



Repositorio Institucional de la Universidad Autónoma de Madrid

<https://repositorio.uam.es>

Esta es la **versión de autor** del artículo publicado en:
This is an **author produced version** of a paper published in:

i
Journal of Conflict Resolution 55.5 (2011): 761-784

DOI: [http://dx.doi.org/ 10.1177/0022002711408010](http://dx.doi.org/10.1177/0022002711408010)

Copyright: © The Author(s) 2011

El acceso a la versión del editor puede requerir la suscripción del recurso
Access to the published version may require subscription

The Dark Side of Altruistic Third-Party Punishment

Andreas Leibbrandt¹ and Raúl López-Pérez²

Abstract

This article experimentally studies punishment from unaffected third parties in ten different games. The authors show that third-party punishment exhibits several features that are arguably undesirable. First, third parties punish strongly a decider if she chooses a socially efficient or a Pareto efficient allocation and becomes the richest party as a result. Interestingly, this form of punishment is especially pro-nounced in women and more left-wing participants. Second, third parties punish strongly a decider if she chooses an equitable allocation and becomes the richest party as a result. Finally, third parties considerably punish passive parties who make no choice, especially if the latter are richer than the third party. Implications of these findings for social theory are discussed.

Keywords

altruistic punishment, efficiency, inequity-aversion, third parties

Introduction

Contemporary theories in the fields of conflict resolution, economics, and political science are often based on the hypothesis that all agents are selfish. This implies that individuals will not damage, harm, or punish another individual if such behavior is costly and provides no future material benefits. Yet a large body of experimental

Corresponding Author:

Raúl López-Pérez, Facultad de Ciencias Económicas, Universidad Autónoma de Madrid, Madrid 28049, Spain

Email: raul.lopez@uam.es

research has challenged this prediction. In the ultimatum game, for instance, responders frequently reject low offers from proposers (Güth, Schmittberger, and Schwarze 1982; Camerer and Thaler 1995), and noncooperators are often punished in social dilemma games with a punishment stage (Ostrom, Walker, and Gardner 1992; Fehr and Gächter 2000). A possible interpretation of this evidence is that some agents are motivated not only by their material interest but also by forces like a desire to retaliate against unfair actions (negative reciprocity; see Gouldner 1960; Rabin 1993; Levine 1998), or a preference to achieve equitable allocations, even at a cost (inequity aversion; see for instance Fehr and Schmidt 1999).

Such nonselfish motives may have important behavioral consequences. In particular, *altruistic punishment*, the costly damaging behavior observed in the previous experiments, may deter people from breaking social norms and informal agreements, thus fostering cooperation and the well functioning of groups (Bowles and Gintis 2000; Fehr and Gächter 2002). On the other hand, egalitarian preferences like those posited by Fehr and Schmidt (1999) may have additional far-reaching implications for the study of conflict and politics. We suggest four possible examples.

First, voters or politicians may support policies that foster equality at the cost of social efficiency, even if those policies do not affect them. An example of such policies is mentioned in Elster (1999, 176), who suggests that those Chinese farmers who were better off because they had fruit trees were ordered to cut them down during the Cultural Revolution. Second, individuals may favor income redistribution even at a cost. On one hand, this implies that inequalitarian societies may be more prone to social discontent and even conflict. This is for instance consistent with the anthropological evidence in Knauff (1991) who argues that traditional societies that followed egalitarian motives were sometimes characterized by large amounts of violence. On the other hand, the achievement of egalitarian outcomes may alleviate conflict—a finding that may be of practical use in conflict mediation (e.g., Thompson 2005). Third, strategies to cope with collective action problems (Olson 1965) may be impaired if they lead to inequitable outcomes when agents have varying interests in the collective good or different levels of income. To avoid the risk of conflicts in such cases, it may be advisable to devise strategies so that agents contribute in proportion to their income, and not only to their benefits from the good. A possible illustration of this kind of problems is the polemic between developed and underdeveloped countries with respect to the size of emission reductions to mitigate global warming. Fourth, egalitarian tendencies may partially account for the development of institutions. For instance, Boehm (1993, 1999) argues that many small-scale societies are egalitarian and not based on strong dominance hierarchies, as leaders are disciplined by potential collective sanctions, and he suggests that such institutional design could be the result of some egalitarian “ethos.”

In this article, we use laboratory experiments with two main objectives in mind: (1) explore potential egalitarian motives to reduce others’ payoffs in order to achieve payoff equality and (2) illustrate several consequences of these motives. Laboratory experiments allow us to control for other motives like material interest and

reciprocity, which are difficult to separate from egalitarian motives in the field. More precisely, our experiment consists of ten different one-shot games with three players and different payoff allocations. In these games, a first party decides between two allocations of payoffs between him and a second, passive party. Then, a third party can reduce the payoff (i.e., punish) of any of the other parties (or both) at a cost to himself (Eckel and Grossman 1996; Fehr and Fischbacher 2004).¹ Since the third party is an unaffected spectator who has not been harmed by the first party, he has no retaliatory reason to punish. Further, he cannot have any self-interested motive to punish: Punishing in our games is a costly behavior that cannot provide any monetary benefit.

Studying third-party punishment seems especially important to understand norm compliance because unaffected third parties 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 affected second parties. Our results are in line with the idea that egalitarian motives play a prominent role in punishment. Indeed, third parties often punish others, and we observe that punishment is significantly correlated with the existence of a disadvantage in payoffs between the third party and the punished party. Hence, our findings offer support for recent economic preference models like Fehr and Schmidt (1999), which assume that a significant proportion of the population dislikes having less than others and is willing to incur personal costs to diminish others' resources in order to reduce inequity.

In addition, our large variety of games allows us to find comprehensive evidence for several implications of these egalitarian motives that qualify the literature on altruistic punishment. While this literature has stressed the potential beneficial effects of altruistic punishment on the well functioning of societies, we offer in this article evidence that it can also exhibit arguably undesirable features. Put illustratively, altruistic punishment can have a "dark side," by which we mean three things.

