

Duersch, Peter; Servátka, Maroš

**Working Paper**

## Punishment with Uncertain Outcomes in the Prisoner's Dilemma

Discussion Paper Series, No. 485

**Provided in Cooperation with:**

Alfred Weber Institute, Department of Economics, University of Heidelberg

*Suggested Citation:* Duersch, Peter; Servátka, Maroš (2009) : Punishment with Uncertain Outcomes in the Prisoner's Dilemma, Discussion Paper Series, No. 485, University of Heidelberg, Department of Economics, Heidelberg

This Version is available at:

<https://hdl.handle.net/10419/127305>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

University of Heidelberg

Department of Economics



Discussion Paper Series | No. 485

**Punishment with Uncertain Outcomes  
in the Prisoner's Dilemma**

---

*Peter Duersch and Maroš Servátka*

---

July 2009

# Punishment with Uncertain Outcomes in the Prisoner's Dilemma

*Peter Duersch<sup>\*</sup> and Maroš Servátka<sup>\*\*§</sup>*

## Abstract

This paper experimentally investigates whether risk-averse individuals punish less if the outcome of punishment is uncertain than when it is certain. Our design includes three treatments: Baseline in which the one-shot prisoner's dilemma game is played; Certain Punishment in which the prisoner's dilemma game is followed by a punishment stage allowing subjects to decrease the other player's payoff by 2 Euros; and Uncertain Punishment in which subjects could decrease the other player's payoff with a 50% probability by 1 Euro and with a 50% probability by 3 Euros. We find that in all cases the risk-averse subjects are equally likely to cooperate in the prisoner's dilemma and equally likely to punish in the second stage in either of the two punishment treatments.

Keywords: experiment, prisoner's dilemma, punishment, risk aversion, uncertainty

JEL Classifications: C72, C91

---

\* Department of Economics, Universität Heidelberg, Bergheimer Strasse 58, 69115 Heidelberg, Germany, [peter.duersch@awi.uni-heidelberg.de](mailto:peter.duersch@awi.uni-heidelberg.de)

\*\* Department of Economics and Finance, University of Canterbury, Private Bag 4800, Christchurch, New Zealand, [maros.servatka@canterbury.ac.nz](mailto:maros.servatka@canterbury.ac.nz)

§ We gratefully acknowledge productive suggestions of the associate editor (Ananish Chaudhuri), three anonymous referees, Astrid Hopfensitz, Nikos Nikiforakis, Joerg Oechssler, Ernesto Reuben, Wendelin Schnedler, and the participants of the Universität Heidelberg departmental seminar. The research funds for this study were provided by Sonderforschungsbereich 504 and the University of Canterbury, College of Business and Economics.

# 1. Introduction

Imagine two researchers working on a joint project. Both can work hard or free-ride on the work of the other researcher. For a certain set of outcomes, this situation resembles a prisoner's dilemma. The researcher's action will eventually be revealed to his co-author, who will then have the opportunity to punish him for slacking off. What could a punishment look like? For example, it could take the form of sharing the bad experience with other colleagues in the profession. However, it is unclear what effect the punishment will have on the flaky researcher. On one hand, it might affect a tenure decision or hiring in a close race for a job or perhaps discourage other colleagues to work with the person in the future. On the other hand, the punishment might be meaningless if other factors already determined the outcome and/or when other colleagues disregard the information about the flaky researcher's input into the project.

Because the co-author who punishes does not necessarily know the flaky researcher's situation, she cannot fully assess the impact of the punishment. Thus, the decision to punish could be viewed as having an uncertain outcome. In fact, there are many real life situations where the punishment might not get through at all and thus be insignificant to the recipient.<sup>1</sup> It is thus natural to ask whether uncertainty associated with the punishment outcome is an important determinant of the punishment decision and whether the punisher's risk attitude sheds any light on her behavior. Understanding the role of risk attitude might turn out to be socially beneficial as it is likely to affect individuals' willingness to cooperate (Becker (1968)) and potentially also punish (in particular in situations where punishment can lead to retaliation and escalation of conflict).

Distributional models of other-regarding preferences (e.g., Fehr and Schmidt (1999), Bolton and Ockenfels (2000), Charness and Rabin (2002)) which are often used to explain the punishing behavior noted elsewhere in the literature, do not consider uncertainty of outcomes. On their own such models cannot provide guidance on whether we should observe more or less punishment in situations when its outcome is uncertain. Moreover, in the above example the uncertainty pertains to the flaky researcher but does not directly affect the (monetary) payoff of the decision maker. Thus, none of existing expected utility theories can be directly applied without making an additional assumption on how uncertainty affects preferences over the payoffs of the other person.

---

<sup>1</sup> Sometimes punishment might also target a wrong person.

In this paper we study how uncertainty of punishment outcomes interacts with a decision to cooperate and also with a decision to decrease the other person's monetary payoff in a punishment stage. In our analysis we assume that a decision maker's risk attitude also determines preferences over expected payoffs of other people in the same manner as it determines preferences over her own expected payoff. We present an experiment in which subjects have an opportunity to punish their counterpart, but the outcome of punishment depends on the realization of a random variable. The subjects' decisions are compared to those in another treatment in which the outcome of punishment is certain. We have embedded our explorations in a prisoner's dilemma game in which the players decide whether to defect and maximize their own monetary payoffs or to cooperate and maximize the joint surplus.

Social dilemma situations have been extensively studied in the economics literature for a long time (see Roth (1988) for an overview). Fehr and Gächter (2000a) show that incentives to free-ride in a voluntary contribution mechanism (VCM) experiment can be counteracted by introducing a second stage which allows for punishment. Despite the punishment being costly, many subjects use that opportunity to deter defection. Initially, the effect on cooperation is small, but the contributions to a public project increase over time in a repeated game. A considerable amount of literature follows this paper and extends the result to non-pecuniary sanctions (Masclot, *et al.* (2003); Noussair and Tucker (2005)) and explores the effectiveness of punishment (Nikiforakis and Normann (2008)) as well as the price of punishment (Anderson and Putterman (2006), Carpenter (2007)).

