

Intentions Matter: Avenging Second- and Third-Party Punishment

MAREK MIHOK*

NYU Abu Dhabi, Class of 2015

marek.mihok@nyu.edu

Abstract

A number of previous experimental studies have shown that punished individuals are willing to counter-punish at a cost, even in one-shot interactions. Counter-punishers target both, second- party punishers (i.e. individuals directly affected by the violation of a social norm) as well as third- party punishers (i.e. individuals not directly affected by the violation of a social norm). Using a modified dictator game I show that the extent of counter-punishment on second parties is significantly greater than on third parties. I explore different possible explanations for the willingness to counter-punish. I find some evidence suggesting that counter-punishment is meted out by the most anti-social of individuals who are also willing to take the most from the second party participants

Keywords: social norm, punishment, counter-punishment, second party, third party, Big Five Markers

I. INTRODUCTION

THE concept of social norms is crucial to a full understanding of human societies (e.g., Berkowitz, 1972; Pepitone, 1976; Triandis, 1977). Social norms have been found to have a strong and regular impact on our behavior (Cialdini, Kallgren, and Reno 1991). Examples of social norms range from norms of reciprocity, which enjoin us to return favors done to us by others, and norms of distribution pertinent to a fair allocation of endowment, to norms of cooperation ordering us to put the greater benefits of the group before our individual gains. However, none of these norms would work without a broad consensus among the agents constituting the society. Social norms are based on a common belief which specifies actions that are regarded as proper or correct, or improper or incorrect (Coleman, 1990). As such, the existence of social norms is dependent on collective cooperation, sustained by an effective enforcement mechanism. But how are these norms maintained and who is the one enforcing them?

Punishment is considered essential for upholding social norms and cooperation, yet the understanding of punishment behavior of different parties and their motivation is still rather limited. Material punishment has been found to increase cooperation in behavioral experiments where subjects can communicate (e.g. Ostrom

et al. 1992), whereas the absence of punishment opportunities has typically resulted in the breakdown of cooperation (Yamagishi, 1986). Second parties, who are directly affected by the norm violation, or third parties, can mete out punishment. Third parties, as opposed to second parties, are not directly involved in the norm violation, but rather they observe interaction between the norm perpetrator and the victim.¹ While second-party punishment has been studied extensively in a range of social experiments using the public goods game (e.g. Fehr and Gächter 2000a), the ultimatum game (e.g., Güth et al., 1982), and the gift-exchange game (Fehr, Gächter, and Riedl 1998) third-party punishment has only recently attracted the attention of social scientists.

As argued in the seminal work of Fehr and Fischbacher (2004), third-party punishment is necessary for social stability as it extends to norm violations beyond the ones with easily identified victims. Second-party punishment is limiting when it comes to imposing sanctions, because the punishment can be imposed only if the violation directly affects someone. In many cases requiring cooperation, i.e., contributing to a public account, a single shirking individual imposes almost no harm on other individuals (Dietz, Ostrom, and Stern 2003). The enforcement of a social norm by a third party enhances the extent to which violating individuals can be held accountable and directly contributes to the enforcement of social norms.

*I would like to thank Professors Nikos Nikiforakis, Max Mihm, Jonathan Rogers, and Arus Haknazaryan as well as Tushi Baul for continued support and help.

¹An example of a second-party punishment would be a direct confrontation of a line jumper by a person standing in the line who is affected by the line-jumper. A third-party punisher would be someone who witnesses the norm violation and is not affected by the line-jumper, but still reprimands him for his actions.

An important insight from the comparison of second- and third-party punishment is that the severity of second-party punishment exceeds that of third-party punishment (Fehr and Fischbacher, 2004). Different motives of the punishing parties have been used to explain the difference in punishment. While second-party punishment can often be rationalized by self-interested reciprocal behavior (Rabin 1993; Fehr and Schimdt 1999; Cox et al. 2007), a third party is thought of as selfless. Third-party punishment has been hypothesized to be driven by a desire to uphold social behavior (Fehr and Fischbacher 2004). These individuals are the ones who do not expect personal gain from imposing sanctions other, perhaps, than the pleasure of standing up for what is right.

Differences in motivation behind second and third party punishment could have an effect on how norm violators retaliate to punishment. Levine (1998) assumes that the type of the person, who can be either spiteful or altruistic, influences the punishment levels, but there is no evidence on how, or if at all, the punished individuals are affected by different intentions of their punishers. In many practical interactions, an agent can take revenge for imposed punishment, an act of retaliation defined as counter-punishment. Agents wishing to impose sanctions on their punishers have been shown to be willing to incur cost to revenge both in a case when the punishing individual is a second party (Nikiforakis 2008) as well as when the punishing individual is a third party (Balafoutas, Grechenig, and Nikiforakis 2014a). Counter-punishment opportunities affect the efficiency of punishment and can lead to breakdown of cooperation as punishers impose lower punishment to violators under the fear of retaliation (Nikiforakis 2008). It is for this reason that counter-punishment needs to be included in the analysis of punishment, as its omission can lead to misguided assumptions about the true levels of cooperation achieved by sanctioning institutions.

As of now, there is no research that would directly compare the extent of counter-punishment to second- and third-party punishment using the same experimental setup. Therefore, this paper aims to provide the answer to the question of how, or if, the retaliation to punishment is affected by the type of the enforcing party. The answer to this research question is interesting for several reasons. First, it contributes to our understanding of the underlying altruistic enforcement mechanisms that govern social norms. Altruistic punishment of third parties is still poorly understood, but it has been shown to be an important mechanism for enforcing cooperation (Fehr and Fischbacher, 2004). Second, the answer can give us a rationale for the emergence of central (impartial) authorities in charge of enforcing so-

cial norms and cooperation. The supporting data in that case would find that punished individuals retaliate for punishment by third parties to a lesser degree than punishment by second parties. Lower counter-punishment would mean higher sustained levels of cooperation.

I compare the extent to which second and third parties are counter-punished in the context of a violation of a distribution norm using an adaptation of a game introduced by Fehr and Fischbacher (2004). In the game, a first participant, the norm perpetrator, indicates how much of another participant's endowment she wishes to appropriate. Then, in the first treatment, the victim (i.e., the second party) can decide to punish the perpetrator by reducing her income. In the last stage, the punished individual (i.e., the perpetrator) can counter-punish the victim using the same punishment mechanism. The second treatment has only one difference — an independent observer (i.e., the third party) can punish the perpetrator instead of the victim, who can do nothing. Similarly, in the last stage, the punished individual must decide whether to retaliate to punishment from the independent observer. By comparing the extent of counter-punishment in the two treatments, I can evaluate whether third or second parties are more likely or equally likely to be counter-punished for enforcing a social norm.

The results from the experiment indicate that the severity of second-party counter-punishment is higher than that of third-party counter-punishment. Punished individuals retaliate more severely to the punishment coming from a second party than to the same punishment coming from a third party. The severity of counter-punishment is 37 percent greater when the punishing party is a second party rather than third party. Similarly, the likelihood that a punished individual will counter-punish is higher for second-party punishers than third-party punishers. Under second-party punishment, punished individuals retaliate in 54.5 percent of the cases, whereas in third-party punishment, punished individuals retaliate only in 32.8 percent of the cases. A closer examination of punishment strategies show that punished individuals are more likely to retaliate for high levels of punishment with significant differences to punishment levels greater than 50 percent of the punished individual's endowment.

Efforts to further understand the counter-punishing behavior by including the Big-Five Markers, a widely used psychological test for assessing humans' personality, prove to be futile. The test comes short of explaining the counter-punishing behavior, as I find no correlation between any of the personality traits and the counter-punishing behavior. Despite this, I find some evidence that counter-punishment is meted out by the most anti-

social of the parties, as evidenced by the fact that the strongest violators of social norms are also the strongest counter-punishers.

The rest of the paper is organized as follows: In the next section I provide a brief overview of the most relevant research to date. In Section 3 the experimental procedure and individual treatments are specified. Section 4 the major results are presented, followed by their discussion and concluding remarks in Section 5.

II. LITERATURE REVIEW

There is vast research on punishment in various social games, but the concept of counter-punishment is still largely understudied. My research builds mainly on three studies: Fehr and Fischbacher (2004); Nikiforakis (2008) and Balafoutas, Grechenig, and Nikiforakis (2014).

Fehr and Fischbacher (2004) introduced third-party sanctions of norm violations into the literature. They studied norm violations in the context of two games and in the presence of second- and third-party punishment. The novel approach consists of adding a third player with a punishing option to the traditional set up of a dictator as well as prisoner's dilemma game. The third player observes the interaction between Player A and Player B, after which he has the option of assigning deduction points to either the dictator (dictator game), or both players (prisoner's dilemma). The reported results indicate that third parties punish violation of the distribution as well as cooperation norms, even when the punishment is costly. The comparison of second- and third-party sanctions further reveals that second-party punishment is significantly higher than third-party punishment. However, the study does not include counter-punishing opportunities, which as Nikiforakis (2008) shows are able to destabilize the cooperation among agents.

The work by Nikiforakis (2008) examines the effect of the existence of counter-punishment opportunities on cooperation in public goods game. The experimental design consists of three treatments — no punishment, punishment, and counter-punishment. In the final treatment participants can assign deduction points to the individuals who punished them during the second stage.² By comparing the results across the treatments, Nikiforakis finds evidence that punished individuals are willing to counter-punish even in one-

shot interactions with the likelihood and severity of counter-punishment being positively correlated to the intensity of preceding punishment. Moreover, in the threat of counter-punishment, individuals punish free riders less due to the fear of retaliation. This leads to the breakdown of cooperation as free riders take advantage of lower punishment levels, and contribution to the public good decreases over time. Yet, the study provides evidence only about the existence of second-party counter-punishment and does not address the issue of retaliation to impartial third-party punishers.

Extending the finding that the threat of counter-punishment shows a significant obstacle to spontaneous cooperation, Balafoutas, Grechenig, and Nikiforakis (2014) use a modified dictator game to answer the question of the willingness to counter-punish impartial third parties. In the game the dictator (i.e., the social norm perpetrator), has to decide whether to take a portion of the receiver's income. Since both players have different endowments in the beginning, the taking option further increases inequality in favor of the dictator. As such, it constitutes a strong violation of distribution, equality norms, and property rights. The second player simply receives the share, and his role is entirely passive.³ A third player (the third party) is given the opportunity to punish the dictator by sacrificing some of their income. The counter-punishment treatment offers the dictator the opportunity to retaliate by removing a lottery ticket from player C, conditional on whether he decided to punish him in the second stage or not. The experiment indicates that 35 percent of individuals are willing to counter-punish even though the interaction is one-shot and that the counter-punishment is severe. Despite proving that counter-punishment opportunities affect cooperation also when the punishing party is unaffected by the norm violation, the results do not allow for a comparison of counter-punishment to second-party punishment (Nikiforakis 2008) as the experimental setup is different.

In summary, these studies show that third parties are important for the enforcement of social norms and that punishment decisions are strongly affected by the presence of counter-punishing opportunities. The findings also indicate that the fear of retaliation to punishment can deter punishment. This paper aims to contribute to this literature by studying how people retaliate to punishment from a self-interested individual (i.e. second party) or an independent observer (i.e.

²In the second stage all participants can punish members of their group, after they have made their contribution to the public account. They are subsequently informed of how much other participants deducted from their account, upon which the decision to counter-punish is presented to them.

³There is vast literature on the transfer levels of Player A in the traditional dictator game. Andreoni and Miller (2002) report that approximately 40 percent of players choose to transfer 0 percent, 20 percent of players gave between 0 and 50 percent, and that around 40 percent gave exactly 50 percent. There are almost no transfers above 50 percent.

third party).

III. EXPERIMENTAL DESIGN

I. Experimental Treatments

The experimental design is based on that of Balafoutas, Grechenig, and Nikiforakis (2014a) and Fehr and Fischbacher (2004). More specifically, I adopt a modified dictator game, as per the design of Balafoutas et al. (2014a). The experiment consists of two treatments between which the only difference is the treatment variable — the counter-punishing party is either an individual directly involved in the social norm violation (2PCP) or an independent observer (3PCP). In its design the game allows for a direct comparison of counter-punishment in second-party treatment (2PCP) and third-party treatment (3PCP).