First, we observe that third parties punish the decider strongly if she chooses a socially efficient allocation (that is, one maximizing the sum of payoffs) and becomes the "richest" party as a result. Perhaps more strikingly, richer deciders are also frequently punished *even* if their choices inflicted no harm on another party; that is, even if they choose a Pareto-efficient allocation. Interestingly, this form of punitive behavior is especially pronounced in women and more left-wing participants. This latter finding is related to and extends the scope of earlier findings on individual differences in redistribution preferences depending on political orientation and gender (e.g., Frohlich et al. 1984; Haidt and Graham 2007; Croson and Gneezy 2009; Graham, Haidt, and Nosek 2009; Alesina and Giuliano 2009). For example, Frohlich et al. (1984) find that females who favor democratic parties are most likely to choose egalitarian outcomes in decision experiments, and Haidt and Graham (2007) find that liberals are less willing to accept outcome inequalities than conservatives. Further, Alesina and Giuliano (2009) use data from the General Social Survey (GSS) and the World Value Survey (WVS) and find that women tend to be more favorable toward redistribution than do men in many different countries and institutional settings.

Second, we observe that third parties punish deciders if they choose the most equitable allocation—that is, the one minimizing the payoff distance between players—and become the “richest” party as a result. This is a subtle but important issue, as it is often argued that people are punished if they behave “unfairly”—for example, low offers from proposers in the ultimatum game might be rejected because they are less equitable than the equal sharing. Contrary to this interpretation, our results indicate that equity in itself is not important for punishment but disadvantageous inequity is.

Third, we observe that third parties frequently reduce the payoff of the passive second party, but mostly if her payoff is higher than the third party’s payoff. Yet passive second parties are relatively less harmed than active first parties showing that third parties take the innocence of second parties at least to some extent into account. In any case, our results suggest that disadvantageous inequity is a key determinant of aggression toward bystanders, an important issue given historical evidence and countless law cases dealing with individuals who harmed bystanders or vandalized their property. For instance, Wilson (1988) studies feuding in nineteenth-century Corsica and argues that envy contributed to the occurrence of behaviors such as the destruction of livestock, the cutting down of trees, or even the killing of humans.

The rest of the paper proceeds as follows. The next section presents our experimental design and procedures, whereas we report the results of our experiment afterwards. The last section concludes with a discussion of these results.

Experimental Design and Procedures

Participants in our study play ten games. All games are three-player games and have the same two-stage structure. In the first stage, one player (the *first party*) chooses between a left-hand and right-hand allocation of payoffs between herself and another player (the second party, or *bystander*). Table 1 shows the two allocations available in each game, presented in points (10 points = 1 Swiss franc).

In the second stage, the third party can punish the first party *or/and* the bystander. Importantly, the third party is endowed with 200 points in each allocation of each game, meaning that the first party’s choice never affects her payoff. The third party can use up to 50 points to punish and each point spent reduces the payoff of the

Table 1. The Allocations in the Ten 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)

Note: At each allocation, the left-hand payoff refers to the first party, and the right-hand payoff to the bystander; the third party always has an endowment of 200 points.

punished player by three points. Hence, if the first party chooses allocation (x_1, x_2) in a game and the third party punishes her with p_1 points and the bystander with p_2 points ($p_1 + p_2 \leq 50$), the payoffs in this game are $x_1 - 3p_1$ for the first party, $x_2 - 3p_2$ for the second party, and $200 - p_1 - p_2$ for the third party. Note well that the bystander makes no payoff relevant decision in any game; that is, she is a passive player (however, we elicited second parties' beliefs about the third party's behavior in each game).²

We selected these ten games to perform a comprehensive study of (1) the punishment of Pareto and socially efficient choices, (2) the punishment of equitable allocations, and (3) the punishment of bystanders. Consider, for instance, game 3 (560/60 vs. 120/140) where the left-hand allocation is socially efficient and game 6 (280/240 vs. 390/240) where the right-hand allocation is Pareto efficient. We were interested in how third parties react to the choice of socially/Pareto-efficient allocations. In addition, these games allow us to study whether third parties refrain from punishing the choice of equitable allocations, like the allocation (280/240) in game 6. As another illustration, consider game 2 (100/100 vs. 50/530) and game 7 (250/80 vs. 80/250) where the bystander gets in the left-hand allocations a payoff lower than 200—that is, lower than the third party's payoff—whereas she gets a larger payoff in the right-hand allocations. Our games allow us to investigate plausible determinants of the punishment of the bystander, like the payoff disadvantage between her and the third party. Table A1 in the appendix shows the main features of each game.

A total of 165 participants took part in this experiment. Each participant was randomly assigned a role and anonymously matched in groups of three. Participants received role-dependent instruction sheets that explained the extensive form of the games but did not provide information about payoff constellations. To make sure that participants understood the rules they had to fill out control questions. We used neutral language and avoided terms such as 'punishment.' Every participant always played the ten games in the same role, and the ten games were presented one at a time. Further, the order in which they were played was randomly predefined for each group. Participants were never told about their counterparts' previous choices to prevent repeated game effects and changes of mood, which could generate history effects. For this reason, we employed the strategy method to elicit the punishment decisions of the third parties, that is, in each game, they had to indicate for both allocations the number of punishment points (0–50) that they wanted to assign to the other participants.³ Only one game was randomly selected for payment at the end of the experiment in order to prevent income effects. These features imply that each game can be treated as a one-shot game—even though subjects were always matched with the same participants.

The experiment lasted approximately 60 minutes and was conducted with the Z-tree software (Fischbacher 2007). Most participants were students from 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). Average earnings were 30 Swiss francs (around \$ 24) including 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 instructions for the third party are provided in the appendix B.

Experimental Results

This section starts with a brief overview of the punishment behavior of third parties. Thereafter, we study how third parties react to socially efficient (Punishment of Socially Efficient and Inefficient Choices section) and equitable choices (section Punishment of Equitable Choices), and whether and when bystanders become victims of punishment (Bystander Damaging section). Overall, we observe that 54 percent of the third parties punish at least once. On average, first parties are punished by 19.3 percent of the third parties and bystanders by 12.1 percent in each allocation. In terms of the strength of punishment, third parties reduce the payoff of the other two players by approximately 380 points, that is, on average by 19 points per allocation. They reduce the payoff of first parties significantly more than the payoff of bystanders (on average 12.9 compared to 6.1 points per allocation, Wilcoxon signed-rank test, $z = 3.672, p = .0002$). Table 2 summarizes the frequency and damage (i.e., point reduction) of third-party punishment, distinguishing between punishment for first parties and bystanders. Table A2 in the appendix shows the behavior of the first parties.