In these experiments (and all other experiments on social dilemmas which we are aware of) the outcome of the punishment stage is certain.<sup>2</sup> The literature on contract enforcement which explores whether an agent will exert effort or shirk when faced with some probability of being monitored is of little help as the underlying game has a sequential rather than simultaneous structure and the experimental investigations (e.g., Fehr and Gächter (2000b), Fehr and List (2004)) focus on the effect of punishment on cooperation but not on the decision whether to punish or not and neither on its determinants.

Experimental subjects are, in general, not risk-neutral (e.g., Harrison (1986), Cox, *et al.* (1988), Holt and Laury (2002), Eckel and Grossman (2008)). If subjects' risk attitudes interact with their other-regarding preferences then a different risk structure present in the laboratory experiments on punishment in social dilemmas could potentially lead to outcomes

---

<sup>2</sup> Experiments which allow for counter-punishment (e.g., Denant-Boemont, *et al.* (2007), Nikiforakis (2008), Engelmann and Nikiforakis (2008)) are an exception, because the recipient of punishment can reciprocate. Thus, because of the strategic uncertainty stemming from unknown moves of other players, it is not obvious what the outcome of the original punishment will be.

which are different from behavior in the outside-the-lab world.<sup>3</sup> Our study addresses the issue for a simple case of one-shot interaction as we feel that the simplicity of the environment is a virtue in exploratory projects such as this one.

Despite a large body of experimental literature on cooperation in the prisoner's dilemma game, we are not aware of any studies exploring the effect of punishment other than ceasing cooperation if the interaction is repeated. Although the prisoner's dilemma game incorporates motivations present also in the VCM (which is sometimes called the  $n$ -person prisoner's dilemma) it is not obvious whether (the possibility of) punishment increases cooperation as observed in VCM experiments. We examine this question by including a prisoner's dilemma game without punishment in our design.<sup>4</sup>

We elicit subjects' risk attitude at an individual level to test whether it explains the behavior in the prisoner's dilemma game and the punishment decisions across our treatments. We find that risk-averse subjects are equally likely to cooperate in all treatments. Moreover, we find no link between risk-aversion and the decision to punish when the outcome of punishment is certain or uncertain.

Next we present the experimental setup and our results, followed by a short discussion. Instructions can be found in the appendix.

## 2. Experimental Setup

The experiment consists of three treatments: *Baseline* (*base* in figures), *Certain Punishment* (*cp*), and *Uncertain Punishment* (*ucp*) implemented in an across subjects design. In each treatment a prisoner's dilemma game is played. The game payoffs (presented in Table 1) are denoted in Euros. The row player chooses Top (cooperation) or Bottom (defection), while the column player chooses Left (cooperation) or Right (defection).

In the punishment stage of Certain Punishment and Uncertain Punishment treatments, after being notified of the result of the prisoner's dilemma game, subjects could engage in

---

<sup>3</sup> Of course, laboratory experiments are, by construction, simplifications of the outside-the-lab world. However, it is important to keep in mind these differences and whenever possible test their influence, because despite being simplifications, laboratory experiments are often (directly or indirectly) used to draw inferences towards outside-the-lab behavior.

<sup>4</sup> Gangadharan and Nikiforakis (2009) study bridges the literature between a repeated prisoner's dilemma game and VCM. They find that subjects behave more cooperatively in the prisoner's dilemma than in the VCM if they are in a group of four players but do not find a difference if the two games are played in pairs.

costly punishment of their partner; in Baseline there was no punishment stage. All information was common knowledge.

**Table 1: Prisoner's Dilemma Payoffs**

	<b>Left</b>	<b>Right</b>
<b>Top</b>	5,5	0,8
<b>Bottom</b>	8,0	2,2

In the punishment stage of the Certain Punishment treatment subjects could decrease their counterpart's payoff by 2 Euros with certainty at the cost of 1 Euro to themselves. In the Uncertain Punishment treatment subjects could decrease the other player's payoff with 50% probability by 1 Euro and with 50% probability by 3 Euros at the cost of 1 Euro to themselves. Thus while the expected punishment was the same in both cases (2 Euros), its outcome depended on the state of the world in the Uncertain Punishment treatment. The subjects were instructed that a coin would be flipped in front of them to determine the punishment outcome. If the coin toss lands on heads, the other player's payoff is decreased by 1 Euro. If the coins toss lands on tails, the other player's payoff is decreased by 3 Euros. If the subject decided to punish, the costs of punishment were incurred irrespective of the outcome of the probability draw.<sup>5</sup>

In our experiment the subjects played the game only once as a repeated environment would allow them to construct portfolios of choices and render answering our research question impossible.<sup>6</sup> It should be noted that most of the experimental designs on punishment allow for repeated interaction. The decision to punish could be explained as an attempt to induce cooperation in the future. However, Fehr and Gächter (2000a) and other studies, provide evidence that subjects punish in the last round and that they punish in a stranger matching when there is no chance of encountering the same subject(s) more than once. Thus, we anticipated observing a relatively large proportion of punishing subjects even in the one-shot scenario.

---

<sup>5</sup> We decided to design the uncertain outcome in such a way that the punished person always learns that he or she is being punished. It is possible that subjects' behavior would differ if there was a chance that the punished person does not learn about the punishment. However, this was not the focus of our study and we leave it for future explorations.

<sup>6</sup> A random rematching after every period could potentially solve the problem. However, we are interested in creating a simple environment where we could study the effects of uncertain outcomes of punishment without having to consider other confounding factors such as being punished by someone else in previous rounds.

The expected effectiveness of punishment (2:1) in our experiment is lower than usually observed in the literature and the expected costs (1 Euro) are relatively high. These two design parameters were driven primarily by the consideration of not allowing subjects to make losses as this was the policy of the laboratory where the experiment was run. An alternative way of avoiding the subjects making losses was to significantly increase the show up fees. However, this could potentially cause the subjects to perceive the game payoffs as relatively small and alter their behavior. In order to get an estimate whether the cost of punishment or the fear of “over-punishment” deterred some of our subjects from punishing, we included a couple of questions pertaining to the demand for punishment in a non-paid questionnaire administered at the end of the experiment.