I.1 Second-party counter-punishment

In the 2PCP treatment, participants are randomly assigned the role of Participant A or B and then paired into a group with a participant of another type. All participants are given the same endowment of 30 EMU⁴ after which they play the following three-stage game once. In the first stage, those who have been assigned the role of Participant A have to decide whether to take 0, 5, or 10 EMU from the individual with whom they have been paired with. Participants B have no decision to make at this stage.

In the second stage, individuals who have been assigned the role of Participant B (the second party) have to decide whether they wish to reduce the earnings of Participant A in their group by a certain amount.⁵ Before they make their decision, all participants B are awarded additional 5 EMU which they can use to reduce Participant A's earnings by any positive percentage in 10 percent increments. The cost of reducing income by any amount greater than zero is 5 EMU, irrespective of the amount by which Participant B reduces Participant A's income. However, if Participant B decides not to reduce Participant A's income, then the 5 EMU are added to his or her final earnings. The rationale behind this mechanism is explained in the next section.

In the third and final stage of the experiment those who have been assigned the role of Participant A are asked whether they wish to reduce the earnings of Participants B in their group. Participants A can retaliate the punishment only of the same individual as the one

who had the opportunity to punish them in the second stage. Income reduction in this stage works exactly as in the second stage. Participants A are awarded additional 5 EMU, which they can use to reduce Participant B's earnings by any positive amount.

I.2 Third-party counter-punishment

The first stage of the second treatment (3PCP) is exactly the same as in the previous treatment. Again, groups of two are formed and the initial roles are retained. However, this time each participant is paired with a new participant. Participant A, who had the option of reducing Participant's B earnings, has again the option to reduce *new* Participant B's earnings. Participant A has to decide whether to take 0, 5, or 10 EMU from the *new* Participant B's initial endowment of 30 EMU. Similarly, Participant C of *another group* can take 0, 5, or 10 EMU from his Participant D. The reduction mechanism is pictured in Figure 1.

Second and third stages of third-party counter-punishment treatment have one change. While in the first treatment (2PCP) Participant B can punish his perpetrator, in 3PCP she can only punish the perpetrator of a different group (Participant C). This model of third-party sanctioning is adopted from Fehr and Fischbacher (2004) and allows for examining the relative strength of third-party counter-punishment holding all else equal. Since Participant C is in a different group, the decision of C, who can appropriate some of Participant's D endowment, has no implication on Participant B's earnings. This design also prevents behavior that follows the principle of "returning favors" — no punishers are strategically involved with other punishers (i.e., the decision of one punisher cannot affect directly the earnings of another punisher). Therefore we can consider Participant B as an unaffected third party with respect to participants C and D.

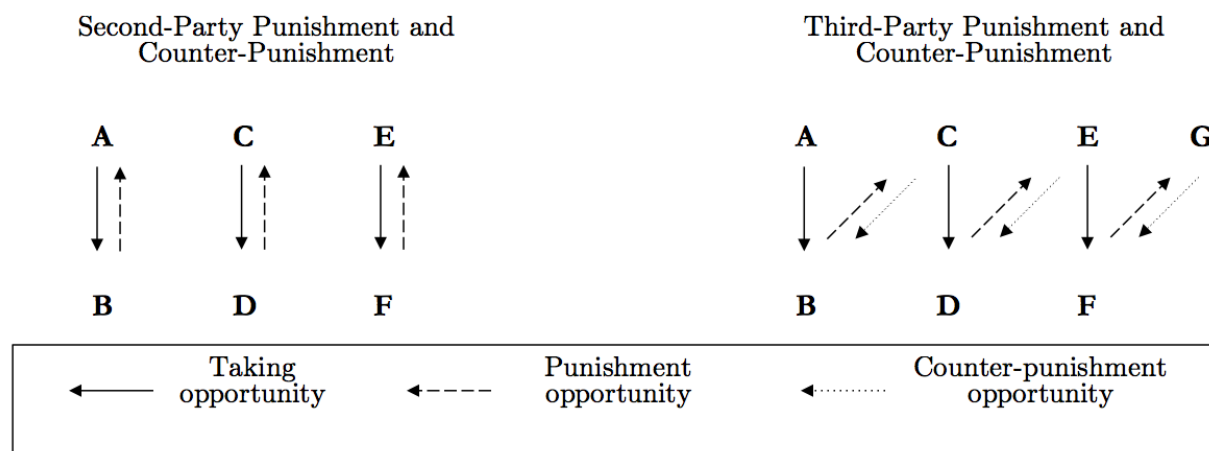
I.3 Big Five Markers test

To better understand what may drive counter-punishment, all participants were asked to complete a standard Big Five Markers test after the completion of the two treatments. Big Five Markers test, developed by Goldberg (1992) is widely regarded as a good measure of human personality and has been used in various types of research previously. It was used to link personality with job performance (Barrick and Mount 1991), with counterproductive behavior (Salgado 2002) and with entrepreneurial status (Zhao and Seibert 2006) just

⁴EMU stands for Experimental Monetary Units and is used as the currency in the experiment: 1 EMU = 2 AED.

⁵The terms: "sanction," "punishment," "revenge," and "counter-punishment" were not used in the experimental instructions. Instead neutral language was used, participants could "deduct" or "reduce" certain percentages of the respective incomes.

Figure 1: *Who can punish and counter-punish whom in the modified dictator game*



to name a few. The survey consists of fifty questions in the form of short statements. The respondent can indicate how much she or he agrees with each statement on a scale from 1-5. The scores are then presented in five broad domains of human personality: openness, which measures imagination, curiosity, and appreciation for adventure; conscientiousness, which measures tendency for self-discipline and aim for achievement; extraversion, measuring the tendency to seek stimulation and company of others; agreeableness, measuring the tendency to be compassionate and cooperative rather than suspicious; and lastly neuroticism, which measures the tendency to easily experience unpleasant emotions like anxiety, depression, or anger.⁶

In addition to the test, all participants were asked to fill-out a short questionnaire with short questions regarding their perception of normative behavior. After the completion of both tests, participants were anonymously paid all the proceedings from the experiment plus the show-up fee.

II. Subject Pool and Procedures

The experiment was conducted in the Social Science Experimental Laboratory at New York University Abu Dhabi in February 2015. All 96 participants were students from the New York University in Abu Dhabi. The experiment was organized and recruited with the software hroot (Bock, Nicklish, and Baetge 2012) and based on the computer software z-Tree (Fischbacher, 1999). Each participant took part in one treatment only, whose

type was predetermined before the start of the experiment. In total 5 complete sessions were conducted with 20 participants in each of the first 4 sessions and 16 in the final session. The experiment lasted roughly 70 minutes and participants earned on average 103 AED (about 28 USD). Each participant was paid a show up fee of 30 AED (about 8 USD). Participants never knew the identities of those with whom they interacted, and they had full knowledge of the experimental instructions.

Upon arrival at the laboratory, participants were seated in partitioned computer terminals and were given a set of instructions specific to their role in the experiment. Each participant had to answer a set of 6 control questions before the commencement of the experiment. The experimenter assured understanding of the experiment by individually checking the answers to all questions. After the completion of the first treatment, another set of instructions was distributed to the participants detailing the procedure of the second treatment. At the end of the second treatment, participants were given a version of the Big-Five Factor Markers, available free to use from International Personality Item Pool website.⁷ Lastly, all participants were asked to complete a short questionnaire created to develop a better understanding of their decision-making process in the experiment.

The willingness to punish and the willingness to counter-punish are elicited using the strategy method. Punishing participants have to indicate the percentage level of punishment for each possible amount that can

⁶For a free version of the survey visit: http://www.sevencounties.org/poc/view_doc.php?type=doc&id=8948.

⁷"Possible Questionnaire Format for Administering the 50-Item Set of IPIP Big-Five Factor Markers." International Personality Item Pool. http://ipip.ori.org/New_IPIP-50-item-scale.htm

be deducted from their income. Similarly, participants in the role of counter-punishers have to state the levels of counter-punishment for all eleven possible levels of incurred punishment. Following this method, the actual choices of the participants are unknown until the final results are presented. There is a significant advantage of using this method instead of the traditional direct-response method. By eliciting responses for all decision sets, we can obtain observations for reactions to decisions, which may not be observed in the game or are observed rarely. This provides us with more robust results in the statistical analysis. Despite the advantage, the method is sometimes criticized for being "cold" as the decisions are stripped off the emotional impact when reacting to actual punishments. Nonetheless, there is substantial evidence that there is no significant difference between the strategy method and the standard direct-response method when examining the presence of a treatment effect.⁸

Another feature common in both treatments is the punishing (counter-punishing) mechanism. Previous experiments have used costly punishment as a way of approximating the real state of the world — punishments are costly not only for the victims, but also for the punishers. In the usual form, the cost of punishment increases proportionally with the severity of punishment (e.g. Fehr and Gächter 2002). The perceived costs of counter-punishing can be higher for an individual who has incurred punishment, due to the reduction in his or her earnings (Nikiforakis 2008). In order to mitigate the effects, I employ a fixed cost of punishment and counter-punishment similar to Nikiforakis, Noussair, and Wilkening (2012). The reduction of other participants' earnings has a fixed cost of 5 EMU and a zero marginal cost, irrespective of the severity of the imposed sanction. This resembles a case where the severity of the sanctions is calculated as a percentage of victim's income (e.g. Fehr and Gächter 2000b; Nikiforakis 2008). Both, the punisher and the counter-punisher have the chance to completely burn each other's endowments. As presented in the results section, there are in fact only a few participants who punish at a hundred percent; the percentage punishment of levels corresponds closely to the anticipated amounts, confirming the effectiveness of such a mechanism.

III. Hypotheses

It is not clear ex ante what the answer to my research question will be. First, I expect to confirm the data

from Nikiforakis (2008), in which demand for counter-punishment increases with increasing punishment incurred irrespective of the treatment. This is predicated on the assumption that people are generally likely to harm more those individuals who harm them.

Hypothesis 1: *The more individuals punish, the higher the counter-punishment will be for both second and third parties.*

There is no clear evidence as to why counter-punishment behavior should differ between second and third parties. On the one hand, punished individuals can perceive third-party punishers as a sort of "impartial judge," whose role is to control the conformity to social norms. If this is the case, counter-punishment of third-party punishment should be less than counter-punishment of second party. On the other hand, third parties could be perceived as "paternalistic" figures interfering in a situation that does not directly affect them. In this case, the results in 2PCP and 3PCP should indicate a greater counter-punishment on third parties than on second parties. However, it could also be the case that punished individuals are not affected by the type of the punishing party and that for a given level of punishment, the retaliation is the same to both parties. The predicted null hypothesis states:

Hypothesis 2: *There will be no differences between the counter-punishment behavior of second and third parties.*

If the above null hypothesis of no differences is true, I should still observe significant differences between punishment levels in the two treatments, given the evidence from Fehr and Fischbacher (2004). They find that second-party punishment is significantly greater than third-party punishment with regard to the same social norm violation. I hypothesize:

Hypothesis 3: *If Hypothesis 2 is true, second-party punishment will be greater than third-party punishment.*

Previous evidence did not find any differences in taking, despite the different levels of punishment (Balafoutas et al. 2014; Fehr and Fischbacher 2004). Even large differences in punishment behavior do not offset the extent of pro-social behavior in one-shot interactions (Nikiforakis and Mitchell 2014). Therefore the changing levels of norm adherence may not lead to a change in the demand for costly punishment.

⁸Brandts and Charness (2011) provide a great survey of existing literature regarding the accuracy of the strategy method compared to the traditional direct-response method. Of the twenty-nine existing comparisons, sixteen find no difference, and nine comparisons find mixed evidence, while only four find differences between the strategy method and the direct-response method.

Hypothesis 4: *There will not be significant differences in the norm violations across treatments.*

IV. RESULTS

I. Counter-Punishment Behavior

The presented results are based on the data from all five completed treatments of the experiment. Figure 2 plots the severity of counter-punishment expressed in percentage of income against the level of punishment also expressed in percentage of income. The amounts are aggregates across individuals for each level of punishment. That is, both the likelihood of punishment happening and the severity of punishment are taken into account. What is immediately noticeable is the trend of counter-punishment, which increases in magnitude with rising levels of punishment for both second and third parties. Punished individuals counter-punish more as punishment increases, which goes in line with the expected outcome from Hypothesis 1.

