*

Evidence for Result 4: If we compare how many points third and second parties spend in total, we find no differences (Mann-Whitney Test, $z = 0.608$, $p = 0.543$). This is likely to be explained by two facts: (i) third parties are more sensitive to payoff differences – i.e. they punish a difference of 100 points more than twice as strongly (see section 4.3), and, (ii) third parties spend part of their money to punish bystanders.

INSERT FIGURE 1 ABOUT HERE

This second point is noteworthy: If we look instead at the punishment of the first party, we observe that third parties spend *overall* fewer points than second parties (Mann-Whitney Test, $z = 3.209$, $p = 0.001$). For a more detailed analysis, Figure 1 breaks down what happens in each game. The dots (squares) indicate the average punishment of third (second) parties for the first party alone in the left- and right-hand allocation in each of the ten games (e.g. 3L = game 3, left-hand allocation) –further, a solid (dotted) line connects the second (third) party observations.³² Note that the punishment pattern of third and second parties in figure 1 is rather accurately anticipated by their co-players - figures A and B in the appendix illustrate the average expectation of punishment in each allocation in 3P and 2P. Furthermore, figure 1 illustrates two important findings.

First, the solid line lies below the dotted line in most allocations, which indicates that third parties tend to punish the first parties more weakly than the second parties. However, the differences are significant only in 4 of the 20 allocations (Mann-Whitney Test on a 10 percent level: game 5 and 9 left, game 7 and 8 right). For instance, hardly any third party punishes in the allocations (80/250) in game 7 and in (50/150) in game 8 (2 and 4 percent), whereas 16

This version thus predicts equal punishment of allocation 250/150 in games 9 and 10, which is not consistent with our data.

³² Note that we connected the points in figure 1 just for illustrative reasons. In particular, this does not suggest a temporal ordering. As explained in the experimental design section, the games were played in random order.

percent of the second parties punish in these two allocations. Further, only 24 and 26 percent of the third parties punish in the allocations (220/260) in game 5 and (250/150) in game 9 compared to 40 percent of the second parties. These significant differences can be attributed to three reasons: (i) Second parties punish more if they have been harmed, as it happens in these allocations, (ii) second parties tend to be more spiteful than third parties, as our previous classification of third and second parties indicated, and (iii) because payoff differences between first and third parties are rather small in these allocations. *Second*, the figure reveals that third party punishment can be as intense as that from second parties. Remarkably, there are no differences in the allocations that are punished strongest on average (game 1 right: $z = -0.054$, $p = 0.956$, game 10 right: $z = 0.327$, $p = 0.743$). In the left-hand allocation of game 3, third party punishment is even slightly stronger ($z = -1.382$, $p = 0.167$). These three allocations have in common that the first party has an income that lies well above the income of the other parties. Hence if large payoff differences exist in our games, second and third parties are equally willing to punish, *even* if the second party has been harmed. ■

RESULT 5: *Third party punishment very closely resembles second party punishment, which suggests that third parties do not act more normatively and impartially than second parties.*

Evidence for Result 5: Figure 1 illustrates that the pattern of second and third party punishment is identical in all of the ten games. Always, when the second party punished one allocation significantly more strongly than the alternative (which is the case in eight of the ten games) the third party behaved accordingly and punished the same allocation more strongly. This latter fact suggests that third parties do not punish in a more normative manner, at least under one possible interpretation of the term ‘normative punishment’. More precisely, we say that someone punishes in a normative manner when she punishes another party because that party deviated from a social norm –i.e., a behavioral rule commending how to behave in a game.³³ For instance, a second (in 2P) or a third party (in 3P) would punish in a normative manner if they punish the first party when she deviates from a norm (the equity, efficiency, and maximin norms cited in 4.2.1 are examples), but not otherwise. If second or third parties followed this kind of behavior, they would never punish the first party in *both* allocations of any game (they are not punishing a deviation from a norm here; otherwise, they would not

³³ Admittedly, this is not the only possible definition of the term. For instance, one could also define it as ‘sanctioning in accordance with any social norm regulating punishment’. An example of such a norm is ‘an eye for an eye, a tooth for a tooth’, based on the idea of retributive justice. Even with this meaning, though, we do not find that third parties punish more normatively than second parties. The punishment of the by-stander, frequently observed in some games, seems particularly at odds with the idea of normative punishment.

punish the alternative choice). As we have noted, however, second and third parties punish in both allocations in several games. In addition, a third party who punishes in a normative manner should never punish the by-stander, who is obviously no wrongdoer. Table 2, however, shows that such behavior is not rare in some games. Finally, the classification analysis in section 4.2.3 indicates that second and third parties do not punish deviations from a norm of equity, efficiency, or maximin..

The overall similarity in punishment also sheds doubt on the assumption that third parties are more impartial and suffer less from an egocentric bias because of their unaffected position in the game (Fehr & Fischbacher, 2004). In this respect, we note that in the literature on Welfare Economics, a social welfare function W is said to be symmetric if $W(u) = W(u')$ whenever the utility vector u constitutes a permutation of vector u' . Based on this, we say that a second/third party punishes in an impartial manner if she does not punish the first party for choosing an allocation that maximizes a symmetric W . With this in mind, the behavior in game 7 (250/80 vs. 80/250) speaks against the idea that third parties punish in an impartial manner. Observe that both allocations in this game are symmetric – i.e., a permutation of each other – so that one should expect that an "impartial" third party who uniformly values each player's welfare regards both allocations as equally fair and punish them less than a second party.³⁴ Contrary to this, we observe that both second and third parties punish the first party equally strongly for choosing the allocation (250/80) (Mann-Whitney Test, $z = 0.954$, $p = 0.340$). ▪