To measure risk attitudes of subjects we used the method developed by Holt and Laury (2002). That is, subjects were repeatedly offered a choice between two lotteries, one involving higher risk than the other. From subjects’ choices between ten lottery pairs it is possible to calculate their individual risk aversion parameter. Further details are provided in the appendix.

## ***2.1 Procedures***

The experiment was conducted at the SonderForschungsBereich 504 laboratory at the University of Mannheim in May and June of 2009. It was run manually under a single blind social distance protocol. The experiment consisted of 13 sessions (with a minimum of 6 subjects per session) that lasted on average 45-50 minutes including the payment of subjects. A total of 184 students of various majors (about half either economics or business), recruited from the laboratory subject pool, participated in our three treatments – Baseline, Certain Punishment, and Uncertain Punishment. Most of the students had previously participated in economics experiments. Each subject only participated in a single session of the study. Subjects earned on average 10.56 Euro including a 5 Euro show up fee.

The sequence of events in a session was the following. (i) Upon entering the laboratory subjects drew a ball from an urn. The number that was indicated on the ball assigned their seat for the experiment and thus determined the matching which was done according to a pre-assigned matching protocol. (ii) The neutrally framed instructions (in German) for the prisoner’s dilemma (and the punishment stage in the two punishment treatments) were handed out. All sheets indicated subjects’ ID number. (iii) The subjects read the instructions and afterwards were encouraged to ask questions. The questions were asked and answered individually. (iv) The subjects simultaneously made their decisions for the prisoner’s



dilemma game. (v) The experimenters collected the decisions forms, transferred the decision information to their anonymous counterparts' decision forms and returned them to subjects. This prevented the exchange of superfluous information and aided in maintaining the anonymity of individual decisions. (vi) After learning the decision of the paired player the subjects made their decisions regarding punishment on a second decision form. (vii) The experimenters collected the decision sheets for prisoner's dilemma and punishment stages.

(viii) Then the instructions and decision forms for risk attitude elicitation task were handed out, filled out by subjects, and collected by the experimenters, one at a time. Subjects were informed beforehand that there would be an additional individual task after the prisoner's dilemma game with punishment (prisoner's dilemma game in Baseline), but not about the nature of this task. (ix) At the end of the session subjects filled out a questionnaire asking for their demand for punishment and demographics.<sup>7</sup> (x) Afterwards all of the subjects were paid privately in cash. Each subject received the following amount: an endowment of 5 Euro plus the earnings in the prisoner's dilemma minus the punishment minus the punishment costs plus a payment for one randomly chosen lottery from the risk attitude elicitation task. All uncertainties and lotteries were resolved by flipping a coin/rolling a 10-sided die.<sup>8</sup>

## ***2.2 Predictions and Hypotheses***

We start off by formulating a hypothesis regarding the subjects' behavior in the prisoner's dilemma game with and without punishment. In the classical solution for self-regarding preferences no punishment will ever be observed because it is costly and players will always choose to defect. However, models of other regarding preferences offer a different prediction. In the following analysis we rely on the Fehr and Schmidt (1999) model of inequality aversion, which has been prominently used in the punishment literature. The model predicts that a player whose payoff is lower than that of her counterpart will be willing to sacrifice some of her payoff to reduce the disadvantageous inequality. That is, a sufficiently inequality-averse cooperator will punish the paired defector if the punishment reduces the defector's payoff more than the cost which the punisher has to bear. Hence, if at least some subjects are inequality-averse, adding a punishment stage to the game structure decreases the expected payoff that players get from defecting, leading to more cooperation in the Certain

---

<sup>7</sup> Because the decision tasks were relatively simple we opted not to include test questions or examples in the instructions in order not to bias the subjects. Answering the questionnaire was not a requirement for payment.

<sup>8</sup> The coin was flipped publicly by a randomly chosen subject. The die was rolled individually by each subject at the time of payment.

Punishment and Uncertain Punishment treatments compared to Baseline.<sup>9</sup> However, the Fehr and Schmidt model does not allow for a prediction whether we should observe more cooperation in Certain Punishment or Uncertain Punishment.

On the other hand, other effects could partly or fully counteract the lower expected payoff. The defection is a dominant strategy, so all cooperation in the one-shot prisoner's dilemma game must stem from other-regarding preferences or intrinsic motivation. The introduction of a punishment stage could diminish the weight put on the payoff of the other person or crowd out the existing intrinsic motivation, thus potentially leading to lower overall cooperation even though the pure monetary benefit of cooperation is relatively higher with punishment.<sup>10</sup>

Our main hypothesis relates to subjects' risk attitudes as predictors of their punishing behavior in the face of uncertain outcomes. It assumes that a decision maker's risk attitude extends also over the payoffs of the other person. Such assumption reflects relatively plausible behavior: When empathizing with another person, the decision maker might try to impose her own preferences and values on that person. For instance, in our case if the decision maker is risk-averse, her behavior might also be risk-averse when making a choice whether to punish or not if the punishment outcome is uncertain from the perspective of her counterpart. Hence, we formulate the null as follows:

*H0: Risk-averse subjects punish less in Uncertain Punishment than in Certain Punishment.*

To see how H0 can be derived from the above assumption and the Fehr and Schmidt model, consider the following 2-player example using payoffs from the prisoner's dilemma game. Suppose an inequality-averse subject who cooperated in the prisoner's dilemma while her paired person defected. The subject's utility function is given by:

$$U_i(x) = x_i - \alpha_i \max[x_j - x_i, 0] - \beta_i \max[x_i - x_j, 0].$$

---

<sup>9</sup> Such prediction is in line with the observed behavior in repeated VCM experiments.