Figure 3 indicates the likelihood of counter-punishment imposed on either second- or third-party punishers. Counter-punishment is a binary variable attaining value 1 for any levels of counter-punishment greater than zero. We can see that for almost all levels of punishment, the second party is much more likely to be counter-punished. This holds true especially for the levels of punishment greater than 50 percent, where the difference ranges between 32 and 40 percentage points. A Two-sample Wilcoxon rank-sum test ($z=1.510$, $p\text{-value}=.13$) reports an insignificant difference between the two treatments. Second-party counter-punishment is, however, weakly significant for values of punishment greater than 50 percent ($z=1.892$, $p=0.0584$). Overall, punished individuals retaliate in 54.5 percent of the cases in 2PCP, whereas in 3PCP, punished individuals retaliate only in 32.8 percent of the cases.

The pairwise comparison is based on each participant representing an independent observation. The levels of counter-punishment are taken as a total average of all participants for a given level of punishment. The reason for the insignificant result is most likely due to a small number of independent observations: 20 in 2PCP treatment and 28 in 3PCP treatment. Yet, the parametric tests provide evidence of significant differences, thus rejecting Hypothesis 2 of no differences.

Table 1 (see page 33 in the Appendix) presents a regression analysis of costly counter-punishment as a function of incurred punishment and amounts of endowment that perpetrators decided to take from the second party participants. The dependent variable is

the percentage level of counter-

Figure 2: *Severity of counter-punishment*

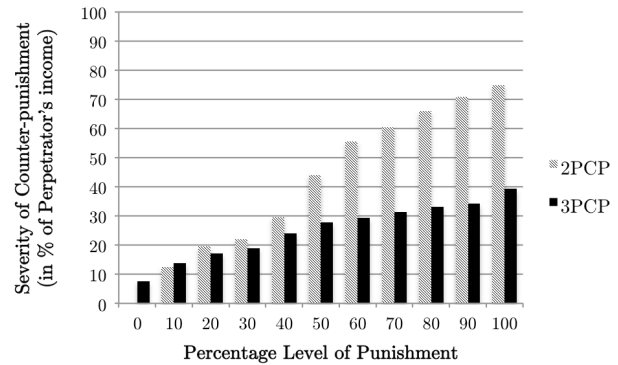
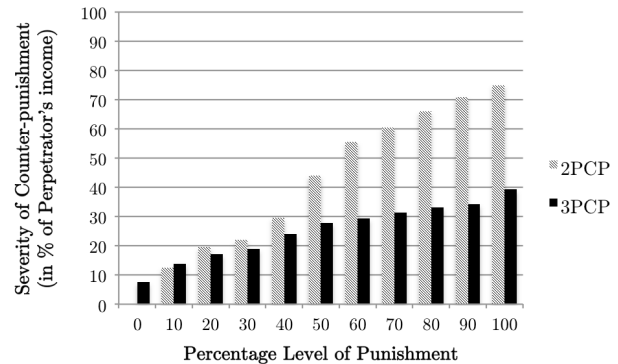


Figure 3: *Likelihood of counter-punishment*



punishment a potential perpetrator assigned to her punisher from Stage 2. The regression further includes a treatment dummy variable and two interaction terms. The treatment's dummy variable takes a value of 1 for data in 2PCP and 0 for data in 3PCP. The interaction term *Third Party*Punishment* is the product of the treatment dummy and the explanatory variable (punishment level). The second interaction term *Third Party*Taking* is the product of the treatment dummy and the amount by which the perpetrator reduced another participant's earnings in Stage 1. Given the use of strategy method, each individual makes 11 decisions. I include random effects at the individual level to account for this fact.

The punishment coefficients from first two linear regressions (1) and (2) suggest that the counter-punishment on second party (2PCP) is significantly greater than the counter-punishment on third party (3PCP) for a given level of punishment. *Punishment*

Table 2: Regression of Punishment

	2PP	3PP	Both	Punishment Level	Punishment Decision
	(1)	(2)	(3)	(4)	(5)
Taken Amount	2.650 (4.23)**	1.786 (4.05)**	2.65 (18.33)**	2.6 (2.77)**	1.177 (2.50)*
Third Party			9.560 (1.17)	7.357 (0.57)	4.601 (1.26)
Third Party * Taken Amount			-0.864 (-1.16)**	-0.6 (-0.49)	-.456 (-1.19)
Constant	0.083 (0.02)	9.643 (1.65)	0.083 (0.01)	.5 (0.05)	-10.282 (-2.36)*
Observations	60	84	144	96	144
Subjects	20	28	48	48	48

(1)–(3) are linear regressions; (4) is a probit regression and (5) is a linear regression with dependent variable $\text{pun} > 0$. All regressions include individual random effects. The regressions include all participants' decisions. * $p < 0.05$; ** $p < 0.01$.

coefficient estimates from (1) and (2) are significant and positive, indicating that an increase in the amount of punishment increases the counter-punishment both in second and third parties, confirming the first Hypothesis. Linear regression in (3) quantifies the extent of the increase in the dependent variable as a result of the treatment effect.

The significant, negative coefficient *Third Party*Punishment* indicates that the slope of 3PCP is -.484 smaller than the slope of 2PCP. The coefficients provide evidence in contrary to the null hypothesis that there are no significant differences between second and third party counter-punishment. Columns (4) and (5) in Table 1 answer the question of what drives counter-punishment.

A probit model captures the likelihood that a participant will counter-punish by a positive amount (4). The interaction term indicates a higher likelihood of retaliating for punishment toward a second party. Reported marginal effects for the probit regression find that the slope of second-party counter-punishment is 37.82 percentage points ($z=7.46$, $p=0.000$), while the slope for third party counter-punishment is 22.95 percentage points ($z=-3.52$, $p=0.000$). Reported difference in the intersects is 9.1 percent ($z=5.53$, $p=0.000$), which represents the decrease in the likelihood of counter-punishment for the treatment where Third party = 0. The likelihood of counter-punishment is modeled separately from the magnitude of counter-punishment conditional on the counter-punishment occurring (Nikiforakis 2008, Nikiforakis and Mitchell 2014). While the increase in punishment level drives more counter-punishment (5), the probability of counter-punishment is significantly greater in 2PCP. Therefore I can con-

clude:

Result 1: *Controlling for the punishment levels, the retaliation of second-party punishment is 37.5 percent higher than the retaliation of third-party punishment. This difference is significant and is driven by the propensity to retaliate more against a second than against third parties.*

The last three columns in Table 1 include the amount that a counter-punishing individual took from his partner in Stage 1 as an independent variable. We could not infer from previous data whether there is a direct relationship between counter-punishment and norm violation (taking from someone else's endowment). Economic models of social preferences predict that punishment should increase as norm violation increases (e.g. Bolton and Ockenfels 2000). At the same time, counter-punishment increases with the severity of punishment (Nikiforakis 2008). Therefore it is unclear whether the reason for any correlation between counter-punishment and norm violation is punishment or whether there is a "deeper" link between the two. Some evidence suggests that people believe strong norm violators are more likely to counter-punish (Balafoutas and Nikiforakis 2012), but this relationship has not yet been confirmed. The strategy method allows me to examine the link between the will to counter-punish and to violate the norm.

Whereas the amount taken has no effects on the level of counter punishment as noted by the insignificant coefficient in (6), the amount taken has a significant effect on the decision to counter-punish. On average,

the more an individual takes in the first stage, the more he or she is willing to counter-punish as seen by the positive value on amount taken in (7). However, there is a big difference in how this amount taken affects counter-punishment when examined separately in 2PCP and 3PCP. If an individual took a positive amount from his partner (either 5 EMU or 10 EMU), his reaction to punishment by second party was much stronger and more severe than the reaction of an individual who decided against taking, as indicated by large positive coefficient on the *Amount taken* in (1). On the other hand, the negative coefficient in the interaction term *Amount taken*Third party* suggests that if a third party punished the "taking" perpetrator, the perpetrator's counter-punishment was significantly lower than if the third party punished a "non-taking" perpetrator. There appears to be evidence that the role of punisher affects greatly the decision to retaliate for punishment. In the next section I will provide support for the argument that the difference is driven by the behavioral characteristics and the perception of the legitimacy of punishment.

Result 2: *Individuals who took away more endowment from their partners in first stage are more likely to retaliate punishment. This provides evidence that counter-punishment is enacted by the most anti-social individuals.*

II. Punishment of Social Norm Violation

As expected in Hypothesis 1, the punishment levels increase with perceived unfairness, as demonstrated by the higher amounts taken (Figure 4). Second-party punishers never punish when the perpetrators do not take any of their endowment. Participant B punished in 45 percent of the cases in which perpetrators stole some of his or her endowment. Surprising evidence from third-party punishers suggests that third party punishers punish in 17.86 percent of the cases even when there is no violation of the social norm.⁹ In the case of social norm violation, third-party punishers punish 44.6 percent of the time, very close to the frequency of second party punishers (Figure 5).

⁹This is most likely due to the outliers in 3PCP treatment. Participant 16 and Participant 19 punished the social norm violator by 100 percent, irrespective of the norm violation. This behavior contradicts the assumptions of punishing behavior and is most likely due to misunderstanding of the experiment.

Figure 4: *Magnitude of punishment*

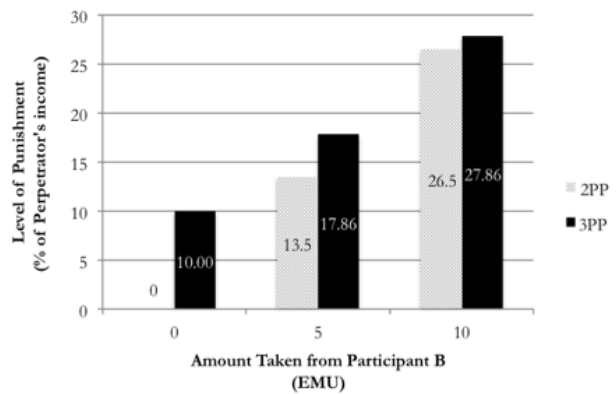
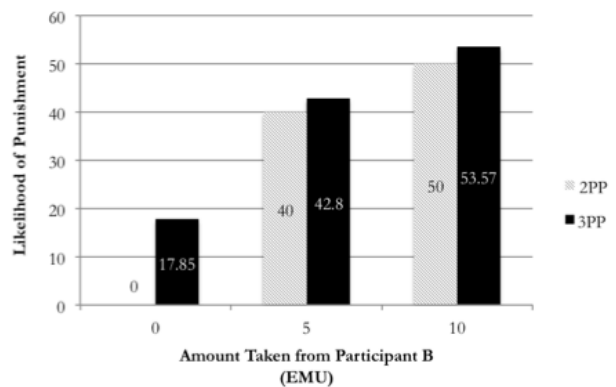


Figure 5: *Likelihood of punishment*



Based on the data from counter-punishment, one would expect to find punishment levels for the second party to be greater than punishment levels of the third party. The assumption is that punished individuals retaliate proportionally to the punishment incurred. However, the average levels and the likelihood of punishment contradict this assumption, as third-party punishment is higher for every possible reduction in Participant's B earnings. A Wilcoxon rank-sum test cannot confirm that second party punishment (2PP) is significantly different from third-party punishment (3PP) ($z=-0.366$, $p\text{-value}=0.714$). This suggests that who the punishing party is, does not have an effect on the likelihood or the level of punishment.

Table 2 documents the linear regression analysis of the relationship between the amount taken and the intensity of punishment. Treatment variable *Third party* (value 0 for 2PP and 1 for 3PP) and interaction term *Third Party*Taken Amount* are also included in the regres-

sion. The results indicate that punishment increases with amount taken; however the demand for 2PP is much steeper (1) than the demand for 3PP (2). Linear regression in (3) shows the change in the effect amount taken has on punishment across treatments. Negative coefficient of the interaction term quantifies the drop in slopes by 0.864 from 2PP to 3PP.

Similarly to the regression of counter-punishment, the likelihood of punishment (4) is modeled separately from the severity of punishment conditional on the punishment occurring (i.e., punishment is greater than zero) (5). Both the likelihood and the magnitude of punishment increase with higher amounts taken, as noted by the positive and significant coefficients. However, the interaction term again reports insignificant differences between second- and third-party punishment, thus I do not find evidence in support of Hypothesis 3.

Result 3: *Punishment increases with increasing unfairness of the norm violation, but the difference between 2PP and 3PP cannot be confirmed due to insignificant results from the parametric and non-parametric tests.*

III. Norm Violation

In the first stage of the experiment, the violation of the distribution norm is captured by the willingness to take some of Participant B's endowment. Recall that each perpetrator had the option to decrease his partner's earnings by 0, 5, or 10 EMU. Since endowments of both participants are equal, taking constitutes the violation of a distribution and equality norms and is likely to trigger punishment.

