Institute for Empirical Research in Economics
University of Zurich

Working Paper Series
ISSN 1424-0459

Working Paper No. 373

The Envious Punisher: Understanding Third and Second Party Punishment with Simple Games

Andreas Leibbrandt and Raúl López-Pérez

May 2008

The Envious Punisher:

Understanding Third and Second Party Punishment with Simple Games

By Andreas Leibbrandt[1] and Raúl López-Pérez[2]

May 27th, 2008

Abstract: We provide a systematic comparison of punishment from unaffected third parties and affected second parties using a within-subject design in ten simple games. We apply the classification analysis by El-Gamal and Grether (1995) and find that a parsimonious model assuming subjects are either envious or selfish best explains the punishment from both third and second parties. Third and second parties punish richer co-players, even if they chose a socially or Pareto-efficient allocation or if they are merely bystanders who made no choice. Despite their unaffected position, we do not find that third parties punish in a more impartial or normative manner.

Keywords: Envy, fairness, inequity aversion, norms, punishment, reciprocity.
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@iew.uzh.ch.
[2] Department of Economic Analysis, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain. E-mail: raul.lopez@uam.es

The authors greatly acknowledge financial support by the EU Research Network ENABLE. We also want to thank Gary Charness, Armin Falk, Ernst Fehr, James H. Fowler, David M. Grether, Daniel Houser, Antonio Martín-Arroyo, Ernesto Reuben, Tatsuyoshi Saijo, Frans van Winden and participants at the Málaga 2007 IMEBE conference, the Shanghai 2007 Asia-Pacific meeting of the ESA, the ENABLE meeting 2007 and 2008 in Amsterdam, the Greifensee Seminar 2007, the FEDEA seminar, and the XXXII symposium of Economic Analysis in Granada.

1. Introduction

Third parties play a crucial role in many institutions: they serve in courts, as referees or arbitrators. The US legal system, for instance, relies on their judgment in juries when it comes to the application of sanctions. 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 (Greif, 1993, 1994), and hence their sanctions are potentially more damaging than those from second parties.

Despite the importance of third parties' actions, little is known about how they 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 that rely on third parties implies that they are likely to 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).

This paper uses laboratory experiments to study and systematically compare the motives for sanctioning from third and second parties, applying a within-subject analysis and the classification method by El-Gamal and Grether (1995). Our experiment consists of a second party (2P) and a third party punishment treatment (3P), and participants in each treatment play ten different games consisting of two stages. In the first stage, one player (the *first party*) chooses between two allocations of payoffs between himself or herself and another player (the *second party*). In the second stage, either the second party can punish the first party (in treatment 2P), or an unaffected *third party* can punish the first party and/or the second party, who is now a bystander (in treatment 3P). In total, we observe 3100 punishment decisions. Our paper provides several important insights into how third and second parties punish.

First, the classification analysis shows that a parsimonious model assuming that subjects are either envious or selfish can capture 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 (in 3P and 2P) and reciprocal (in 2P) types are slightly more accurate; they come at the cost of increased complexity. We find that both third and second-party punishment is often directed towards richer co-players, *even* if (i) they are merely bystanders who make no choice (in 3P), or if (ii) they chose a socially efficient allocation – i.e. one maximizing the sum of payoffs – even in case of Pareto efficiency: Richer players are punished even if their previous choices inflicted no harm on another party. *Second*, we also observe that the *strength* of punishment depends heavily on the size of the payoff disadvantage (in 3P and 2P) and the sole existence of harm (in 2P). *Third*, and contrary to the idea that third parties are less "infected" by egocentric or "non-normative" motives than second parties, we observe that third party punishment generally resembles second party punishment. Furthermore, we find no support for the conjecture that third parties are impartial, in the sense of valuing any co-player's welfare equally.

The existing experimental literature on punishment overwhelmingly examines the decisions of personally affected second parties. 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 make 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 (Fehr and Gächter 2000). However, this experimental literature on punishment has a *completely* different focus than ours because it is restricted to second party punishment and also because the analyzed games are not well suited for discriminating the motives behind punishment. In the ultimatum game, responders might reject offers due to envy, 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 envy, inequity-aversion, reciprocity, spite, or as a reaction to a transgression of cooperation norms.[3]

Because we study the motives behind third and second party punishment, our paper is most related to Fehr and Fischbacher (2004), Falk et al. (2005), and Dawes et al. (2007). The first of these papers constitutes one of the few studies on third party punishment (Carpenter and Matthews, 2005; Charness et al., forthcoming; Engle-Warnick and Leibbrandt, 2006),

[3] Theories of inequity aversion (Fehr and Schmidt 1999, Falk and Fischbacher 2006) predict punishment of richer co-players if that reduces the payoff distance (hence they model a particular form of envy), while reciprocity theories predict punishment of an individual who previously harmed the aggressor (Rabin 1993, Dufwenberg and Kirchsteiger 2004; Cox et al., 2007). Further, Bolton and Ockenfels (2000) predict punishment of any co-player when the aggressor's *relative* payoff is lower than the average one, Levine (1998) posits the existence of spiteful types who punish indiscriminately and type-reciprocal agents who punish selfish or spiteful co-players, and López-Pérez (forthcoming) predicts punishment of norm deviators.

showing 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 envy, 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). The study by Falk et al. (2005) investigates whether inequity aversion, spite, or reciprocity models better account for second party punishment in two variants of a prisoner's dilemma game with a punishment possibility. They observe that punishment is mainly targeted towards previous defectors, regardless of whether punishment is cheap or expensive – i.e. when payoff differences cannot be reduced – thus concluding that "retaliation seems to be the most important motive behind fairness-driven informal sanctions" (*ibid*, p. 2017).[4] In contrast, Dawes et al. (2007) show that the "egalitarian motive", which often coincides with envy, explains most punishment in a modified public goods game.

While these studies (and all other studies about punishment we are aware of) draw their inferences from a between-subjects design, we provide a detailed analysis using a within-subjects design. This helps reveal whether subjects consistently follow one motivation, hence offering an arguably more systematic and accurate picture of heterogeneity in individual behavior, which can be used to classify subjects as envious, reciprocal, etc. This classification analysis is especially worthwhile in view of the abundant experimental evidence suggesting that subjects have heterogeneous social preferences. Falk et al. (2005), for example, report the existence of different types of punishers, as some defectors punished cooperators (especially when punishment was cheap), which cannot be reconciled with retaliation or inequity-aversion. Charness and Rabin's (2002) large study of many different experimental games stresses the existence of subjects who follow different motivations (maximizing social welfare, minimizing income differences, etc).

To the best of our knowledge, the present study is the first to investigate whether third parties are impartial and less prone to punish in an egocentric or "non-normative" manner than second parties. Another new feature of this study is that we analyze punishment of socially efficient but inequitable choices, a surprisingly under-studied topic in view of the great interest that exists on finding out whether deciders choose socially efficient allocations

[4] However, their results are also consistent with a model of envy predicting punishment of richer parties even when that does *not* reduce the payoff distance (note that unilateral defectors get the largest payoff). Our experimental design allows us to discriminate between such a model of envy and models of reciprocity.

even at their own material disadvantage (Charness and Rabin 2002, Engelmann and Strobel 2004; Fehr et al. 2006). We also are first to provide a *systematic* study about the punishment (or damage) of bystanders, an important topic given the casual evidence that bystanders often become victims of punishment.