<sup>10</sup> See Frey and Jegen (2001) for a survey on crowding out intrinsic motivation. Alternatively, the punishment stage could be seen as a form of control which leads to lower cooperation because subjects dislike being controlled as has been observed by Falk and Kosfeld (2006).

If there is no punishment, the subject's payoff is  $U_i = 5 - \alpha_i 8$ . In the Certain Punishment treatment the subject can punish and obtain a payoff of  $U_i = 4 - \alpha_i 6$ . Thus, she will prefer to punish for values of  $\alpha_i > 0.5$ . In the Uncertain Punishment treatment she gets a payoff of  $U_i = 4 - \alpha_i 7$  with probability 1/2 and  $U_i = 4 - \alpha_i 5$  also with probability 1/2, yielding an expected utility (assuming risk neutrality) of  $U_i = 4 - \alpha_i 6$ ; the same as in Certain Punishment. However, under the assumption that the decision maker's risk attitude extends over the payoffs of the other person, a risk-averse subject will have a lower utility derived from punishment in Uncertain Punishment than in Certain Punishment.

### 3. Results

As our hypothesis is connected to subjects' risk aversion, we start off by describing the distribution of risk attitudes in our sample. The risk attitudes were elicited after the punishment decisions had been made (but before the decision of the paired player or the punishment outcome were revealed). In this method the risk attitude is determined by the number of safe choices made while choosing between a safe and a risky lottery. Never choosing the safe lottery corresponds to an extremely risk-loving subject. On the other hand, the higher the number of safe choices, the more risk-averse the subject is. The distribution of safe choices is shown in Figure 1. The risk neutrality corresponds to choosing the safe lottery exactly four times.<sup>11</sup> A Median test confirms that the random allocation of subjects to treatments yielded three subject groups which do not differ in their distributions of risk attitude ( $p$ -value = 0.833). In line with other experiments, our subjects show considerable amount of risk aversion, while only few are risk-loving.

---

<sup>11</sup> Definite statements about the risk attitude are only possible if the choices are monotonically ordered, that is when there is one lottery such that the subject always chooses the safe lottery for lower ranked lottery pairs and the more risky lottery for higher ranked lottery pairs; 93.5% of our subjects display such monotonic choice behavior. In the analysis we use the data on all subjects but control for those whose choices were non-monotonic and also for those who chose the dominated safe option in the last row of the risk attitude elicitation task.

Figure 1: Distribution of Risk Attitudes

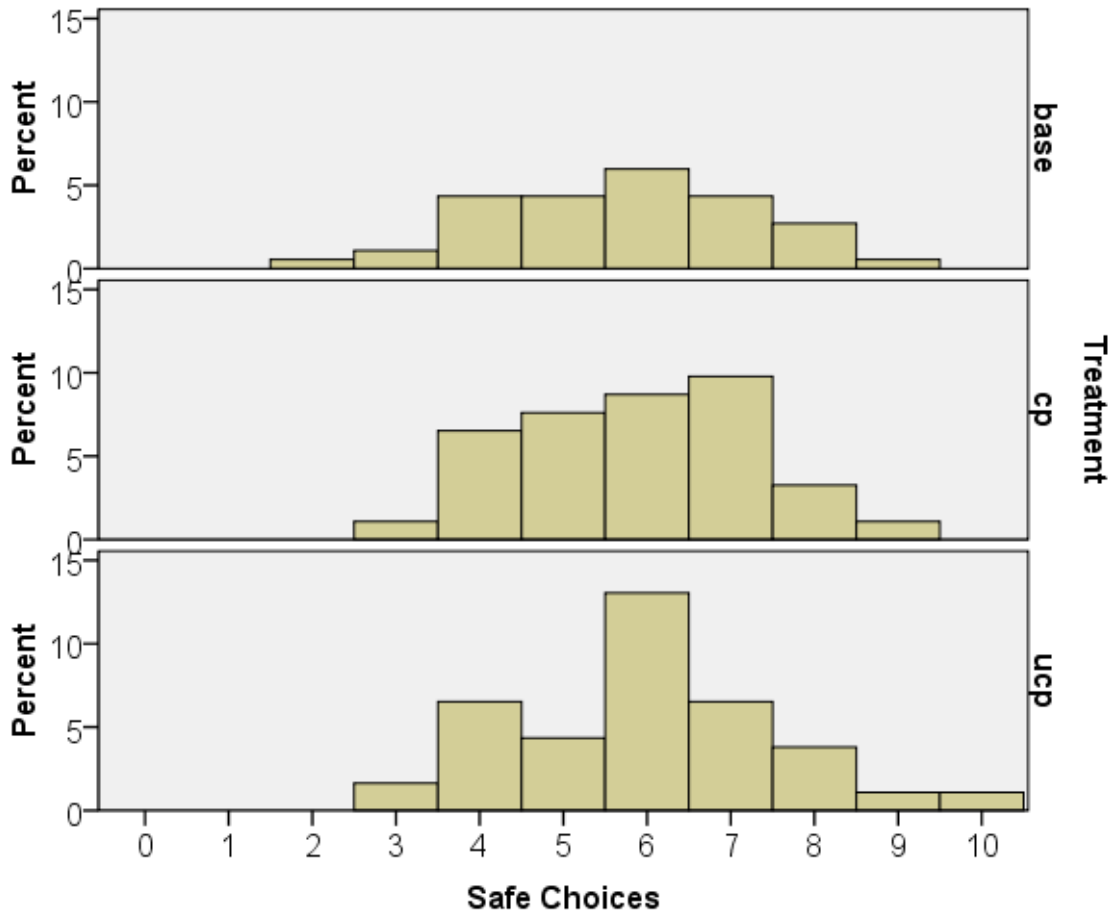


Table 2 presents subjects' behavior in the prisoner's dilemma game and in the punishment stage across treatments. In the Baseline treatment 43.2% of subjects chose to cooperate with their partner. The cooperation rate decreases to 32.9% with the introduction of certain punishment. However, this decrease is not statistically significant (1-sided Fisher exact test  $p$ -value = 0.180). If the punishment is uncertain the cooperation rate equals 44.3% which is approximately the same as in Baseline (1-sided Fisher exact test  $p$ -value = 0.532). Similarly, we find no difference in cooperation rates between Certain Punishment and Uncertain Punishment at the standard significance level (1-sided Fisher exact test  $p$ -value = 0.112). The results do not change when we perform the tests with risk-averse subjects only.

