

# TAKING PUNISHMENT INTO YOUR OWN HANDS: AN EXPERIMENT

JULIA MÜLLER

*Erasmus University Rotterdam & Tinbergen Institute*

PETER DUERSCH

*University of Heidelberg*

ABSTRACT. In a punishment experiment, we separate the demand for punishment in general from the demand to conduct punishment personally. Subjects experience an unfair split of their earnings from a real effort task and have to decide on the punishment of the person who determines the distribution. First, it is established whether the allocator's payoff is reduced and, afterwards, subjects take part in a second price auction for the right to (physically) carry out the act of payoff reduction themselves. Subjects bid positive amounts and are happier if they get to punish personally.

**Keywords:** personal punishment; real effort task; experiment; auction.

**JEL-Classification Code:** C92, D03.

**PsycINFO-Classification Code:** 2360, 3020.

---

*E-mail addresses:* `jmuller@ese.eur.nl`, `peter.duersch@awi.uni-heidelberg.de`.

*Date:* May 21, 2013.

Corresponding Author: Julia Müller, Erasmus School of Economics, H13-15, Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, The Netherlands.

We thank for valuable comments: Dirk Engelmann, Wolfgang Hoechtl, Nikos Nikiforakis, Jörg Oechssler, Christiane Schwieren, Maroš Servátka, the associate editor Simon Gächter and two anonymous referees. Financial support by the DFG through SFB 504 is gratefully acknowledged.

*If the person who had done us some great injury, who had murdered our father or our brother, for example, should soon afterwards die of a fever, or even be brought to the scaffold upon account of some other crime, though it might sooth our hatred, it would not fully gratify our resentment. Resentment would prompt us to desire, not only that he should be punished, but that he should be punished by our means, and upon account of that particular injury which he had done to us. (Adam Smith<sup>1</sup>)*

## 1. INTRODUCTION

The desire for revenge, to punish those who did wrong upon oneself, is a strong motivation for humans. From ancient Greek dramas to modern movies, it is ubiquitous in storylines. It has also been the focus of extensive research in economics, both in the form of experiments which find that, indeed, subjects are willing to forgo monetary gains to exert punishment, and in the form of theoretical models that seek to explain such behavior. However, both the quote by Adam Smith above and many prominent fictional works<sup>2</sup> feature a very specific form of punishment: According to Adam Smith, humans not only care about punishment being inflicted on the perpetrator of a crime against them, but they also value carrying out that punishment themselves, in person. It is this, personal, characteristic of punishment that we try to isolate in the laboratory. Our experiment is designed to exclude other possible reasons why one would be willing to give up money to punish. In particular, subjects do not have to spend money to assure punishment is carried out, they only spend money to assure it is carried out by them personally.

In our experiment we first establish a situation where punishment is possible: One group of subjects (type *A*) take part in a real effort task to create an endowment, which is then distributed by the non-working subjects of type *B*. Subjects *B* can either leave the complete

---

<sup>1</sup>In *The Theory of Moral Sentiments*, page 113.

<sup>2</sup>To use two well known movies as examples: In *Pulp Fiction*, after being rescued from a rapist by Butch, Marsellus tells Butch, who is about to kill the rapist, to move aside, so he can shoot the rapist himself. Similarly, in *Dogville*, Grace, after ordering her father's men to torch the village which enslaved her, kills the man who hurt her most personally, telling her father: "Some things, you have to do yourself".

endowment to  $A$ , or take away 80% of it. After learning the distribution chosen by  $B$ , subjects  $A$  are allowed to punish  $B$  by destroying a part of the money that  $B$  allocated to herself.<sup>3</sup> We use two separate decisions to split the demand for personal punishment from the demand for punishment in general. In a first question, subjects  $A$  get to decide whether  $B$  will be punished. Then, knowing that  $B$  will be punished for sure, we ask subjects  $A$  whether they want to be the one to personally and physically execute the punishment. We do this by means of a second price auction where we auction off the right to be the one to execute the punishment. Since the auction has a winner in any case, punishment is always ensured. The bids in the auction elicit subject  $A$ 's willingness to pay for personal punishment.

More than a third of our subjects bid positive amounts in the personal punishment auction. Bidding for personal punishment is higher by subjects who also wanted any, potentially non-personal, punishment in the first question. We interpret bidding in the auction as a desire, among our subjects, to actively punish, as opposed to having the perpetrators payout reduced by a third party.

This is in line with models which include actions, along with payoffs, in the utility function, as in Andreoni (1990). He examines the private provision of a public good and models the utility of individuals as a function not only of the amount of the public good but also of the own gift to the public good. This individual donation produces what Andreoni calls a “warm glow”, utility derived from the act of giving. In a fMRI experiment Harbaugh, Mayr, and Burghart (2007) identify this joy of giving within the brain.

In the same vein as joy of giving, there is also a joy of punishing: Direct neuroeconomic evidence that subjects “like” to punish was found by de Quervain, Fischbacher, Treyer, Schellhammer, Schnyder, Buck, and Fehr (2004), who use PET recordings of brain activation to investigate the mechanisms in the brain involved in punishment. Subjects played a trust game where cooperating players could punish defecting partners. In the punishment condition activation of the dorsal striatum was found, which is well known for its reward processing properties. This could either be due to the fact that the defecting partner lost money or it could be pleasure derived from the act of punishing. This

---

<sup>3</sup>For ease of exposition, let  $A$  be male and  $B$  be female.

	real effort	punishment opportunity	auction for personal punishment	auction for dummy envelope	participants
1A	yes	yes	yes	no	76
2A	yes	yes	yes	yes	40
NC	no	no	no	yes	33

TABLE 1. Experimental designs

is what we disentangle in our experiment as the decision to punish is separated from the decision to punish personally.

Further evidence comes from a measurement of affective happiness, which we conduct before and after the experiment: Subjects who win the personal punishment auction and thus have to pay the second highest bid, but get to be the ones to execute the punishment, become happier.