The size of the (disadvantageous) payoff difference between the third party and the other parties seems to be a major factor in explaining punishment in our games. To illustrate this issue, Figure 1 depicts the individual punishment decisions from the third parties who punish at least once dependent on the size of the payoff differences. The horizontal axis indicates the payoff difference between the third and the punished party (the dashed vertical line corresponds to zero distance;

Table 2. Frequency and Damage of Punishment—Third Parties

				First party				Second p. (bystander)			
Game				Left		Right		Left		Right	
1	(150,150)	vs.	(590,60)	.06	(0.9)	.44	(44.1)	.09	(1.5)	.04	(0.9)
2	(100,100)	vs.	(50,530)	.11	(8.7)	.06	(1.2)	.04	(2.7)	.26	(27.9)
3	(560,60)	vs.	(120,140)	.45	(44.1)	.07	(2.4)	.06	(0.9)	.15	(4.5)
4	(150,90)	vs.	(50,630)	.29	(11.4)	.07	(3.6)	.04	(2.1)	.26	(20.7)
5	(220,260)	vs.	(220,400)	.24	(9.6)	.09	(2.7)	.13	(4.5)	.22	(16.5)
6	(280,240)	vs.	(390,240)	.22	(10.8)	.33	(23.1)	.11	(2.7)	.13	(3.0)
7	(250,80)	vs.	(80,250)	.29	(19.8)	.02	(0.3)	.02	(0.3)	.24	(12.3)
8	(100,100)	vs.	(50,150)	.06	(1.2)	.04	(1.5)	.06	(1.2)	.18	(6.0)
9	(250,150)	vs.	(110,290)	.26	(15.0)	.06	(3.9)	.02	(1.2)	.22	(13.5)
10	(250,150)	vs.	(330,70)	.26	(15.0)	.44	(38.4)	.11	(1.5)	.04	(0.3)

Note: Numbers outside the parentheses indicate the frequency of punishment. Average damage (i.e., point reduction) by all participants in parentheses. The endowment of the third party is always 200 points. Fifty-five observations were made in each allocation of each game.

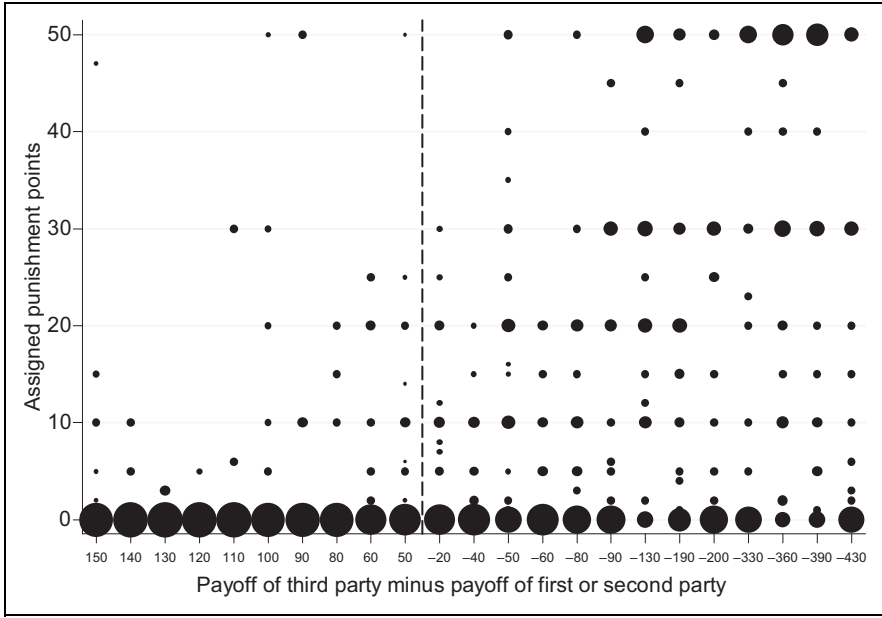


Figure 1. Third-party punishment depending on payoff differences

to the right of this line the third party gets less than the punished party). The location of the dots indicates the number of points spent from third parties depending on the distance. The size of the dots is proportional to the number of observations, that is, given a payoff distance between another player and the third party, a dot becomes larger as more and more third parties spend the same amount to punish that player.⁴ The figure suggests that payoff differences play an important role in the decision to punish. In fact, third parties rarely punish if their payoffs are higher than that of the first/second party but often and severely if their payoffs are lower.

Figure 1 also suggests that much of the observed punishment is in line with the model of inequity-aversion by Fehr and Schmidt (1999). They assume that people derive disutility from disadvantageous inequity toward any other player (weighted by a factor of $\alpha/n - 1$, where n is the number of players and $\alpha > 0$), and from advantageous inequity toward any other player (analogously weighted by a factor $\beta/n - 1$, where $\beta > 0$ is the corresponding coefficient). Taking into account that our games can be treated as one-shot games due to our experimental design, their model hence predicts that an inequity-averse third party with $\alpha > 2$ would punish the richer coplayer at any allocation, but never a poorer coplayer (if both coplayers are richer, the third party would be indifferent regarding whom to punish). The precise predictions by Fehr and Schmidt (1999) regarding the occurrence of punishment in our games are shown in Table A3 in the appendix. In Leibbrandt and

López-Pérez (2009), we apply to our data a classification analysis that provides further evidence that the punishment pattern of third parties is largely in line with inequity-aversion.⁵

Punishment of Socially Efficient and Inefficient Choices

In this section, we concentrate our attention on games 1 through 6, which have a unique allocation maximizing the sum of payoffs—that is, a socially efficient allocation. These games allow us to explore two questions: (1) Do third parties react punitively when the first party chooses a socially efficient allocation, thus generating a great payoff gain to one of the parties at a small or null payoff loss for the other party? and (2) Do third parties punish the first party for choosing socially inefficient allocations? The answer to the first question is affirmative, under one condition.

- *Result 1a:* Third parties punish the first party frequently and strongly if she chooses the socially efficient allocation, but only if she becomes the richest party as a result. This is true even if no one is harmed by the first party's choice.

Evidence for Result 1a. We focus first on games 1 (150/150 vs. 590/60), 3 (560/60 vs. 120/140), and 6 (280/240 vs. 390/240), where the first party gets the highest payoff in the socially efficient allocation. In these games, the frequency of punishment for the first party when she chooses this allocation is 44, 45, and 33 percent, respectively. These percentages are substantially above the average frequency of punishment for the first party (19.3 percent).