Despite a relatively large number of subjects in both punishment treatments ( $n = 70$ ) we observe few instances of punishment: there was only 1 subject who punished in Certain Punishment and 8 subjects who punished in Uncertain Punishment. It appears that given the cost and effectiveness of punishment, the subjects were unwilling to spend resources to decrease the payoff of the other person in a one-shot game because the punishment could not

lead to more cooperation in the future. However, at this stage we are not able to provide a definitive answer why our subjects did not punish.<sup>12</sup>

**Table 2: Subjects' Behavior in the Prisoner's Dilemma Game and Punishment Stage across Treatments**

	# observations	Prisoner's Dilemma		Punishment Stage
		Cooperate	Defect	Punish
<b>Baseline</b>	44	19 43.2%	25 56.8%	-
<b>Certain Punishment</b>	70	23 32.9%	47 68.1%	1 1.4%
<b>Uncertain Punishment</b>	70	31 44.3%	39 55.7%	8 11.4%

To verify the effect of risk aversion on the decision to cooperate, we run a probit regression and report the results in Table 3. Certain Punishment and Uncertain Punishment are dummy variables for the respective treatments; Risk aversion is a dummy for risk-averse subjects; *Inconsistent* is a dummy which takes on the value of 1 if the subject “jumped” back and forth between lotteries in the risk attitude elicitation task; *Confused* is a dummy which takes on the value of 1 if the subject chose the lower paying lottery in the last row of the risk attitude elicitation task;<sup>13</sup> *Age* and *Male* are the subject's age and gender as reported in the post-experimental questionnaire. The estimated coefficients are presented in the first column: Risk aversion as well as all other variables is statistically insignificant, supporting our previous finding that risk aversion does not influence subjects' willingness to cooperate. This result is robust to excluding demographic variables (results presented on the right hand side of Table 3), representing the risk attitude by the number of safe choices instead of a dummy, looking at differences from Certain Punishment or Uncertain Punishment as opposed to Baseline, and dropping *Inconsistent*, *Confused*, *Age*, and *Male* variables from the regression. The lack of statistical significance should not be interpreted as evidence against the expected payoff hypothesis or the crowding out hypothesis, because it is possible that both effects are present, but cancel each other out.

<sup>12</sup> It is possible that the observed behavior is driven by the large proportion of business and economics majors in our sample. Marwell and Ames (1981), Carter and Irons (1991), Frank, *et al.* (1993), and Rubinstein (2006) show that economics students behave in accordance with the predictions of neoclassical theory.

<sup>13</sup> In total, there are 10 inconsistent subjects (1 in Baseline, 5 in Certain Punishment, and 4 in Uncertain Punishment) and 4 confused ones (1 in Baseline and 3 in Uncertain Punishment).

**Table 3: Probit Regression Estimates for the Cooperation Behavior**

<i>Cooperate<sub>t</sub></i>	Coefficient	St. Error	Z	$p >  z $	Coefficient	St. Error	Z	$p >  z $
<b>Certain Punishment</b>	-0.30	0.25	-1.20	0.230	-0.26	0.25	-1.07	0.284
<b>Uncertain Punishment</b>	0.01	0.25	0.05	0.962	0.03	0.25	0.14	0.888
<b>Risk Aversion</b>	-0.21	0.23	-0.92	0.358	-0.16	0.23	-0.73	0.468
<b>Inconsistent</b>	0.65	0.44	1.47	0.142				
<b>Confused</b>	-0.10	0.71	-0.14	0.892				
<b>Age</b>	-0.0004	0.02	-0.02	0.987				
<b>Male</b>	-0.04	0.20	-0.23	0.822				
<b>Constant</b>	0.01	0.60	0.01	0.992	-0.05	0.25	-0.19	0.848

Number of observations = 184

Our main hypothesis asserts that risk-averse subjects punish less if the punishment outcome is uncertain. At the same time, the punishment decision likely depends on the history of play: When making their choice, subjects who cooperated might decide differently from those who defected themselves and probably behave differently towards cooperators than towards defectors (data presented in Appendix 1). To address this issue we include a dummy variable for a subject's own behavior in the prisoner's dilemma (*OwnPD*) and for the behavior of his or her counterpart (*OtherPD*) in another probit model (reported in Table 4) exploring the determinants of punishment.<sup>14</sup> While we observe that both one's own and the paired person's behavior in the prisoner's dilemma game are important factors of the punishment decision, risk aversion, age, and gender come out insignificant. If we exclude socio-demographic variables the statistical significance of *OwnPD* and *OtherPD* decreases (presented on the right hand side of Table 4). As before, the results are robust to including the number of safe choices in the regression instead of the risk aversion dummy and also to including an interaction variable *UCP*  $\times$  *RA*. Hence, we reject hypothesis H0.<sup>15</sup>

<sup>14</sup> We dropped Inconsistent variable from the specification because no inconsistent subject punished.

<sup>15</sup> This result is consistent with our finding in a classroom experiment where only several students were randomly chosen to be paid. The experiment involved a slightly different setup which allowed for punishment as well as for reward (Duersch and Servátka (2007)).

