Why people punish in public goods games, and on whom punishment is most effective: Conventional wisdoms reconsidered

Ramzi Suleiman & Yuval Samid

University of Haifa

Please address all correspondence to Dr. Ramzi Suleiman, Dept. of Psychology, University of Haifa, Haifa 31509, Israel. Email: suleiman@psy.haifa.ac.il, Mobile: + 972-(0)50-5474- 215, Homepage: http://suleiman.haifa.ac.il.

Acknowledgments: This research was supported by the Israeli Science Foundation (grant No. 1213/11). We thank Simon Gächter, Christian Thöni, and Benedikt Herrmann for providing the raw data of the Hermann et al. (2008) Science paper.

Why people punish in public goods games, and on whom punishment is most effective: Conventional wisdoms reconsidered

Abstract

Theoretical and experimental research underscores the role of punishment in the evolution of cooperation between humans. Experiments using the public goods game have repeatedly shown that in cooperative social environments, punishment makes cooperation flourish, and that withholding punishment makes cooperation collapse. In less cooperative social environments, where antisocial punishment has been detected, punishment was detrimental to cooperation. The success of punishment in enhancing cooperation is explained as deterrence of free riders by cooperative strong reciprocators, who were willing to pay the cost of punishing them, whereas in environments in which punishment diminished cooperation, antisocial punishment was explained as *revenge* by low cooperators against high cooperators suspected of punishing them in previous rounds.

The present paper reconsiders the generality of both explanations. Using data from a novel public good experiment with punishment and from 16 public goods experiments from countries around the world, we report results showing that revenge alone does not drive antisocial punishment of cooperators, and that such punishment is predominantly part of an upward and downward punishment strategy, presumably aimed at punishing those who deviate from the punisher's aspired cooperation norm. More interestingly, we show that the effect of punishment on the emergence of cooperation is mainly due to *contributors* increasing their cooperation, more than free riders being deterred. We also show that the anticipation of being punished is more effective in enhancing cooperation than the actual punishment itself, and that the ratio of strong reciprocators in a given social group is a potent predictor of the group's level of cooperation and success in providing public goods.

Keyword: Cooperation, Punishment, Antisocial Punishment, Public Goods, Strong Reciprocity

Introduction

Previous research shows that altruistic punishment is fundamental for maintaining cooperation between strangers, when reciprocity and reputation effects are absent (Fehr & Gächter, 2002). Several experiments using the public goods game as a workhorse have demonstrated that many individuals punish free riders, altruistically, when punishment is costly to them and yields no material gain (see, e.g., Fehr & Gächter, 2002; Fehr & Fischbacher, 2004; Egas & Riedl, 2008; Sutter et al., 2010). A similar positive effect for punishment was also detected in public goods experiments with fixed groups' composition (Fehr & Fischbacher, 2004). In all experiments, when punishment was an option, group members punished harshly and cooperation increased dramatically, whereas when it was absent, cooperation decreased at similar rates (Fehr & Gächter, 2002). Remarkably, research has also shown that in some countries, individuals playing public goods games in groups with fixed composition (hereby "partners") punish cooperators and defectors similarly (Herrmann, Thöni, & Gächter, 2008; Gintis, 2008). The widespread phenomenon of punishing cooperators, termed "antisocial punishment," was interpreted as costly *revenge* by low contributors against cooperators suspected of punishing them in previous rounds (Herrmann et al., 2008; Gintis, 2008, O'Gorman, Henrich, and Van Vugt, 2009). Studies have found that antisocial punishment is associated with the country's level of political rights, civil liberties, freedom of the press, and corruption as indexed by the World Democracy Audit (WDA). Participants from countries with poor performance on the WDA, such as Oman, Saudi Arabia, Greece, and Russia, showed high levels of antisocial punishment, whereas participants from countries with high performance on the WDA, such as the United States, the UK, Germany, and Australia, showed low levels of antisocial punishment (for details, see Herrmann et al., 2008).

In the present paper, we reconsider the generality of both explanations. Using data from a novel public good experiment with punishment, conducted on Israeli students, and from a study published in *Science*, on 16 public goods experiments from countries around the world (Hermann et al., 2008), we show the emergence of cooperation is mainly due to *contributors*, rather than non-contributors, increasing their cooperation. We also show that the bulk of what is considered antisocial punishment is not revenge by low-cooperators against high cooperators, but rather part of an *upward and downward* punishment strategy, presumably aimed at punishing deviations from what the punisher considered norm violators.

Do high cooperators get punished as revenge?

As noted above, the punishment of high cooperators in explained in as revenge by low cooperators who suspected high cooperators had punished them in previous rounds (e.g.,

Hermann et al., 2008; Gächter & Herrmann, 2009). In Hermann et al. (2008), *across all 16 participant pools*, a significant relationship between the punishment points assigned in time t and the punishment points received in period t-1 substantiated this explanation. A closer look at the regression coefficients of the separate participants pools, reported in Table S3B in Hermann et al.'s supporting material, reveals that among eight participant pools that demonstrated low to moderate contribution levels and high 'antisocial' punishment (Muscat, Athens, Riyadh, Samara, Minsk, Istanbul, Seoul, and Dnipropetrovs'k), the reported Tobit regression estimates between the punishment points assigned in period t and the punishment received in period t-1 were significant at $p < 0.01$ *only* for two participant pools (Samara and Seoul). On the other hand, the reported regressions for five participant pools that demonstrated high contribution levels and low 'antisocial' punishment (Boston, Nottingham, Bonn, Zurich, and St. Gallen), the regression estimates were significant (at $p < 0.01$). Thus, while the 'revenge' explanation is supported by the aggregate data, it seems more suitable to account for the punishment of cooperators in high contributing pools, than in low contributing pools, where punishment of cooperators was particularly prominent.