Our results indicate that it can be misleading to assume that third parties are impartial, make less egocentric choices, and enforce (informal) rules in a normative manner. In addition, this study gives important implications for the further development of recent theories of social preferences. Envy appears to be an indispensable factor in explaining the occurrence and strength of second and third party punishment, while reciprocity and spite play an important, although *relatively* minor role, in explaining the occurrence of second party punishment. In particular, we believe that reciprocity should not be used alone in predicting punishment in general, as it fails to explain any third party punishment.

The rest of the paper proceeds as follows. The next section presents the experimental design and procedure. Section three provides an overview of the punishment behavior and an analysis of the factors affecting the occurrence and strength of second and third party punishment. In section four, we compare third to second party punishment, report reactions to socially efficient choices, and study further topics like the punishment of bystanders and the impact of strictly equal allocations on punishment. The fifth section concludes.

2. 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 right-hand allocation of payoffs between herself and another player (the *second party*). Table 1 shows the two allocations available in each game. They are identical in 2P and 3P and presented in points (10 points = 1 SFR).

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)	(120,140)	(150,90)	(220,260)	(280,240)	(80,250)	(100,100)	(250,150)	(250,150)
	Right	(590,60)	(50,530)	(560,60)	(50,630)	(220,400)	(390,240)	(250,80)	(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 parties, while the second party in 3P makes no decision, i.e. she is a "bystander". 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 $0 \le p \le 50$ points, the first party's payoff 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 with p_2 points ($p_1 + p_2 \le 50$), 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 picked the various allocations in our ten different games for four main reasons. First of all, our selection allows us to discriminate between recent models of social preferences that provide a rationale for costly punishment, including some which have not yet been studied before in the literature. Take game 6 (280/240 vs. 390/240) for instance. An envious individual punishes both allocations because she has a lower payoff than her counterpart (no matter whether she is a second or third party), while a reciprocal individual punishes neither allocation because she cannot be harmed in this game. In contrast, an individual who punishes deviations from a norm of equity would punish the choice of the allocation (390/240) because it is more inegalitarian than the alternative (280/240), and an individual who punishes deviations from a norm of social efficiency would punish the allocation (280/240) because the joint payoff is bigger in the alternative allocation (390/240). In table 5 of the appendix we present the predictions in each of the games for the different motives, and we describe the different theories considered in section 3.2.1.

Second, we chose games 1-6 to close a gap in the literature on punishment which has until now neglected to investigate the reactions of second and third parties to the choice of socially efficient allocations, i.e. those maximizing the sum of the players' payoffs. Consider, for instance, game 3 (560/60 vs. 120/140) where the left-hand allocation is socially efficient. We were interested whether second parties were willing to accept a small disadvantage and refrain from punishing if this small disadvantage is associated with a comparatively large advantage for their counterpart. Moreover, we wanted to find out whether third parties react differently to the choice of socially efficient allocations like (560/60) in game 3. Are they less willing to punish since they are unaffected by the allocation choice? This question is also related to the next point.

Third, our ten games enable a thorough comparison of second and third party punishment. Game 7 (250/80 vs. 80/250), for instance, provides the opportunity to investigate whether third parties are less prone to "self-serving" arguments than second parties. To be precise, the allocations are symmetric – i.e., a permutation of each other – so that an "impartial" party who uniformly values each player's welfare should regard both allocations as equally fair and punish them less than a second party.[5] Finally, we wanted to analyze whether second and third party punishment is influenced by the availability of a strictly equal allocation, as in game 8 (100/100 vs. 50/150). More precisely, we investigate whether deviations from strictly equal allocations are more heavily punished than deviations from slightly unequal allocations and whether choices for strictly equal allocations are less punished compared to choices for slightly unequal allocations.

We ran eight sessions and each 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 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 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. 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.[6]

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 whether

[5] We add two remarks. First, 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’. We have this kind of idea in mind when we refer to impartiality. Second, an impartial spectator might still consider the choice of allocation (250/80) unfair because it fails to be courteous –i.e. by choosing this, the first party signals that she cares more for herself than for the second party. We take care of this issue and investigate later whether third parties follow such a kind of reasoning and punish "greedy" first parties but we do not find evidence in favor of it.

[6] 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.

and they wanted to punish the other subject(s) and if so, by how much. 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.[7] 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 present evidence suggests that the strategy method does not affect the 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.[8] In our view, the use of the strategy method seems unavoidable for the study of punishment behavior with a within-subjects design and a large set of games (unless the researcher has access to huge samples in order to control for order effects). Additionally, it maximizes the amount of statistical data gathered.

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 SFR (around $ 24) which included a show-up fee of 10 SFR (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.

---

[7] For other decisions than punishment, there is evidence of no systematic differences in behaviour between the strategy and specific response method (Cason and Mui, 1998; Brandts and Charness, 2000; Falk and Kosfeld, 2006).

[8] 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.

3. Experimental Results I

This section starts with a brief analysis of the occurrence of third and second party punishment on an aggregate level. The major part of this section is, however, devoted to the analysis of third and second party punishment on the *individual* level, where we present a classification procedure and then report its results. We finish this section with an analysis of the determinants of the strength of third and second party punishment.

3.1 The Occurrence of Third and Second Party Punishment: *Aggregate Analysis*

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. Table 2 summarizes the frequency and strength of third party punishment, distinguishing between punishment for first and second parties. 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. In eight of the ten games, third (second) parties punish the first party more strongly in one allocation ($p<0.05$; Wilcoxon-Signed Rank-Test). The behavior of the first parties in 3P and 2P can be seen in Table 4 in the appendix.

Table 2— PROBABILITY AND STRENGTH OF PUNISHMENT
THIRD PARTIES

				A-player				B-player (By-stander)			
	Game			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.

Table 3— PROBABILITY 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. The endowment of the third party is always 200 points.

RESULT 1: *Disadvantageous inequity (envy) seems to be the key factor in explaining the occurrence of third party punishment. Disadvantageous inequity as well as harm (reciprocity) seem to be the key factors in explaining the occurrence of second party punishment.*

In 3P, the envy theories of Fehr and Schmidt (1999) and Falk and Fischbacher (2006) predict that third parties punish another player only if she gets a larger material payoff. This prediction can be reconciled with the data from 14 out of the 15 allocations where a first or a second party is punished by more than 20 percent of the third parties (and/or their average punishment is larger than 3 points). Figure 1 illustrates the individual punishment decisions dependent on the size of the payoff differences between the third and the other two parties. On the left side of the horizontal axis are the allocations where the third party has a higher payoff than the first or second party and on the right side are the allocations where the first or the second party has a higher payoff than the third party. The location of the dots indicates the average amount of money spend in an allocation from a third party that punishes at least once. The size of the dots is proportional to the quantity of observations. Hence, a dot becomes larger if more than one third party spends the same amount to punish in the same allocation. We can see that payoff differences appear to play an important role in the decision to punish. In fact, third parties rarely punish if their payoff is higher than that of the first/second party, but often and severely if their payoff is lower.