**Table 4: Probit Regression Estimates for the Punishment Behavior**

<i>Punish<sub>t</sub></i>	Coefficient	St. Error	Z	$p >  z $	Coefficient	St. Error	Z	$p >  z $
<b>Own PD</b>	0.87	0.42	2.07	0.038	0.66	0.38	1.74	0.083
<b>Other PD</b>	-1.11	0.55	-2.00	0.045	-0.65	0.42	-1.57	0.117
<b>Risk Aversion</b>	-0.49	0.45	-1.11	0.267	-0.30	0.42	-0.72	0.470
<b>Confused</b>	1.91	0.94	2.03	0.043				
<b>Age</b>	-0.03	0.06	-0.05	0.961				
<b>Male</b>	-0.19	0.41	-0.46	0.647				
<b>Uncertain Punishment</b>	0.97	0.51	1.88	0.060	1.03	0.48	2.17	0.030
<b>Constant</b>	-2.02	1.49	-1.35	0.176	-2.12	0.58	-3.68	0.000

Number of observations = 140

Recall that in order to keep our design simple we chose to restrict the punishment to only one (expected) option. However, it is possible that some subjects would like to punish more or less and thus the size of desired punishment might vary across treatments. In order to get at least partially at this issue, in the post-experimental questionnaire we asked our subjects the following question:

*If you could decide how much to destroy of the other player's payoff how much would you like to destroy?*

About 20% of subjects provided a positive number as their answer. The answers do not significantly vary between treatments, suggesting that the size of the desired punishment does not interact with uncertainty of its outcome. When we treat the subjects' answers as observations, a tobit regression with a lower bound of zero (presented in Appendix 2) shows a similar pattern: Risk-averse subjects did not destroy more or less in the Uncertain Punishment than in the Certain Punishment, thus providing further support for rejecting H0.

## 4. Discussion

This paper reports an experiment designed to study the role of risk attitude in punishing behavior when the outcome of punishment is uncertain. We assume that if the decision maker is risk-averse, her behavior will exhibit risk-aversion also over the payoffs of the person who is being punished. We observe a relatively small number of punishment instances in our setup and find that including a punishment stage (whether with certain or uncertain outcomes) had no significant effect on the cooperation rate in the prisoner's dilemma game. Moreover, we do not find any evidence that risk aversion is a factor when making a decision to punish.

There are several potential explanations and implications for our findings. We discuss three of them which are directly related to our design. The first one is that our assumption about risk aversion does not reflect reality. Based on models of conditional other-regarding preferences (e.g. Dufwenberg and Kirchsteiger (2004), Falk and Fischbacher (2006), Cox, *et al.* (2007), Cox, *et al.* (2008)) if the decision maker decides to punish, it can be argued that the aim of such action is not to benefit the other person but to hurt him. Therefore, it is possible that a risk-averse decision maker does not take into account the uncertainty which only affects the punished person. In our setup this might imply that the term describing the other player's payoff enters the utility function of a risk-averse decision maker in a risk-neutral way. The second explanation is connected with the use of Holt and Laury's measure of risk attitudes which might be not appropriate for punishment decisions such as the one presented in this paper. Studies by Isaac and James (2000) and Dave, *et al.* (2007) point out that different methods measuring risk attitudes yield significantly different estimates and a recent paper by Deck, *et al.* (2009) finds that their subjects behave as though Holt and Laury task was an investment decision. Therefore, a robustness check of our findings with respect to a different risk attitude elicitation method seems warranted. Third, in our experiment the punished person always learns that he or she is being punished. It is possible that our results do not directly apply to environments in which there is a chance that the punished person does not find out about the punishment.

Finally, the lack of punishment effect on the cooperation rate seems to be at odds with the literature on public goods VCM where punishment successfully deters free riding. However, because of the obvious differences in designs and cost structures we do not offer an explanation and leave this line of research for future explorations.



## References

- ANDERSON, C. M. and PUTTERMAN, L. (2006) "Do Non-Strategic Sanctions Obey the Law of Demand? The Demand for Punishment in the Voluntary Contribution Mechanism," *Games and Economic Behavior*, 54, 1-24.
- BECKER, G. (1968) "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 76(2), 169-217.
- BOLTON, G. and OCKENFELS, A. (2000) "ERC: A Theory of Equity, Reciprocity, and Cooperation," *American Economic Review* 90, 166-193.
- CARPENTER, J. (2007): "The Demand for Punishment," *Journal of Economic Behavior and Organization*, 62(4), 522-542.
- CARTER, J. and IRONS, M. (1991) "Are Economists Different, and If So, Why?" *Journal of Economic Perspectives*, 5(2), 171-177.
- CHARNESS, G. and RABIN, M. (2002) "Understanding Social Preferences with Simple Tests," *Quarterly Journal of Economics*, 117(3), 817-869.
- COX, J.C., FRIEDMAN, D., and GJERSTAD, S. (2007) "A Tractable Model of Reciprocity and Fairness," *Games and Economic Behavior*, 59, 17-45.
- COX, J.C., FRIEDMAN, D. and SADIRAJ, V. (2008) "Revealed Altruism," *Econometrica*, 76, 31-69.
- COX, J.C., SMITH, V., and WALKER, J. (1988) "Theory and Individual Behavior of First Price Auctions," *Journal of Risk and Uncertainty*, 61-99.
- DAVE, C., ECKEL, C., JOHNSON, C., and ROJAS, C. (2007) "On the Heterogeneity, Stability, and Validity of Risk Preference Measures," University of Texas at Dallas working paper.
- DECK, C., LEE, J., REYES, J., and ROSEN, C. (2009) "Measuring Risk Attitudes Controlling for Personality Traits" University of Arkansas working paper.
- DENANT-BOEMONT, L., MASCLET, D., and NOUSSAIR, C. (2007) "Punishment, Counterpunishment and Sanction Enforcement in a Social Dilemma Experiment," *Economic Theory*, 33: 145–167.
- DUERSCH, P. and SERVÁTKA, M. (2007) "Risky Punishment and Reward in the Prisoner's Dilemma," University of Heidelberg working paper.
- DUFWENBERG, M. and KIRCHSTEIGER, G. (2004) "A Theory of Sequential Reciprocity," *Games and Economic Behavior*, 47, 268-298.