In the next section, we introduce the design of the experiment and discuss related literature. Section 3 presents our hypotheses and section 4 the results. Finally, in section 5, we conclude.

## 2. EXPERIMENT

**2.1. Designs.** To test the demand for personal punishment, we use three related experimental designs, 1A (one auction), 2A (two auctions) and NC (no context).<sup>4</sup> We start by describing 1A.

Design 1A. Subjects were matched in groups of four; each group consists of three subjects *A* and one subject *B*. The experiment was anonymous, so subjects never learned about the identity of the other subjects they were matched with. Instructions for the experiment, which fully described the experiment for both type *A* and type *B*, were handed to subjects at the start of the experiment. After reading the instructions, subjects had to answer a series of detailed questions in order to make sure that they understood the experimental instructions. Only when all subjects had correctly answered these test questions, did the experiment proceed.<sup>5</sup>

<sup>4</sup>See the online-appendix for translations of all instruction material: <http://www.uni-heidelberg.de/md/awi/professuren/with2/pdjm-pp-appendix.zip>

<sup>5</sup>Subjects who were not able to answer the test questions correctly were replaced by extra participants (who were otherwise dismissed with a flat payment after reading the instructions).

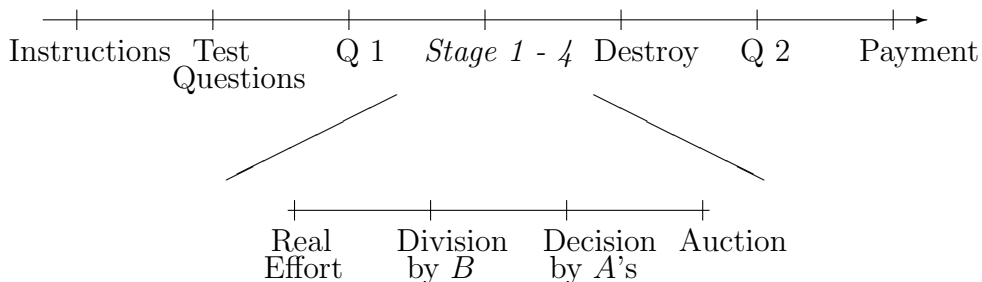


FIGURE 1. Timing

In the first stage,<sup>6</sup> all subjects  $A$  participated in a real effort task where they could earn EUR 10. They were asked to fill a sheet of graph paper (A5,  $148 \times 210$  mm, about 1260 squares) with alternating  $o$  and  $+$  signs. The allocated time frame was 25 minutes. Subjects who did not finish the task in time dropped out of the experiment and received no money apart from the show up fee. We chose this particular task for two reasons: First, it is simple and does not require any special abilities, so all subjects should be equally fit for the task. Second, as we found out in previous tests, the task is considerably more exhausting than it appears. We wanted to induce a feeling of ownership towards the money in those subjects who completed the task. On the other hand, it was to look easy to the non-participating subjects  $B$ . During the task, all subjects  $B$  were sitting in the same room as the subjects  $A$ , but without any assignment.

After the task, the experimenters collected the sheets and informed each subject  $B$  how many subjects  $A$  in her group had succeeded. Upon learning that information, in stage two, subject  $B$  had to decide on an allocation of the money earned by the subjects  $A$  in the previous stage. The only two allocations available were (2,8): 2 for  $A$ , 8 for  $B$ , or (10,0): 10 for  $A$ , 0 for  $B$ .<sup>7</sup> Subject  $B$  could only implement the same

<sup>6</sup>The instructions use a different numbering, since we subdivided some stages for clarity. We also handed to all subjects a flow chart as an overview what happens in each stage. The flow charts are included in the Online-Appendix.

<sup>7</sup>The distributional choice of subjects  $B$  is similar to the one in a dictator game (Kahneman, Knetsch, and Thaler (1986)) or ultimatum game (Güth, Schmittberger, and Schwarze (1982).) with a restricted choice domain. Falk, Fehr, and Fischbacher (2003) use similar distributional choices in a series of restricted ultimatum games.

allocation for all three subjects  $A$  she was matched with, not different allocations for different subjects  $A$ . So in the case of three successful subjects  $A$  subject  $B$  had to decide between 24 for herself and 2 for each  $A$  or 0 for herself and 10 for each  $A$ .

Before stage three, the experimenters informed all subjects  $A$  about the decision of their matched  $B$ . The money that subject  $B$  allocated to  $A$  was handed to subject  $A$ . The money that subject  $B$  allocated to herself was *also handed to subject A*, however it was put in an envelope. Then all subjects  $A$  had to decide whether they wanted to reduce subject  $B$ 's payoff by destroying one of the three envelopes designated for  $B$ . If all  $A$ 's decided not to reduce, the envelopes were collected by the experimenters, handed to subject  $B$  and stage four did not take place.

If at least one subject  $A$  decided to reduce, the entire group entered stage four.<sup>8</sup> Here, all subjects  $A$  of the group took part in a sealed bid second price auction. The highest bidder won the right to destroy the envelope lying in front of him. Only the envelope of the winner was destroyed.<sup>9</sup> Note that subjects  $B$ 's payoff depends entirely on stages 1 to 3. The auction only selected the subject  $A$  who would destroy the envelope, it did not affect subject  $B$ 's payoff. The auction provides a non-arbitrary way to separate the decision to punish from the decision to punish personally. Since, in a second price auction, no participant has a reason to misrepresent his preferences, subjects are incentivized to truthfully state the value they attach to personal punishment.

The auction winner was informed that he won the auction and about the second highest bid he had to pay. He could then proceed to destroy the envelope of subject  $B$ . The instructions did not specify any mode of destruction; however a small metal bin was present on the tables of

---

<sup>8</sup>Subjects were informed that stage four had been reached, but not informed about the number of subjects  $A$  who had decided to reduce.