Punishment of the socially efficient choice in these games is not only frequent but also in terms of strength clearly above average. In fact, the socially efficient allocations in games 1 and 3 are the most strongly punished allocations in our experiment (44.1 points reduction in both). Even in game 6, the punishment of the first party for choosing the Pareto-dominant allocation is clearly above average (23.1 points reduction; recall that average point reduction for the first party across all allocations is 12.9 points). This latter result shows that third parties punish a socially efficient choice even if it harms nobody. Further, we note that the socially efficient choice is punished significantly more than the alternative choice in all of these games (game 1: Wilcoxon signed-rank tests, $z = -4.628$, $p < .0001$; game 3: $z = 4.934$, $p < .0001$; game 6: $z = -2.590$, $p < .001$).

Consider now games 2 (100/100 vs. 50/530), 4 (150/90 vs. 50/630), and 5 (220/260 vs. 220/400). In these games, the first party does not get the highest payoff if she chooses the socially efficient allocation. In line with Result 1a, we observe little and modest punishment for the first party if she chooses the efficient allocation: The frequency of punishment (6, 7, and 9 percent, respectively) is always well below the 19 percent average frequency of punishment across allocations, and the average reduction in points in these efficient allocations (1.2, 3.6, and 2.7, respectively) is also clearly below the average point reduction across

allocations (12.9). Note that these results are also below the above cited values of frequency and severity of punishment in games 1, 3, and 6.

In order to obtain a more accurate picture of this issue, we pass now to consider a natural, related question, that is, do third parties punish socially *inefficient* choices? The answer is again positive, under some conditions.

- *Result 1b*: If the first party chooses a socially inefficient allocation in which the first and the second party do not get the same payoff, third parties punish the first party considerably, provided that at least one of the following conditions holds: (i) the inefficient choice reduces the bystander's payoff (compared to the alternative) and (ii) the first party gets a larger payoff than the third party.

Evidence for Result 1b. Consider first games 4 (150/90 vs. 50/630), 5 (220/260 vs. 220/400), and 6 (280/240 vs. 390/240). In these games, there is no strict equality of payoffs. Moreover, at least one of the above mentioned conditions (i) and (ii) is satisfied. We observe that punishment of the socially inefficient choice is relatively frequent (29, 24, and 22 percent, respectively) and moderately harsh (11.4, 9.6, and 10.8 points reduction, respectively). Furthermore, games 4 and 6 suggest that conditions (i) and (ii) are each on their own sufficient for the occurrence of considerable punishment (in game 5, both conditions hold). Observe also that third parties punish the socially inefficient choice significantly more than the alternative choice in game 4 (Wilcoxon signed-rank test, $z = 2.510$, $p = .012$) and game 5 ($z = 2.909$, $p = .0036$). This does not happen in game 6, where our previous Result 1a applies.

That conditions (i) and (ii) are both important is further suggested by game 3 (560/60 vs. 120/140), where none of these conditions hold. Indeed, punishment of the first party in the socially inefficient allocation is infrequent (7 percent) and modest (2.4 points reduction). Finally, games 1 (150/150 vs. 590/60) and 2 (100/100 vs. 50/530) suggest that strict equality of payoffs has an alleviating effect on punishment, even if condition (i) holds, as in game 2. In these games, the frequency of punishment of the socially inefficient choice is well below average (6 and 11 percent, respectively), and the same is true for the strength of punishment (0.9 and 8.7 points reduction, respectively). We further clarify the effect of strict payoff equality in the Punishment of Equitable Choices section.

We emphasize again that the evidence in Results 1a and 1b (especially Result 1a) is largely consistent with inequity-aversion as modeled in Fehr and Schmidt (1999). This model does not only account well for the occurrence of punishment but also for the observed intensity of punishment. The latter point is nicely illustrated by the above mentioned evidence from game 6 (280/240 vs. 390/240), where we observe that the first party gets significantly more punished if she increases the payoff distance to the other players by choosing the *Pareto-dominant* allocation 390/240. Yet it is worthy to stress that payoff differences are not the only significant factor behind third-party punishment. Indeed, we have noted before that the choice of the socially inefficient allocation in game 4 (150/90 vs. 50/630) is punished frequently and

considerably, which is obviously at odds with inequity aversion, and points out the significance of other motivations. For instance, third parties might punish the first party if her choice harms the passive party, that is, if the decider chooses in a game the allocation with the smallest bystander's payoff—that is, third parties may be motivated by indirect reciprocity (Seinen and Schram 2006).

In contrast, other recent economic models on nonselfish preferences do not seem to provide much additional insights. Thus, the model by Bolton and Ockenfels (2000) cannot rationalize much of the observed third-party punishment, as it predicts no punishment in games 1 (150/150 vs. 590/60), 3 (560/60 vs. 120/140), and 4 (150/90 vs. 50/630), where we observe considerable punishment.⁶ To understand these predictions, note that this model predicts punishment of the first *and/or* the second party in any allocation if the third party can use punishment to bring her relative payoff closer to $1/3$, the equitable relative payoff in three-player games. More precisely, let $\sigma = x_1 + x_2 + 200$ denote the sum of players' payoffs at any allocation and hence $200/\sigma$ denote the third party's relative payoff. Clearly, this relative payoff is smaller than $1/3$ if $400 < x_1 + x_2$, whereas p units of punishment (which imply a reduction of $3p$ points in the payoff of the sanctioned party) increase the relative payoff if $\frac{200-p}{\sigma-3p-p} > \frac{200}{\sigma} \Leftrightarrow \sigma < 800 \Leftrightarrow x_1 + x_2 < 600$. Hence, Bolton and Ockenfels (2000) predict punishment only if $400 < x_1 + x_2 < 600$, which is not satisfied at any allocation in games 1, 3, and 4.

Further, a theory assuming that people punish anyone who does *not* choose a socially efficient allocation but never someone who chooses a socially efficient allocation 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 harshly. 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.⁷

We finish this section with a discussion of whether the punishment of socially efficient choices is gender specific and related to the individual political orientation. While there is experimental evidence that there are gender differences in pro-social behavior more generally (for a survey see Croson and Gneezy [2009]), we are not aware of experimental studies relating gender and political orientation to the punishment of socially efficient choices.⁸ Thus, our following two findings provide new suggestive evidence that still has to stand the test of replication.

- *Result 1c:* On average, women punish socially efficient choices—even if they are Pareto-efficient—more frequently and significantly more than do men.