Experiment

To re-examine the revenge explanation, we ran a public goods experiment with punishment on 192 Israeli participants, all students at the University of Haifa. The experiment included a "partners" treatment and a "strangers" treatment. In the 'partners' treatment, the composition of the group was fixed for the all rounds of the game, whereas in the 'strangers' treatment, the groups' composition was changed, such that no subject ever met another subject more than once in all the rounds. If punishment of high contributors is indeed an act of revenge by low contributors who suspected the former of punishing them in previous rounds, we would expect to find considerably lower antisocial punishment in the 'strangers' treatment, compared to the 'partners' treatment, if any.

Method

In both treatments, groups of four members played a phase of six rounds of a public goods game with punishment, followed by another phase of six rounds of a similar game without punishment. In each round, participants received an endowment of 20 monetary units (MUs), each, from which they could contribute any amount between 0 and 20 MUs to a group project. Each MU kept by a group member yielded one MU, and each MU contributed yielded 0.4 MUs for each group member. Because the cost of contributing one MU to the project was exactly one MU, whereas the return on that MU was only 0.4 MUs, keeping all one's MUs was always in the participant's material self-interest, irrespective of how much the other three group

members contributed. Yet if each group member retained all of his or her MUs, there were no earnings to be shared. On the other hand, each member would earn 0.4 × 80 = 32 MUs if all group members invested their entire 20 MU endowment in the project.

In the second stage of each period (the punishment stage), after all group members were informed about their investments, each group member could punish each of the other group members. A punishment decision was implemented by assigning the punished member between zero and 10 deduction points. Each deduction point assigned reduced the punished member's earnings by three MUs and cost the punishing member one MU. All punishment decisions were made simultaneously, and participants were not informed who had punished them.

We ran eight experimental sessions of 24 subjects each, three in the 'partners' treatment and five in the 'strangers' treatment. All the experimental sessions were computer mediated. In running the sessions, we used the "Zurich toolbox for ready-made economic experiments" (z-Tree) software (Fischbacher, 2007).

In the 'partners' treatment, participants interacted only within their groups; thus, the unit of analysis for this treatment is the group (18 observations). On the other hand, in the 'strangers' treatment, all members of a session interacted with each other; thus, the unit of analysis for this treatment is the entire session (five observations).

Cooperation

Before turning to the main question of testing the revenge explanation, we briefly summarize the contribution results. The mean contributions in each trial, under the two treatments, are depicted in Figures 1a and 1b. The efficacy of punishment in enhancing cooperation in the two treatments is quite moderate, compared to the rates detected in cities from Western countries such as Boston, Copenhagen, and Zurich, and is closest to the cooperation levels observed in non-western cities, such as Samara and Minsk. On average, in the partners treatment, punishment increased the contribution level from 10.63 MUs in the first two periods to 12.40 MUs in the last two periods (the difference is significant at $p = 0.02$ using the signed-rank test). In the strangers treatment, punishment increased the mean contribution from 8.33 NUs in the first two periods to 9.67 MUs in the last two periods (the difference is not significant, $p = 0.63$ using the signed-rank test). Interestingly, whereas the removal of punishment in the partners treatment was associated with only a slight drop of 0.94 MUs in the mean contribution (from 11.67 to 10.73 MUs, the difference is not significant at $p = 0.52$ using the signed-rank test), the removal of punishment in the strangers treatment was associated with a drop of 2.98 MUs in the mean contribution (from 9.27 to 6.29 MUs, difference is significant at $p = 0.0625$ using the signed-rank test).