<sup>9</sup>The minimum feasible bid was zero, the maximum feasible bid 10 and subjects could bid in increments of 0.01 (one cent). If there was a tie, the experimenters randomly chose a winner. This also applies to the special case of all three subjects  $A$  choosing a bid of zero. The upper limit of 10 for the bids ensured that losses were not possible. At this stage of the experiment, subjects of type  $A$  had received the show-up fee of 8, in the division stage the matched subject  $B$  could give to  $A$  either 2 or 10, so when the auction happens all type  $A$  subject will have at least 10.

each subject  $A$ .<sup>10</sup> The envelopes in front of the non-winning subjects  $A$  were collected by the experimenters and delivered to the respective subject  $B$ .

Between the test questions and the real effort task we asked some demographics from our subjects and two questions about their happiness (“how happy are you in general” / “how happy are you right now”). After stage four and before paying, we presented subjects with a second questionnaire asking their happiness again (only “how happy are you right now”), their perception of subject  $B$ ’s behavior and several attitude questions<sup>11</sup>. All subjects received a EUR 8 “show up fee” for answering the questionnaires. If a subject  $A$  had won the auction and had to pay more money than he earned in the experiment, he had to use a part of those EUR 8 to pay for his bid.

Design 2A. The 2A design is similar to design 1A, with the difference that it uses two auctions instead of one. Stages one to three are identical to 1A. However, in the auction stage, subjects had to make two bids. Bid one was for the auction as described above. For the second auction, the experimenters placed a second envelope in front of the each subject  $A$ . The instructions stated that this envelope would be, unless destroyed, collected again by the experimenters and would never have any influence on the payoff of subjects  $A$  or subjects  $B$ . That is, the second envelope is a dummy, intended to test whether subjects would be willing to pay for destruction of *any* envelope. After bids were made, the experimenters threw a coin in public to determine whether auction one or auction two would be enacted. Only the bids from the chosen auction did count, and only the envelope from the auction chosen was destroyed by the winning subject  $A$ . If auction one was chosen, the winner destroyed his envelope from auction one, the other envelopes were handed to subject  $B$ , and all three envelopes from auction two were collected by experimenters. If auction two was chosen, the winner destroyed his envelope from auction two, the other envelopes from auction two were collected by the experimenters. In this case, the experimenters also randomly retained one of the envelopes from auction

---

<sup>10</sup>The subjects chose different methods to “destroy”: Most ripped the envelope apart – some ripped just once, some ripped until only small pieces of paper were left – and deposited the pieces inside the metal bin. Some just folded the envelope.

<sup>11</sup>See Online-Appendix.

one, such that subject  $B$  received the same amount of envelopes, no matter whether auction one or auction two was chosen by the coin-flip. Design NC. Finally, we used the NC condition to separate the auction stage from the rest of our experiment. To insure that conditions remained comparable, we conducted the control subsequent to another, unrelated and about 1 hour long, experiment, where the subjects earned on average EUR 10.60.<sup>12</sup> This money was used to pay for bids in the control auction. After the end of the other experiment, we distributed the instructions for NC. Instructions and test questions were as close as possible to those in the main experiment, but only included the auction stage.<sup>13</sup>

Subjects were placed in groups of three (corresponding to our group size of three subjects  $A$ , who did participate in the auction). The highest bidder won the right to destroy an envelope lying in front of him (the envelope was not payoff relevant, as in auction two of 2A). The winner of the auction could destroy the envelope, all others were collected by the experimenters. Auction winners were paid what they earned in the prior experiment minus the second highest bid in their group.

**2.2. Related literature.** In the last 20 years, the topic of punishment occurring in situations where no direct monetary incentive to punish exists has been studied extensively in the economic literature. Some recent surveys can be found in the papers by Balliet, Mulder, and Van Lange (2011), Chaudhuri (2011), and Gächter and Herrmann (2009).

In all previous experimental designs that we know of, achieving punishment (leading to a reduced payoff for the punished person) and punishing personally are not separated. We introduce a separate decision for personal punishment, after the decision whether to reduce the payoff is already taken. Carpenter (2007) and Anderson and Puterman (2006) find that the demand for punishment is decreasing in the price of punishment. Separating a demand for personal punishment from the overall demand for punishment is only feasible if some

---

<sup>12</sup>This is close to the average earnings of EUR 10.81 that subjects of type  $A$  had accumulated in the other conditions (2A and 1A) before the auction was conducted.

<sup>13</sup>We used both envelopes filled with paper money and empty envelopes (the unrelated prior experiment did not use paper money), but did not find any difference and pooled the data.



demand for punishment in general exists. Therefore, in the first punishment decision, we set the price for reducing the perpetrators payoff equal to the lowest possible value of zero. That is, a payoff reduction can be achieved without any cost. Only the second decision, to punish personally, is costly.

In a series of papers (Falk, Fehr, and Fischbacher, 2003, 2005, 2008), Falk, Fehr, and Fischbacher investigate the connection between different fairness norms and sanctioning behavior. In the 2003 paper, they use a mini-ultimatum game, which features payoff distributions similar to those that our subjects  $B$  can implement in stage 2. Different to the mini-ultimatum game, our subjects  $A$  do not get an option to “reject”. Instead, they can, without cost, reduce the payoff of  $B$ , however not to the point that  $B$ 's payoff is reduced to zero. As their main result, the authors find that in terms of sanctioning, both payoff differences to the person being sanctioned and intentions of that person matter. This is a sign that theories of reciprocity based exclusively on payoffs do not fully capture sanctioning. In our research we also look for evidence that preferences of punishers go beyond payoffs, but we proceed in a different direction. It is not the intentions of the punished person we are interested in, but the evaluation of the action of punishment by the punisher. A positive demand for personal punishment can be seen as evidence for an action-based utility function.