Evidence for Result 1c. Seventeen out of our 55 third parties are women. Focusing our analysis on those games with a unique socially efficient choice—that is, games 1 through 6—we find that the fraction of women who punish at least once such choice is larger than that of men (59 percent women and 45 percent men) and that women reduce the first party’s payoff more than twice compared to men (195.15 vs. 84.54 points reduction in all efficient allocations, Wilcoxon-Rank sum test, $z = 1.848$, $p = .064$; T -test, $t = -2.54$, $p = .0138$). Interestingly, women punish also the first party stronger in game 6 when she chooses the Pareto-efficient allocation 390/240 ($z = -2.086$, $p = .037$). Note, however, that women do not punish socially inefficient choices in these games more than do men (women = 12.06 points, men = 15.71 points, $t = 0.274$, $p = .783$).

- *Result 1d:* Left-wing participants punish socially efficient and inefficient choices significantly more than do right-wing participants.

Evidence for Result 1d. At the end of the experiment, we asked third parties about their political orientation on a nine-point scale (0 = *left-wing*, 9 = *right-wing*). The most centered political orientations (3, 4, 5, 6, and 7) were each one crossed by more than 10 percent of the participants. Further, we observe that the more left-wing participants consider themselves on the scale, the more they spend on punishing the first party for choosing socially efficient allocations (Spearman rank correlation; $r = -0.294$, $p = .029$) in games 1 through 6. Note in addition, that participants who consider themselves more left-wing tend also to punish socially inefficient choices more (Spearman rank correlation; $r = -.0429$, $p = .001$).

Punishment of Equitable Choices

In our games, we say that the allocation with the smallest distance between the lowest and the highest payoff is the equitable allocation of the game. Do third parties punish the first party if she makes an equitable choice?

- *Result 2a:* Third parties punish the first party considerably if she chooses the equitable allocation, but only if at least one of two conditions holds: (i) this choice damages the bystander’s payoff and (ii) the first party gets a larger payoff than the third party. Strict payoff equality, however, alleviates punishment.

Evidence for Result 2a. We already offered evidence in this regard for games 1 through 6 when commenting Result 1b, as the equitable choice in these games happens to be at the same time the socially inefficient choice. Hence, we concentrate on the remaining games. Consider first games 9 (250/150 vs. 110/290) and 10 (250/150 vs. 330/70): in the equitable allocation (250/150), the first party gets always a larger payoff than does the third party. In line with our result, the first party gets frequently and relatively harshly punished (the frequency and strength of punishment is above

average in both games). Observe that this is true even when the equitable choice in game 10 does not damage the bystander's payoff: third parties punish a richer first party, even if she makes an equitable choice that is "kind" toward the bystander.

Given the previous evidence, it seems problematic to argue that third parties punish first parties mainly because their choices are unfair—that is, less fair than the alternative choice.⁹ Further evidence in this respect comes from game 7 (250/80 vs. 80/250), where we observe that third parties punish *even if* both allocations are arguably equally fair. In fact, game 7 shows something more: as both allocations in this game are equally fair from the viewpoint of any impartial criterion of distributive justice (not only efficient or egalitarian ones), it seems that punishment cannot be explained alone as a reaction to a transgression of any reasonable norm of distributive justice. In contrast, the previous evidence is again in line with models that assume some form of inequity aversion (like Fehr and Schmidt 1999). As further evidence in this respect, we observe in game 10 that the punishment of the equitable choice (250/150), although relatively high, is smaller (in frequency and strength) than the punishment of the alternative allocation (330/70). This could be explained by the fact that the payoff disadvantage increases if the alternative, inequitable allocation is chosen—note that the same occurs in game 6 (280/240 vs. 390/240). Finally, game 8 (100/100 vs. 50/150) provides evidence for the last sentence in Result 2a: equitable allocations in which both first and second parties get the same payoff are rarely and mildly punished. In game 8, for instance, allocation 100/100 is punished by 6 percent of the third parties, and the average point reduction is just 1.2 points. This raises the question whether strictly equal allocations are differently treated compared to slightly unequal allocations.

- *Result 2b:* The choice of a strictly equal allocation is punished relatively less, especially when that choice harms the bystander.

Evidence for Result 2b. We focus first in the games where a strictly equal allocation is available, that is, games 1 (150/150 vs. 590/60), 2 (100/100 vs. 50/530), and 8 (100/100 vs. 50/150). Indeed, that choice is rarely and mildly punished in these games: frequency is 6 percent in games 1 and 8, and 11 percent in game 2, and intensity is highest in game 2, with an average of 8.7 points reduction. In a comparison across games, the choice of the strictly equal allocation in game 2 is less frequently punished (Wilcoxon signed-rank test, $z = -2.018$, $p = .043$) than the choice of allocation (150/90) in game 4 (150/90 vs. 50/630). Observe that the comparison between games 2 and 4 seems justified because the alternative to the equitable allocation is very similar in both. For similar reasons, we can also compare games 1 and 3 (560/60 vs. 120/140). In these games, the choice of the equitable allocation does not harm the bystander. In line with the qualification stated in Result 2b, 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 ($z = 1.350$, $p = .177$).

We now consider a related question: Does the existence of strictly equal allocations affect punishment decisions at *alternative* allocations? In particular, are first parties punished more if they deviate from a strictly equal allocation than if they deviate from a slightly unequal allocation? We are motivated in this respect by Güth, Huck, and Müller (2001) who report that the rate of rejection of the (proposer and responder) sharing proposal (17, 3) in three mini-ultimatum games depends on the alternative, which differs across games—that is (11, 9), (9, 11), and (10, 10). The authors conclude from their analysis that “The general message of these results [...] is that fairness concerns may be less pronounced in settings where splitting equally is impossible” (p. 166). With regard to the punishment of inequitable allocations by third parties, we do not find such phenomenon.¹⁰

- *Result 2c:* A deviation from an strictly equal allocation is not punished relatively more.

Evidence for Result 2c. Consider again the comparison between games 1 (150/150 vs. 590/60) and 3 (560/60 vs. 120/140). We observe that the alternative allocation is not punished more in game 1, where it constitutes a deviation from strict equality, than in game 3 ($z = 0.023$, $p = .981$). The same is true when comparing games 2 and 4.

Bystander Damaging

Until now, we concentrated on the punishment of the active first parties. We now turn to the investigation of the punishment of the passive second parties, that is, the bystanders.

- *Result 3a:* Third parties damage the bystander considerably, but almost exclusively if the bystander is richer than the third party.