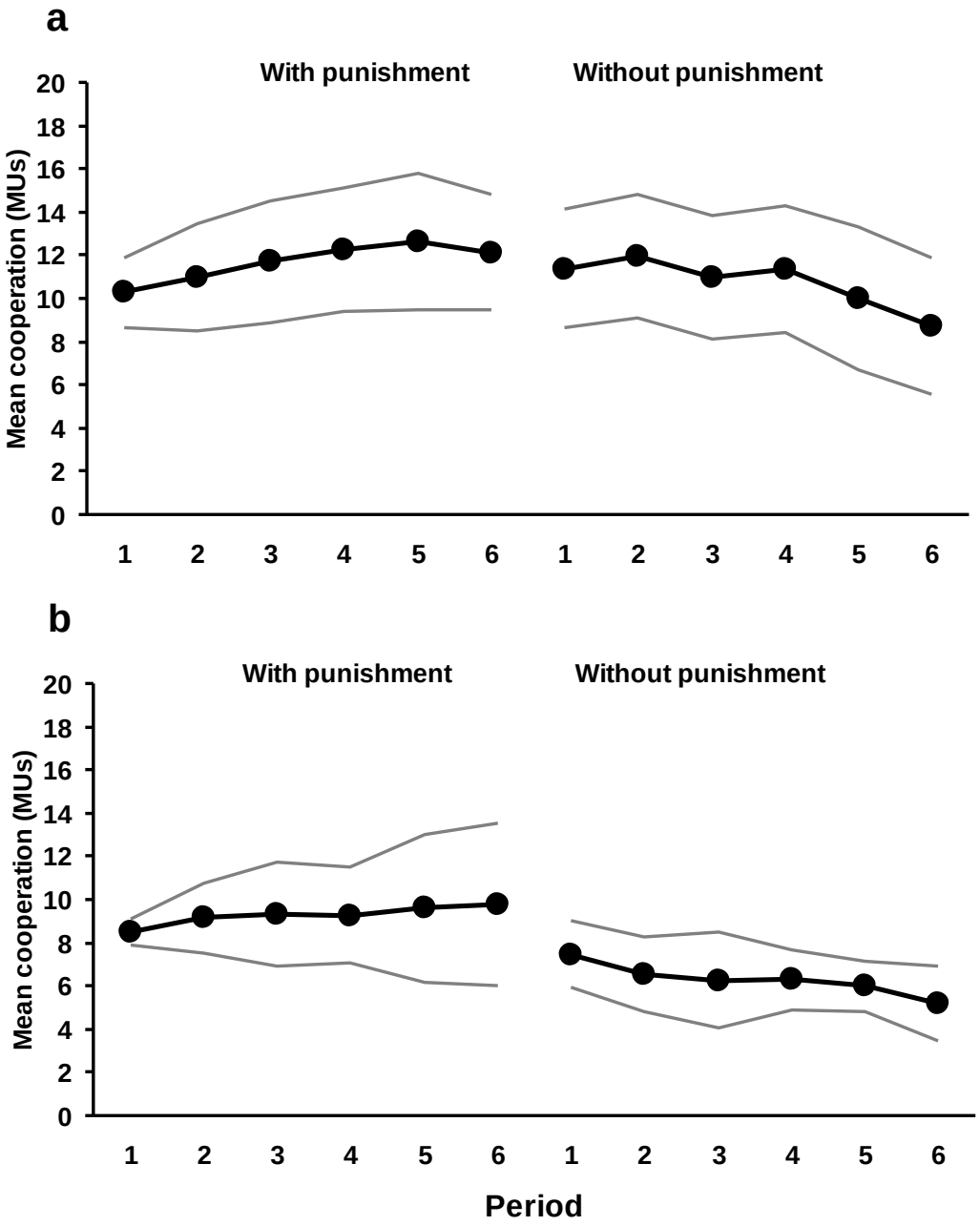


Figure 1: Mean contribution for partners (Figure a) and strangers (Figure b) as functions of period in the punishment and the no-punishment phases

Punishment

The punishment expenditures in the two treatments are shown in Figure 2, which depicts the expenditures on punishment as functions of the deviation of the contribution of the punished group member from the contribution of the punishing member.

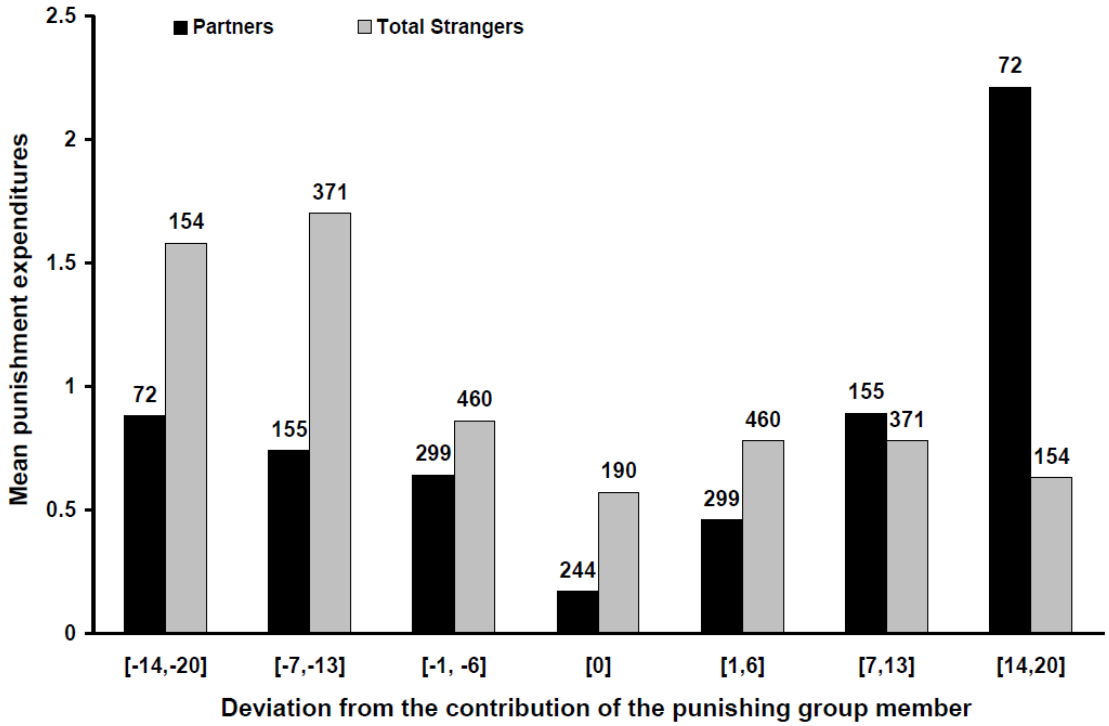


Figure 2: Mean expenditures on punishment (in points/round) as functions of the deviation of the contribution of the punished group member from the contribution of the punishing member. The bars corresponding to negative deviations depict mean points allocated to punish free riders, and the bars corresponding to non-negative deviations depict mean points allocated to antisocial punishment.