To finish, we address two possible objections to our claim that punishment from third parties is not generally weaker than punishment from second parties. *First*, one might argue that this is an artifact of our setting because second parties have a lower endowment in comparison to third parties in some allocations (regardless of the punishment technology which is the same for third and second parties in all allocations). Indeed, if the marginal utility of money is decreasing in our games, parties with a small endowment should be relatively more reluctant to spend money from their already low endowment. *Second*, the use of the strategy method might have an asymmetric effect on the strength of punishment from third and second parties. In principle, a "hotter" environment induced for instance by the specific

³⁴ As an aside, note that an impartial spectator might still consider the choice of allocation (250/80) unfair because it fails to be courteous –i.e., that choice signals that the first party cares more for herself than for the second party. Recall, however, that the classification analysis does not find evidence in favor of this anti-greedy behavior.

response method could increase the strength of punishment from second parties (as in Falk et al., 2005) but not from third parties (since they are unaffected, their reactions might be more independent of the environment).

We can exclude the first objection in our games. Second parties are not more reluctant to spend money if their balance is low. We can see this in an OLS regression analysis, where the endowment of the second party is an uninformative variable to predict punishment in all the cases where the second party endowment is lower than the third party endowment, controlling for the difference in payoffs ($t = -0.77$, $p = 0.446$). In fact, second parties punish especially in games 1, 3 and 10, where their balance is lowest. In summary, the intensity of punishment in our games does not decrease when second parties have a lower balance than third parties.

To address the second objection we conducted an additional experiment in which both third and second parties played *only* one of our games, now using the specific-response method.³⁵ We chose game 1 (150/150 vs. 590/60) because, given our results from the 2P and the 3P treatments, we expected a large amount of punishment, and also because third and second parties punished the allocation (590/60) equally strongly. The experiment was conducted at the Autonomous University of Madrid, and the participants were students from different disciplines (60 subjects participated in the 2P and 75 subjects in the 3P treatment).³⁶ Our data shows that when comparing the behavior of the strategy method with the specific response method, neither second parties nor third parties punish the choice of the allocation (590/60) significantly stronger when using the specific response method (in 3P: $z = -1.544$, $p = 0.123$); in fact, second parties punish even slightly less when using the specific response method (in 2P: $z = 1.857$, $p = 0.063$). Interestingly, in this “hotter” environment (because of the applied specific response method and maybe the location), third parties punish the allocation (590/60) significantly stronger than second parties ($z = 3.596$, $p < 0.01$). This suggests that the use of the strategy method in our games did not lead to an underestimation of second compared to third party punishment.

³⁵ We are thankful for helpful suggestions by Gary Charness.

³⁶ The experimental protocol and the instructions were as similar as possible to those of the Zurich sessions, except that the experiment was not computerized. More information on this experiment is available upon request.

5. Conclusion

We investigate third and second party punishment in a set of ten different games to find out more about the individual motivations behind both types of punishment and to provide insights into the different existing theoretical approaches. The results suggest that inequity-aversion is the crucial (although not the only) cause of third and second party punishment. Our data also suggests that third parties do not act more “normatively” or less egocentrically than second parties, casting doubt on the idea that third parties are more impartial.

The evidence from our experiment has implications for the different theoretical models. To start, models that incorporate inequity-aversion like Fehr and Schmidt (1999) and Falk and Fischbacher (2006) fare relatively better in explaining the occurrence of punishment in 3P and give also rather good predictions in 2P. These models (especially Falk and Fischbacher, 2006) are also more accurate in predicting the strength of punishment. Reciprocity models like Dufwenberg and Kirchsteiger (2004), and Cox et al. (2007) are less accurate, especially with regard to third party punishment. Levine (1998) is inconsistent with the heavy punishment of socially and Pareto efficient actions, and with the role that strict equality plays in reducing punishment (reciprocity also faces this problem). In turn, norm approaches face an unanticipated problem in 2P and 3P: There seems to be no way to explain punishers’ choices as a reaction to a prior deviation from any *sensible* norm of distributive justice (taking standard concepts like social efficiency, equity, or maximin into account). A clear illustration of this is that both allocations are punished in some games or that bystanders are damaged by third parties. Under one possible interpretation of the term, our data suggests that third party punishment is not more “normative”. This is not to say, though, that norms are unimportant in explaining punishment, as many third and second parties (even if inequity-averse) might rationalize their punishment as a reaction to a prior violation of a norm, as the classical philosopher Seneca noted: “Reason wishes the decision that it gives to be just; anger wishes to have the decision which it has given seem the just decision”. People might not punish normatively, but they are likely to believe that they do so.

The results from our classification analysis can be used for predictive purposes. For example, it is a very natural question how a change in the third party endowment could affect punishment. In this regard, our analysis suggests that one group of third parties (the inequity-averse) will probably stop punishing if their endowment rises enough, that is, if they are richer than the other parties. In contrast, other groups (the envy-perspective and spiteful ones)

might punish even if they are richer than the other parties (the envy-perspective group would punish only if the first party is richer than the second party). Our evidence provides support in this regard, as we observe less, albeit still some punishment in those allocations where the third party is richer than their co-players (as in some allocations in games 1, 2, 3, 4, and 8), *especially* if in addition the first party is richer than the second party; 29 percent of the third parties punish the first party if she chooses allocation (150, 90) in game 4.