- ENGELMANN, D. and NIKIFORAKIS, N. (2008) "Feuds in the Laboratory? A Social Dilemma Experiment," Dept. Economics, University of Melbourne, Research Paper 1058.
- ECKEL, C. C., and GROSSMAN, P.J. (2008) "Forecasting Risk Attitudes: An Experimental Study Using Actual and Forecast Gamble Choices," *Journal of Economic Behavior and Organization*, 68(1), 1-17.
- FALK, A. and FISCHBACHER, U. (2006) "A Theory of Reciprocity," *Games and Economic Behavior*, 54, 293-315.
- FALK, A. and KOSFELD, M. (2006) "The Hidden Costs of Control," *American Economic Review*, 96, 1611-1630.
- FEHR, E. and GÄCHTER, S. (2000a) "Cooperation and Punishment in Public Goods Experiments," *The American Economic Review*, 90, 980-994.
- FEHR, E. and GÄCHTER, S. (2000b) "Fairness and Retaliation: The Economics of Reciprocity," *Journal of Economic Perspectives*, 14, 159-181.
- FEHR, E. and LIST, J. A. (2004) "The Hidden Costs and Returns of Incentives – Trust and Trustworthiness among CEOs," *Journal of the European Economic Association*, 2(5), 743-771.
- FEHR, E. and SCHMIDT, K. (1999) "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics*, 114, 817-868.
- FRANK, R., GILOVICH, T. and REGAN, D. (1993) "Does Studying Economics Inhibit Cooperation?" *Journal of Economic Perspectives*, 7(2), 159-171.
- FREY, B.S. and JEGEN, R. (2001) "Motivation Crowding Theory," *Journal of Economic Surveys*, 15, 589-611.
- GANGADHARAN, L. and NIKIFORAKIS, N. (2009) "Does the Size of the Action Set Matter for Cooperation?" *Economics Letters*, 104, 115-117.
- HARRISON, G. (1986) "An Experimental Test for Risk Aversion," *Economics Letters*, 21: 7-11.
- HOLT, C. A. and LAURY, S. K. (2002) "Risk Aversion and Incentive Effects," *The American Economic Review*, 92, 1644-1655.
- ISAAC, R. M. and JAMES, D. (2000) "Just Who Are You Calling Risk Averse," *Journal of Risk and Uncertainty*, 20(2), 177-87.
- MARWELL, G. and AMES, R. (1981) "Economists Free Ride, Does Anyone Else? Experiments on the Provision of Public Goods," *Journal of Public Economics*, 15(3), 295-310.

- MASCLET, D., NOUSSAIR, C, TUCKER, S., and VILLEVAL, M.C., (2003): "Monetary and Non-monetary Punishment in the Voluntary Contributions Mechanism," *American Economic Review*, 93, 366-380.
- NIKIFORAKIS, N. (2008) "Punishment and Counter-Punishment in Public Good Games: Can We Really Govern Ourselves?" *Journal of Public Economics*, 92, 91-112.
- NIKIFORAKIS, N. and NORMAN, H.-T. (2008) "A Comparative Statics Analysis of Punishment in Public-Good Experiments," *Experimental Economics*, 11, 358-369.
- NOUSSAIR, C. and TUCKER, S. (2005) "Combining Monetary and Social Sanctions to Promote Cooperation," *Economic Inquiry*, Vol. 43, 649-660.
- ROTH, A. (1988) "Laboratory Experimentation in Economics: A Methodological Overview," *The Economic Journal*, 98, 974-1031.
- RUBINSTEIN, A. (2006) "A Sceptic's Comment on Studying Economics," *Economic Journal*, 116, c1-c9.

## Appendix 1.

**Table 5: Subjects' Punishment Decisions Based on the History of Play**

	<b>Own Behavior</b>	<b>Other Person's Behavior</b>	<b>Punishment</b>
<b>Certain Punishment</b>	Cooperate	Cooperate Defect	1
	Defect	Cooperate Defect	
<b>Uncertain Punishment</b>	Cooperate	Cooperate Defect	1 4
	Defect	Cooperate Defect	1 2

## Appendix 2.

**Table 6: Tobit Regression Estimates for Destroy**

<i>Destroy<sub>t</sub></i>	<b>Coefficient</b>	<b>St. Error</b>	<i>t</i>	<i>p</i> >   <i>t</i>
<b>Own PD</b>	7.28	1.77	4.11	0.000
<b>Other PD</b>	-6.30	1.95	-3.23	0.001
<b>Risk aversion</b>	0.25	1.89	0.13	0.896
<b>Inconsistent</b>	1.78	3.70	0.48	0.630
<b>Confused</b>	4.74	5.33	0.89	0.375
<b>Age</b>	-0.22	0.21	-1.02	0.310
<b>Male</b>	2.38	1.65	1.44	0.152
<b>Baseline</b>	3.44	2.17	1.59	0.115
<b>Uncertain Punishment</b>	2.94	1.97	1.49	0.137
<b>Constant</b>	-5.82	5.78	-1.01	0.315

Number of observations = 182, left-censored at 0.

## Appendix 3. UNCERTAIN PUNISHMENT INSTRUCTIONS

(Translation from German, for column players. The decision forms were printed on separate sheets. The original instructions are available from the authors upon request.)

### *GENERAL INSTRUCTIONS*

**No talking:** Now that the experiment has begun, we ask that you do not talk or communicate any longer with each other. If you have a question after we finish reading the instructions, please raise your hand and the experimenter will approach you and answer your question in private.

**Monetary payments:** The experiment will consist of two stages and will be followed by a separate decision problem for which you will get paid as well. The amount of money you make will depend on the choices made (as described below). Each participant will receive a lump sum payment of **5 Euro**. This one-off payment can be used to pay for eventual losses. Your earnings will be paid to you in cash individually and privately at the end of the session.

**Matching:** During the session you will be matched with another person. However, no participant will ever know the identity of the person he or she is matched with.

**Roles:** In the experiment a "row" player is always randomly matched with a "column" player. You are the **column player**.