The papers most closely related to our design are Casari and Luini (2009, 2012). They test whether punishment is treated as a second-order public good. After a public goods game, participants have the opportunity to assign punishment points to each other. In one treatment, punishment is sequential: Later punishers assign their punishment points knowing the previously assigned punishment points. As it turns out, subjects mostly ignore the second-order public goods aspect of punishment and assign points irrespective of being first or second punisher, again in line with an utility function based on the action of punishment. Different to our design, Casari and Luini do not separate the decision to personally punish from the decision to reduce the offenders payoff. Both movers can assure personal punishment by assigning punishment points, but that implies, at the same time, a reduction in payoff. In a recent working paper, Ouss and Peysakhovich (2013) have a treatment similar to Casari and Luini, with equivalent results. Both pairs of authors extend the model of Andreoni (1990) into the

negative domain (positive utility derived from punishment instead of giving) and argue in favor of action-based utility.

Naturally, punishment is also a research topic in psychology, where the focus was mostly on the person being punished, but more recently psychologists also try to understand the motives of the person who punishes. Looking for subjective punishment goals of victims, Orth (2003) found five punishment goals: retaliation, recognition of victim status, confirmation of societal values, victim security, and societal security. These goals fall into the two broad classical justifications for punishment: retribution and utilitarian motives. The retribution perspective is sometimes also referred to as “just deserts”, and it summarizes the idea that a offender shall be punished to suffer in proportion to his malefaction. Here punishment is justiciable in itself, the utilitarian perspective focuses on its possible positive future consequences, like deterrence and incapacitation. Studies using hypothetical cases and questionnaires have found that people answer that they have both kind of motivations, retribution and utilitarian motives, for punishment. By examining what kind of information people use when deciding on punishment recent studies found evidence that people punish primarily on the basis of retribution (e.g. Darley, Carlsmith, and Robinson (2000); Carlsmith, Darley, and Robinson (2002); Carlsmith (2006)). As utilitarian motives can also be fulfilled if someone else punishes an offender, only a motivation of retribution could explain a demand for personal punishment.

**2.3. Procedures.** The experiment was conducted the laboratory of the Sonderforschungsbereich 504 in Mannheim and in the laboratory of the economics department at the University of Heidelberg. In total, 149 subjects participated in the experiment (40% male, 60% female). Subjects were students of various fields at the University of Heidelberg and the University of Mannheim. The experiment consisted of eight sessions; no subject participated twice. The 2A treatment was conducted in Heidelberg, the 1A and the NC treatments in Mannheim.<sup>14</sup> All recruitment was done via ORSEE (Greiner (2004)).

In both laboratories seats are separated by dividers. In Heidelberg subjects are seated in two long rows facing the walls, in Mannheim there

---

<sup>14</sup>In 1A and 2A there were usually 20 subjects in a session, in one session in 1A only 16 subjects participated. In the two NC-sessions we had 24 and 9 participants.

are a several rows, each providing four seats. Subjects could not see each other unless they deliberately leaned away from their workspace. In interaction with the subjects, e.g. when announcing auction winners, the experimenters approached every seat, regardless of that subject being a winner or not, to avoid giving away additional information.

In total, the experiment lasted slightly less than 2 hours, for which we paid an average of EUR 13.79 (only averaging over subjects in 1A and 2A.) The full experiment was conducted via pen and paper. During the experiment, we used an experimental currency unit called “Thaler”. Thaler were a printed play money handed to subjects during the experiment. At the end of the experiment, we exchanged all Thaler into Euro at a rate of 1:1.<sup>15</sup> All subjects were paid in cash and private.

### 3. HYPOTHESES

If we assume a purely money-maximising selfish individual, such an individual could choose punishment in our experiment (as it is free of cost), but should not care about the way in which subject  $B$ 's payoff is reduced, i.e. personal punishment cannot be explained.

Spurred on by the experimental observation that people do not always act purely selfish, theories of other-regarding preferences have been put forward. Inequity aversion models (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) add to the utility derived from own income a term that represents concern about the payoff distribution. These kind of models can explain why people punish as this usually reduces inequality in the payoffs, but can not explain a demand for personal punishment.<sup>16</sup>

Other theories (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006) develop techniques to embed concerns for reciprocity and capture intentions. Using these models, one can again explain punishment: an unkind action of type  $B$  (choosing  $(2, 8)$  instead of  $(10, 0)$ ) is reciprocated by an unkind action of type  $A$  (punishment reduces the payoff of subject  $B$ ), but not personal punishment.

---

<sup>15</sup>The main reason for using play money was that we did not want subjects to worry about destroying legal tender.

<sup>16</sup>Inequity is reduced by punishment in our design (the higher payoff of  $B$  is reduced at no cost to  $A$ ), but inequality is increased by personal punishment (bidding does not further decrease  $B$ 's payoff, but reduces the payoff of the winning  $A$ ).

We are not primarily interested in the fact that the payoff of an offender is reduced, but especially in *who* will derive satisfaction from punishing. Only the person who conducted the punishment? Or everyone who saw the offender being punished, even if the punishment was not conducted “personally”?

Perhaps the theory closest to our design is the warm glow theory by Andreoni (1990) (see sections 1 & 2.2). If one assumes, in a similar manner to utility being derived from the act of giving, that the act of punishing enters the utility function positively, one would arrive at a theory that could account for a demand to punish personally.

Following Andreoni and the reasoning of Adam Smith, we believe that our subjects want to take punishment into their own hands. This is also connected to the idea of self-efficacy which is defined as the belief of having the capability to reach desired effects with one’s own actions and to reach goals (Bandura, 1977). If subjects desire a higher self-efficacy, they should bid in the personal punishment auction.

These considerations lead to our main hypothesis:

**Hypothesis 1.** *Personal punishment: Subjects A bid more in the personal punishment auction than in the dummy auction.*

Connected to hypothesis 1 we would also expect those subjects who punish personally to have some emotional payoff from doing so that makes their monetary loss worthwhile.