Evidence for Result 3a. There are five allocations in which more than 20 percent of the third parties damage the bystander (the right-hand allocation of games 2, 4, 5, 7, and 9; recall that the average frequency of punishment of the bystander across allocations is 12.1 percent). In all these allocations the bystander’s payoff exceeds the third party’s payoff. Importantly, the intensity of punishment is above the average of 6.1 points only in these allocations, and it positively depends on the size of the payoff distance between the bystander and the third party.

To clarify this latter point, we first classify our allocations in three groups, depending on whether the bystander is (i) considerably richer (at least 100 points), (ii) richer, and (iii) poorer than the third party at that allocation. We observe that the average punishment of the bystander in these groups is respectively (i) 21.54, (ii) 12.60, and (iii) 1.86. Richer bystanders are significantly more damaged than poorer bystanders by the third parties (Wilcoxon signed-rank test, $z = 3.120$, $p = .0018$).

These results are further confirmed by an OLS regression analysis, which shows that the strength of punishment significantly increases with the payoff distance. For each 100 points in payoff difference, the third party reduces the bystander’s payoff by 5.79 points ($t = 3.59, p < .001$).

- *Result 3b:* Bystanders are treated differently than are active parties.

Evidence for Result 3b. To start, note that bystanders are overall punished significantly less strongly than first parties ($z = 3.672, p = .0002$). This is nicely illustrated by 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 first parties and bystanders are damaged when they are advantaged (29 vs. 24 percent; chi-square test, $p = .516$), there are significant differences in the average strength of punishment (19.8 vs. 12.3 points, $z = -2.076; p = .0378$). Table 3 depicts the results from an OLS regression that further clarifies this issue. In this regression, the dependent variable is the strength of punishment toward any party at each allocation and the independent variables are the payoff difference between the third and other party and a bystander dummy (this variable takes the value one if the punishment is directed toward a bystander). The significantly negative coefficient for the bystander dummy shows that bystanders are punished by 2.05 points less after controlling for payoff differences. This provides further evidence that bystanders are treated differently than first parties. In addition, we find suggestive evidence that women tend to punish bystanders more strongly than men (Wilcoxon-Rank sum test, $z = 1.814, p = .068$; but T -test, $t = 1.163, p = .250$) and that more left-wing participants have a tendency to punish bystanders more strongly, which is, however, not yet significant at the 10 percent level (Spearman rank correlation, $p = .204, p = .134$).

Result 3b is in line with previous empirical literature that stresses the role of intentions (Falk, Fehr, and Fischbacher 2003, 2008): As bystanders are not responsible for any outcome, they are less affected by punishment as compared to active parties. Note that models of inequity-aversion (Fehr and Schmidt 1999; Bolton and Ockenfels 2000), which do not incorporate intentions, cannot explain this different punishment pattern.

Table 3. Determinants of Third-Party Punishment

Payoff difference between third and other party	0.022*** (0.003)
Bystander dummy	−2.051*** (0.700)
Constant	3.960*** (0.516)
R^2	0.126

Note: $N = 2,200$. Standard errors (in parentheses) are clustered on individual level.

*** signifies $p < .01$.

Conclusion

This article studies altruistic punishment from unaffected third parties and finds that they punish deciders strongly for choosing the socially efficient allocation and becoming the “richer” party as a result, *even* if their choices inflicted no harm on another party. Interestingly, this form of punitive behavior seems especially pronounced in women and more left-wing participants. We also find considerable punishment of deciders if they choose an equitable allocation and get the highest payoff or harm the bystander—however, strict equality alleviates this latter type of punishment. Furthermore, with regard to the damaging of bystanders we find that their payoff is frequently diminished by third parties but only if they are richer than the third party. Overall, we observe that a disadvantage in payoffs between the third party and another party is positively correlated with the frequency and strength of punishment toward the latter party. For this reason, our data is consistent to a large extent with models that integrate inequity-aversion as a motivation for punishment (Fehr and Schmidt 1999; Falk and Fischbacher 2006).

To the best of our knowledge, our experiment is the first to study how third parties react to the choice of socially efficient or equitable allocations and how they react toward bystanders. Moreover, our study complements previous studies that report some punishment from affected second parties, which is not motivated to establish cooperation (Herrmann, Thöni, and Gächter 2008; Nikiforakis 2008; Reuben and Riedl 2009) and the handicapping of receivers in binary dictator games (Charness and Grosskopf 2001; Charness and Rabin 2002) and in money-burning games (Zizzo 2003; Dawes et al. 2007).

Our findings might have several implications for policy makers or scholars interested in conflict resolution. Because participants punish socially efficient allocations even if they are Pareto-dominant, 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. In this respect, our results also suggest that many people may have a preference for the redistribution of resources in order to achieve fair allocations. Finally, the role of third parties in acting as mediators and moderators of conflict should be taken with care, among other things because they even damage bystanders. Our results suggest that many third parties are not impartial and hence can hardly be regarded as ideal or “neutral mediators” (Princen 1992), at least if one considers that impartiality is an important attribute.¹¹ This relates to evidence on the behavior of jury members in law cases showing that third parties can make inappropriate decisions in the context of sanctioning (e.g., Kennedy 1997) and studies that also questions the usefulness of third parties as mediators in civil conflicts (e.g., Regan 2002).

Appendix A

Table A1. Main Characteristics of the Ten Games

Game		In the unique socially efficient allocation, whose is higher?	In the most equitable allocation whose is higher?	Strictly equal allocation available?	Pareto-dominant allocation available?	Second party's > 200 at any allocation?
1	(150,150) vs. (590,60)	First party (FP)	—	YES	NO	NO
2	(100,100) vs. (50,530)	Second party (SP)	—	YES	NO	YES
3	(560,60) vs. (120,140)	FP	SP	NO	NO	NO
4	(150,90) vs. (50,630)	SP	FP	NO	NO	YES
5	(220,260) vs. (220,400)	SP	SP	NO	YES	YES
6	(280,240) vs. (390,240)	FP	FP	NO	YES	YES
7	(250,80) vs. (80,250)	—	—	NO	NO	YES
8	(100,100) vs. (50,150)	—	—	YES	NO	NO
9	(250,150) vs. (110,290)	—	FP	NO	NO	YES
10	(250,150) vs. (330,70)	—	FP	NO	NO	NO