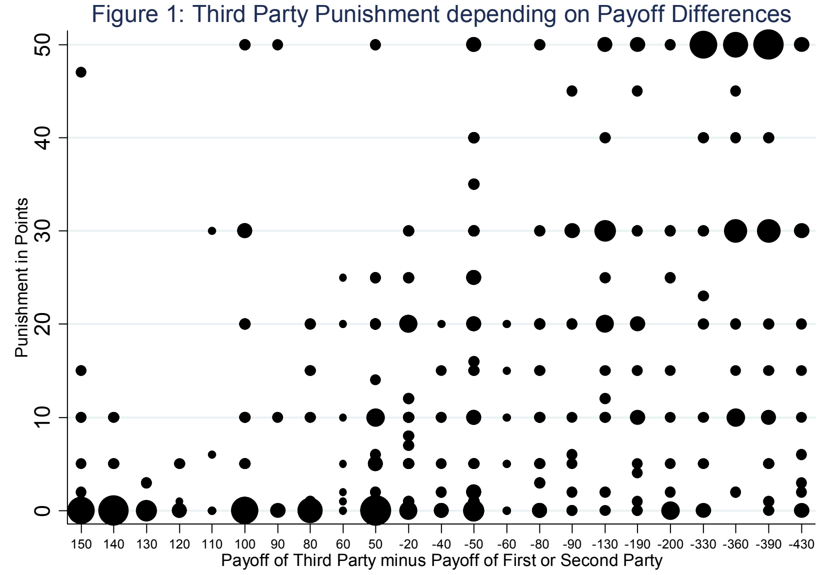


Figure 1: Third Party Punishment depending on Payoff Differences

Explanations other than envy seem to play a much less important role in the occurrence of third party punishment. To start, note that reciprocity cannot play any role in 3P because the third party is unaffected by the first and second party, i.e. cannot be harmed. In turn, the model by Bolton and Ockenfels (2000), which predicts punishment of any other player only if the third party's relative payoff is smaller than 1/3 of the total, i.e. the equitable relative payoff in three-player games, seems to be missing an important point. For example, this model never predicts punishment in games 7 (250/80 vs. 80/250), 9 (250/150 vs. 110/290), and 10 (250/150 vs. 330/70), where we observe considerable punishment. Spite (an important ingredient of Levine, 1998) cannot account for why there are so many allocations where there is hardly any or no punishment. It also seems that third parties do not punish deviations from a social norm. For instance, the evidence from game 6 (280/240 vs. 390/240), where we observe punishment of the first player in *both* allocations, suggests that third parties do not punish deviations from a norm like an equity, maximin or social efficiency norm.

In 2P, we frequently observe considerable punishment in the allocations where it is either predicted by envy or reciprocity theories. In many allocations, the behavior points to the importance of envy. In game 6 (280/240 vs. 390/240), for instance, about one third of the second parties punish the first party *in either allocation,* even though the first party's choice did not harm them, i.e. the second party's payoff is the same in either allocation. There is also considerable punishment in both allocations of game 10 (250/150 vs. 330/70), even when the choice (250/150) is clearly "very kind" towards the second party – the first party actually sacrifices money to increase the second party's payoff. However, reciprocity also seems important. For instance, envy predicts that second parties should never punish in game 5 (220/260 vs. 220/400), while reciprocity predicts punishment in the allocation (220/260). Consistent with this, 40 percent of the second parties punish considerably in the allocation (220/260). There is also some punishment of the allocation (100/100) in game 2 (100/100 vs. 50/530). Other explanations seem not to add much to the understanding of punishment in 2P. In fact, average punishment in any 2P game allocation is stronger than 4 points *if and only if* it is predicted by envy or reciprocity.[9]

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

The previous section provided an aggregate and therefore rather imprecise picture of the motives behind third and second party punishment. We now turn to a more 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). More precisely, we posit 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. This classification procedure has several favorable attributes. By selecting the decision rule that best fits each subject's behavior, we can classify subjects in types. It also helps us 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 an accurate but parsimonious explanation of punishment in our games. Importantly, the procedure circumvents the multicollinearity problems that would appear in a classical regression analysis if the decision rules entered as independent variables and allows appropriate inferences even when testing all possible decision rules –no matter how similar their predictions are– at the same time.[10]

[9] The only exception is game 8. In section 4.4 we provide a possible explanation.

[10] Multicollinearity problems may occur as soon as decision rules share predictions in some allocations (a very common thing in our games). For instance, this is the case for the decision rules that predict no punishment of the second party in 3P and hence share predictions in 20 out of 40 allocations.

3.2.1 Decision Rules in 2P and 3P

In this section, we specify the decision rules that we tested for 2P and 3P. For simplicity, we restrict our analysis in 2P and 3P to some appealing decision rules and only use *binary* decision rules, that is, rules indicating only whether the subject punishes and not the strength of punishment. 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 only considered eight decision rules in 2P that seem to be especially appealing (which we denoted as the "selfish", "envy", "reciprocity", "spite", "greed", "efficiency", "equity" and "maximin" decision rule).[11] Letting $(x_{FP}^{L},\ x_{SP}^{L})$ refer to the left-hand and $(x_{FP}^{R},\ x_{SP}^{R})$ to the right-hand allocation at any game (with FP denoting first party and SP denoting second party), they are defined as follows: (1) the "selfish" rule consists of a vector of 20 zeros and predicts never punishment, (2) the "envy" rule (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Falk and Fischbacher, 2006) predicts punishment only at those allocations where $x_{FP}^{i} > x_{SP}^{i}$, for $i = (L,\ R)$, (3) the "reciprocity" rule (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004; Cox et al., 2007) predicts punishment at any allocation $j = (L,\ R)$ such that $x_{SP}^{j} < x_{SP}^{i} \quad (j \neq i)$, that is, if the second party was harmed by the first party, (4) the "spite" rule (inspired by Levine, 1998) consists of a vector of 20 ones, i.e. predicts punishment at all allocations[12]-, (5) the "greed" rule (also inspired by Levine, 1998) predicts punishment if $x_{FP}^{i} > x_{FP}^{j}$, for $i = (L,\ R)$; subjects who follow this rule punish the first party if she chooses the allocation maximizing her own money payoff, the intuition being that subjects enjoy punishing selfish or greedy first parties, (6) the "efficiency" rule (inspired by López-Pérez, forthcoming) predicts punishment in allocation $(x_{FP}^{j},\ x_{SP}^{j})$ if $x_{FP}^{j} + x_{SP}^{j} < x_{FP}^{i} + x_{SP}^{i} \quad (j \neq i)$ and no punishment otherwise; the intuition being that subjects punish deviations from a norm of efficiency, (7) the "equity" rule (inspired by Elster, 1989 and López-Pérez, forthcoming) predicts punishment in allocation $(x_{FP}^{j},\ x_{SP}^{j})$ only if $\left|x_{FP}^{j} - x_{SP}^{j}\right| < \left|x_{FP}^{i} - x_{SP}^{i}\right| \quad (j \neq i)$; that is, in this case subjects punish deviations from a norm of equity, and (8) the "maximin" rule (inspired by Charness and

[11] We do not report the complete analysis here. For instance, we tested a large number of "hybrid" decision rules like an "envy and reciprocity" rule. Including such "hybrid" rules did not significantly improve the model. The results are available upon request.

[12] This might seem a very stringent rule, 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.