We finish with some possible ideas for future research. First, all the models we have considered deal exclusively with monetary punishment. Hence, none of them is consistent with the idea that people can punish others by non-monetary means (insults, humiliating speech, etc). When do second and third parties use this kind of sanctions and when are they useful in preventing undesirable behavior? Second, since our main objective in this paper was studying and comparing the motives for second and third party punishment, our games have just one sanctioning party. However, it could be interesting to study what happens when there are multiple parties who can punish, as they might be less willing to punish, on the idea that "others will do it" – this could have to do with the phenomenon of responsibility alleviation reported in Charness (2000). Finally, Falk et al. (2005) report that some defectors in a prisoner's dilemma punish other players, in particular when the cost of sanctioning is cheap. This correlation between punishment and its cost is not as pronounced for the punishment of defectors by cooperators, and suggests that some type of punishment (spiteful?) is more sensitive to its cost than others (inequity-averse, reciprocal?), a topic that deserves also further study.

Tables

TABLE 4— THEORETICAL PUNISHMENT PREDICTIONS AND GAME CHARACTERISTICS

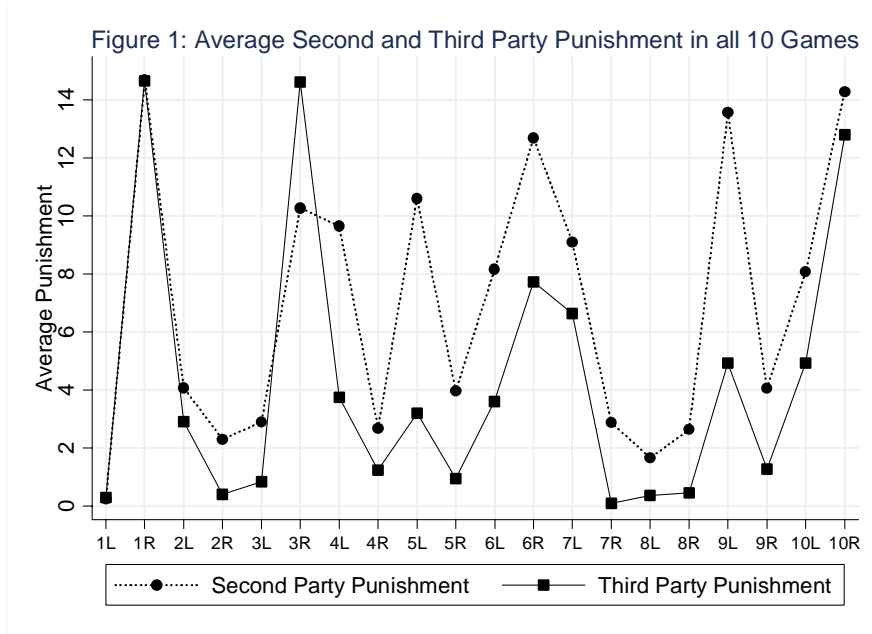
Game	Theories predicting punishment in 2P		Theories predicting punishment for the first party in 3P		Theories predicting punishment for the by-stander in 3P		Game Characteristics					
	Allocation		Allocation		Allocation		Allocation		JPM	Pareto	Strict	Equity
	Left	Right	Left	Right	Left	Right	Left	Right				
1	(150,150)	vs. (590,60)	EF, C	IA, RE, EQ, AG	EF, C	IA, EQ, AG	C	C	yes	no	y	y
2	(100,100)	vs. (50,530)	RE, EF, AG, C	EQ, C	EF, AG, C	ERC, EQ, C	C	IA, ERC	y	n	y	n
3	(560,60)	vs. (120,140)	IA, RE, EQ, AG	EF, C	IA, EQ, AG	EF, C	C	C	y	n	n	y
4	(150,90)	vs. (50,630)	IA, RE, EF, AG	EQ, C	EF, AG, C	EQ, C	C	IA	y	n	n	n
5	(220,260)	vs. (220,400)	RE, EF, C	EQ, C	IA, ERC, EF	IA, EQ	IA, ERC	IA	y	y	n	n
6	(280,240)	vs. (390,240)	IA, EF	IA, EQ, AG	IA, ERC, EF	IA, EQ, AG	IA, ERC	IA	y	y	n	y
7	(250,80)	vs. (80,250)	IA, RE, AG	C	IA, AG	C	C	IA	n	n	n	-
8	(100,100)	vs. (50,150)	RE, AG, C	EQ, C	AG, C	EQ, C	C	C	n	n	y	n
9	(250,150)	vs. (110,290)	IA, RE, AG	EQ, C	IA, AG	EQ, C	C	IA	n	n	n	n
10	(250,150)	vs. (330,70)	IA	IA, RE, EQ, AG	IA	IA, EQ, AG	C	C	n	n	n	y

IA = Inequity-aversion, RE = Reciprocity, AG = Anti-Greed, EQ = Equity rule, EF = Efficiency rule, ERC = Bolton-Ockenfels (in 3P), C = Competitiveness.

JPM = Is there a joint payoff maximizing allocation available in the respective game? Pareto = Is there a Pareto-dominant allocation available in the respective game? Strict = Is there a strictly equal allocation available in the respective game? Equity = If the less equal allocation is chosen, has the second party a lower payoff than A?