### *INSTRUCTIONS – STAGE 1*

**Your decision in Stage 1:** On the DECISION FORM 1 you will see a payoff table. The row player decides between Top and Bottom rows, and the column player decides between the Left and Right columns. The intersection of the designated row and column determines which part of the payoff matrix is relevant (Top Left, Top Right, Bottom Left, Bottom Right) and thus determines the earnings for each person. In each cell, the row player's payoff is shown first and the column player's payoff is shown second. Your payoff is printed in bold.

After you have made the decision, we will collect the decision forms and inform you about the decision of your matched row player.

### **DECISION FORM – STAGE 1**

	<b>Left</b>	<b>Right</b>
Top	\$5, <b>\$5</b>	\$0, <b>\$8</b>
Bottom	\$8, <b>\$0</b>	\$2, <b>\$2</b>

Please **circle** either the **Left** or the **Right** column.

## ***INSTRUCTIONS – STAGE 2***

**Your Decision in Stage 2:** After learning the other player's decision in Stage 1, you can **decrease** the other player's payoff with 50% probability by 1 Euro and with 50% probability by 3 Euro at the cost of 1 Euro to you. We will flip a coin in front of you to determine the outcome. If the *heads* comes up, the other player's payoff will decrease by 1 Euro. If the *tails* comes up, the other player's payoff will decrease by 3 Euro.

If you decide to **decrease** the other player's payoff, you will **circle** the words "**I want to decrease the other player's payoff.**"

If you decide to **not change** the other player's payoff, you will **circle** the words "**I do not want to change the other player's payoff.**"

The other player can also decrease your payoff or leave it unchanged.

## **DECISION FORM – STAGE 2**

Do you want to decrease the other player's payoff by 1 Euro with probability 50% and by 3 Euros with probability 50% at the cost of 1 Euro to you? Please circle.

**I want to decrease the other player's payoff.**

OR

**I do not want to change the other player's payoff.**

## RISK ATTITUDE ELICITATION

The next page shows ten decision questions. Each decision is a paired choice between "Option A" and "Option B."

You will make ten choices and record these in the box to the left of the option. That is, if you prefer option A to option B, you will mark an X by option A. Only one of the ten decisions will be used in the end to determine your earnings for this part of the experiment.

A ten-sided die will be used to determine payoffs; the faces are numbered from 1 to 10 (the "0" face of the die will serve as 10.) After you have made all of your choices, you will throw this die twice, once to select one of the ten decisions to be used, and a second time to determine what your payoff is for the option you chose, A or B, for the particular decision selected. Even though you will make ten decisions, only one of these will end up affecting your earnings, but you will not know in advance which decision will be used. Obviously, each decision has an equal chance of being used in the end.

Now, please look at Decision 1 at the top. Option A pays \$2.00 if the throw of the ten sided die is 1, and it pays \$1.60 if the throw is 2-10. Option B yields \$3.85 if the throw of the die is 1, and it pays \$0.10 if the throw is 2-10. The other Decisions are similar, except that as you move down the table, the chances of the higher payoff for each option increase. In fact, for Decision 10 in the bottom row, the die will not be needed since each option pays the highest payoff for sure, so your choice here is between \$2.00 Euro or \$3.85.

To summarize, you will make ten choices: for each decision row you will have to choose between Option A and Option B. You may choose A for some decision rows and B for other rows, and you may change your decisions and make them in any order.

When you are finished, we will collect your decision sheet. Again, two persons from the class will be randomly selected to receive the monetary payoffs. To determine the payoffs from this task you will throw the ten-sided die to select which of the ten Decisions will be used. Then you will throw the die again to determine the money earnings for the Option you chose for that Decision. If you are selected, earnings (in \$) for this choice will be paid to you in cash when we finish.

So now please look at the empty boxes on the record sheet. You will have to mark an X in one and only one of the boxes in each row, depending whether you prefer option A or option B. Then the die throw will determine which of the ten decisions is going to count. We will look at the decision that you made for the one that counts, and circle it, before throwing the die again to determine your earnings for this part.



## DECISION FORM

### Option A

2.00€ with probability of 1/10,  
1.60€ with probability of 9/10 *OR*

---

2.00€ with probability of 2/10, 1.60€  
with probability of 8/10 *OR*

---

2.00€ with probability of 3/10,  
1.60€ with probability of 7/10 *OR*

---

2.00€ with probability of 4/10,  
1.60€ with probability of 6/10 *OR*

---

2.00€ with probability of 5/10,  
1.60€ with probability of 5/10 *OR*

---

2.00€ with probability of 6/10,  
1.60€ with probability of 4/10 *OR*

---

2.00€ with probability of 7/10,  
1.60€ with probability of 3/10 *OR*

---

2.00€ with probability of 8/10,  
1.60€ with probability of 2/10 *OR*

---

2.00€ with probability of 9/10,  
1.60€ with probability of 1/10 *OR*

---

2.00€ with probability of 10/10,  
1.60€ with probability of 0/10 *OR*

### Option B

3.85€ with probability of 1/10,  
0.10€ with probability of 9/10

---

3.85€ with probability of 2/10,  
0.10€ with probability of 8/10

---

3.85€ with probability of 3/10,  
0.10€ with probability of 7/10

---

3.85€ with probability of 4/10,  
0.10€ with probability of 6/10

---

3.85€ with probability of 5/10,  
0.10€ with probability of 5/10

---

3.85€ with probability of 6/10,  
0.10€ with probability of 4/10

---

3.85€ with probability of 7/10,  
0.10€ with probability of 3/10

---

3.85€ with probability of 8/10,  
0.10€ with probability of 2/10

---

3.85€ with probability of 9/10,  
0.10€ with probability of 1/10

---

3.85€ with probability of 10/10,  
0.10€ with probability of 0/10

## QUESTIONNAIRE

Thank you for participating in the experiment. Finally, please answer the following questions. Your answers will have no impact on your final payoff.

1. If you could decide how much to destroy of the other player's payoff how much would you like to destroy?
2. How much of your own payoff would you be willing to pay for it?
3. How old are you?
4. What is your gender?
5. What is your major?
6. In which country were you born/raised?