Rabin, 2002) predicts punishment in allocation $(x^j_{FP},\ x^j_{SP})$ if $\min\{x^j_{FP}, x^j_{SP}\} < \min\{x^i_{FP}, x^i_{SP}\}$ $(j \neq i)$.[13] Table 5 in the appendix indicates the predictions of these rules (and of some 3P ones) in our ten games.

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 consider eleven decision rules to investigate third party punishment (which we denoted as the "selfish", "envy", "spite", "greed", "efficiency", "equity", "maximin", "indirect reciprocity", "ERC", "envy-active" and "envy-perspective" decision rule). The first seven of them are based on the 2P rules mentioned above: (1) The "selfish" rule consists of a vector of 40 zeros, (2) the "envy" rule (Fehr and Schmidt, 1999; Falk and Fischbacher, 2006) is a logical extension of the envy rule in 2P predicting punishment of the first *and/or* second party when they have a larger payoff than the third party, (3) the "spite" rule is a vector of 40 ones, and the (4) "greed", (5) "efficiency", (6) "equity" and (7) "maximin" rules are defined like in 2P (they never predict punishment of the second party in 3P). In addition, we include (8) an "indirect reciprocity" rule (inspired by Nowak and Sigmund, 2005 and Seinen and Schram, 2006) predicting punishment of the first party if $x^j_{SP} < x^i_{SP}$ $(j \neq i)$ and no punishment of the second party,[14] (9) an "ERC" rule (Bolton and Ockenfels, 2000) predicting punishment of the first *and/or* the second party in allocation $(x^j_{FP},\ x^j_{SP},\ 200)$ if $\frac{200}{x^j_{FP} + x^j_{SP} + 200} < 1/3 \Leftrightarrow 400 < x^j_{FP} + x^j_{SP}$, (10) an "envy-active" rule predicting punishment of the first party in the same conditions as the envy 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, and finally, (11) the "envy-perspective" rule predicting punishment of the first party if $x^i_{FP} > x^i_{SP}$, and no punishment of the second party (third parties who follow this rule put themselves in the shoes of an envious second party).

3.2.2 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

[13] In other words, this rule predicts punishment for the first party if she does not choose the maximin allocation, maybe because that constitutes a "maximin norm" transgression. Charness and Rabin (2002) report that dictators are often willing to sacrifice part of their own material payoff to increase the payoff of all recipients, *especially* that of low-payoff recipients.

[14] Important: This rule is not in line with pure reciprocity models, e.g. Rabin (1993), which predict no punishment in the 3P treatment.

each allocation with probability $\varepsilon > 0$, in which case it is assumed that they randomize with equal probability between punishing or not punishing.[15] Consequently, the probability that a subject s deviates from her rule at any allocation is $\frac{\varepsilon}{2}$, while the probability that she follows her rule X_s times out of her *d* choices (20 in 2P, 40 in 3P) is:

$$(1-\frac{\varepsilon}{2})^{X_s} \times (\frac{\varepsilon}{2})^{d-X_s} .$$ [16]

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

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

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 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,

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

---

[15] 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.

[16] 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.

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.

3.2.3 Results of the Classification Procedure

Tables 6 and 7 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 number 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 $\hat{\varepsilon}/2$. 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.[17] 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 $(d + n)\cdot k$.[18] 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.

RESULT 2: *A combination of envious and selfish types can sufficiently capture third parties' punishment patterns. If we allow for more complexity, a combination of two different envious types, selfish and spiteful types, best explains the third parties' punishment pattern.*

As Table 6 shows, the classification procedure detected 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, and 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. envy 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 envy rule. Then 22 percent of the third parties are classified as

[17] 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.

[18] 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 $n\cdot k$).

envious and the error rate drops to 18.2 percent.[19] (3) Adding a third rule is suboptimal according to the AIC, which suggests that by the assumption that there are just selfish and envious types can sufficiently capture the punishment pattern in 3P. (4) If we nevertheless add a third rule, the algorithm picks the envy-perspective rule, and 34 percent of the third parties are classified as envious (20 percent envious and 14 percent envy-perspective). (5) If we add a fourth rule, spite is chosen.

TABLE 6— RESULTS OF CLASSIFICATION PROCEDURE IN 3P (THIRD PARTIES)

Number of rules	Rule(s) chosen	Number of third parties for each rule	ε	AIC	Chi-squared (p-Value)
1	selfish	55	0.309	-1042.2	
2	selfish, envy	43, 12	0.182	-862.5	549.31 (0)
3	selfish, envy, envy-perspective	36.5, 11, 7.5	0.162	-905.6	653.08 (0)
4	selfish, envy, envy-perspective, spite	36.5, 9, 7.5, 2	0.151	-971	712.24 (0)

We make two additional remarks; first, the relative success of the envy-perspective rule is somewhat surprising but also an illustration of how this classification procedure can be used to provide new intuitions on punishment.[20] From our knowledge, no experimental paper has provided evidence on this rule before. We 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.

Second, the previous results can be clearly used for predictive purposes. To provide an 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 envious ones) will probably stop punishing if their endowment rises enough, that is, if they are richer than the other parties, while other groups (the envy-perspective and spiteful ones) might punish even if they are richer than the other parties, while the envy-perspective group would punish only if the first party is richer than the second party. Our evidence provides support in

[19] 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.

[20] Subjects following this rule punish as an envious 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.

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.

RESULT 3: *A combination of envious and selfish types can sufficiently capture second parties' punishment patterns. If we allow for more complexity, a combination of envious, selfish, spiteful and reciprocal types best explains second parties' punishment patterns.*

Table 7 indicates the following behavioral patterns for second parties: (1) If we force the algorithm to choose only one rule, unsurprisingly 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 envy-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 envy rule. We can also see that a considerable fraction of 42 percent is then best classified as envious. 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 envious 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 envious and 13 percent as spiteful. (5) If we add a fourth rule, reciprocity is chosen (22 percent envious, 13 percent spiteful and 11 percent reciprocal). Note that our results are in line with Charness and Rabin (2002) who suggest that, considering distributional preferences alone (i.e., no reciprocity) and when no self-interest is at stake, about 20 percent of choices can be attributed to difference aversion (i.e., envy), and 10 percent to competitiveness (i.e., spite), whereas the remaining 70 percent can be attributed to social-welfare-maximization (that is, the kind of people who would never punish).

TABLE 7— RESULTS OF CLASSIFICATION PROCEDURE IN 2P (SECOND PARTIES)

Number of rules	Rule(s) chosen	Number of second parties for each rule	ε	AIC	Chi-squared (p-Value)
1	selfish	45	0.502	-572.2	
2	selfish, envy	26,19	0.290	-503.4	267.52 (0)
3	selfish, envy, spite	26, 13, 6	0.220	-506.9	390.66 (0)
4	selfish, envy, spite, reciprocity	24, 10, 6, 5	0.193	-545.9	442.58 (0)

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.[21]

To sum up, our classification analysis shows that a model assuming two types of players (selfish and envious) 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 envy can account for the occurrence of all punishment in our games: As we have seen in the previous section, reciprocity plays also an important role in 2P, and other minor variables affect third and second party punishment. Indeed, 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. However, the analysis also indicates that envy is a key explanatory motivation in our games.