Figures

Figure 1: Average Second and Third Party Punishment in all 10 Games

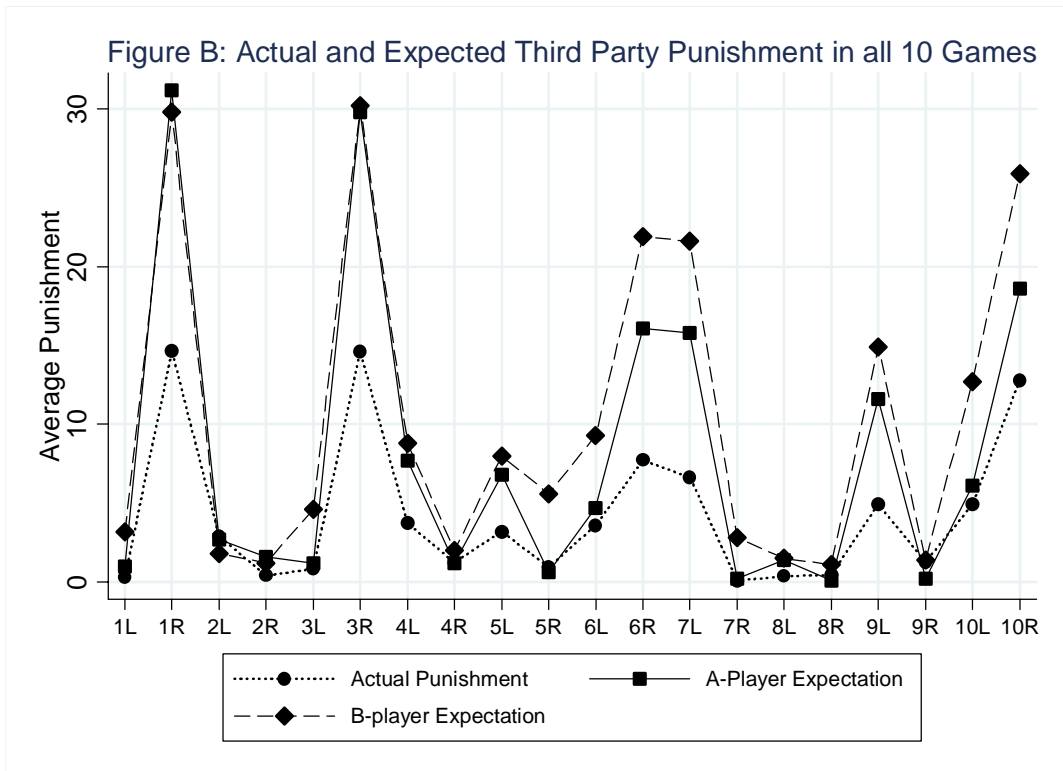
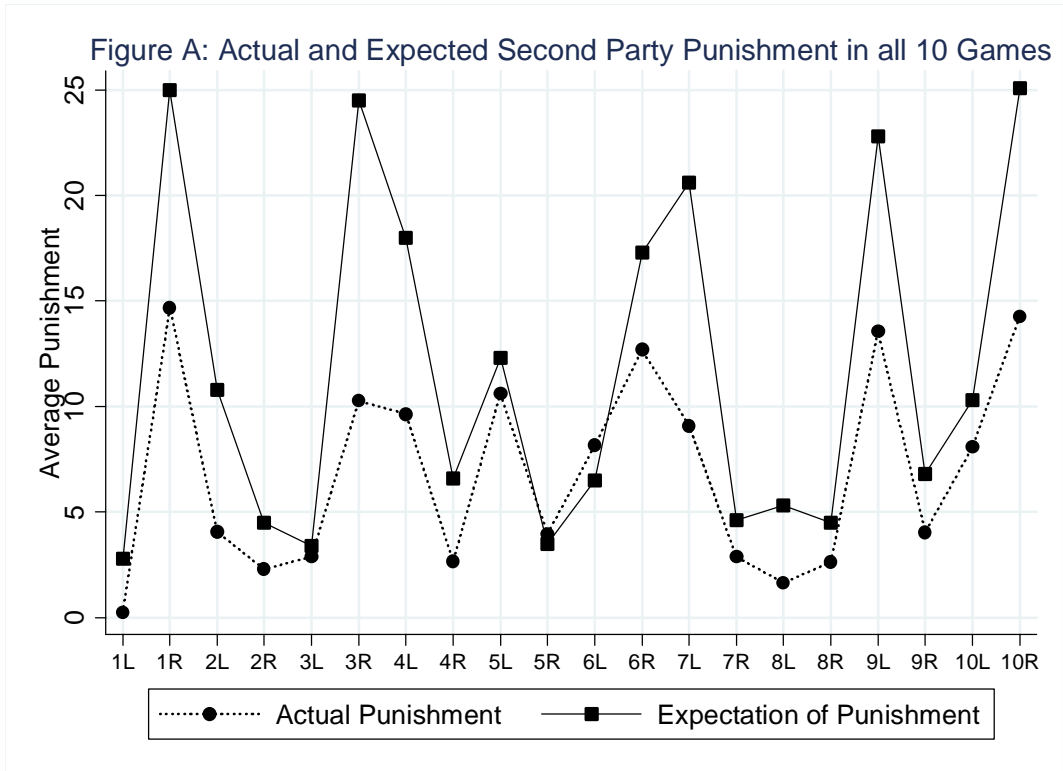


Appendix Tables

TABLE A—OBSERVED FREQUENCIES OF CHOICES FROM FIRST PARTIES IN 2P & 3P

Game				Left		Right	
		vs.		2P	3P	2P	3P
1	(150,150)	vs.	(590,60)	.09	.11	.91	.89
2	(100,100)	vs.	(50,530)	.80	.82	.20	.18
3	(560,60)	vs.	(120,140)	1	.98	0	.02
4	(150,90)	vs.	(50,630)	.73	.87	.27	.13
5	(220,260)	vs.	(220,400)	.20	.18	.80	.82
6	(280,240)	vs.	(390,240)	.13	.07	.87	.93
7	(250,80)	vs.	(80,250)	1	1	0	0
8	(100,100)	vs.	(50,150)	.96	1	.04	0
9	(250,150)	vs.	(110,290)	.93	.96	.07	.04
10	(250,150)	vs.	(330,70)	.40	.18	.60	.82

Appendix Figures



Appendix: Derivation of the maximum likelihood estimate of ε .