To measure this, we asked the subjects about their happiness. We used two different questions ”How happy are you in general?” and ”How happy are you at the moment?” (see section 2.1).

Research on happiness uses the term *subjective well-being*. This term refers to people’s evaluations of their life. Often it is separated into *evaluative happiness* and *affective happiness*, where the former refers to the assessment of the life as a whole and the latter to the daily feelings or moods (Diener, 2000). As an event of little account, i.e. punishment in the experiment, should not influence someone’s life satisfaction, we asked our subjects about their happiness *at the moment* to assess their affective happiness. There exist various measures of subjective-well being, but for simplicity we use a single question about happiness. Such a single question self-report measurement is used, for example, in the United Kingdom Office for National Statistics (ONS) survey.

treatment (choice by B)	groups	punishment		bid observations <sup>17</sup>	
		yes	no	punishment auction	dummy auction
1A (10,0)	3	2	7	9	-
(2,8)	16	28	20	45	-
2A (10,0)	0	0	0	0	0
(2,8)	10	18	12	30	30
NC	11	-	-	-	33
<b>total</b>	29	48	39	84	63

Column 2: Each group includes 1 subject  $B$  and 3 matched subjects  $A$ . Column 3&4: decisions by subjects  $A$ . Column 5&6: number of bid observations from subjects'  $A$  resulting from groups where the auction was happening.

TABLE 2. Overview punishment and auction stage

How should personal punishment influence someone's mood? Self-efficacy plays a role in happiness, as stated by Maddox (2002): "Most philosophers and psychological theorists agree that a sense of control over our behavior, our environment, and our own thoughts and feelings is essential for happiness and a sense of well-being." This leads to the following hypothesis.

**Hypothesis 2.** *Happiness: Subjects  $A$  who punish personally are happier than those who do not.*

#### 4. RESULTS

Stages one and two of our treatments 1A and 2A were constructed to produce a large number of observations where punishment could possibly occur. A first look at the data confirms that this goal is achieved. All 87 subjects  $A$  in 1A and 2A did complete the real effort task, therefore all 29 subjects  $B$  had to make their decision for three matched successful subjects  $A$ . Out of the 29 subjects  $B$ , all but three implemented the allocation (2, 8), which was worse for subjects  $A$ . All three subjects  $B$  implementing (10, 0) played in design 1A.

Trying to find personal punishment is only viable if there is some punishment in the first place. Given the allocation of their matched subject  $B$ , all subjects  $A$  could chose to have the auction in stage four implemented. Demanding the auction is equivalent to subject  $B$  being

punished, since this ensures that subject  $B$ 's payoff will be reduced by 8.

Subjects  $A$  who faced the “bad” (2,8) split demand the auction significantly more often than those who got the “nice” (10,0) allocation ( $p = 0.040$ , 87 obs., one-sided Fisher-exact test). See table 2 for an overview of which groups voted for punishment. In total, 55% of subjects  $A$  demanded the auction to happen in stage 3. However, since the auction is implemented if at least one subject  $A$  demands it, the auction takes place in almost all of our groups.<sup>17</sup>

Table 3 shows the percentage of subjects  $A$  who bid a positive amount in the auctions. Recall that bids in the auction are not payoff relevant for subject  $B$ , only whether the auction happens or not influences the payoff of subject  $B$ . Subjects  $A$  who are either strict money maximizers or only interested in the monetary consequences of punishment for the matched subject  $B$  have no incentive to bid larger than zero. In contrast to that, we find that, in total, 36% of our subjects bid positive amounts of money in the punishment auction (line 1). So a substantial minority of subjects is interested enough in punishing personally to be willing to sacrifice some of their own money to achieve this.

We also find some positive bids of subjects who did not want the auction to happen in the previous stage (lines 3), but the average bid by subjects who wanted the auction (line 2) is significantly higher ( $p = 0.043$ , 84 obs., two-sided MWU test). One possible explanation for bids from subjects  $A$  who did not previously demand punishment is counter-punishment: By bidding the second highest amount in the auction, subjects  $A$  can (counter-)punish the auction winner for punishing  $B$ . If the possibility of counter-punishment leads to an overall reduction of bids, then the counter-punishment effect runs opposite to the personal punishment effect and should bias our estimates for personal punishment towards zero, making our estimates conservative.<sup>18</sup>

<sup>17</sup>Whether the auction happens depends on the random matching of subjects  $A$  into groups of 3. The auction did not happen in one group in 1A, which had seen the distribution (2,8). In 2 additional groups (one which saw (2,8) and one which saw (10,0)), all subjects did not want the auction to happen, but none-the-less, due to a procedural error, subjects entered bids (assuming an auction would take place). These groups are included in the bids data, even though no auction took place.

<sup>18</sup>We thank an anonymous referee for pointing out the possibility of counter-punishment.

bids by subjects $A$	treat.	obs.	avg.(SD)	max	bid>0	bid=0
<i>punishment auction</i>	1A,2A	84	0.43 (1.11)	5.50	36%	64%
wanted auction: yes	1A,2A	48	0.51 (1.24)	5.50	46%	54%
wanted auction: no	1A,2A	36	0.32 (0.93)	4.00	22%	78%
saw (2, 8)	1A,2A	75	0.48 (1.17)	5.50	39%	61%
saw (10, 0)	1A,2A	9	0.00 (0.00)	0.01	11%	89%
dummy auction present	2A	30	0.34 (0.98)	4.00	37%	63%
DA not present	1A	54	0.48 (1.18)	5.50	35%	65%
<i>dummy auction</i>	2A, NC	63	0.36 (1.32)	6.50	35%	65%
context	2A	30	0.03 (0.10)	0.50	17%	83%
no context	NC	33	0.67 (1.78)	6.50	52%	48%

TABLE 3. Bids

In contrast to subjects who could feel “wronged” by the distribution (2, 8), subjects who got the beneficial distribution (10, 0) almost never bid (lines 4-5).<sup>19</sup> Finally, bids are similar ( $p = 0.913$ , 84 obs., two-sided MWU test) in the punishment auctions of treatments 1A and 2A (lines 6-7), but very different ( $p = 0.007$ , 63 obs., two-sided MWU test) in the dummy auctions of treatments 2A, where another auction with context is present, and NC, where no context is given at all (lines 9-10).