The bars corresponding to negative deviations in the histogram indicate mean expenditures on punishing free riders, and the bars corresponding to non-negative deviations indicate mean expenditures on punishing cooperators. The figure above each bar indicates the number of observations in the respective interval. As the figure shows, the patterns of punishment in the two treatments are similar. In both treatments, the higher the absolute difference between the contributions of the punished player and the punishing player, the more punishment points participants assigned.

Figure 3 depicts the overall mean expenditures on punishing free riders and on punishing cooperators in the two treatments. As the figure shows, the expenditure on punishing free riders was significantly higher in the partners than in the complete-strangers" treatment (3.87 and 2.17 MUs per punishment opportunity, respectively, $p = 0.019$, two-sided Wilcoxon exact test). We attribute this difference to the fact that, on average, contribution was lower in the 'strangers' than in the 'partners' treatment. On the other hand, the figure reveals that the mean expenditures on punishing cooperators in the strangers treatment was almost equal to expenditure of

punishing cooperators in the partners treatment (the difference insignificant, p = 0.108, two-sided Wilcoxon exact test).

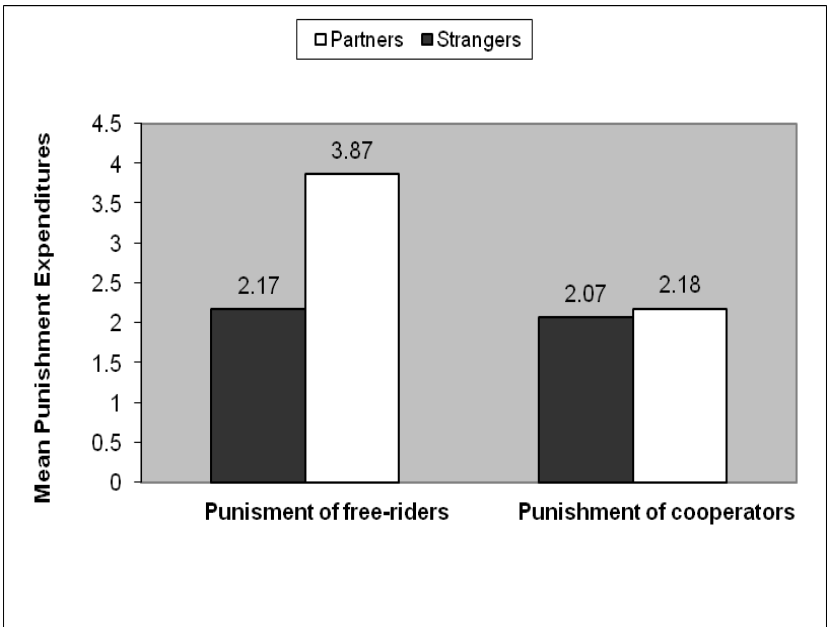


Figure 3: Mean expenditures on punishing free riders and cooperators

We also tested the revenge explanation directly. For this purpose, we conducted a series of Tobit regression analyses on the points assigned to punishment in period t as a function of the punishment points received in period t-1, the contribution of the punished group member at time t, the contribution of the punishing group member at time t, and the average contribution of the other two group members at time t. We used a Tobit estimation procedure to account for the fact that the dependent variable is censored at 0 and 10 punishment points (Tobin, 1958; Schnedler, 2005). Table 1 depicts the results of the Tobit analyses. Inspection of the table reveals that, contrary to the revenge explanation, in the two treatments, punishment of cooperators in period t was not correlated with the punishment received in period t-1.

The table also shows that in both treatments, the contribution of the punished group member best explains the points assigned to punishment. In the partners treatment, the regression coefficient between the assigned punishment points and the punished contribution is – 0.203 ($p < 0.0001$), and in the strangers treatment, the regression coefficient is – 0.366 ($p < 0.0001$). The negative coefficients indicate that the less a group member contributes, the more punishment points he or she receives. Interestingly, in the partners treatment, the punishment of free riders is significantly correlated with the mean contribution of the other group members (regression coefficient of 0.121, $p < 0.0001$) but not with the punisher's contribution ($p > 0.6$), whereas in

the strangers treatment, the punishment of free riders is significantly correlated with the punisher's contribution (regression coefficient of 0 .101, $p<0.05$) but not with the mean contribution of the other group members ($p > 0.5$). These findings suggest that for situations in which the groups' composition is kept fixed, a collective benchmark partly explains the punishment of free riders. Conversely, when the group composition is varied from period to period, an individualistic benchmark partly explains the punishment of free riders.