3.3 The Strength of Second and Third 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. While this is not a problem when testing most theoretical models, we may lose some information concerning models of envy and reciprocity which respectively forecast a positive relation between the strength of punishment and the difference in payoffs and the size of the harm, i.e. the net payoff loss of the second party. We first take a look at third parties. An OLS

[21] 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.

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). That is, the bigger the difference in payoffs, the more the third party punishes the first and second parties.

In table 8, we investigate whether the difference in payoffs and the size or existence of harm predicts the strength of second party punishment. Column (1) reports 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. In column (2), 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 (1), we see that the dummy for the existence of harm 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 (3) 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 in 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 8—DETERMINANTS OF SECOND PARTY PUNISHMENT (*OLS)*

Dependent Variable	Strength of Punishment for the First Party		
Column	(1)	(2)	(3)
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 4: *Models that combine envious with reciprocal motives, like Falk and Fischbacher (2006), perform well in predicting the strength of second and third party punishment.*

The predictions in column (3) 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).[22]

4. Experimental Results II

We first devote this section to compare third and second party punishment and to study how they react to socially efficient choices. In addition, we shed light on whether and when bystanders become victims of punishment, and whether the availability of an alternative with a strictly equal allocation influences punishment behavior.

4.1 Comparing Second and Third Party Punishment

The previous section showed that envy plays an essential role in explaining the occurrence of both second and third party punishment, and that the difference in payoffs was crucial in understanding the strength of their punishment. However, we did not yet compare the strength of second and third party punishment. Since second parties are affected by the

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

first party's choice, and because we have shown that the mere existence of harm affects the intensity of the punishment, one might expect that second parties *overall* should punish more harshly than third parties.[23] We can only partly confirm this conjecture.

RESULT 5***: 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.***

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 can 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 3.3), and, (ii) third parties spend part of their money to punish bystanders. Thus, if we only look at the punishment of the first party, we observe that third parties spend *overall* less points than second parties (Mann-Whitney Test, z = 3.209, p = 0.001). For a more detailed analysis, figure 2 breaks down what happens in each game. The dots (squares) indicate the average punishment of third (second) parties for the first party in the left- and right-hand allocation in each of the ten games –further, a solid (dotted) line connects the second (third) party observations.[24] Figure 2 illustrates two important findings.

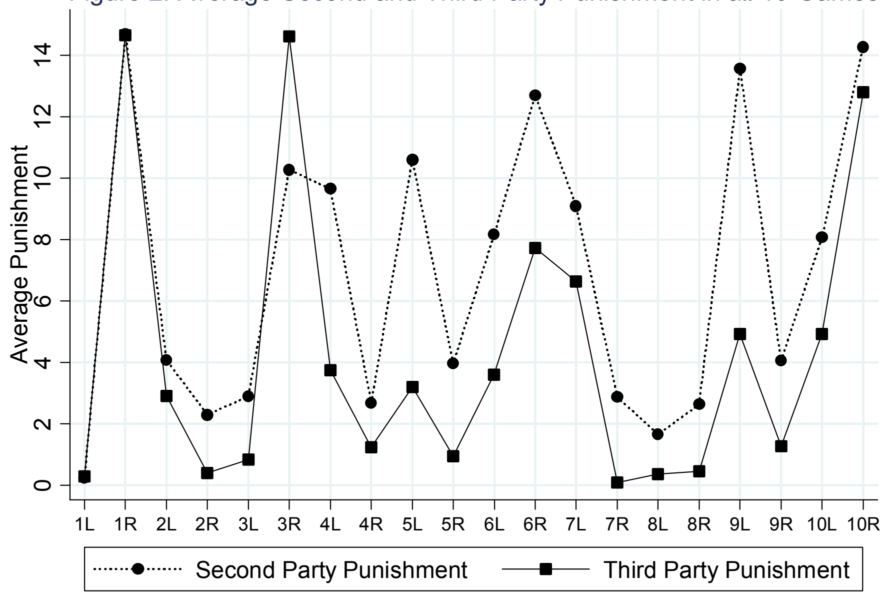
First, the solid line lies below the dotted line in most allocations, which indicates that third parties tend to punish the first parties more weakly than the second parties. However, the differences are significant only in 4 of the 20 allocations (Mann-Whitney Test on a 10 percent level: game 5 and 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 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

[23] Fehr and Fischbacher (2004, page 80) report that second parties punish much more strongly than do third parties in their experiments. However, this could be just an artefact of their design as third parties were consistently less disadvantaged than second parties and, as we saw before, this should affect the strength of their punishment.

[24] Note that we connected the points in figure 2 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.

classification of third and second parties indicated, and (iii) 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.

Figure 2: Average Second and Third Party Punishment in all 10 Games



RESULT 6: *Third party punishment very closely resembles second party punishment, which implies that third parties do not act more normatively and impartially than second parties. Third and second party punishment is correctly anticipated.*

Figure 2 also 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 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 is also the case for third parties in the two remaining games, where second parties punish both allocations equally. This latter fact provides additional evidence that third parties do not punish deviations from a norm of equity, efficiency, or maximin (otherwise they would not punish *both* choices), something supported as well by the classification analysis in section 3.2. Further, the overall similarity in punishment 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). The behavior in game 7 (250/80 vs. 80/250) also speaks against this assumption, as both parties punish the first party equally strongly for choosing the allocation (250/80) (Mann-Whitney Test, $z = 0.954$, $p = 0.340$).

We were not only interested in the actual punishment but also whether subjects anticipated it. For this reason we asked first parties in 2P and 3P (and second parties in 3P) about their expectations of punishment in each allocation. The pattern of actual punishment is very well anticipated in 2P and 3P. For instance, first parties correctly expect to be punished more strongly in the one allocation in the eight games where this actually happens ($p < 0.01$; Wilcoxon-Signed Rank-Test). However, subjects often expect to be punished somewhat more strongly than they are (3P: $z = 1.416$, $p = 0.156$; 2P: $z = 2.162$, $p = 0.030$). Figures 3 and 4 in the appendix show the average expectation of punishment in each allocation in 3P and 2P.

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 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. In fact, they punish especially in games 1, 3 and 10, where their balance is lowest. Further, in an OLS regression analysis, where we use the size of the payoff differences and the endowment of the second party to predict the strength of the punishment in all the cases where the second party endowment is lower than the third party endowment (< 200 points), we find that the second party endowment is an uninformative variable (t = -0.77, p = 0.446). 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. We chose game 1 (150/150 vs. 590/60) because we expected a large amount of punishment from our results and also because third and second parties punished the allocation (590/60) equally strongly. The experiment was conducted in Madrid in different university classes, with subjects from different disciplines (60 subjects participated in the 2P and 75 subjects in the 3P treatment).[25] 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: $z = 1.857$, $p = 0.063$). Therefore, this seems to contradict the idea that the specific response method should foster second but not third party punishment in our games.

4.2 Punishment of Socially Efficient and Inefficient Choices

Are second parties willing to accept a small disadvantage for a great advantage of the other player? Do third parties value social efficiency differently? The answer is no.