The positive bids in the punishment auctions indicate that our subjects want personal punishment, but a more direct test for the existence of personal punishment is the comparison of the results for the two auctions in design 2A (lines 6 and 9). Here, within subject, are two identical auctions, leading to a similar result (an envelope gets destroyed and subject  $B$  loses a payoff of 8), the only difference is whether subjects get to destroy an unrelated envelope or the envelope belonging to subject  $B$ . Table 3 reveals a large difference in average bids, and

<sup>19</sup>The difference is almost significant ( $p = 0.076$ , 84 obs., two-sided MWU test). However, we have only few observations where groups saw (10, 0).

frequency of positive bids.<sup>20</sup> A Wilcoxon Signed Ranks test shows that this difference is significant ( $p = 0.015$ , 30 obs.).<sup>21</sup>

**Result 1.** *a) In the context of an experienced unfair split of earnings, more than one third of subjects bid in an auction for personal punishment.*

*b) Bids for personal punishment are higher by subjects who wanted punishment than by those who did not.*

*c) Bids for personal punishment are higher than bids in a simultaneous dummy auction, but not higher than in a dummy auction without earnings context.*

The second part of result 1c was surprising to us. The bids for the dummy auction are very different in 2A and NC. This points out the importance of giving subjects a context in which to evaluate an auction. Without the preceding stages, the auction must have made little sense to subjects in NC.<sup>22</sup> This could very well be an experimenter demand effect (Zizzo, 2010). Interestingly, this effect is reduced when subjects have to simultaneously participate in two auctions: Even though the price (destroy an unrelated envelope) in the dummy auction in 2A is identical to the price in the single auction in NC, the bids are much lower in 2A. It seems as if subjects only feel compelled to bid – to possibly please the experimenter – if the bidding is the only action in the experiment and if there is no morale for or against bidding possible. In 2A, when making multiple decisions, not bidding becomes acceptable.

Finally, we look at the result of the physical destruction carried out by the winners of the auction. Do they enjoy the act of destroying

---

<sup>20</sup>While very infrequent, there is some bidding in the dummy auction in 2A. The answers from the subject with the highest bid in the dummy auction to an open ended question about motivation for bidding are perhaps revealing:

(personal punishment auction): “Even though subject  $B$  is in no way affected (since he always gets 2 envelopes), it feels good to release some pressure this way”

(dummy auction): “To erase the feeling of anger, that, even though I did the whole work, candidate  $B$  will earn 3x as much”

<sup>21</sup>A sign test leads to a similar result ( $p = 0.016$ , 30 obs., one-sided sign test).

<sup>22</sup>In all designs, subjects had to correctly answer a set of test questions before the experiment proceeded. However, the test questions only related to the mechanism of the auction (and the previous stages for 1A and 2A), not any possible rationale behind holding it.



	(1)	(2)	(3)	(4)
<i>auction winner</i>	0.692** (0.262)	0.692** (0.263)	0.706** (0.293)	0.659** (0.281)
<i>wanted auction</i>		-0.171 (0.255)	-0.169 (0.257)	-0.108 (0.256)
<i>bid</i>			-0.013 (0.121)	0.048 (0.120)
<i>(10,0)-distribution</i>				1.150** (0.479)
<i>age</i>				-0.006 (0.044)
<i>female</i>				-0.301 (0.270)
<i>constant</i>	-0.269 (0.151)	-0.164 (0.218)	-0.163 (0.220)	0.020 (1.039)
<i>N</i>	78	78	78	78
<i>R</i> <sup>2</sup>	0.084	0.089	0.090	0.195
adj. <i>R</i> <sup>2</sup>	0.072	0.065	0.053	0.127

Notes: dependent variable: *happiness difference*; \*\* denotes significance at 5%-level; standard errors in parentheses; *bid*: the bids from the punishment auction (1A and 2A)[in both sessions of 2A, the coin flip chose the punishment auction, therefore the punishment auction was resolved and the data is used in the regressions]

TABLE 4. Regression on happiness difference for the punishment auctions 1A, 2A

subject *B*'s money? We asked all participants to report their subjective *affective happiness* on a seven point scale at the start and at the end of the experiment.<sup>23</sup> While the absolute level might depend on a number of causes we can not control, we can use the difference in affective happiness between the start and end of the experiment. Let the *happiness difference* be the amount of affective happiness reported at the end of our experiment minus the amount reported at the start. So subjects with a positive happiness difference felt better after our experiment than before.

<sup>23</sup>See Online-Appendix for the translated questionnaires.

Subjects who actually punished subject  $B$  were happier than those who did not ( $p = 0.02$ , 78 obs., two-sided MWU test).<sup>24</sup>

Table 4 reports regressions on happiness difference for the punishment auctions in treatments 1A and 2A. The regression shows that subjects  $A$  who went on to win the auction are happier than those who did not win. So despite being paid less money in the end, subjects who personally destroyed subject  $B$ 's money leave the experiment happier than those who do not, in line with hypothesis 2. The other significant predictor is the  $(10,0)$ -distribution. Not surprisingly, subjects  $A$  who encountered the allocation  $(10,0)$  felt happier compared to those who received only EUR 2 from allocation  $(2,8)$ .

**Result 2.** *Subjects who won the auction for personal punishment are happier than those who did not.*

While in treatments 1A and 2A winning the auction is equivalent to actually punishing subject  $B$ , this is not the case in treatment NC, where winning the auction has no such meaning. In treatment NC there is no significant difference in happiness between those subjects who won the auction and those who did not ( $p = 0.26$ , 33 obs., two-sided MWU test).