Table 1

Results of Tobit regression analyses on the points assigned to punishment as the dependent variable

Explanatory variable	Punishment of free riders (Negative deviations from the punisher's contribution)		Punishment of cooperators (Non-negative deviations from the punisher's contribution)	
	Partners	Strangers	Partners	Strangers
Received punishment in t-1	-0.124 (0.069)	0.126** (0.044)	-0.057 (0.068)	0.084 (0.052)
Punished contribution	-0.203**** (0.033)	-0.366**** (0.046)	0.087* (0.038)	0.017 (0.037)
Punisher Contribution	-0.015 (0.034)	0.101** (0.036)	-0.344**** (0.034)	-0.0803 † (0.042)
Mean contribution of others	0.121**** (0.031)	0.021 (0.037)	0.307**** (0.037)	-0.031 (0.041)
Constant	0.194 (0.526)	0.015 (0.631)	-2.634**** (0.616)	-1.458* (0.612)
P	0.000	0.000	0.000	0.002
N	152	342	145	252

† p=0.054, * $p < 0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$

Other Punishment Schemes

In search of an alternative explanation to the observed punishment of cooperators, we looked at participants who consistently used the same punishment scheme in all rounds of the punishment phase. This classification yields four types: group members who cooperated but never punished ('non-punishers'); group members who cooperated and punished low contributors (strong reciprocators); group members who punished upward and downward relative to their own contribution, and group members who punished only cooperators (antisocial punishers). Table 2 depicts the relative frequencies of each type, together with the means (and SDs) of the punishment points, and the contribution of each type.

Table 2
Relative frequencies, punishment investments and contributions by punisher's type

Variable	Punishers' Types							
	Non-punishers		**Punishers of Free riders**		**Punishers of cooperators**		**Punishers of free riders & cooperators**	
	Partners n =21	Strangers n=27	Partners n =20	Strangers n=38	Partners n =2	Strangers n=6	Partners n =22	Strangers n=42
Relative Frequency	32.3%	23.9%	30.8%	33.6%	3.1%	5.3%	33.8%	37.2%
Punishment (MU/period)	-	-	1.60 (1.66)	2.46 (2.62)	3.92 (1.77)	1.75 (1.26)	2.92 **(*)** (1.96)	5.91 **(**)** (5.31)
Contribution (MU)	14.56 (5.96)	7.64 (5.38)	12.94 (5.96)	12.88 (3.92)	12.50 (4.01)	8.72 (5.01)	9.96 (5.34)	8.14 (3.41)

(*) The 2.92 MUs/period consists of 1.34 MUs/period for punishing free riders plus 1.58 MUs/period for punishing cooperators.
(**) The 5.91 MUs/period consists of 2.81 MUs/period for punishing free riders plus 3.10 MUs/period for punishing cooperators.
Seven observations in each treatment, corresponding to cases in which a group member had no opportunity to use either type of punishment, were deleted.

The table shows that, with the exception of a slightly lower ratio of 'strangers' who never punished, participants in each treatment are distributed almost evenly between non-punishers, strong reciprocators, and participants, who punished free riders and cooperators alike. In both

treatments, the ratio of participants who punished only cooperators was negligible (3.1% and 5.3% for partners and strangers, respectively). Thus, what was previously interpreted as antisocial punishment appears to be predominantly part of an upward-downward punishment strategy that some participants have used against deviants from their own contributions.

We also compared the mean contributions of non-punishers, strong reciprocators and upward-downward punishers. For the 'partners' treatment, a generalized estimating equations (GEE) analysis (Liang and Zeger, 1986; Rotnitzky and Jewell, 1990) revealed that cooperators who did not punish and strong reciprocators contributed more than upward-downward punishers ($\chi^2_{(1)}$=8.22, p=0.0041 and $\chi^2_{(1)}$=4.27, p=0.039, respectively). A similar analysis for the 'strangers' treatment revealed that strong reciprocators contributed more than cooperators who did not punish ($\chi^2_{(1)}$=8.32, p=0.0039) and more than the upward-downward punishers ($\chi^2_{(1)}$=89.43, p<.0001).

We also reanalyzed the Hermann et al. data using the above detailed typology. Table 3 depicts the results for two participant pools with the highest mean contribution (Boston and Nottingham), and two participant pools with the lowest mean contribution (Athens and Riyadh). The complete table for the 16 investigated participant pools is depicted in Table 1c, Appendix C. The top figure in each cell in Table 3 indicates the percentage of participants who used the respective strategy. The figure below indicates the participants' numbers, and bottom figure indicates the mean contribution of participants using the respective strategy. Table 3 and the complete table (see Appendix C) reveal stark differences between high- and low-contributing pools. For example, out of the participants in Boston (who contributed, on average, 18 MUs out of 20), 41% contributed 18.37 MUs and never punished, 45% contributed 18.37 MUs and punished free riders, 10% contributed 16.35 MUs and punished low and high contributors, and 4% contributed 12.55 and punished only high contributors. By contrast, the results for Athens show that the percentages of non-punishers and punishers of free riders were considerably low (11% and 10%, respectively), with a majority of participants (55%) punishing both low and high contributors.

Overall, Tables 3 and 1c clearly show that participant pools in which punishment was effective in enhancing cooperation, are characterized by high percentages of strong reciprocators, and of cooperators who did not punish, along with low percentages of punishers of cooperators. Conversely, participant pools in which punishment was ineffective in enhancing cooperation are characterized by low percentages of strong reciprocators and cooperators who did not punish, along with high percentages of participants who punished both cooperators and non-cooperators. Also, consistent with the findings of the above reported experiment, the

percentage of participants who punished only cooperators was relatively low in most participant pools, regardless of their mean contribution levels.