We obtain $\hat{\varepsilon}$ by solving the following maximization problem (assuming that an interior solution $\hat{\varepsilon} \in (0, 1)$ exists)

$$\max_{\varepsilon \geq 0} \prod_{s=1}^n \left(1 - \frac{\varepsilon}{2}\right)^{X_s} \times \left(\frac{\varepsilon}{2}\right)^{d - X_s},$$

or, equivalently,

$$\max_{\varepsilon \geq 0} \sum_{s=1}^n \left[X_s \cdot \text{Ln} \left(1 - \frac{\varepsilon}{2}\right) + (d - X_s) \cdot \text{Ln} \left(\frac{\varepsilon}{2}\right) \right]. \quad (1)$$

Computing the first derivative of (1) and after some algebra, one gets the first order condition of problem (1), that is,

$$\sum_{s=1}^n \left[\frac{d(2 - \varepsilon) - 2X_s}{\varepsilon(2 - \varepsilon)} \right] = \frac{dn}{\varepsilon} - \frac{2 \sum X_s}{\varepsilon(2 - \varepsilon)} = 0 \Leftrightarrow \hat{\varepsilon} = \frac{2[dn - \sum X_s]}{dn}. \quad (2)$$

Further, since the second derivative of (1) $-\frac{dn}{\varepsilon^2} + \frac{4(1 - \varepsilon) \sum X_s}{[\varepsilon(2 - \varepsilon)]^2}$ is negative because $\sum X_s < dn$ and $\frac{4(1 - \varepsilon)}{(2 - \varepsilon)^2} < 1$, it follows that indeed (2) is a maximum. Finally, and in case expression (2) takes a value equal or larger than 1 so that an interior solution does not exist, the optimum is clearly $\hat{\varepsilon} = 1$. ■

Acknowledgements

The authors greatly acknowledge financial support by the EU Research Network ENABLE. We also want to thank Gary Charness, Martin Dufwenberg, Armin Falk, Ernst Fehr, James H. Fowler, David M. Grether, Daniel Houser, Antonio Martín-Arroyo, Ernesto Reuben, Tatsuyoshi Saijo, Frans van Winden and numerous participants at several conferences who provided valuable feedback.

References

- Babcock, Linda; Loewenstein, George; Issacharoff, Samuel and Camerer, Colin.** Biased Judgments of Fairness in Bargaining, *American Economic Review*, 1995, 85(5), 1337-1343.
- Bendor, Jonathan and Swistak, Piotr.** The Evolution of Norms, *American Journal of Sociology*, 106, 1493-1545.
- Bolton, Gary E., and Ockenfels, Axel.** ERC: A Theory of Equity, Reciprocity, and Competition, *American Economic Review*, 2000, 90(1), 166-93.
- Brandts, Jordi and Charness, Gary.** Hot vs. Cold: Sequential Responses and Preference Stability in Experimental Games, *Experimental Economics*, 2000, 2(3), 227-238.
- Cason, Timothy and Mui, Vai-Lam.** Social Influence in the Sequential Dictator Game, *Journal of Mathematical Psychology*, 1998, 42, 248-465.
- Camerer, Colin and Hogarth, Robin M.** The Effect of Financial Incentives in Experiments, *Journal of Risk and Uncertainty*, 1999, 19, pp. 7-42.
- Camerer, Colin and Thaler, Richard.** Ultimatums, Dictators, and Manners, *Journal of Economic Perspectives*, 1995, 9, 209-219.
- Carpenter, Jeffrey P. and Matthews, Peter H.** Norm Enforcement: Anger, Indignation or Reciprocity?, *IZA Discussion Paper Series* No. 1583, 2005.
- Charness, Gary.** Responsibility and Effort in an Experimental Labor Market, *Journal of Economic Behavior and Organization*, 2000, 42(3), pp. 375-384.
- Charness, Gary; Cobo-Reyes, Ramón and Jiménez, Natalia.** An Investment Game with Third-party Intervention, *Journal of Economic Behavior and Organization*, 2008, 68(1), 18-28.
- Charness, Gary and Grosskopf, Brit.** Relative Payoffs and Happiness: An Experimental Study, *Journal of Economic Behavior and Organization*, 2001, 45, 301-328.
- Charness, Gary, and Rabin, Matthew.** Understanding Social Preferences with Simple Tests, *Quarterly Journal of Economics*, 2002, 117, 817-869.
- Cox, James C.; Friedman, Daniel and Gjerstad, Steven.** A Tractable Model of Reciprocity and Fairness, *Games and Economic Behavior*, 2007, 59, 17-45.
- Dawes, Christopher; Fowler, James H.; Johnson, Tim; McElreath, Richard and Smirnov, Oleg.** Egalitarian Motives in Humans, *Nature*, 2007, 446, 794-796.
- Dufwenberg, Martin, and Kirchsteiger, Georg.** A Theory of Sequential Reciprocity, *Games and Economic Behavior*, 2004, 47, 268-98.