Table 5 reports a regression analysis of happiness difference for subjects<sup>25</sup> in treatment NC (following a similar approach as for the 1A and 2A treatments, compare table 4). In the regression analysis auction winner only has a weakly significant effect in the full model (column 3), where it is significant at the 10%-level.

## 5. CONCLUSION

In an experiment designed to separate the decision to punish personally from the more general decision to punish, we find that many subjects bid positive amounts in a second price auction that auctions off

---

<sup>24</sup>Observations for all groups where the punishment auction took place. In all sessions of treatment 2A, the coin flip to determine which auction would take place selected the punishment auction and not the dummy auction. Therefore we have auction winner data for the punishment auctions in 1A and 2A and auction winner data for the dummy auction in NC. See footnote 17.

<sup>25</sup>Obviously, subjects in NC did not see allocations and did not decide on conducting the auction.

	(1)	(2)	(3)
<i>auction winner</i>	0.364 (0.344)	0.538 (0.396)	0.755* (0.432)
<i>wanted auction</i>			
<i>bid</i>		-0.96 (0.106)	-0.162 0.136
<i>(10, 0)-distribution</i>			
<i>age</i>			-0.003 (0.036)
<i>female</i>			0.494 (0.385)
<i>constant</i>	0.273 (0.199)	0.278 (0.199)	0.086 (1.024)
<i>N</i>	33	33	33
<i>R</i> <sup>2</sup>	0.035	0.060	0.119
adj. <i>R</i> <sup>2</sup>	0.004	-0.002	-0.007

Notes: dependent variable: *happiness difference*; \* denotes significance at 10%-level; standard errors in parentheses; *bid*: the bids from the dummy auction (NC).

TABLE 5. Regression on happiness difference for the dummy auction without context

the right to punish personally. Some of these subjects bid substantial amounts.

The experimental designs are constructed to eliminate a range of other effects, which might have an influence on subjects decisions in more general settings. Due to the one-shot nature of the experiment, it is not possible to use bids as a signaling device for future play. Furthermore, seats in the experiment were separated by blinds, so the act of punishing was hard to use to express disapproval as in Masclet, Noussair, Tucker, and Villeval (2003). Since punishment is the physical act of destroying (paper) money, it might be a worry that subjects like to destroy money. However, the results of our questionnaire suggest otherwise.<sup>26</sup> The act of destroying the envelope is a punishment

<sup>26</sup>The final questionnaire included the question “Do you like destroying money?”. Not one of the subjects answered with yes. Additionally, subjects were given the

of subject  $B$ , not money burning as in Zizzo (2003), where no strategic component was involved. Most importantly, the bids in the auction, and thus the willingness to pay for personal punishment, have no influence on the payoff of the offending subject  $B$ . Subject  $B$ 's payoff is completely determined in stages 1 to 3. One of the mandatory test questions covered this point to make it clear to every subject. Our decision to use a second price sealed bid auction stems from the previous considerations. It is a fast and incentive compatible method that lets us elicit a very fine grained willingness to pay for personal punishment. Since the auction always has a winner, it emphasizes the point that punishment will always occur, regardless of the bids of subjects  $A$ .

The personal punishment we address in this paper differs from anti-social punishment as in Herrmann, Thöni, and Gächter (2008), which is punishing people that behaved pro-socially. In our case, when subjects  $B$  decided on the distribution, they (mostly) chose the unfair (2, 8)-split; they therefore do not behave pro-social. When we look for antisocial punishment in our data, we find that only 2 out of 9 subjects (22.2%), who were confronted with the fair or pro-social (10, 0)-split, voted for punishment.

We further find that winning the auction has a positive effect on affective happiness. While we can not exclude the possibility that subjects happiness is only due to winning the auction, the result is also consistent with subjects enjoying the personal punishment they achieved.

Using an auction might introduce a motivation to bid due to a “desire to win”. Van den Bos, Li, Lau, Maskin, Cohen, Montague, and McClure (2008) find evidence for this in a sealed bid first price auction. In one of their treatments, the opponents are other human subjects (similar to our NC design), while in two other treatments, subjects bid against computerized agents. Furthermore, all subjects are taught to calculate the (risk-neutral) Nash-equilibrium strategy, to rule out a winner’s curse effect stemming from limited cognitive ability. They find that subjects playing against humans overbid significantly more often than those playing against computers. There is also evidence from a fMRI experiment by Delgado, Schotter, Ozbay, and Phelps (2008) who compare subjects’ reactions to losing a lottery versus losing an auction opportunity to destroy some of their own remaining money during the final questionnaire. Again, none took this opportunity.

to conclude that “The fear of losing the social competition inherent in an auction may lead people to pay too high a price for the good for sale”. It is possible that, in a similar vein, our subjects did not want to “lose” the auction and therefore bid positive amounts. Our results in NC can be seen as further evidence for such an effect. However, in 2A, we directly compare the results of two auctions. If a desire to win exists, it should influence both auctions in a similar way, yet we find a significant difference between the two.

Overall, the effects we observe are significant, but not huge. This is not surprising, since we exclude many other effects which would otherwise work in a similar direction. In many real life examples, the demand for punishment and the demand for personal punishment will be measured simultaneously. Additionally, the personal punishment, as Adam Smith describes it, is punishment for a grave offense. For obvious reasons, laboratory experiments can only implement minor offenses.