On average, in 2PCP treatment, individuals with the taking options reduced other participants' earnings 40 percent of the time. In 3PCP treatment, perpetrators took a positive amount in 46.4 percent of the cases. The average amount taken in the presence of 2PCP by Participant A is 2.75 EMU, while in the presence of 3PCP it is 4.11 EMU — Mann-Whitney rank-sum test ($z=-0.901$, $p\text{-value}=0.367$). The result from the non-parametric and Fisher's exact test ($p\text{-value}=0.207$) are not statistically significant. This may be due to the fact that the tests were run on a limited number of independent observations.

Figure 6: *Magnitude of punishment*

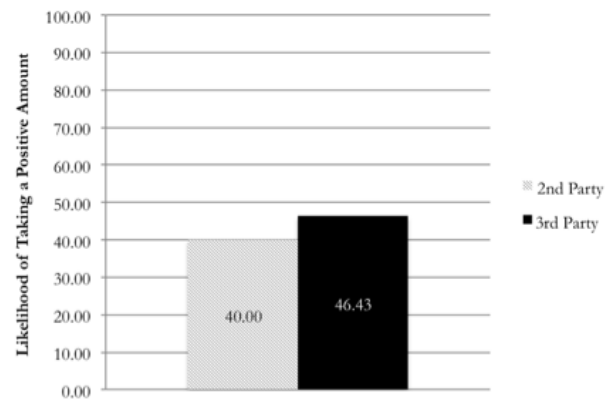


Figure 7: *Likelihood of punishment*

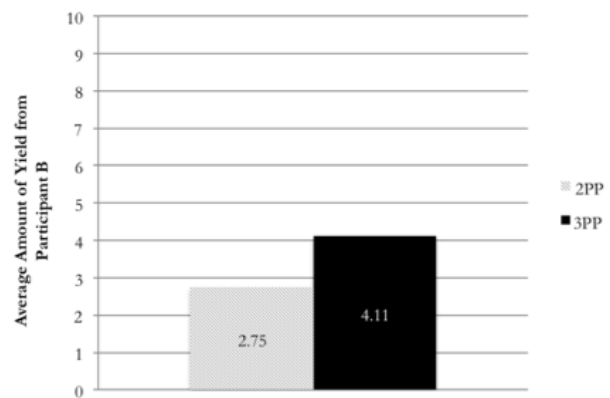


Table 3: *Expected earnings with or without norm violation*

Taken Amount Earnings Stage 1	
0 EMU 30 EMU	
5 EMU 35 EMU	
Expected Punishment Expected Earnings 2PP	
0 EMU 30 EMU	
4.73 EMU 33.11 EMU	
Expected Punishment Expected Earnings 3PP	
3 EMU 29.46 EMU	
6.25 EMU 32.32 EMU	

Table 3 indicates the expected earnings of participants A at the end of the second stage, that is, after the punishment has been imposed. Expected earnings after punishment are calculated based on the likelihood and severity of punishment for each level of amount taken,

¹⁰Expected earnings = $(30+10)\text{EMU} - .5(40 \times .265)$

¹¹Expected earnings = $40 - 0.5357(40 \times 0.2786)$

at the average level. Expected punishment is the average amount imposed by the sanctioning individuals. The results suggest that it is economically beneficial for Participant A to take the maximum amount of 10 EMU in the first stage, as their expected earnings are 34.7 EMU.¹⁰ For 3PP the maximum yield of 10 EMU would result in an expected earnings 34.03 EMU of at the end of first stage.¹¹

Result 4: *The amount taken by participants A in Stage 1 is not statistically different across treatments.*

IV. Understanding Counter-punishment

IV.1 Big Five Markers test

Despite the recent interest of social scientists in the role of third parties in the maintenance of social norms, there is still no satisfactory explanation or an economic model for the existence of counter-punishment. However, counter-punishment is observed not only in many laboratory experiments, but also in the field (Balafoutas and Nikiforakis 2012; Balafoutas, Nikiforakis, and Rockenbach 2014). I included the Big Five Markers test to better understand the motivation behind counter-punishment behavior at the end of the experiment.

Table 4 (see page 34 in the Appendix) summarizes the findings of linear regressions of counter-punishment with each of the five personality domains included in the regression separately. Each of the independent variables is coded as the total score of each participant in the respective areas as indicated by the available answer key (Goldberg 1992). None of the personality traits has a statistically significant relationship with the coefficients on punishment, and in fact the coefficients on punishment and the interaction term are not affected by any of the personality traits at all. This result is further supported by the uncorrelated distribution of personality traits to the counter-punishment behavior (see Appendix).

Using the Big Five Markers to explain punishment behavior was of no avail. Just as in the case of counter-punishment, I find no significant results (see Appendix for the full regression of punishment on Big Five). Yet, we would expect to find people who are selfish to be more likely to retaliate punishment. As mentioned earlier, the positive relationship between the norm violation and counter-punishment suggests that individuals that counter-punish are the selfish, anti-social ones.

Regression analysis of the taking behavior with respect to Big Five markers is presented in Table 5 and offers some evidence that the counter-punishing individuals are indeed the anti-social ones. Out of the five

personality traits, two show significance: *agreeableness* and *extraversion*. The more agreeable an individual is, the smaller the amount he or she takes from the other participant. Put differently, the kinder and more considerate individual violates the given social norm to a lesser degree. This makes sense, as we would expect kind and cooperative people to be indeed less likely to violate social norms. In contrast to agreeableness, high extraversion increases the propensity to reduce others' endowments as manifested in the positive coefficient in Table 5. Extroverted people are typically known to be enthusiastic, assertive and sociable, which hardly explains why they should take more. In conclusion, I find little evidence that Big Five Markers test is a good predictor of counter-punishment behavior:

Result 5: *The results from the regression analysis of the norm violation suggest that the most anti-social of individuals are more likely to counter-punish, but I find no supporting evidence from the regression of counter-punishment on the Big Five.*

Table 5: *Big Five Markers and the taking behavior*

	Amount taken
Extraversion	0.156 (2.06)*
Agreeableness	-0.262 (3.10)**
Conscientiousness	0.028 (0.33)
Neuroticism	0.029 (0.38)
Intellect	-0.087 (0.83)
Constant	10.334 (1.85)
R ²	0.24
N	48

*The dependent variable in the regression is Amount taken as measured by the deduction that Perpetrators took from other participants. The linear regression contains independent variables, which are the scores of the participants as coded from the Big Five Markers Test. *p<0.05 **p<0.01.*

IV.2 Survey Questions

In addition to the Big Five personality test, a survey consisting of 7 questions was administered at the end of the experiment. In this survey various questions were given to the participants asking them about their preferences and attitudes regarding social norms and

their violations. The questions' forms were that of hypothetical scenarios in which participants were actively involved in the situation. The questions, for example, asked about how the individual would feel if reprimanded by a family member for inconsiderate behavior or how the person would react to witnessing someone jumping a line in public.¹²

Only one of the seven questions turned out to be statistically significant when included as an independent variable in the linear regression model (Question 3 asked participants what they would do if they were to witness an example of second-party punishment).¹³ The results suggest that the more socially acceptable people perceive this kind of action to be, the greater is their counter-punishment (Table 6). After restricting the observations to the treatment 2PCP first and after that restricting the observations to 3PCP treatment, the coefficient on the question drops from 9.22** to 8.60* respectively. This makes little sense, as individuals who condone second-party punishment should be less willing to retaliate for it. I did not find any other significant results that would point out that there is any relationship between the way people perceive exemplary violations of social norms and their behavior in our experiment.

Result 6: *Social preferences, as elicited through a questionnaire file do not seem to have any predictive effects on the counter-punishing behavior. Individuals, who agree with second party punishment, seem to retaliate it much stronger than individuals, who disagree with second party punishment.*

V. CONCLUDING REMARKS

In this paper I studied counter-punishing behavior as an act of retaliation for incurred punishment. I examined the extent to which second and third parties counter-punish in a context of social-norm violation concerning the fairness of distribution and cooperation. The findings of the experiment point out significant differences between how individuals react to punishment from a self-interested individual and from an objective individual.

So far, the theory of reciprocity has been one way of explaining punishing behavior (e.g. Fehr et al. 2002). The principle of people hurting those who hurt them applies to counter-punishing perpetrator, who is also willing to retaliate punishment to a second as well as to a third party (Nikiforakis 2008; Balafoutas et al. 2014). However, the theory of reciprocity does not explain why punished individuals retaliate differently to the same punishment respective of who the punisher is.

There are other motives that have been used to explain punishing behavior beyond the economic ones, but they also fail to predict our recorded differences between 2PCP and 3PCP. The design of the experiment does not allow for any direct economic benefits from counter-punishment, in fact, punishers are strictly better off if they do not punish. Even though some models have offered explanations as to why people punish despite the monetary cost, these theories cannot help us understand the differences in counter-punishment of a second party and a third party.¹⁴

What then is the driving factor behind counter-punishing individuals? The notion of legitimacy is a fundamental element governing social interactions (Andrighetto et al. 2014). Andrighetto et al. argue that if punishment is perceived as legitimate, "it will be more effective and less costly in boosting voluntary cooperation than if it has to rely on coercion only" (2014: 4). Other studies have also shown that increasing the number of punishers increases the legitimacy of punishment and that the recipients of punishment view it as a way for the group to express norms governing their behavior (Villatoro et al. 2014). Higher compliance of punished individuals, as manifested by lower counter-punishment rates in 3PCP, could be therefore explained by how the punished individuals view third-party punisher.

By analyzing the strategies of punishers in 2PCP and 3PCP, I hypothesize that the impartial motivation of a third party to uphold social norms is perceived as a form of legitimizing punishment. Individuals who took a positive amount from their partners in the first stage were more likely to take revenge if the punishment came from a second party. In comparison, the same high-taking individuals' retaliation to punishment from a third-party was significantly smaller. The reason is that the perpetrators recognize the punishing behav-

¹²See Appendix for a complete list of the survey questions.

¹³Question 3: *Imagine that a person jumps the line in a supermarket. A person standing in line confronts the line jumper and asks them to leave by saying: "What do you think you are doing? Can't you see we are waiting here? Go to the back or leave!". On a scale from 1-7 (1 being not acceptable at all, 7 being completely socially acceptable and 4 neither acceptable nor unacceptable); how socially acceptable do you consider the reaction of the person standing in the line?*

¹⁴Levine (1998) and Fehr and Schmidt (1999) suggest that individuals should behave consistently when faced with the same strategic decisions. In the case of this paper, individuals should retaliate to the same punishment consistently. However, this is not the case given the difference in 2PCP and 3PCP.

Table 6: The Big Five Markers and counter-punishment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Punishment	0.774 (16.15)**	0.774 (18.33)**	0.774 (18.33)**	0.774 (18.33)**	0.774 (18.33)**	0.774 (18.33)**	0.774 (18.33)**
Third party		7.961 (0.79)	8.209 (0.81)	8.053 (0.80)	6.444 (0.64)	7.616 (0.75)	10.404 (1.01)
Third-party*Punishment		-0.484 (8.76)**	-0.484 (8.76)**	-0.484 (8.76)**	-0.484 (8.76)**	-0.484 (8.76)**	-0.484 (8.76)**
Extroversion			-0.419 (0.73)				
Neuroticism				0.676 (1.10)			
Agreeableness					-0.789 (1.27)		
Conscientiousness						0.452 (0.70)	
Openness							-0.900 (1.06)
Constant	2.705 (0.39)	2.705 (0.35)	15.705 (0.81)	-18.829 (0.89)	34.358 (1.32)	-13.425 (0.56)	36.454 (1.11)
Observations	220	528	528	528	528	528	528

The dependent variable in all regressions is the counter-punishment level as measured by the percentage reduction of the violator's income. All regressions include individual random effects. The regressions include all participants' decisions. * $p < 0.05$; ** $p < 0.01$.

ior of a third party as ruled not by self-interest, but rather by normative reasons — that is, by the efforts to uphold a social norm. Because of that, punished individuals were less likely to retaliate. This goes in line with studies demonstrating that punishment promotes compliance if the violators are made aware of what is right and what is wrong (Ostrom et al. 1992; Posner 2000).