Table 3

Relative frequencies of punisher's types in participant poolsfrom representative cities with high and low cooperation

(Data source: Hermann et al., Science, 2008)

City	Punisher's Types				Total
	Non Punishers	Punishers of free-riders	Punishers of cooperators	Upward/downward punishers	
Boston (high cooperation)	41% (23) 18.46	45% (25) 18.37	4% (2) 12.55	10% (6) 16.35	100% (56) 18
Copenhagen (high cooperation)	32.35% (22) 19.06	47.06% (32) 17.99	7.35% (5) 18.22	13.24% (9) 13.42	100% (68) 17.7
Riyadh (low cooperation)	20.83% 10 6.82	**18.75%** 9 8.79	8.33% 4 2.86	**52.08%** **25** 6.93	100% 48 6.9
Athens (low cooperation)	11.36% 5 7.00	**22.73%** 10 8.03	11.36% 5 0	**54.55**% **24** 5.65	100% 44 5.7

Punishment as deterrence

The common view holds that the importance of punishment in promoting cooperation lies in the fact that it deters free riders from exploiting the cooperation of others (Fehr and Gächter, 2002; Fehr and Fischbacher, 2004). We tested this argument by looking at the effect of punishment on participants who used different types of punishment strategies. For this purpose, we reanalyzed the raw data collected by Hermann et al. (2008), and compared the mean contributions of the four punisher types across all 16 participant pools. Figure 4 depicts mean contributions of the four punisher types in the no-punishment and punishment rounds of the

game (corresponding figures for representative pools from cities with high, moderate and low contributions, are included in Appendix D).

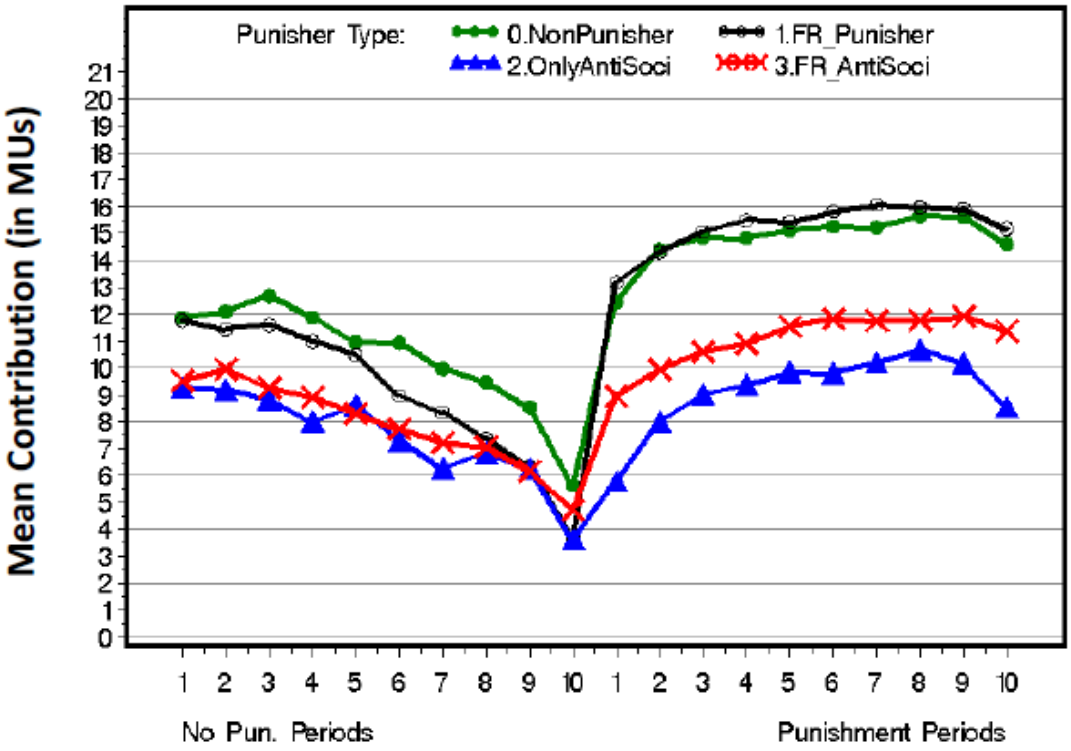
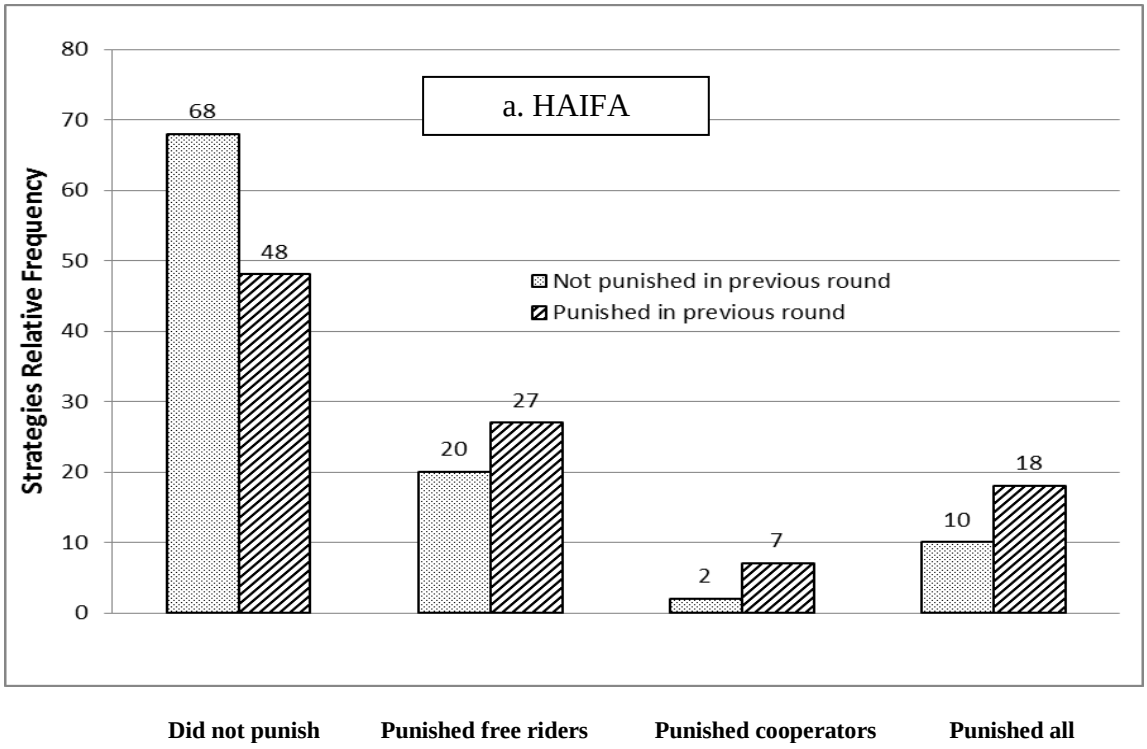