## REFERENCES

- ANDERSON, C. M., AND L. PUTTERMAN (2006): “Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism,” *Games and Economic Behavior*, 54(1), 1–24.
- ANDREONI, J. (1990): “Impure altruism and donations to public goods: a theory of warm-glow giving,” *The Economic Journal*, 100(401), 464–477.
- BALLIET, D., L. B. MULDER, AND P. A. M. VAN LANGE (2011): “Reward, Punishment, and Cooperation: A Meta-Analysis,” *Psychological Bulletin*, 137(4), 594–615.
- BANDURA, A. (1977): “Self-efficacy: Toward a Unifying Theory of Behavioral Change,” *Psychological Review*, 84(2), 191–215.
- BOLTON, G. E., AND A. OCKENFELS (2000): “ERC: A theory of equity, reciprocity, and competition,” *American Economic Review*, 90(1), 166–193.
- CARLSMITH, K. M. (2006): “The roles of retribution and utility in determining punishment,” *Journal of Experimental Social Psychology*, 42(4), 437–451.
- CARLSMITH, K. M., J. M. DARLEY, AND P. H. ROBINSON (2002): “Why do we punish?: Deterrence and just deserts as motives for punishment..,” *Journal of Personality and Social Psychology*, 83(2), 284–299.
- CARPENTER, J. P. (2007): “The demand for punishment,” *Journal of Economic Behavior & Organization*, 62(4), 522–542.
- CASARI, M., AND L. LUINI (2009): “Cooperation under alternative punishment institutions: An experiment,” *Journal of Economic Behavior & Organization*, 71(2), 273–282.
- (2012): “Peer punishment in teams: expressive or instrumental choice?,” *Experimental Economics*, 15(2), 241–259.
- CHAUDHURI, A. (2011): “Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature,” *Experimental Economics*, 14(1), 47–83.
- DARLEY, J. M., K. M. CARLSMITH, AND P. H. ROBINSON (2000): “Incapacitation and just deserts as motives for punishment,” *Law and human behavior*, 24(6), 659–83.

- DE QUERVAIN, D. J.-F., U. FISCHBACHER, V. TREYER, M. SCHELLHAMMER, U. SCHNYDER, A. BUCK, AND E. FEHR (2004): "The neural basis of altruistic punishment.," *Science*, 305(5688), 1254–8.
- DELGADO, M. R., A. SCHOTTER, E. Y. OZBAY, AND E. A. PHELPS (2008): "Understanding overbidding: using the neural circuitry of reward to design economic auctions," *Science*, 321(5897), 1849–1852.
- DIENER, E. (2000): "Subjective well-being: The science of happiness and a proposal for a national index.," *American Psychologist*, 55(1), 34–43.
- DUFWENBERG, M., AND G. KIRCHSTEIGER (2004): "A theory of sequential reciprocity," *Games and Economic Behavior*, 47(2), 268–298.
- FALK, A., E. FEHR, AND U. FISCHBACHER (2003): "On the Nature of Fair Behavior," *Economic Inquiry*, 41(1), 20–26.
- (2005): "Notes and comments driving forces behind informal sanctions," *Econometrica*, 73(6), 2017–2030.
- (2008): "Testing theories of fairness—Intentions matter," *Games and Economic Behavior*, 62(1), 287–303.
- FALK, A., AND U. FISCHBACHER (2006): "A theory of reciprocity," *Games and Economic Behavior*, 54(2), 293–315.
- FEHR, E., AND K. M. SCHMIDT (1999): "A theory of fairness, competition, and cooperation," *Quarterly journal of Economics*, 114(3), 817–868.
- GÄCHTER, S., AND B. HERRMANN (2009): "Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 364(1518), 791–806.
- GREINER, B. (2004): "An online recruitment system for economic experiments," in *Forschung und wissenschaftliches Rechnen 2003*, ed. by K. Kremer, and V. Macho, chap. GWD Berich, pp. 79–93. Gesellschaft für Wissenschaftliche Datenverarbeitung, Göttingen.
- GÜTH, W., R. SCHMITTBERGER, AND B. SCHWARZE (1982): "An experimental analysis of ultimatum bargaining," *Journal of Economic Behavior & Organization*, 3(4), 367–388.
- HARBAUGH, W. T., U. MAYR, AND D. R. BURGHART (2007): "Neural responses to taxation and voluntary giving reveal motives for charitable donations," *Science*, 316, 1622–1625.

- HERRMANN, B., C. THÖNI, AND S. GÄCHTER (2008): “Antisocial punishment across societies,” *Science*, 319(5868), 1362–7.
- KAHNEMAN, D., J. L. KNETSCH, AND R. THALER (1986): “Fairness as a constraint on profit seeking: Entitlements in the market,” *The American economic review*, 76(4), 728–741.
- MADDOX, J. E. (2002): “Self-efficacy: the power of believing you can,” in *Handbook of positive psychology*, ed. by C. Snyder, and S. J. Lopez. Oxford University Press, New York.
- MASCLET, D., C. NOUSSAIR, S. TUCKER, AND M.-C. VILLEVAL (2003): “Monetary and nonmonetary punishment in the voluntary contributions mechanism,” *American Economic Review*, 93(1), 366–380.
- ORTH, U. (2003): “Punishment goals of crime victims,” *Law and Human Behavior*, 27(2), 173–186.
- OUSS, A., AND A. PEYSAKHOVICH (2013): “When Punishment Doesn’t Pay: “Cold Glow” and Decisions to Punish,” *mimeo*.
- RABIN, M. (1993): “Incorporating fairness into game theory and economics,” *The American Economic Review*, 151(3712), 867–868.
- SMITH, A. (1759): *The Theory of Moral Sentiments*. A. Millar, London.
- VAN DEN BOS, W., J. LI, T. LAU, E. MASKIN, J. D. COHEN, P. R. MONTAGUE, AND S. M. McCLURE (2008): “The value of victory: social origins of the winner’s curse in common value auctions,” *Judgment and decision making*, 3(7), 483–492.
- ZIZZO, D. J. (2003): “Inequality and Procedural Fairness in a Money Burning and Stealing Experiment,” *Research on Economic Inequality*, 11(155), 215–247.
- (2010): “Experimenter demand effects in economic experiments,” *Experimental Economics*, 13(1), 75–98.