Table A2. Observed Frequencies of Choices from First Parties

Game				Left	Right
1	(150,150)	vs.	(590,60)	0.11	0.89
2	(100,100)	vs.	(50,530)	0.82	0.18
3	(560,60)	vs.	(120,140)	0.98	0.02
4	(150,90)	vs.	(50,630)	0.87	0.13
5	(220,260)	vs.	(220,400)	0.18	0.82
6	(280,240)	vs.	(390,240)	0.07	0.93
7	(250,80)	vs.	(80,250)	1	0
8	(100,100)	vs.	(50,150)	1	0
9	(250,150)	vs.	(110,290)	0.96	0.04
10	(250,150)	vs.	(330,70)	0.18	0.82

Table A3. Theoretical Punishment Predictions

Game	Theories predicting punishment for the first party				Theories predicting punishment for the bystander	
	Allocation		Allocation		Allocation	
	Left	Right	Left	Right	Left	Right
1	(150,150)	vs. (590,60)	—	IA	—	—
2	(100,100)	vs. (50,530)	—	ERC	—	IA, ERC
3	(560,60)	vs. (120,140)	IA	—	—	—
4	(150,90)	vs. (50,630)	—	—	—	IA
5	(220,260)	vs. (220,400)	IA, ERC	IA	IA, ERC	IA
6	(280,240)	vs. (390,240)	IA, ERC	IA	IA, ERC	IA
7	(250,80)	vs. (80,250)	IA	—	—	IA
8	(100,100)	vs. (50,150)	—	—	—	—
9	(250,150)	vs. (110,290)	IA	—	—	IA
10	(250,150)	vs. (330,70)	IA	IA	—	—

Note: IA = inequity-aversion (as in Fehr and Schmidt 1999); ERC = Bolton and Ockenfels (2000).

Appendix B

General Instructions for the Third Party

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.

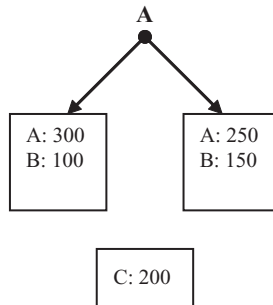
There are three types of participants in this experiment: participant A, participant B and participant C. **You are participant C.** You will never get to know the identity of any participant A (B or C), nor will any participant get to know who you are. The payment at the end of the experiment is also 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 you get 100 points. If he decides for the allocation on the right side, he gets 250 points and you get 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
- $3 \times$ 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
- $3 \times$ 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 participant A in the game that was chosen by the dice
- The deduction points you assigned to participant 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 of the first situation, you will have to decide how many deduction points you assign. 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.

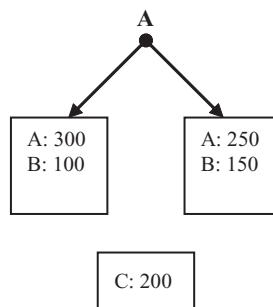
Please answer now the following control questions and raise your hand if you have answered them. The experiment starts as soon as all participants have correctly filled out the control questions.

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?.....



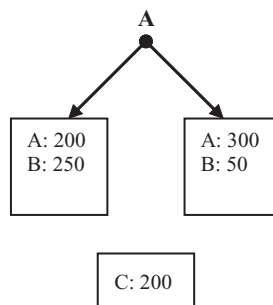
b) Participant C assigns 30 deduction points to participant B.

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?.....



Acknowledgement

We would like to thank Ernst Fehr, Jimmy Walker, seminar participants at the Workshop for Political Theory and Policy Analysis and the University of Zurich, as well as an editor and two anonymous referees for their helpful comments and valuable feedback.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors greatly acknowledge financial support by the EU Research Network ENABLE.

Notes

1. Two terminological caveats are due. First, as is usual in the experimental literature, we use the terms *punishment* or *sanction* as synonyms of “damaging behavior” and not only to refer to that damaging behavior, which is directed toward a wrongdoer, as is most usual in common use. Second, 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 the material payoff of A may depend on the decisions of C; for instance, it could be the case that C punishes A and thus reduces her material payoff).
2. As we report in Leibbrandt and López-Pérez (2009), these beliefs were on average rather accurate.
3. In principle, this method might induce a different behavior than the specific response method, where participants know the choice made by the other player. 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. One further advantage of the strategy method is that it maximizes the amount of statistical data gathered.
4. If there is more than one allocation with the same payoff difference, we weight the behavior in these allocations uniformly.
5. This companion paper also provides data from an experiment on second party punishment. The phenomena that we report here for third-party punishment also extend to second party punishment, with just one exception (see the Punishment of Equitable Choices section).
6. The same is also true in games 7 (250/80 vs. 80/250), 9 (250/150 vs. 110/290), and 10 (250/150 vs. 330/70). Table A3 in the appendix shows the predictions of Bolton and Ockenfels (2000) in our games.
7. Finally, a pure model of reciprocity like Dufwenberg and Kirchsteiger (2004) predicts multiple equilibria in our games depending on second-order beliefs (a proof of this can be requested to the authors). In any game, however, this model predicts at least one equilibrium in which the third party never punishes. This equilibrium is clearly at odds with our data.

8. Eckel and Grossman (1996) also report gender differences in punishing. They find that women are more sensitive to the price of punishment than are men.
9. However, third parties might *believe* that their behavior is compatible with such a fairness argument. Talking on envy, Elster (1999, 165) notes that “When it leads to action, it is usually mediated by the prior transmutation into a more acceptable emotion, such as righteous anger or righteous indignation.” This is relevant in our setting, where some form of envy or inequity aversion seems to account for a large part of the punishment.
10. In another experiment, we have studied *second party* punishment in a set of 10 games analogous to our games. In this setting, we observe *sometimes* that deviations from strictly equal allocations are punished more. The data are available upon request from the authors.
11. There is some debate in this respect in the literature on mediation—see Kleiboer (1996) for a survey that considers this issue.