- El-Gamal, Mahmoud, and David Grether.** Are People Bayesian? Uncovering Behavioral Strategies, *Journal of the American Statistical Association*, 90(432), December 1995, 1137-1145.
- Elster, Jon.** Social Norms and Economic Theory, *Journal of Economic Perspectives*, 1989, 3(4), 99-117.
- Engelmann, Dirk, and Strobel, Martin.** Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments, 2004, *American Economic Review*, 94(4), 857-869.
- Engle-Warnick, Jim.** Inferring Strategies from Observed Actions: A Nonparametric, Binary Tree Classification Approach, *Journal of Economic Dynamics and Control*, 2003, 27(11-12), 2151-2170.
- Falk, Armin; Fehr, Ernst and Fischbacher, Urs.** Driving Forces behind Informal Sanctions, *Econometrica*, 2005, 7(6), 2017-30.
- Falk, Armin and Fischbacher, Urs.** A Theory of Reciprocity, *Games and Economic Behavior*, 2006, 54, 293-315.
- Fehr, Ernst and Fischbacher, Urs.** Third Party Punishment and Social Norms, *Evolution and Human Behavior*, 2004, 25, 63-87.
- Fehr, Ernst and Gächter, Simon.** Cooperation and Punishment in Public Goods Experiments, *American Economic Review*, 2000, 90, 980-994.
- Fehr, Ernst and Schmidt, Klaus.** A Theory of Fairness, Competition and Cooperation, *Quarterly Journal of Economics*, 1999, 114(3), 817-68.
- Fehr, Ernst; Näf, Michael and Schmidt, Klaus.** Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments: Comment, *American Economic Review*, 2006, 96, 1912-1917.
- Fischbacher, Urs.** z-Tree: Zurich Toolbox for Ready-made Economic Experiments, *Experimental Economics*, 2007, 10 (2), 171-178.
- Greiner, Ben.** An Online Recruitment System for Economic Experiments, in: *Forschung und wissenschaftliches Rechnen 2003*, ed. by Kurt Kremer and Volker Macho, GWDG Bericht 63, Göttingen, Ges. für Wiss. Datenverarbeitung.
- Güth, Werner; Schmittberger, Rolf and Schwarze, Bernd.** An Experimental Analysis of Ultimatum Bargaining, *Journal of Economic Behavior and Organization*, 1982, 3, 367-388.
- Güth, Werner; Huck, Steffen and Müller, Wieland.** The Relevance of Equal Splits in Ultimatum Games, *Games and Economic Behavior*, 2001, 37, 161-169.

- Homans, George.** Social Behavior, 1961, New York: Harcourt, Brace & World.
- Kennedy, Randall.** Race, Crime and the Law, 1997, pp. 168-255, New York: Pantheon.
- Kirchsteiger, Georg.** The Role of Envy in Ultimatum Games, *Journal of Economic Behavior and Organization*, 1994, 25(3), 373-389.
- Kritikos, Alexander and Bolle, Friedel.** Distributional Concerns: Equity- or Efficiency-Oriented? *Economics Letters*, 2001, 73, 333-338.
- Levine, David K.** Modeling Altruism and Spitefulness in Experiments. *Review of Economic Dynamics*, 1998, 1, 593-622.
- López-Pérez, Raúl.** Aversion to Norm-Breaking: A Model, *Games and Economic Behavior*, 2008, 64, 237-267.
- Nikiforakis, Nikos.** Punishment and Counter-punishment in Public Good Games: Can we really govern ourselves. *Journal of Public Economics*. 2008, 92(1-2), 91-112.
- Ostrom, Elinor; Walker, James and Gardner, Roy.** “Covenants with and without a Sword: Self-Governance is Possible”, *American Political Science Review*, 1992, 86(2), 404-417.
- Rabin, Matthew.** Incorporating Fairness into Game Theory and Economics, *American Economic Review*, December 1993, 83(5), 1281-1302.
- Ross, Lee; Greene, David and House, Pamela.** The False Consensus Effect: An Egocentric Bias in Social Perception and Attribution Processes, *Journal of Experimental Social Psychology*, 1977, 13(3), 279-301.
- Roth, Alvin E.** Bargaining Experiments, in J. Kagel and A. Roth (eds.): *Handbook of Experimental Economics*, 1995, Princeton, Princeton University Press.
- Seinen, Ingrid and Schram, Arthur.** Social Status and Group Norms: Indirect Reciprocity in a Repeated Helping Experiment, *European Economic Review*, 2006, 50, 581-602.
- Zizzo, Daniel J.** Money Burning and Rank Egalitarianism with Random Dictators, *Economics Letters*, 2003, 81, pp. 263-66.

Appendix: Instructions

General Instructions for the second party in Treatment 2P

We welcome you to our experiment. If you read the following instructions carefully you will be able to earn money in addition to your show up fee of 10 Swiss Francs – depending on your decisions and the decisions of the other participants. Therefore it is very important, that you read the following instructions carefully. If you have any question, please address them to us.

During the experiment you are not allowed to talk to other participants. If you do not follow this rule we will have to exclude you from the experiment and you will not be able to earn money.

In this experiment you will have to make one decision in ten different situations that can influence your payoff. The order of these ten situations is randomly determined. At the end of the experiment, a ten-sided dice will be thrown to determine which of the ten situations becomes relevant for your payment.

In this experiment we always speak of points. 10 points are worth 1 Swiss Franc.

10 points = 1 Swiss Franc.

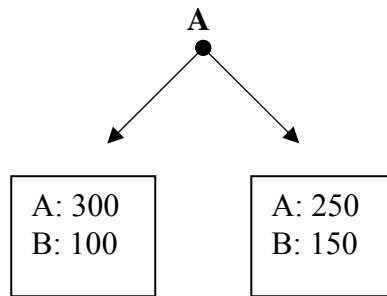
There are two types of participants in this experiment: Participant A and participant B. **You are participant B.** You will never get to know the identity of any participant A (or B), nor will any participant A get to know who you are. The payment at the end of the experiment is also anonymous, that is, no other participant will know how much you earned in this experiment.