In its present form, my experimental model cannot exhaustively explain counter-punishment. There are several limitations, which I have identified. First, the impartiality of the punishing player, as adopted from the matching mechanism of Fehr and Fischbacher (2004), is not perfect. Punishing Participant D, though not involved in the norm violation between Participant A and Participant B, is still involved in the norm violation of Participant C. In order to have perfect impartiality, the punishing player would not be involved in any other interaction than the decision to punish the violator. However, the problem with that is that a second-party punisher would experience income effects as a result of the norm violation, which would no longer be the case for the third-party punishers. In order to mitigate these effects I therefore decided to keep the matching mechanism of Fehr and Fischbacher. Third, the Big Five personality test did not indicate any behavioral charac-

teristics that would explain counter-punishing behavior. This, along with the inability to explain the difference in treatments by other motivation models, suggests that we still understand little about what drives counter-punishment. Lastly, in order to observe any differential effects in punishment coming from counter-punishment as anticipated by potential enforcers, an additional two treatments without a counter-punishment stage would have to be included. As of now, we cannot compare how the levels of punishment were affected by the introduction of counter-punishment opportunities.

Furthermore, as for the possible extensions, it would have been interesting to employ a third treatment in which both the second and the third party can punish. From a hypothetical standpoint, two scenarios could occur. First, punished individuals would choose the efficient punishment institution (i.e., third party punishers), as they are less likely to counter-punish. On the other hand, punished individuals could seek personal revenge and let their emotions get in the way. If this is the case, second-party punishment should be the prevalent choice.

This study has offered some insight into the mechanism behind counter-punishment, but there is still much to be done in the research on this topic. Despite the fact that we were unable to pinpoint the rationale

behind counter-punishment, our findings support the idea that impartial observers such as central authorities may be efficient enforcers of social norms. In terms of policy implications, this could be used as an argument for organizations resolving conflicts between disputed parties or even as an argument to encourage members of the general community to interfere when they observe behavior in violation with social norms.

REFERENCES

- Andreoni, James, and John Miller. "Giving according to GARP: An experimental test of the consistency of preferences for altruism." *Econometrica* 70, no. 2 (2002): 737-753.
- Andrighetto, Giulia, Jordi Brandts, Rosaria Conte, Jordi Sabater-Mir, Hector Solaz, and Daniel Villatoro. "Counter-punishment, Norm Communication and Accountability." *PLoS ONE* 8(6): e64941 (2013). doi:dx.doi.org/10.1371/journal.pone.0064941
- Balafoutas, Loukas, Kristoffel Grechenig, and Nikos Nikiforakis. "Third-party Punishment and Counter-punishment in One-shot Interactions." *Economics Letters* 122, no. 2 (2014): 308-10.
- Balafoutas, Loukas, and Nikos Nikiforakis. "Norm Enforcement in the City: A Natural Field Experiment." *European Economic Review* 56.8 (2012): 1773-785.
- Balafoutas, Loukas, Nikos Nikiforakis, and Bettina Rockenbach. "Direct and indirect punishment among strangers in the field." *Proceedings of the National Academy of Sciences* 111, no. 45 (2014): 15924-15927.
- Barrick, Murray R., and Michael K. Mount. "The Big Five Personality Dimensions And Job Performance: A Meta-Analysis." *Personnel Psychology* 44, no. 1 (1991): 1-26. doi:dx.doi.org/10.1111/j.1744-6570.1991.tb00688.x.
- Berkowitz, Leonard. "Social norms, feelings, and other factors affecting helping and altruism." *Advances in Experimental Social Psychology* 6 (1972): 63-108.
- "Big Five Personality Traits — Psychological Testing." Big Five Personality Traits — Psychological Testing. Accessed April 05, 2015. http://www.sevencounties.org/poc/view_doc.php?type=doc&id=8948
- Bock, Olaf, Andreas Nicklisch, and Ingmar Baetge (2012): hroot: Hamburg registration and organization online tool. WiSo-HH Working Paper Series No. 1, 2012.
- Bolton, Gary E., and Axel Ockenfels. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review* 90, no. 1 (2000): 166-93. doi:dx.doi.org/10.1257/aer.90.1.166
- Brandts, Jordi, and Gary Charness. "The Strategy versus the Direct-response Method: A First Survey of Experimental Comparisons." *Experimental Economics* 14, no. 3 (2011): 375-98. doi:dx.doi.org/10.1007/s10683-011-9272-x
- Cialdini, Robert B., Carl A. Kallgren, and Raymond R. Reno. "A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior." *Advances in Experimental Social Psychology* 24, no. 20 (1991): 1-243.
- Coleman, James S. *Foundations of Social Theory*. Cambridge, MA: Belknap Press of Harvard University Press, 1990.
- Cox, James C., Daniel Friedman, and Steven Gjerstad. "A Tractable Model of Reciprocity and Fairness." *Games and Economic Behavior* 59, no. 1 (2007): 17-45. doi:dx.doi.org/10.1016/j.geb.2006.05.001
- Dietz, Thomas, Elinor Ostrom, and Paul C. Stern. "The struggle to govern the commons." *Science* 302, no. 5652 (2003): 1907-1912.
- Fehr, E., and K. M. Schmidt. "A Theory of Fairness, Competition, and Cooperation." *The Quarterly Journal of Economics* 114, no. 3 (1999): 817-68. doi:dx.doi.org/10.1162/003355399556151
- Fehr, Ernst, and Simon Gächter. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review* 90, no. 4 (2000a): 980-94. doi:dx.doi.org/10.1257/aer.90.4.980
- Fehr, Ernst, and Simon Gächter. "Fairness and Retaliation: The Economics of Reciprocity." *Journal of Economic Perspectives* 14, no. 3 (2000b): 159-82. doi:dx.doi.org/10.1257/jep.14.3.159
- Fehr, Ernst, and Simon Gächter. "Altruistic punishment in humans." *Nature* 415, no. 6868 (2002): 137-140.
- Fehr, Ernst, and Urs Fischbacher. "Third-party Punishment and Social Norms." *Evolution and Human Behavior* 25, no. 2 (2004): 63-87. doi:dx.doi.org/10.1016/S1090-5138(04)00005-4
- Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl. "Gift Exchange and Reciprocity in Competitive Experimental Markets." *European Economic Review* 42, no. 1 (1998): 1-34. dx.doi.org/doi:10.1016/S0014-2921(96)00051-7

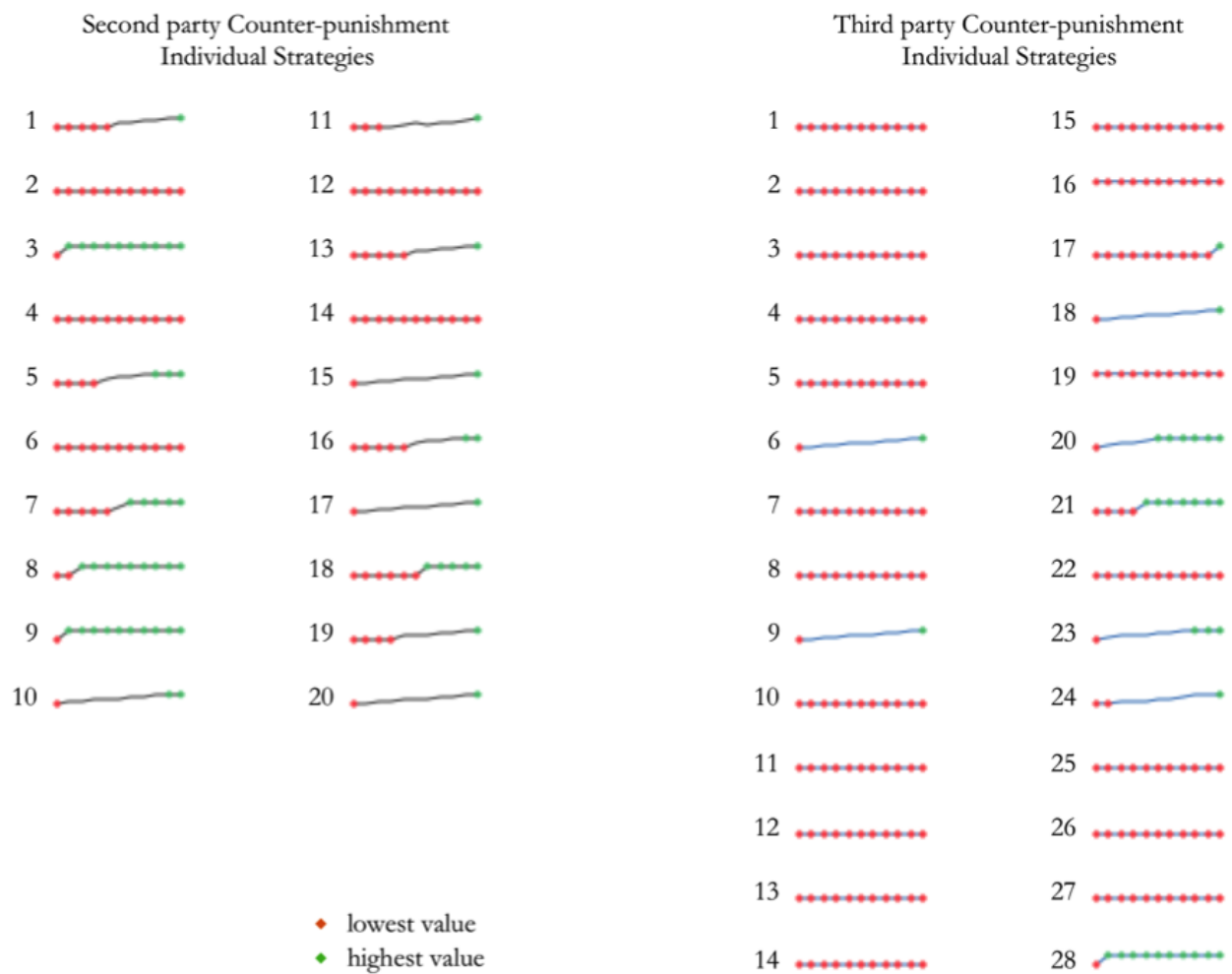
- Goldberg, Lewis R. "The Development of Markers for the Big-Five Factor Structure." *Psychological Assessment* 4, no. 1 (1992): 26-42. doi:dx.doi.org/10.1037/1040-3590.4.1.26
- Güth, Werner, Rolf Schmittberger, and Bernd Schwarze. "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior & Organization* 3, no. 4 (1982): 367-88. doi:dx.doi.org/10.1016/0167-2681(82)90011-7
- Levine, David K. "Modeling Altruism and Spitefulness in Experiments." *Review of Economic Dynamics* 1, no. 3 (1998): 593-622. dx.doi.org/doi:dx.doi.org/10.1006/redy.1998.0023
- Nikiforakis, Nikos. "Punishment and Counter-punishment in Public Good Games: Can We Really Govern Ourselves?" *Journal of Public Economics* 92, no. 1-2 (2008): 91-112. doi:dx.doi.org/10.1016/j.jpubeco.2007.04.008
- Nikiforakis, Nikos, Helen Mitchell. Mixing the Carrots with the Sticks: Third Party Punishment and Reward. Working paper GATE 2013-38. 2014.
- Nikiforakis, Nikos, Charles N. Noussair, and Tom Wilkening. "Normative conflict and feuds: The limits of self-enforcement." *Journal of Public Economics* Vol. 96, no. 9 (2012): 797-807.
- Ostrom, Elinor, James Walker, and Roy Gardner. "Covenants with and without a Sword: Self-governance Is Possible." *American Political Science Review* Vol. 86, no. 02 (1992): 404-417. doi:dx.doi.org/10.2307/1964229
- Pepitone, Albert. "Toward a normative and comparative biocultural social psychology." *Journal of Personality and Social Psychology* Vol. 34, no.4 (1976): 641.
- Rabin, Matthew, "Incorporating Fairness into Game Theory and Economics," *American Economic Review* Vol. 83, no. 5 (1993), 1281-1302.
- Salgado, Jesus F. "The Big Five Personality Dimensions and Counterproductive Behaviors." *International Journal of Selection and Assessment* Vol. 10, no. 1&2 (2002): 117-25. doi:dx.doi.org/10.1111/1468-2389.00198.
- Triandis, Harry Charalambos. *Interpersonal Behavior*. Monterey, CA: Brooks/Cole Pub., 1977.
- Villatoro, D., Andrighetto, G., Brandts, J., Nardin, L.G., Sabater-Mir, J., & Conte, R. (2014). "The Norm-Signaling Effects of Group Punishment: Combining Agent-Based Simulation and Laboratory Experiments." *Social Science Computer Review*, Vol. 32 no. 3, 334-353. doi:dx.doi.org/10.1177/0894439313511396
- Yamagishi, Toshio. "The provision of a sanctioning system as a public good." *Journal of Personality and Social Psychology* Vol. 51, no. 1 (1986): 110.
- Zhao, Hao, and Scott E. Seibert. "The Big Five Personality Dimensions and Entrepreneurial Status: A Meta-analytical Review." *Journal of Applied Psychology* Vol. 91, no. 2 (2006): 259-71. doi:dx.doi.org/10.1037/0021-9010.91.2.259