References

- Alesina, Alberto, and Giuliano, Paola. 2011. “Preferences for Redistribution”. In Jess Benhabib, Matthew O. Jackson and Alberto Bisin editors: *Handbook of Social Economics*, Vol. 1A, The Netherlands: North Holland, 93-131.
- Bendor, Jonathan, and Piotr Swistak. 2001. “The Evolution of Norms.” *American Journal of Sociology* 106:1493-545.
- Boehm, Christopher. 1993. “Egalitarian Behavior and Reverse Dominance Hierarchy.” *Current Anthropology* 34 (3): 228-54.
- Boehm, Christopher. 1999. *Hierarchy in the Forest: The Evolution of Egalitarian Behavior*. Cambridge, MA: Harvard University Press.
- Bolton, Gary E., and Axel Ockenfels. 2000. “ERC: A Theory of Equity, Reciprocity, and Competition.” *American Economic Review* 90 (1): 166-93.
- Bowles, Samuel, and Herbert Gintis. 2000. “Social Capital and Community Governance.” *The Economic Journal* 112 (483): 419-36.
- Camerer, Colin, and Richard Thaler. 1995. “Ultimatums, Dictators, and Manners.” *Journal of Economic Perspectives* 9:209-19.
- Charness, Gary, and Brit Grosskopf. 2001. “Relative Payoffs and Happiness: An Experimental Study.” *Journal of Economic Behavior and Organization* 45:301-28.
- Charness, Gary, and Matthew Rabin. 2002. “Understanding Social Preferences with Simple Tests.” *Quarterly Journal of Economics* 117:817-69.
- Croson, Rachel, and Uri Gneezy. 2009. “Gender Differences in Preferences.” *Journal of Economic Literature* 47:448-74.
- Dawes, Christopher, James H. Fowler, Tim Johnson, Richard McElreath, and Oleg Smirnov. 2007. “Egalitarian Motives in Humans.” *Nature* 446:794-96.
- Dufwenberg, Martin, and Georg Kirchsteiger. 2004. “A Theory of Sequential Reciprocity.” *Games and Economic Behavior* 47:268-98.
- Eckel, Catherine C., and Philip J. Grossman. 1996. “The Relative Price of Fairness: Gender Differences in a Punishment Game.” *Journal of Economic Behavior and Organization* 30:143-58.

- Elster, Jon. 1999. *Alchemies of the Mind: Rationality and the Emotions*. Cambridge: Cambridge University Press.
- Falk, Armin, Ernst Fehr, and Urs Fischbacher. 2003. "On the Nature of Fair Behavior." *Economic Inquiry* 41:20-26.
- Falk, Armin, Ernst Fehr, and Urs Fischbacher. 2005. "Driving Forces behind Informal Sanctions", *Econometrica*, 7(6), 2017-30.
- Falk, Armin, Ernst Fehr, and Urs Fischbacher. 2008. "Testing Theories of Fairness—Intentions Matter." *Games and Economic Behavior* 62:287-303.
- Falk, Armin, and Urs Fischbacher. 2006. "A Theory of Reciprocity." *Games and Economic Behavior* 54:293-315.
- Fehr, Ernst, and Urs Fischbacher. 2004. "Third Party Punishment and Social Norms." *Evolution and Human Behavior* 25:63-87.
- Fehr, Ernst, and Simon Gächter. 2000. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review* 90:980-94.
- Fehr, Ernst, and Simon Gächter. 2002. "Altruistic Punishment in Humans." *Nature* 415:137-40.
- Fehr, Ernst, and Klaus Schmidt. 1999. "A Theory of Fairness, Competition and Cooperation." *Quarterly Journal of Economics* 114 (3): 817-68.
- Fischbacher, Urs. 2007. "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics* 10 (2): 171-8.
- Frohlich, Norman, Joe Oppenheim, Pat Bond, and Irvin Boschman. 1984. "Beyond Economic Man: Altruism, Egalitarianism, and Difference Maximizing." *Journal of Conflict Resolution* 28 (1): 3-24.
- Gouldner, Alvin W. 1960. "The Norm of Reciprocity: A Preliminary Statement." *American Sociological Review* 25:161-78.
- Graham, Jesse, Jonathan Haidt, and Brian A. Nosek. 2009. "Liberals and Conservatives Rely on Different Sets of Moral Foundations." *Journal of Personality and Social Psychology* 96 (5): 1029-46.
- Güth, Werner, Steffen Huck, and Wieland Müller. 2001. "The Relevance of Equal Splits in Ultimatum Games." *Games and Economic Behavior* 37:161-69.
- Güth, Werner, Rolf Schmittberger, and Bernd Schwarze. 1982. "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization* 3:367-88.
- Haidt, Jonathan, and Jeff Graham. 2007. "When Morality Opposes Justice: Conservatives Have Moral Intuitions that Liberals may not Recognize." *Social Justice Research* 20 (1): 98-116.
- Herrmann, Benedikt, Christian Thöni, and Simon Gächter. 2008. "Antisocial Punishment Across Societies." *Science* 319:1362-67.
- Kennedy, Randall. 1997. *Race, Crime and the Law*, 168-255. New York: Pantheon.
- Kleiboer, Marieke. 1996. "Understanding Success and Failure of International Mediation." *Journal of Conflict Resolution* 40:360-89.
- Knauff, Bruce. 1991. "Violence and Sociality in Human Evolution." *Current Anthropology* 32:391-428.
- Leibbrandt, Andreas, and Raúl López-Pérez. 2009. "An Exploration of Third and Second Party Punishment in Ten Simple Games." Mimeo. Available at <http://www.uam.es/raul.lopez>.

- Levine, David K. 1998. "Modeling Altruism and Spitefulness in Experiments." *Review of Economic Dynamics* 1:593-622.
- Nikiforakis, Nikos. 2008. "Punishment and Counter-Punishment in Public Good Games: Can we Really Govern Ourselves?" *Journal of Public Economics* 92:91-112.
- Olson, Mancur. 1965. *The Logic of Collective Action*. Cambridge, MA: Harvard University Press.
- Ostrom, Elinor, James Walker, and Roy Gardner. 1992. "Covenants with and without a Sword: Self-Governance is Possible." *American Political Science Review* 86 (2): 404-17.
- Princen, Thomas. 1992. *Intermediaries in International Conflict*. Princeton: Princeton University Press.
- Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review* 83 (5): 1281-302.
- Reuben, Ernesto, and Arno Riedl. 2009. "Public Goods Provision and Sanctioning in Privileged Groups." *Journal of Conflict Resolution* 53:72-93.
- Seinen, Ingrid, and Arthur Schram. 2006. "Social Status and Group Norms: Indirect Reciprocity in a Repeated Helping Experiment." *European Economic Review* 50:581-602.
- Thompson, Leigh. 2005. *The Mind and Heart of the Negotiator*. Englewood Cliffs, NJ: Prentice Hall.
- Wilson, Stephen. 1988. *Feuding, Conflict and Banditry in Nineteenth-Century Corsica*. Cambridge University Press.
- Zizzo, Daniel J. 2003. "Money Burning and Rank Egalitarianism with Random Dictators." *Economics Letters* 81:263-66.