Exact Description of the Experiment for Participant B in Treatment 2P

Each of the ten situations consists of two stages. In the following, we will explain these two stages.

The first stage

In the first stage, participant A makes his decision. He can decide between two allocations. Take for instance the following example. If he decides for the allocation on the left side, he gets 300 points and you get 100 points. If he decides for the allocation on the right side, he gets 250 points and you get 150 points.



The second stage

In the second stage, you make your decision. You can assign deduction points to participant A. Every deduction point you assign, reduces your payoff by 1 point and the payoff of participant B by 3 points. You can assign between 0 and 50 deduction points. For instance, if you assign 50 deduction points, your payoff is reduced by 50 points and the payoff of participant B by 150 points. If you assign 25 deduction points, your income is reduced by 25 points and the payoff of participant B by 75 points.

Time Line of the Experiment

You will have to decide how many deduction points you assign in all ten different situations before you know which allocation participant A has chosen in the first stage of the situations. We will present you the two different allocations in each situation and you will have to decide how many deduction points you assign in each allocation. While you are making your decisions, participant A will choose one allocation in each situation. After all participants A and B have made their decisions in the ten situations, the experiment is over. One situation will be randomly determined by a ten-sided dice and you will be paid according to your and participant A's decision in this situation.

Calculation of Payoffs

The payoffs of participant A and B are calculated as follows:

The payoff of participant A =

- + Points for A in the allocation participant A has chosen in the game that was chosen by the dice
- 3x the deduction points you assigned to participant A in the game that was chosen by the dice

Your payoff (participant B) =

- + Points for B in the allocation participant A has chosen in the game that was chosen by the dice
- The deduction points you assigned to participant A in the game that was chosen by the dice

Exact Description of the Experiment on the Computer Screen for Participant B

In the second stage, you will have to decide how many deduction points you assign in the first situation. The following computer screen will appear:

Sie sind Teilnehmer B
Entscheidungssituation 1

Teilnehmer A hat sich für eine der beiden folgenden Verteilungen entschieden:

A: 100 Punkte
B: 100 Punkte

oder

A: 50 Punkte
B: 530 Punkte

Stellen Sie sich vor, Teilnehmer A wählt die Verteilung A: 100 Punkte, B: 100 Punkte. Wie viele Abzugspunkte werden Sie ihm zuweisen?

Stellen Sie sich vor, Teilnehmer A wählt die Verteilung A: 50 Punkte, B: 530 Punkte. Wie viele Abzugspunkte werden Sie ihm zuweisen?

Hilfe
Bitte erinnern Sie sich daran, dass Sie bis zu 50 Abzugspunkte zuweisen können. Jeder Abzugspunkt kostet Sie 1 Punkt und reduziert das Einkommen von Teilnehmer A um 3 Punkte.
Der OK-Button erscheint in kurzer Zeit!

OK

After that, you will make your decision in the second, third, ..., tenth situation.

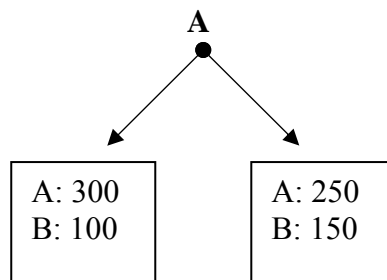
You can take as much time as you need. The OK-Button appears with a little time delay.

If you have made your decisions in all ten situations, you will be informed about your payoff.

Please answer now the following control questions and raise your hand if you have answered them. The experiment starts as soon as all participants have correctly filled out the control questions.

Control Questions

1.



Participant A chooses the allocation on the left side.

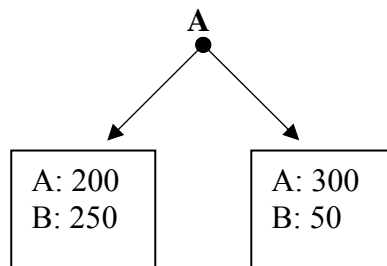
a) Participant B assigns 0 deduction points to participant A.

What is the payoff of participant A? B?.....

b) Participant B assigns 30 deduction points to participant A.

What is the payoff of participant A? B?.....

2.



Participant A chooses the allocation on the right side.

a) Participant B assigns 0 deduction points to participant A.

What is the payoff of participant A? B?.....

b) Participant B assigns 50 deduction points to participant A.

What is the payoff of participant A? B?.....

Do you have any further questions?

General Instructions for the third party in Treatment 3P

We welcome you to our experiment. If you read the following instructions carefully you will be able to earn money in addition to your show up fee of 10 Swiss Francs – depending on your decisions and the decisions of the other participants. Therefore it is very important, that you read the following instructions carefully. If you have any question, please address them to us.

During the experiment you are not allowed to talk to other participants. If you do not follow this rule we will have to exclude you from the experiment and you will not be able to earn money.

In this experiment you will have to make one decision in ten different situations that can influence your payoff. The order of these ten situations is randomly determined. At the end of the experiment, a ten-sided dice will be thrown to determine which of the ten situations becomes relevant for your payment.

In this experiment we always speak of points. 10 points are worth 1 Swiss Franc.

10 points = 1 Swiss Franc.