VI. APPENDIX

I. Order of Treatment

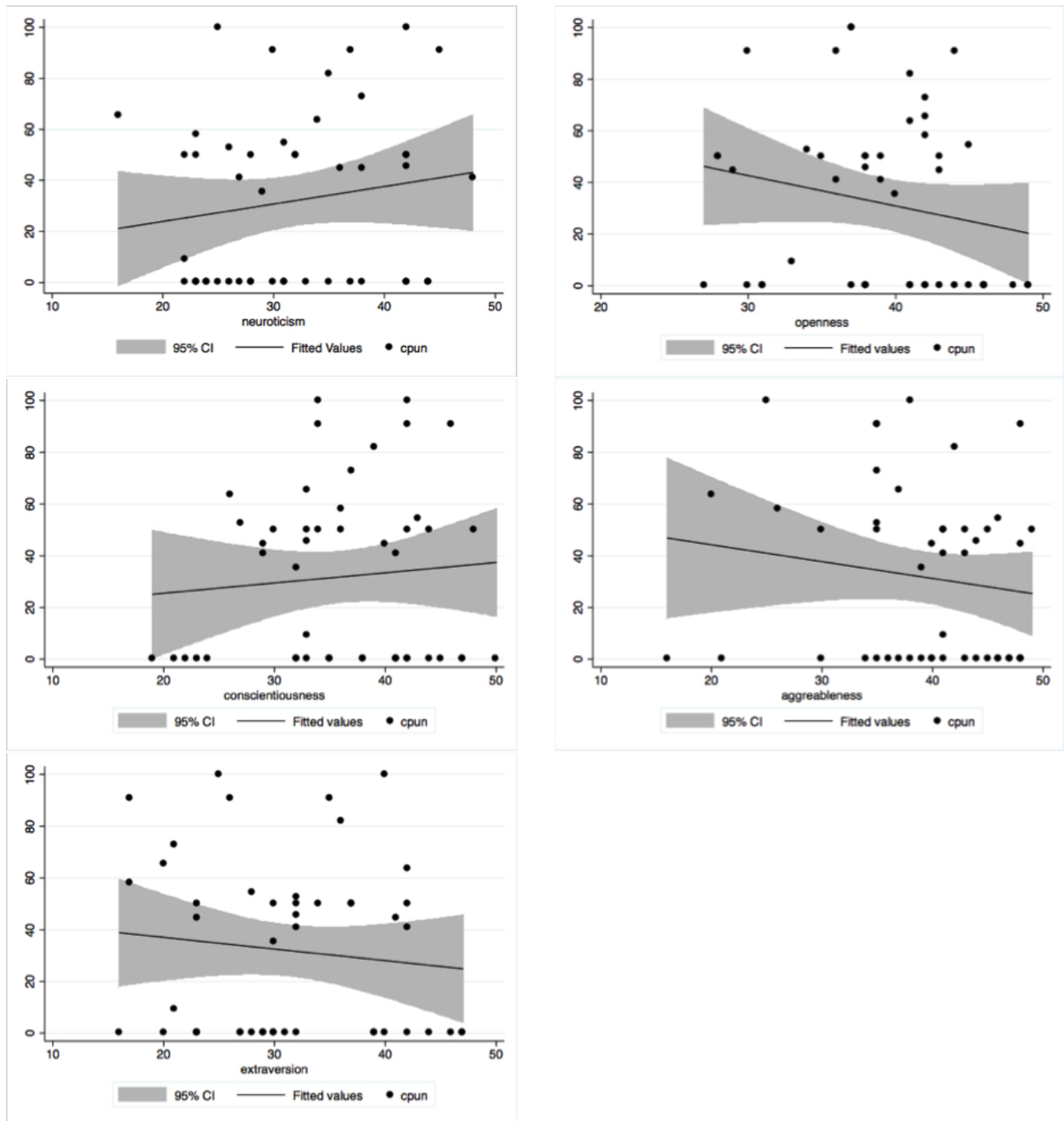
1. 2PCP (20 participants)
2. 3PCP (20 participants)
3. 2PCP (20 participants)
4. 3PCP (20 participants)
5. 3PCP (16 participants)

II. Individual Counter-punishment Strategies



Legend: Individual points represent the level of counter-punishment for each possible level of punishment (0%, 10%, 20% ... 100%). Lowest and highest value are measured for each participant separately.

III. Correlation between counter-punishing behavior and Big Five personality traits



IV. Survey Questions

1. On a scale from 1-7 (1 being not hurt at all, 7 being very hurt), please state how hurt you would be if someone reprimands you in front of your family and friends by saying that you're being inconsiderate of others
2. On a scale from 1-7 (1 being not happy at all, 7 being very happy), please state how happy you would be if someone compliments you in front of your family and friends by saying that you show concern for others.
3. Imagine that a person jumps the line in a supermarket. A person standing in line confronts the line jumper and asks them to leave by saying: "What do you think you are doing? Can't you see we are waiting here? Go to the back or leave!" On a scale from 1-7 (1 being not acceptable at all, 7 being completely socially acceptable and 4 neither acceptable nor unacceptable); How socially acceptable do you consider the reaction of the person standing in the line?
4. Consider a scenario similar to the previous. This time the person that confronts the line jumper is not someone standing in the line, but a bystander still shopping for their groceries. On a scale from 1-7 (1 being not acceptable at all, 7 being completely socially acceptable and 4 neither acceptable nor unacceptable); How socially acceptable do you consider the reaction of the bystander?
5. Imagine you are the one jumping the line. On a scale from 1-7 (1 being not likely at all, 7 being very likely) how likely are you to argue if the person standing in line confronted you by saying: "What do you think you are doing? Can't you see we are waiting here? Go to the back or leave!"
6. Consider a scenario similar to the one above, but this time the person that confronts you is not someone standing in line, but a bystander. On a scale from 1-7 (1 being not likely at all, 7 being very likely) how likely are you to argue if the bystander confronted you?
7. On a scale from 1-7 (1 - not likely at all, 7 - very likely), how likely would you say you are to stick up for what is socially right in your daily life?

V. Punishment and Big Five Markers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Amount taken	2.650 (4.67)**	2.650 (4.67)**	2.650 (4.67)**	2.650 (4.67)**	2.650 (4.67)**	2.650 (4.67)**	2.650 (4.67)**
Third party		9.581 (1.14)	9.463 (1.15)	7.724 (0.93)	9.487 (1.15)	11.189 (1.37)	10.780 (1.21)
Third party*Amount taken		-0.864 (1.16)**	-0.864 (1.16)**	-0.864 (1.16)**	-0.864 (1.16)**	-0.864 (1.16)**	-0.864 (1.16)**
Extroversion			-0.007 (0.01)				
Neuroticism				-0.375 (0.70)			
Agreeableness					-0.602 (1.26)		
Conscientiousness						0.065 (0.15)	
Openness							-0.935 (1.59)
Constant	0.286 (0.02)	0.286 (0.02)	15.856 (0.68)	22.084 (1.20)	2.179 (1.00)	-10.983 (0.88)	32.454 (1.05)
Observations	144	144	144	144	144	144	144

*The dependent variable in all regressions is the punishment level as measured by the percentage reduction of the violator's income. All regressions include individual random effects. The regressions include all participants' decisions. * $p < 0.05$; ** $p < 0.01$.*

VI. Instructions

VI.1 Instructions for Participant A in 2PCP

General Instructions for Participants A

You are now taking part in an economic experiment. Please do not communicate with other participants during the experiment. Should you have any questions, please raise your hand and one of us will come to assist you. The amount of money you earn from this experiment depends on your decisions and those made by others in the experiment. It is therefore important that you take your time to understand the instructions.

This experiment consists of two parts. You will receive instructions for the second part once the first part of the experiment has been completed. At the end of the experiment you will be asked to fill out a questionnaire.

Earnings

During this experiment **we do not deal with Emirati dirhams, but with Experimental Monetary Units (EMU)**. At the end of the experiment, the total amount of EMU you earned during the experiment will be converted into Emirati Dirhams at the rate of

$$1 \text{ EMU} = 2 \text{ AED}$$

The earned sum will consist of the show-up fee of 30 AED plus any individual gains from the experiment. This total will be paid to you in cash in private, immediately after the completion of the experiment.

Should you have any questions, please raise your hand and an experimenter will come to assist you.

Specific Instructions for Participant A

Overview of Part 1

This experiment comprises 2 different types of Participants: Participants A and Participants B. **You are a Participant A**. At the start of the experiment, the computer will randomly place you into a group with an individual who has been assigned the role of Participant B.

The experiment consists of **three stages**. In the first stage, those who have been assigned the role of Participant A have to decide whether to take 0, 5 or 10 EMU from the individual with whom they have been paired with. In the second stage, individuals who have been assigned the role of Participant B will have to decide whether they wish to **reduce the earnings** of Participant A by a certain amount. In the third and final stage of the experiment, those who have been assigned the role of Participant A will be asked whether they wish to reduce the earnings of Participant B.

Neither during nor after the experiment will you be aware of the identities of any of the participants. Equally, other participants will never know with whom they were dealing. All earnings will be paid out anonymously at the experiment. No participant will ever learn how much you earned in the experiment.

Detailed description of Part 1

Stage One

In stage one only individuals assigned the role of **Participant A** can make decisions. As mentioned, in stage one, you are grouped together with another participant. As a Participant A you get an endowment of **30 EMU** at the beginning of stage one. Participant B gets the same endowment of 30 EMU.

You must decide whether or not to take a portion of Participant's B income. **You can take 0, 5 or 10 EMU from Participant B**. If you decide to take 0 EMU, your earnings and the earnings of Participant B at the end of the first

stage will be the same as the original endowment you received. If you take 5 EMU, your earnings at the end of the first stage will be 35 EMU, while the earnings of Participant B will be 25 EMU. If you decide to take 10 EMU, your earnings at the end of the first stage will be 40 EMU, while the earnings of Participant B will be 20 EMU.

You will make your decision as displayed on the screenshot below:

STAGE 1

You are Participant A. Your initial endowment is 30 EMU.

You are randomly paired with a Participant B whose initial endowment is the same as yours: 30 EMU
You now have the option to take some of Participant B's endowment.

How many EMU, if any, do you wish to take from Participant B?

(Select one of the following)

☐ 0 EMU
☐ 5 EMU
☐ 10 EMU

OK

Stage Two

In stage two, only individuals assigned the role of **Participant B** can make decisions. In particular, **Participant B from your group** will be asked whether or not s/he wishes to reduce your income. Participant B will be informed of your decision in stage one. S/he will be awarded an additional **5 EMU**, which s/he can use to reduce your income by 10%, 20%, 30% ... 90%, or 100%. The cost of reducing your income by a positive amount (i.e., more than 0%) is 5 EMU irrespective of the amount by which Participant B reduces your income (i.e., 10%, 20%, ... or 100%). If Participant B decides not to reduce your income, then the 5 EMU are added to his or her final earnings.

Stage Three

Your task in the third stage will be to **decide whether you want to reduce or leave unaffected the earnings of Participant B**; that is, the participant who was asked in stage two whether or not to reduce your earnings from stage one. Income reduction in this stage works exactly as in the second stage. You will be awarded **5 EMU**, which you can use to reduce Participant's B earnings by 10%, 20%, 30% ... 90%, 100%. If you do not reduce Participant B's earnings, then the 5 EMU will be added to your final earnings.

Note that you will have to make your decision about how much you want to decrease Participant B's earnings **before you find out his or her decision in stage two**. You will have to state whether and by how much you wish to reduce B's earnings for each possible decision Participant B may have taken in the second stage. To make your decision you will use a screen as the one shown below.

On the screen you must indicate whether and by how much you want to reduce Participant B's earnings in the case he or she reduced your earnings by 0%, 10%, 20% ... 100%.