RESULT 7: *Second and third parties punish the first party strongly if she chooses the socially efficient allocation and becomes the "richer" party as a result, or if she chooses the inefficient allocation and that choice harms the second party.*

Games 1 (150/150 vs. 590/60) and 3 (560/60 vs. 120/140) provide evidence for the first part of our result, as third and second parties punish the socially efficient choice significantly more (Wilcoxon signed-rank test, 3P: $p < 0.001$; 2P: $p < 0.05$). Perhaps more surprisingly, third and second parties punish the richer first party significantly more even if she chooses the *Pareto-dominant* right-hand allocation in game 6 (280/240 vs. 390/240) (3P: $z = -2.590$, $p < 0.001$; 2P: $z = -2.534$, $p = 0.011$). Evidence for the second part of result 7 comes from games 4 (150/90 vs. 50/630) and 5 (220/260 vs. 220/400). The first party is

[25] The experimental protocol and the instructions were as similar as possible to those of the Zurich sessions. More information on this experiment is available upon request.

punished more strongly in game 4 by third parties (z = 2.510, p = 0.012) and second parties (z = 2.980, p = 0.003) if she is not willing to give up 100 points to increase the payoff of the second party by 540 points, and the choice of the *Pareto-dominated* allocation in game 5 is punished significantly more by third (z = 2.909, p = 0.004) and second parties (z = 3.092, p = 0.002).

The previous evidence has important consequences for some theories of social preferences. To start, a theory assuming that people punish anyone who does *not* choose a socially efficient allocation but never someone who chooses a socially efficient allocation – maybe because they punish deviations from an "efficiency norm" – is clearly at odds with the data. The model by Levine (1998) also fares badly because it assumes that some types of people punish others if they believe them to be selfish or spiteful, but not if they are altruistic – in other words, if they care about social efficiency. Now, the choice of the efficient allocation in the above mentioned games 1, 3, and 6 is hardly a clear signal of being selfish (an altruistic type would also choose it), but it is punished relatively more. Further, Levine (1998) predicts that the choice of the socially inefficient allocation in game 2 (100/100 vs. 50/530) should be a clear signal of selfishness and hence should be harshly punished, which is again at odds with the observed behavior.

4.3 Bystander Damaging

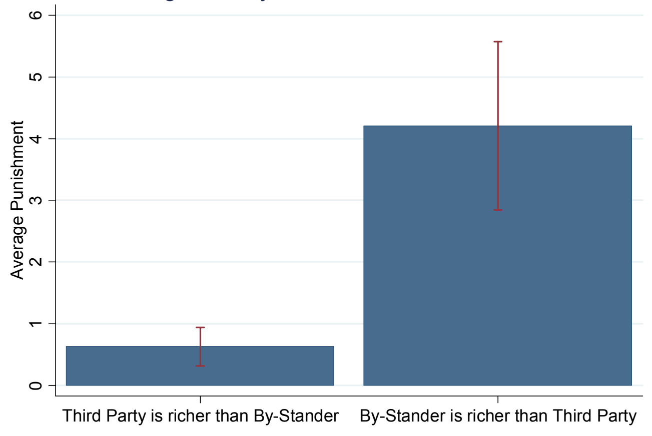
We allowed third parties to reduce the payoff of the second parties, who make no choice in the 3P, in order to study bystander damaging systematically, an important topic that has not received much attention until now. Some studies show that in dictator games the recipient is sometimes handicapped (Charness and Grosskopf, 2000; Kritikos and Bolle, 2001; Charness and Rabin, 2002; Zizzo, 2003; Dawes et al., 2007). In Charness and Grosskopf (2000), for instance, 34% of the dictators preferred the (dictator, dummy) allocation (600, 600) over (600, 900), and 12% of the dictators preferred the allocation (600, 600) over (625, 1200). Although the results from these studies indicate that bystanders are most often handicapped when they are richer, that could happen because of envy, spite, a concern for a norm of equity or, as Dawes et al. (2007) suggest, an "egalitarian" motive that predicts punishment (in our games) of those co-players getting a payoff larger than the *average* one.[26] Our evidence indicates that envy is the main factor behind the damaging of passive parties.

[26] More generally, this motive predicts punishment that reduces the standard deviation from the group mean. This coincides always with envy in two-player games and in most three-player games. In our 3P games, envy and the egalitarian motive can be discriminated in game 8. While envious third parties should not damage the

RESULT 8: *Third parties frequently reduce the payoff of bystanders; but only if they are richer than the third party. However, bystanders are treated differently than first parties.*

Figure 5 distinguishes average punishment in allocations where the bystander's payoff is smaller or bigger than the third party's endowment. We observe that bystanders are almost exclusively damaged when they are richer than third parties (Wilcoxon signed-rank test, z = -4.384, p < 0.001, two-sided). Every time we observe more than 20 percent of the third parties damaging the second party – i.e. in the right-hand allocation of games 2, 4, 5, 7, 9 – the bystander's payoff exceeds the third party's. For instance, the bystander is considerably more damaged in the Pareto-dominant allocation (z = - 2.568, p = 0.010) in game 5 (220/260 vs. 220/400).

Figure 5: By-Stander Punishment in 3P



Yet, third parties treat bystanders differently than first parties, as they damage bystanders overall less (z = 5.225, p < 0.001). This can be also nicely observed in game 7 (250/80 vs. 80/250), where the third party gets 50 points less than the first (second) party in the left- (right-) hand allocation. While about the same percentage of the first party and bystanders are damaged (29 vs. 24 percent), there are differences in the average strength of punishment (6.6 vs. 4.1 points) which leads to the result that bystanders are damaged less (z = -2.076; p = 0.037). Hence, our results demonstrate that bystanders frequently become victims if they are richer than a party who can damage them, although they are somewhat less affected by punishment compared to active parties.

---

bystander in the allocation (50/150), third parties following the egalitarian motive should do so. We find some support for the egalitarian motive since 18 percent damage the bystander in this allocation. However, in the classification analysis for third parties the envy rule outperforms the egalitarian motive.

4.4 The Impact of Strictly Equal Allocations on Punishment

This section finishes with the observation whether strictly equal allocations are treated differently compared to slightly unequal allocations. Abundant investigations in negotiation and mediation recommend reaching strictly equal outcomes because they serve as one goal the involved parties often accept and do not want to deviate from (e.g. Thompson 2005). Further, Güth et al. (2001) suggest that the availability of a strictly equal offer in an ultimatum mini-game can play a role since an unfair offer is more likely to be rejected if the alternative is a strictly equal instead of a slightly unequal offer. Motivated by this evidence, we investigate whether strictly equal allocations are less punished and whether deviations from strictly equal allocations are punished more, even if the alternative is socially efficient.

RESULT 9: *Deviations from strictly equal allocations are punished more, and the choice of strictly equal allocations is punished less, especially in 2P.*

To illustrate the first part of this result, consider games 1 (150/150 vs. 590/60) and 3 (560/60 vs. 120/140) *in 2P*. We observe that deviations from (150/150) in game 1 are punished more strongly than deviations from (120/140) in game 3 ($z = 2.869$, $p = 0.004$), a fact that is difficult to explain by envy, reciprocity, or any other theory. As evidence relating the second part of result 9, the choice of allocation (150/150) in game 1 is significantly less punished than the choice of (120/140) in game 3 (Wilcoxon signed-rank test, $z = -2.235$, $p = 0.025$). This phenomenon is perhaps even more puzzling from a theoretical point of view, as neither envy nor reciprocity predicts punishment in any of these two allocations. Reciprocity theories and Levine (1998) have similar problems to give a coherent account of the evidence from games 2 (100/100 vs. 50/530), 4 (150/90 vs. 50/630), and 8 (100/100 vs. 50/150), where harmful choices are punished much more if they are not strictly egalitarian (as in game 4).