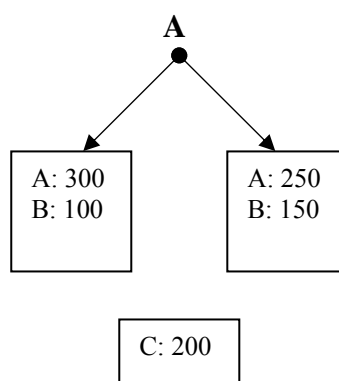
There are three types of participants in this experiment: Participant A, participant B and participant C. **You are participant C.** You will never get to know the identity of any participant, nor will any participant get to know who you are. The payment at the end of the experiment is also anonymous, that is, no other participant will know how much you earned in this experiment.

Exact Description of the Experiment for Participant C

Each of the ten situations consists of two stages. In the following, we will explain these two stages.

The first stage

In the first stage, participant A makes his decision. He can decide between two allocations. Take for instance the following example. If he decides for the allocation on the left side, he gets 300 points and B gets 100 points. If he decides for the allocation on the right side, he gets 250 points and B gets 150 points. Independently of participant A's choice, you will always get 200 points in all ten situations.



The second stage

In the second stage, you make your decision. You can assign deduction points to participant A and/or participant B. Every deduction point you assign to participant A (or participant B), reduces your payoff by 1 point and the payoff of participant A (or B) by 3 points. You can assign in total between 0 and 50 deduction points. For instance, if you assign 50 deduction points to participant A, your payoff is reduced by 50 points and the payoff of participant A by 150 points. If you assign 30 deduction points to participant B, your income is reduced by 30 points and the payoff of participant B by 90 points.

Time Line of the Experiment

You will have to decide how many deduction points you assign in all ten different situations before you know which allocation participant A has chosen in the first stage of the situations. We will present you the two different allocations in each situation and you will have to decide how many deduction points you assign in each allocation. While you are making your decisions, participant A will choose one allocation in each situation and participant B will be asked how many deduction points you will assign. After all participants have made their decisions in the ten situations, the experiment is over. One situation will be randomly determined by a ten-sided dice and you will be paid according to your decision in this situation.

Calculation of Payoffs

The payoffs of participant A, B and C are calculated as follows:

The payoff of participant A =

- + Points for A in the allocation participant A has chosen in the game that was chosen by the dice
- 3x the deduction points you assigned to participant A in the game that was chosen by the dice

The payoff of participant B =

- + Points for B in the allocation participant A has chosen in the game that was chosen by the dice
- 3x the deduction points you assigned to participant B in the game that was chosen by the dice

Your payoff (participant C) =

- + 200 Points (your endowment)
- The deduction points you assigned to participants A and/or B in the game that was chosen by the dice

Exact Description of the Experiment on the Computer Screen for Participant C

In the second stage, you will have to decide how many deduction points you assign in the first situation. The following computer screen will appear:

Sie sind Teilnehmer C
Entscheidungssituation 1

Teilnehmer A hat sich für eine der beiden folgenden Verteilungen entschieden:

A: 150 Punkte
B: 90 Punkte

oder

A: 50 Punkte
B: 630 Punkte

Stellen Sie sich vor, Teilnehmer A wählt die Verteilung A: 150 Punkte, B: 90 Punkte. Wie viele Abzugspunkte werden Sie Teilnehmer A zuweisen?

Und wie viele Abzugspunkte weisen Sie Teilnehmer B zu?

Stellen Sie sich vor, Teilnehmer A wählt die Verteilung A: 50 Punkte, B: 630 Punkte. Wie viele Abzugspunkte werden Sie Teilnehmer A zuweisen?

Und wie viele Abzugspunkte weisen Sie Teilnehmer B zu?

Hilfe
Sie haben 200 Punkte zur Verfügung. Bitte erinnern Sie sich daran, dass Sie bis zu 50 Abzugspunkte zuweisen können. Jeder Abzugspunkt kostet Sie 1 Punkt und reduziert das Einkommen des anderen Teilnehmers um 3 Punkte.
Der OK-Button erscheint in kurzer Zeit!

OK

After that, you will make your decision in the second, third,..., tenth situation.

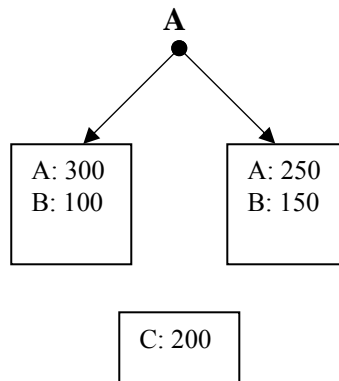
You can take as much time as you need. The OK-Button appears with a little time delay.

If you have made your decisions in all ten situations, you will be informed about your payoff.

Please answer now the following control questions and raise your hand if you have answered them. The experiment starts as soon as all participants have correctly filled out the control questions.

Control Questions

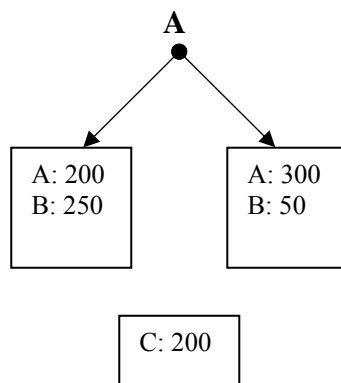
1.



Participant A chooses the allocation on the left side.

- a) Participant C assigns 0 deduction points to participant B.
What is the payoff of participant A? B?..... C?.....
- b) Participant C assigns 30 deduction points to participant B.
What is the payoff of participant A? B?..... C?.....

2.



Participant A chooses the allocation on the right side.

- a) Participant C assigns 0 deduction points to participant A.
What is the payoff of participant A? B?..... C?.....
- b) Participant C assigns 50 deduction points to participant A.
What is the payoff of participant A? B?..... C?.....

Do you have any further questions?