Given that you will not know which one of the possible actions Participant B has decided on, any of your decisions might be used to determine your final earnings and the earnings of Participant B. Which of your decisions is actually realized depends on the actual decision of Participant B in stage two.

STAGE 3

Your task in this stage is to decide whether you want to reduce the income of Participant B **in your group**, who may have reduced your income in Stage 2.

Please indicate for every possible action of Participant B in Stage 2 by how much, if at all, you would like to reduce his or her income.

If Participant B in my group reduced my income in Stage 2 by:	I will reduce Participant B's income by the following percent. Do not use the "%" sign, and use increments of 10
0%	<input type="text"/>
10%	<input type="text"/>
20%	<input type="text"/>
30%	<input type="text"/>
40%	<input type="text"/>
50%	<input type="text"/>
60%	<input type="text"/>
70%	<input type="text"/>
80%	<input type="text"/>
90%	<input type="text"/>
100%	<input type="text"/>

After you have made your decisions, you will be informed of your earnings from the experiment.

CONTROL QUESTIONS

Before the experiment can commence, all participants need to answer the following questions correctly. Once you have answered all the questions, please raise your hand to attract the attention of the experimenter. The experiment will begin as soon as all of the participants have answered the questions correctly. As with all examples in the instructions, the numbers should not be taken as a guide for behavior in the experiment.

1. You decide to take away 5 EMU from Participant B's initial endowment. Participant B decides to reduce your income by 40%. What are your final earnings if you do not reduce Participant B's income in stage three?
_____EMU
2. You decided to take 5 EMU from Participant B's initial endowment. What are your earnings at the end of the second stage, if Participant B reduces your income by
 - (a) 0%? _____EMU
 - (b) 10%? _____EMU
 - (c) 50%? _____EMU
 - (d) 100%? _____EMU
3. What is the cost for Participant B for reducing your earnings in stage two by
 - (a) 0%? _____EMU
 - (b) 10%? _____EMU
 - (c) 50%? _____EMU
 - (d) 100%? _____EMU

4. You decided to take 10 from Participant's B initial endowment. Participant B then reduced your income by 50%. What are your final earnings if you decide

- (a) Not to reduce B's earnings: _____ EMU
- (b) To reduce B's earnings by 10%: _____ EMU
- (c) To reduce B's earnings by 100%: _____ EMU

VI.2 Instructions for Participant B in 2PCP

General Instructions for Participants B

You are now taking part in an economic experiment. Please do not communicate with other participants during the experiment. Should you have any questions, please raise your hand and one of us will come to assist you. The amount of money you earn from this experiment depends on your decisions and those made by others in the experiment. It is therefore important that you take your time to understand the instructions.

This experiment consists of two parts. You will receive instructions for the second part once the first part of the experiment has been completed. At the end of the experiment you will be asked to fill out a questionnaire.

Earnings

During this experiment **we do not deal with Emirati dirhams, but with Experimental Monetary Units (EMU)**. At the end of the experiment, the total amount of EMU you earned during the experiment will be converted into Emirati Dirhams at the rate of

$$1 \text{ EMU} = 2 \text{ AED}$$

The earned sum will consist of the show-up fee of **30 AED** plus any individual gains from the experiment. This total will be paid to you in cash in private, immediately after the completion of the experiment.

Should you have any questions, please raise your hand and an experimenter will come to assist you.

Specific Instructions for Participant B

Overview of Part 1

This experiment comprises 2 different types of Participants: Participants A and Participants B. **You are a Participant B**. At the start of the experiment, the computer will randomly place you into a group with an individual who has been assigned the role of Participant A.

The experiment consists of **three stages**. In the first stage, those who have been assigned the role of Participant A have to decide whether to take 0, 5 or 10 EMU from the individual with whom they have been paired with. In the second stage, individuals who have been assigned the role of Participant B will have to decide whether they wish to **reduce the earnings** of Participant A by a certain amount. In the third and final stage of the experiment, those who have been assigned the role of Participant A will be asked whether they wish to reduce the earnings of Participant B.

Neither during nor after the experiment will you be aware of the identities of any of the participants. Equally, other participants will never know with whom they were dealing. All earnings will be paid out anonymously at the experiment. No participant will ever learn how much you earned in the experiment.

Detailed description of Part 1

Stage One

As mentioned, in stage one, you are grouped together with another participant — Participant A. As a Participant B you get an endowment of **30 EMU** at the beginning of stage one. Participant A gets the same endowment of 30 EMU.

Participant A must decide whether or not to take a portion of your income. **Participant A can take 0, 5 or 10 EMU from you.** If s/he decides to take 0 EMU, your earnings and the earnings of Participant A at the end of the first stage will be the same as the original endowment you received. If s/he takes 5 EMU, your earnings at the end of the first stage will be 25 EMU, while the earnings of Participant A will be 35 EMU. If s/he decides to take 10 EMU, your earnings at the end of the first stage will be 20 EMU, while the earnings of Participant A will be 40 EMU.

Stage Two

Your task in the second stage will be to **decide whether you want to reduce or leave unaffected the earnings of Participant A.** You will be awarded an additional 5 EMU, which you can use to reduce Participant's A earnings by 10%, 20%, 30% ... 90%, 100%. The cost of reducing Participant A's income by a positive amount (i.e., more than 0%) is 5 EMU irrespective of the amount by which you reduce his/her income (i.e., 10%, 20%, ... or 100%). If you do not reduce Participant A's earnings, then the 5 EMU will be added to your final earnings.

Note that you will have to make your decision about how much you want to decrease Participant A's earnings **before you find out his or her decision in stage one.** You will have to state whether and by how much you wish to reduce A's earnings for each possible decision Participant A may have taken in the first stage. To make your decision you will use a screen as the one shown below.

STAGE 2

You are Participant B. Your initial endowment is 30 EMU.

Your task in this stage is to decide whether you want to reduce the income of Participant A in your group.

Participant A's initial endowment was the same as yours: 30 EMU.

Participant A had the option to take some of your initial endowment in Stage 1.

Please indicate for every possible action of Participant A in Stage 1 by how much, if at all, you would like to reduce his or her income.

If Participant A took away from me in Stage 1:	I will reduce Participant A's income by the following percent. Do not use the "%" sign, and use increments of 10
0 EMU	<input style="width: 100%;" type="text"/>
5 EMU	<input style="width: 100%;" type="text"/>
10 EMU	<input style="width: 100%;" type="text"/>

Given that you will not know which one of the three actions Participant A has decided on, any of the three decisions might be used to determine final earnings of Participant A. Which of your decisions is actually realized depends on the actual decision of Participant A in stage one.

Stage Three

In stage three, **Participant A** will be asked whether or not s/he wishes to reduce your income. Income reduction in this stage works exactly as in the second stage. S/he will be awarded an additional **5 EMU**, which s/he can use to reduce your income by 10%, 20%, 30%... 90%, or 100%. If Participant A decides not to reduce your income, then the 5 EMU are added to his or her final earnings.

After you have made your decisions, you will be informed of the conclusion of the experiment.

CONTROL QUESTIONS

Before the experiment can commence, all participants need to answer the following questions correctly. Once you have answered all the questions, please raise your hand to attract the attention of the experimenter. The experiment will begin as soon as all of the participants have answered the questions correctly. As with all examples in the instructions, the numbers should not be taken as a guide for behavior in the experiment.

1. Participant A decides to take away 5 EMU from your initial endowment. You decide to reduce Participant A's income by 40%. Participant A does not reduce your income in stage three. What are your final earnings?
_____ EMU
2. Participant A decides to take 5 EMU from your initial endowment. What are your earnings at the end of the second stage, if you reduce Participant A's income by
 - (a) 0%? _____ EMU
 - (b) 10%? _____ EMU
 - (c) 50%? _____ EMU
 - (d) 100%? _____ EMU
3. What is the cost for Participant A for reducing your earnings in stage three by
 - (a) 0%? _____ EMU
 - (b) 10%? _____ EMU
 - (c) 50%? _____ EMU
 - (d) 100%? _____ EMU
4. Participant A decided to take 10 EMU from your initial endowment. You then reduced Participant A's income by 50% in stage two. What are your final earnings if Participant A decides:
 - (a) Not to reduce your earnings: _____ EMU
 - (b) To reduce your earnings by 10%: _____ EMU
 - (c) To reduce your earnings by 100%: _____ EMU

VI.3 Instructions for Participant A in 3PCP

General Instructions for Participants A

You are now taking part in an economic experiment. Please do not communicate with other participants during the experiment. Should you have any questions, please raise your hand and one of us will come to assist you. The amount of money you earn from this experiment depends on your decisions and those made by others in the experiment. It is therefore important that you take your time to understand the instructions.

This experiment consists of two parts. You will receive instructions for the second part once the first part of the experiment has been completed. At the end of the experiment you will be asked to fill out a questionnaire.

Earnings

During this experiment **we do not deal with Emirati dirhams, but with Experimental Monetary Units (EMU)**. At the end of the experiment, the total amount of EMU you earned during the experiment will be converted into Emirati Dirhams at the rate of

$$1 \text{ EMU} = 2 \text{ AED}$$

The earned sum will consist of the show-up fee of **30 AED** plus any individual gains from the experiment. This total will be paid to you in cash in private, immediately after the completion of the experiment.

Should you have any questions, please raise your hand and an experimenter will come to assist you.

Specific Instructions for Participant A

Overview of Part 1

This experiment comprises different types of Participants: Participant A, Participant B, Participant C, Participant D etc. **You are a Participant A.** At the start of the experiment, the computer will randomly place you into a group with an individual who has been assigned the role of Participant B. Participant C will be placed into a group with Participant D, Participant E will be placed into a group with Participant F etc.

Neither during nor after the experiment will you be aware of the identities of any of the participants. Equally, other participants will never know with whom they were dealing. All earnings will be paid out anonymously at the experiment. No participant will ever learn how much you earned in the experiment. The experiment consists of **three stages**, whose procedure is described in detail below.

Detailed description of Part 1

Stage One

In stage one, you as **Participant A** will be able to make decisions. As mentioned, in stage one, you are grouped together with another Participant B. As Participant A you get an endowment of **30 EMU** at the beginning of stage one. All the other participants get the same endowment of 30 EMU.

You must decide whether or not to take a portion of Participant B's income. **You can take 0, 5 or 10 EMU from Participant B.** If you decide to take 0 EMU, your earnings and the earnings of Participant B at the end of the first stage will be the same as the original endowment you received. If you take 5 EMU, your earnings at the end of the first stage will be 35 EMU, while the earnings of Participant B will be 25 EMU. If you decide to take 10 EMU, your earnings at the end of the first stage will be 40 EMU, while the earnings of Participant B will be 20 EMU.

You will be able to make a decision as presented on the screenshot:

STAGE 1

You are Participant A. Your initial endowment is 30 EMU.

You are randomly paired with a Participant B whose initial endowment is the same as yours: 30 EMU
You now have the option to take some of Participant B's endowment.

How many EMU, if any, do you wish to take from Participant B?

(Select one of the following) ☐ 0
☐ 5
☐ 10

OK

Participant C from another pair also has the option to reduce or leave unaffected the endowment of Participant D by the same amounts — 0, 5 or 10 EMU. Similarly, Participant E has the option to reduce or leave unaffected the endowment of Participant F and so forth.

Stage Two

In stage two, **Participant D** from another group will be asked whether or not s/he wishes to reduce your income. Participant D will be informed of your decision in stage one. S/he will be awarded an additional **5 EMU**, which s/he can use to reduce your income by 10%, 20%, 30% ... 90%, or 100%. The cost of reducing your income by a positive amount (i.e., more than 0%) is 5 EMU irrespective of the amount by which Participant D reduces your income (i.e., 10%, 20%, ... or 100%). If Participant D decides not to reduce your income, then the 5 EMU are added to his or her final earnings.

Participant B, from your group, will be asked whether s/he wishes to reduce the earnings of a Participant E in a third group. **Participant B cannot reduce your income**, but rather the income of Participant E in a third group.

Stage Three

Your task in the third stage will be to **decide whether you want to reduce or leave unaffected the earnings of Participant D**; that is, the participant from the second pair who was asked in stage two whether or not to reduce your earnings from stage one. Income reduction in this stage works exactly as in the second stage. You will be awarded **5 EMU**, which you can use to reduce Participant D's earnings by 10%, 20%, 30% ... 90%, 100%. If you do not reduce Participant D's earnings, then the 5 EMU will be added to your final earnings.