Interestingly, the picture is less clear for third parties. On one side, only 11 percent punish the egalitarian choice (100/100) in game 2, whereas 29 percent punish choice (150/90) in game 4, which cannot be reconciled with envy. This significant difference ($z = -2.018$, $p = 0.043$) could in principle be attributed to the existence of the strictly equal allocation in game 2 (but also to players following an envy-perspective rule). However, the evidence from games 1 and 3 seems to be at odds with this interpretation. To start, the strictly equal allocation in game 1 appears to have no alleviating impact on punishment, as it is not punished less than the slightly unequal allocation in game 3 ($z = 1.350$, $p = 0.177$). Further, the alternative allocation is not punished more in game 1 (z = 0.023, p = 0.981), where it constitutes a deviation from strict equality. One explanation for this mixed evidence might be that third parties, who get always a payoff of 200 points, are uninfluenced by strict equality between first and second party payoffs if they do not get the same payoff. More experimental evidence is therefore required to settle this point.

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 envy is a crucial cause of third and second party punishment. As a result, both parties often punish socially efficient choices and bystanders frequently become victims if they are in a better material situation. Our data also shows 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, pure reciprocity models like Rabin (1993), Dufwenberg and Kirchsteiger (2004), and Cox et al. (2007) fail to account for third party punishment. For this reason, we believe that reciprocity alone should not be applied to explain punishment. In contrast, inequity-aversion models like Fehr and Schmidt (1999) and Falk and Fischbacher (2006) fare much 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 how strong the punishment will be. Levine (1998) is inconsistent with the fact that many choices are not punished, the heavy punishment of socially and Pareto efficient actions, and with the role that strict equality plays in reducing punishment (reciprocity faces also 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. Indeed, our data indicates 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 envious ones) 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.

We gain several important findings for policy makers. Subjects punish socially efficient but disadvantageous choices severely even if they are Pareto-dominant. Therefore, policies and institutions that create larger inequalities may have severe negative side effects, even if they improve everyone’s income; it can therefore be beneficial to complicate or avoid punishment altogether, or alternatively look for strictly equal outcomes, which induce much less destruction of the pie. Further, the role of third parties in acting as mediators and moderators of conflict should be taken with care, as they are equally “infected” by envy.

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 third 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). Third, Falk et al. (2005) report that defectors in a prisoner’s dilemma punish cooperators, 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 (envious, reciprocal?), a topic that deserves also further study. In addition, our results and those from Falk et al. (2005) hint that people mostly punish not because they are inequity averse, but because they dislike those who are richer. We believe that this issue requires further empirical and theoretical understanding.

Appendix

TABLE 4—PROBABILISTIC CHOICES OF FIRST PARTIES IN 2P & 3P

Game				Left		Right	
				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

TABLE 5—PREDICTION OF PUNISHMENT IN 2P & 3P

Game				Predictions Left			Predictions Right		
	Left		Right	Punishment of first party in 2P	Punishment of first party in 3P	Punishment of By-Stander (3P)	Punishment of first party in 2P	Punishment of first party in 3P	Punishment of By-Stander (3P)
1	(150,150)	vs.	(590,60)	EF	EF		EN, R, EQ, G	EN, ERC, EQ, G	ERC
2	(100,100)	vs.	(50,530)	R, EF, G	EF, G		EQ	ERC, EQ	EN, ERC
3	(560,60)	vs.	(120,140)	EN, R, EQ, G	EN, ERC, EQ, G	ERC	EF	EF	
4	(150,90)	vs.	(50,630)	EN, R, EF, G	EF, G		EQ	ERC, EQ	EN, ERC
5	(220,260)	vs.	(220,400)	R, EF	EN, ERC, EF	EN, ERC	EQ	EN, ERC, EQ	EN, ERC
6	(280,240)	vs.	(390,240)	EN, EF	EN, ERC, EF	EN, ERC	EN, EQ, G	EN, ERC, EQ, G	EN, ERC
7	(250,80)	vs.	(80,250)	EN, R, G	EN, G				EN
8	(100,100)	vs.	(50,150)	R, G	G		EQ	EQ	
9	(250,150)	vs.	(110,290)	EN, R, G	EN, G		EQ	EQ	EN
10	(250,150)	vs.	(330,70)	EN	EN		EN, R, EQ, G	EN, EQ, G	

EN = Envy, R = Reciprocity, G = Greed, EQ = Equity rule, EF = Efficiency rule, ERC = Bolton-Ockenfels Model (in 3P).

Figure 3: Actual and Expected Second Party Punishment in all 10 Games

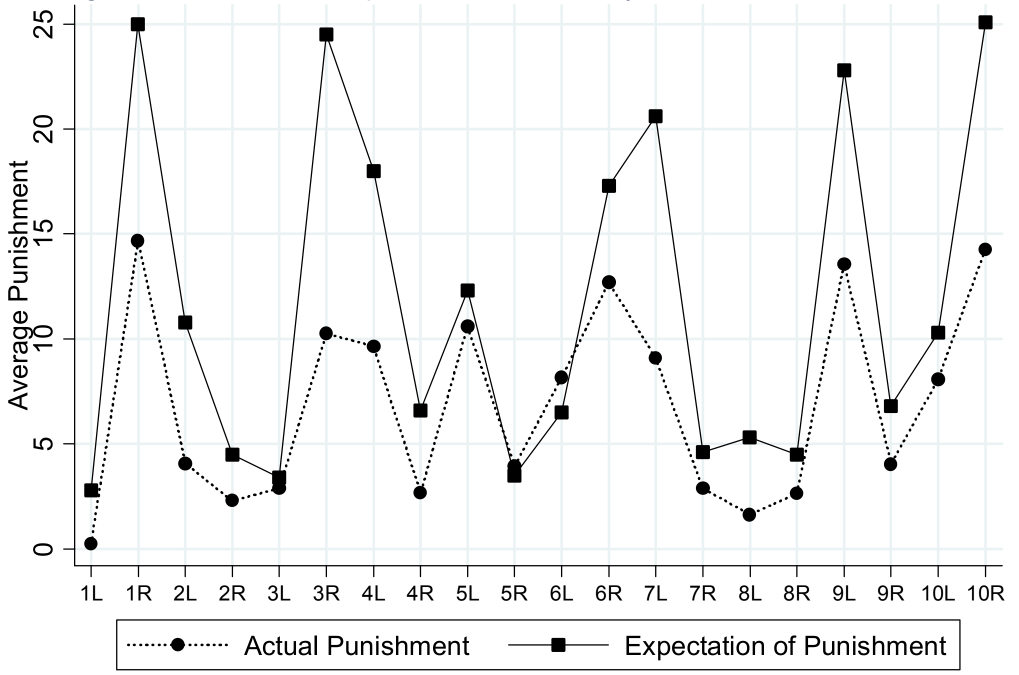
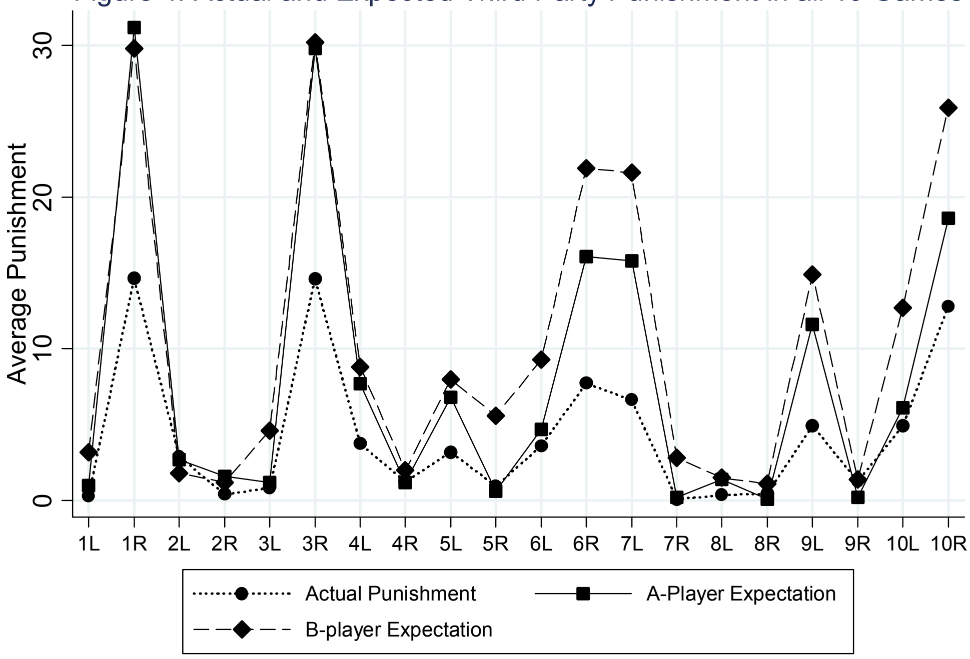