Figure 4: Mean contributions of punisher types in the no-punishment and punishment phases

Strikingly, the figure reveals that more than being instrumental in increasing the contribution of upward-downward punishers and of antisocial punishers, the possibility of sanctions was instrumental in increasing the contributions of strong reciprocators and cooperators who did not punish. The figure also shows that for strong reciprocators and non-punishing cooperators, the increase in cooperation in the first round of the punishment phase is more than the total increase in cooperation in all the subsequent rounds. Because in the experiment the contribution decisions were made prior to the punishment decisions, the significant increase in cooperation in the first round of the punishment phase is due to an anticipatory effect of punishment, rather than to the punishment itself, implying that the possibility of being punished was more effective than the actual punishment.

We also looked at the effect of participants being punished in a given round on their investment in punishing others in the subsequent round. Rather than looking at types of punishers, we looked at the sample space of all *punishment strategies* played at each round of the game. For a given round t, each player could have played one of four strategies: do not punish, punish only

low contributors, punish only high contributor, or punish both high and low contributors. Figures 5a and 5b depict the relative frequencies of the four punishment strategies in the Haifa's participant pool, in which the effect of punishment on contribution was moderate (Figure 5a), and the corresponding frequencies in Zurich, in which the effect of punishment on contribution was significant (Figure 5b). The light shaded bars in each figure correspond to the relative frequencies of the various punishment strategies in cases in which the focal player was not punished in the previous round, and the dark shaded bars correspond to the frequencies of the same strategy when the focal player was punished in the previous round. The figures show that in the two participant pools, the frequencies of the "cooperated but didn't punish" strategy was significantly lower when participants were punished in the previous round, compared to when they were not. In Haifa's pool, the relative frequency of "cooperated but didn't punish" was 20% less (48% compared to 68%). In Zurich's pool, the frequency of the same strategy was 26% less following punishment than following no-punishment (47% compared to 73%). Concurrently, the percentage of strong reciprocation (punish only free riders) increased by7% in Haifa and by 10% in Zurich. Also in Haifa, the frequency of "norm keeping" (punish high and low contributors) increased by 8%, and in Zurich, the antisocial punishment increased by 14%. In the total analysis, punishment seems to have caused a sizable minority of contributors who did not punish in a given round (about 30% in Haifa and 36% in Zurich) to switch to strong reciprocation in the subsequent round. Others seem to have responded to being punished either by punishing antisocially (in Zurich) or punishing high and low contributors alike (in Haifa). The intriguing possibility that these results raise is that despite its dark sides, punishment of cooperators has a bright side, because it causes a fraction of cooperators who did not punish to behave as strong reciprocators.