Note that you will have to make your decision about how much you want to decrease Participant D's earnings **before you find out his or her decision in stage two**. You will have to state whether and by how much you wish to reduce D's earnings for each possible decision Participant D may have taken in the second stage. To make your decision you will use a screen as the one shown below.

On the screen you must indicate whether and by how much you want to reduce Participant D's earnings in the case he or she reduced your earnings by 0%, 10%, 20% ... 100%.

STAGE 3

Your task in this stage is to decide whether you want to reduce the income of Participant D **in another group**.
Participant D from the other group may have reduced your income in Stage 2.
Please indicate for every possible action of Participant D in Stage 2 by how much, if at all, you would like to reduce his or her income.

If Participant D from the other group reduced my income in Stage 2 by:	I will reduce Participant D's income by the following percent. Do not use the "%" sign, and use increments of 10
0%	<input type="text"/>
10%	<input type="text"/>
20%	<input type="text"/>
30%	<input type="text"/>
40%	<input type="text"/>
50%	<input type="text"/>
60%	<input type="text"/>
70%	<input type="text"/>
80%	<input type="text"/>
90%	<input type="text"/>
100%	<input type="text"/>

Given that you will not know which one of the possible actions Participant D has decided on, any of your decisions might be used to determine your final earnings and the earnings of Participant D. Which of your decisions is actually realized depends on the actual decision of Participant D in stage two.

After you have made your decisions, you will be informed of your earnings from the experiment.

CONTROL QUESTIONS

Before the experiment can commence, all participants need to answer the following questions correctly. Once you have answered all the questions, please raise your hand to attract the attention of the experimenter. The experiment will begin as soon as all of the participants have answered the questions correctly. As with all examples in the instructions, the numbers should not be taken as a guide for behavior in the experiment.

1. You decide to take away 10 EMU from Participant B's initial endowment. Participant D decides to reduce your income by 40%. What are your final earnings if you decide not to reduce Participant D's income in stage three? _____ EMU
2. You decided to take 5 EMU from Participant B's initial endowment. What are your earnings at the end of the second stage, if Participant D reduces your income by
 - (a) 0%? _____ EMU
 - (b) 10%? _____ EMU
 - (c) 50%? _____ EMU
 - (d) 100%? _____ EMU
3. What is the cost for Participant D for reducing your earnings in stage two by
 - (a) 0%? _____ EMU

- (b) 10%? _____ EMU
 (c) 50%? _____ EMU
 (d) 100%? _____ EMU
4. You decide to take 10 points from Participant's B initial endowment. Who will be informed about this decision?
 _____ and _____
5. Which one of the above in question 4 will have the chance to reduce your earnings in Stage 2?

6. You decided to take 10 EMU from Participant's B initial endowment. Participant D then reduced your income by 50%. What are your final earnings if you decide
- (a) Not to reduce D's earnings: _____ EMU
 (b) To reduce D's earnings by 10%: _____ EMU
 (c) To reduce D's earnings by 100%: _____ EMU

VI.4 Instructions for Participant B in 3PCP

General Instructions for Participants B

You are now taking part in an economic experiment. Please do not communicate with other participants during the experiment. Should you have any questions, please raise your hand and one of us will come to assist you. The amount of money you earn from this experiment depends on your decisions and those made by others in the experiment. It is therefore important that you take your time to understand the instructions.

This experiment consists of two parts. You will receive instructions for the second part once the first part of the experiment has been completed. At the end of the experiment you will be asked to fill out a questionnaire.

Earnings

During this experiment **we do not deal with Emirati dirhams, but with Experimental Monetary Units (EMU)**. At the end of the experiment, the total amount of EMU you earned during the experiment will be converted into Emirati Dirhams at the rate of

$$1 \text{ EMU} = 2 \text{ AED}$$

The earned sum will consist of the show-up fee of **30 AED** plus any individual gains from the experiment. This total will be paid to you in cash in private, immediately after the completion of the experiment.

Should you have any questions, please raise your hand and an experimenter will come to assist you.

Specific Instructions for Participant B

Overview of Part 1

This experiment comprises different types of Participants: Participant A, Participant B, Participant C, Participant D etc. **You are a Participant B**. At the start of the experiment, the computer will randomly place you into a group with an individual who has been assigned the role of Participant A. Participant C will be placed into a group with Participant D, Participant E will be placed into a group with Participant F etc.

Neither during nor after the experiment will you be aware of the identities of any of the participants. Equally, other participants will never know with whom they were dealing. All earnings will be paid out anonymously at the experiment. No participant will ever learn how much you earned in the experiment. The experiment consists of **three stages**, whose procedure is described in detail below.

Detailed description of Part 1

Stage One

In stage individuals assigned the role of **Participant A** can make decisions. As mentioned, in stage one, you are grouped together with Participant A. As Participant B you get an endowment of **30 EMU** at the beginning of stage one. All the other participants get the same endowment of 30 EMU.

Participant A must decide whether or not to take a portion of your income. **Participant A can take 0, 5 or 10 EMU from you.** If s/he decides to take 0 EMU, your earnings and the earnings of Participant A at the end of the first stage will be the same as the original endowment you received. If s/he takes 5 EMU, your earnings at the end of the first stage will be 25 EMU, while the earnings of Participant A will be 35 EMU. If s/he decides to take 10 EMU, your earnings at the end of the first stage will be 20 EMU, while the earnings of Participant A will be 40 EMU.

Participant C from another pair also has the option to reduce or leave unaffected the endowment of Participant D by the same amounts — 0, 5 or 10 EMU. Similarly, Participant E has the option to reduce or leave unaffected the endowment of Participant F and so forth.

Stage Two

Your task in the second stage will be to **decide whether you want to reduce or leave unaffected the earnings of Participant C**; that is, the participant from **another group** who was asked in stage one whether or not to reduce **Participant D's** income. You will be awarded an additional **5 EMU**, which you can use to reduce Participant C's earnings by 10%, 20%, 30% ... 90%, 100%. The cost of reducing Participant C's income by a positive amount (i.e., more than 0%) is 5 EMU irrespective of the amount by which you reduce his/her income (i.e., 10%, 20%, ... or 100%). If you do not reduce Participant C's earnings, then the 5 EMU will be added to your final earnings.

Note that you will have to make your decision about how much you want to decrease Participant C's earnings **before you find out his or her decision in stage one**. You will have to state whether and by how much you wish to reduce C's earnings for each possible decision Participant C may have taken in the first stage. To make your decision you will use a screen as the one shown below.

STAGE 2

You are Participant B. Your initial endowment is 30 EMU. Participant A took away 5 EMU from you in Stage 1.

Your task in this stage is to decide whether you want to reduce the income of Participant C from another group.

Participant C's initial endowment was the same as yours: 30 EMU.

Participant C had the option to take some of Participant D's initial endowment in Stage 1.

Please indicate for every possible action of Participant C in Stage 1 by how much, if at all, you would like to reduce his or her income.

If Participant C took away from Participant D in Stage 1:	I will reduce Participant C's income by the following percent. Do not use the "%" sign, and use increments of 10
0 EMU	<input style="width: 100%;" type="text"/>
5 EMU	<input style="width: 100%;" type="text"/>
10 EMU	<input style="width: 100%;" type="text"/>

Given that you will not know which one of the three actions Participant C has decided on, any of the three decisions might be used to determine final earnings of Participant C. Which of your decisions is actually realized depends on the actual decision of Participant C in Stage One.

Stage Three

In stage three, **Participant C** will be asked whether or not s/he wishes to reduce your income. Income reduction in this stage works exactly as in the second stage. S/he will be awarded an additional **5 EMU**, which s/he can use to reduce your income by 10%, 20%, 30% ... 90%, or 100%. If Participant C decides not to reduce your income, then the 5 EMU are added to his or her final earnings.

Also, Participant A from your pair can deduct income of a Participant in a third group — F. Participant **A cannot deduct your income**, but rather an income of another Participant in a third group.

After you have made your decisions, you will be informed of the conclusion of the experiment.

CONTROL QUESTIONS

Before the experiment can commence, all participants need to answer the following questions correctly. Once you have answered all the questions, please raise your hand to attract the attention of the experimenter. The experiment will begin as soon as all of the participants have answered the questions correctly. As with all examples in the instructions, the numbers should not be taken as a guide for behavior in the experiment.

1. Participant C decides to take away 10 EMU from Participant D's initial endowment. You decide to reduce Participant C income by 40%. Participant C does not reduce your income in Stage three. What are his/her final earnings? _____EMU
2. Participant A decides to take 5 EMU from your initial endowment. What are your earnings at the end of the second stage, if you reduce Participant C's earnings by
 - (a) 0%? _____EMU

- (b) 10%? _____ EMU
 (c) 50%? _____ EMU
 (d) 100%? _____ EMU
3. What is the cost for Participant C for reducing your earnings in stage three by
- (a) 0%? _____ EMU
 (b) 10%? _____ EMU
 (c) 50%? _____ EMU
 (d) 100%? _____ EMU
4. Participant A decided to take 10 EMU from your initial endowment. Participant C decided to take 5 EMU from Participant D's initial endowment. You then reduced Participant C's income by 50% in Stage two. What are your final earnings if Participant C decides:
- (a) Not to reduce your earnings: _____ EMU
 (b) To reduce your earnings by 10%: _____ EMU
 (c) To reduce your earnings by 100%: _____ EMU

Table 1: Counter-punishment Regressions

	2PCP (1)	3PCP (2)	Both (3)	CPUN Decision (4)	CPUN Level (5)	Both (6)	CPUN Decision (7)	CPUN Level (8)
Punishment ^a	0.774 (16.15)**	0.290 (9.12)**	0.774 (18.33)**	0.240 (14.61)**	0.663 (12.60)**	0.774 (18.33)**	0.170 (7.74)**	0.665 (12.63)**
Third party			7.961 (0.79)	-5.780 (-3.39)**	10.880 (1.22)	9.682 (0.75)	2.045 (1.06)**	-2.161 (-0.20)
Punishment* Third party			-0.484 (-8.76)**	-1.145 (-6.65)**	-0.084 (-1.17)	-0.484 (-8.76)**	-1.109 (-4.57)**	-0.084 (-1.17)
Amount taken						1.261 (0.536)	0.933 (2.02)*	-1.248 (-0.99)
Amount taken* third party						-0.836 (-0.34)	-0.963 (-2.01)*	3.518
Constant	2.705 (0.39)	10.666 (1.54)	2.705 (0.35)	-8.964 (-8.20)**	34.010 (5.59)**	-0.763 (-0.08)	-12.344 (-9.05)**	38.470 (5.18)**
Observations	220	308	528	528	221	528	528	221
Subjects	20	28	48	48	20	48	48	20

^aThe dependent variable in all regressions is the counter-punishment level as measured by the percentage reduction of the violator's income. (1)-(3) are linear regressions; (4) is a probit regression and (5) is a linear regression with dependent variable > 0. All regressions include individual random effects. The regressions include all participants' decisions. CPUN = Counter-punishment; * p<0.05; ** p<0.01

Table 4: *The Big Five Markers and Counter-punishment*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Punishment ^a	0.774 (16.15)**	0.774 (18.33)**	0.774 (18.33)**	0.774 (18.33)**	0.774 (18.33)**	0.774 (18.33)**	0.774 (18.33)**
Third party		7.961 (0.79)	8.209 (0.81)	8.053 (0.80)	6.444 (0.64)	7.616 (0.75)	10.404 (1.01)
Third party*Punishment		-0.484 (8.76)**	-0.484 (8.76)**	-0.484 (8.76)**	-0.484 (8.76)**	-0.484 (8.76)**	-0.484 (8.76)**
Extroversion			-0.419 (0.73)				
Neuroticism				0.676 (1.10)			
Agreeableness					-0.789 (1.27)		
Conscientiousness						0.452 (0.70)	
Openness							-0.900 (1.06)
Constant	2.705 (0.39)	2.705 (0.35)	15.705 (0.81)	-18.829 (0.89)	34.358 (1.32)	-13.425 (0.56)	36.454 (1.11)
Observations	220	528	528	528	528	528	528

^aThe dependent variable in all regressions is the counter-punishment level as measured by the percentage reduction of the violator's income. (1)-(3) are linear regressions; (4) is a probit regression and (5) is a linear regression with dependent variable > 0. All regressions include individual random effects. The regressions include all participants' decisions. CPUN = Counter-punishment; * p<0.05; ** p<0.01