Figure 4: Actual and Expected Third Party Punishment in all 10 Games



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)

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

or, equivalently,

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

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

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

Further, since the second derivative of (1) $-\frac{dn}{\varepsilon^2} + \frac{4(1-\varepsilon)\sum X_s}{\left[\varepsilon(2-\varepsilon)\right]^2}$ 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 larger than 1, the optimum is clearly $\hat{\varepsilon} = 1$.▪

References

Babcock, Linda, Loewenstein, George, Issacharoff, Samuel, and Camerer, Colin. "Biased Judgments of Fairness in Bargaining", *American Economic Review*, 1995, 85(5), pp. 1337-1343.

Bendor, Jonathan, and Swistak, Piotr. "The Evolution of Norms", *American Journal of Sociology*, 106, pp. 1493-1545.

Bolton, Gary E., and Ockenfels, Axel. "ERC: A Theory of Equity, Reciprocity, and Competition", *American Economic Review*, 2000, 90(1), pp. 166-93.

Brandts, Jordy and Charness, Gary. "Hot vs. Cold: Sequential Responses and Preference Stability in Experimental Games", *Experimental Economics*, 2000, 2(3), pp. 227-238.

Cason, Timothy and Mui, Vai-Lam. "Social Influence in the Sequential Dictator Game", *Journal of Mathematical Psychology*, 1998, 42, pp. 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, pp. 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*, forthcoming.

Charness, Gary and Grosskopf, Brit. "Relative Payoffs and Happiness: An Experimental Study", *Journal of Economic Behavior and Organization*, 2001, 45, pp. 301-328.

Charness, Gary, and Rabin, Matthew. "Understanding Social Preferences with Simple Tests", *Quarterly Journal of Economics*, 2002, 117, pp. 817-869.

Cox, James C., Friedman, Daniel, and Gjerstad, Steven. "A Tractable Model of Reciprocity and Fairness", *Games and Economic Behavior*, 2007, 59, pp. 17-45.

Dawes, Christopher, Fowler, James H., Johnson, Tim, McElreath, Richard and Smirnov, Oleg. "Egalitarian Motives in Humans", *Nature*, 446, pp. 794-796.

Dufwenberg, Martin, and Kirchsteiger, Georg. "A Theory of Sequential Reciprocity", *Games and Economic Behavior*, 2004, 47, pp. 268-98.

El-Gamal, Mahmoud, and David Grether. "Are People Bayesian? Uncovering Behavioral Strategies," *Journal of the American Statistical Association*, 90(432), December 1995, pp. 1137-1145.

Elster, Jon. "Social Norms and Economic Theory", *Journal of Economic Perspectives*, 1989, 3(4), pp. 99-117.

Engelmann, Dirk, and Strobel, Martin. "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments", 2004, *American Economic Review*, 94(4), pp. 857-869.

Engle-Warnick, Jim and Leibbrandt, Andreas. "Who gets the Last Word? An Experimental Study of the Effect of a Peer Review Process on the Expression of Social Norms", Cirano Working paper series, 2006.

Falk, Armin; Fehr, Ernst and Fischbacher, Urs. "Driving Forces behind Informal Sanctions" *Econometrica*, 2005, 7(6), pp. 2017-30.

Falk, Armin and Fischbacher, Urs. "A Theory of Reciprocity", *Games and Economic Behavior*, 2006, 54, pp. 293-315.

Falk, Armin and Kosfeld, Michael. "The Hidden Cost of Control", *American Economic Review*, 2006, 96, pp. 1611-1630

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

Fehr, Ernst and Gächter, Simon. "Cooperation and Punishment in Public Goods Experiments", *American Economic Review*, 2000, 90, pp. 980-994.

Fehr, Ernst and Schmidt, Klaus. "A Theory of Fairness, Competition and Cooperation", *Quarterly Journal of Economics*, 1999, 114(3), pp. 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, pp. 1912-1917.

Fischbacher, Urs. "z-Tree: Zurich Toolbox for Ready-made Economic Experiments", *Experimental Economics*, forthcoming.

Greif, Avner. "Contract enforceability and economic institutions in early trade: The Maghribi traders` coalition", *American Economic Review*, 1993, 83, pp. 525-548.

Greif, Avner. "Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies", *Journal of Political Economy*, 1994, 102, pp. 912-950.

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, pp.367-388.

Güth, Werner, Huck, Stefan and Müller. "The Relevance of Equal Splits in Ultimatum Games", *Games and Economic Behavior*, 2001, 37, pp. 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.

Kritikos, Alexander and Bolle, Friedel. "Distributional Concerns: Equity- or Efficiency-Oriented?" *Economics Letters*, 2001, 73, pp. 333-338.

Levine, David K. "Modeling Altruism and Spitefulness in Experiments." *Review of Economic Dynamics*, 1998, 1, pp. 593-622.

López-Pérez, Raúl. "Aversion to Norm-Breaking: A Model", *Games and Economic Behavior*, forthcoming.

Rabin, Matthew. "Incorporating Fairness into Game Theory and Economics", *American Economic Review*, December 1993, 83(5), pp. 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), pp. 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, pp. 581-602.

Thompson, Leigh. "The Mind and Heart of the Negotiator", 2005, Englewood Cliffs, NJ: Prentice Hall.

Zizzo, Daniel J. "Money Burning and Rank Egalitarianism with Random Dictators", *Economics Letters*, 2003, 81, pp. 263-66.