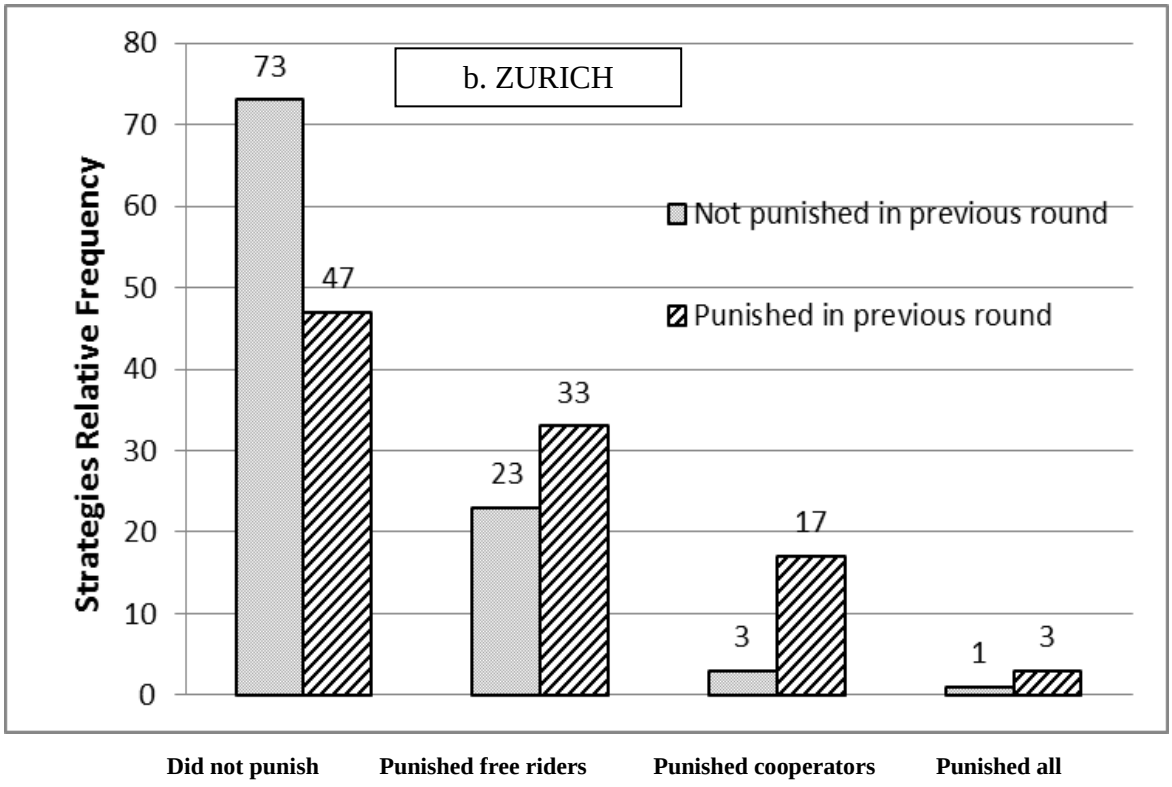


Figure 5: Relative frequencies of the four punishment strategies in Haifa (Figure 5a) and in Zurich (Figure 5b)

Strong reciprocators and cooperation

Research on evolutionary and experimental economics studies stresses the role of strong reciprocation in the emergence and sustenance of cooperation (e.g., Gintis, 2000; Gintis et al., 2003; Fehr, Fischbacher, & Gächter, 2002; Bowles & Gintis, 2004). Our typology of participants into various punisher types enabled a direct test of the role of strong reciprocators is promoting cooperation in public goods games. For each participant pool investigated in Hermann et al. (2008), we calculated the ratio of strong reciprocators relative to the total number of punishers in the respective pool, along with the contribution (and net profit) in the punishment phase relative to contribution (and net profit) in the no-punishment phase. Figures 6a and 6b depict the relative contributions and relative net profits in each participant pool as functions of the ratios of strong reciprocators in the respective pool. The resulting Pearson correlation coefficients between the ratio of strong reciprocators and the relative contributions was r=0.858 p= 0.0001), and the correlation between the ratio of strong reciprocators and the relative net earnings was r=0.962 (p= 0.0001).

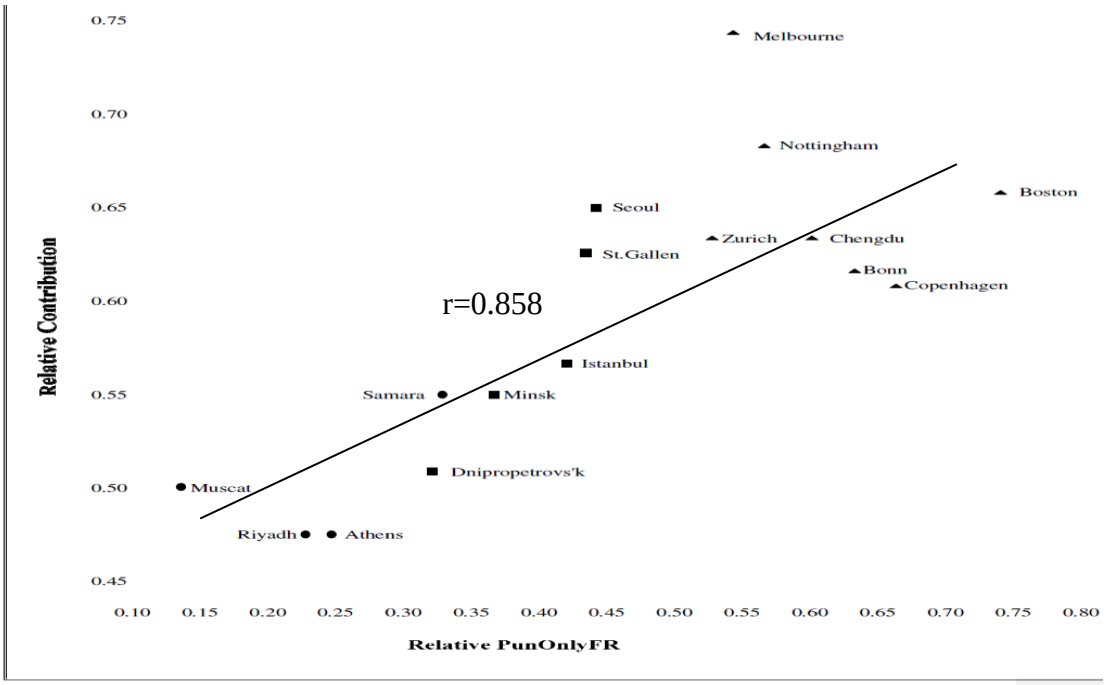


Figure 6a: Relationship between the relative contributions in sixteen participant pools from cities around the world, and the respective relative frequencies of strong reciprocators

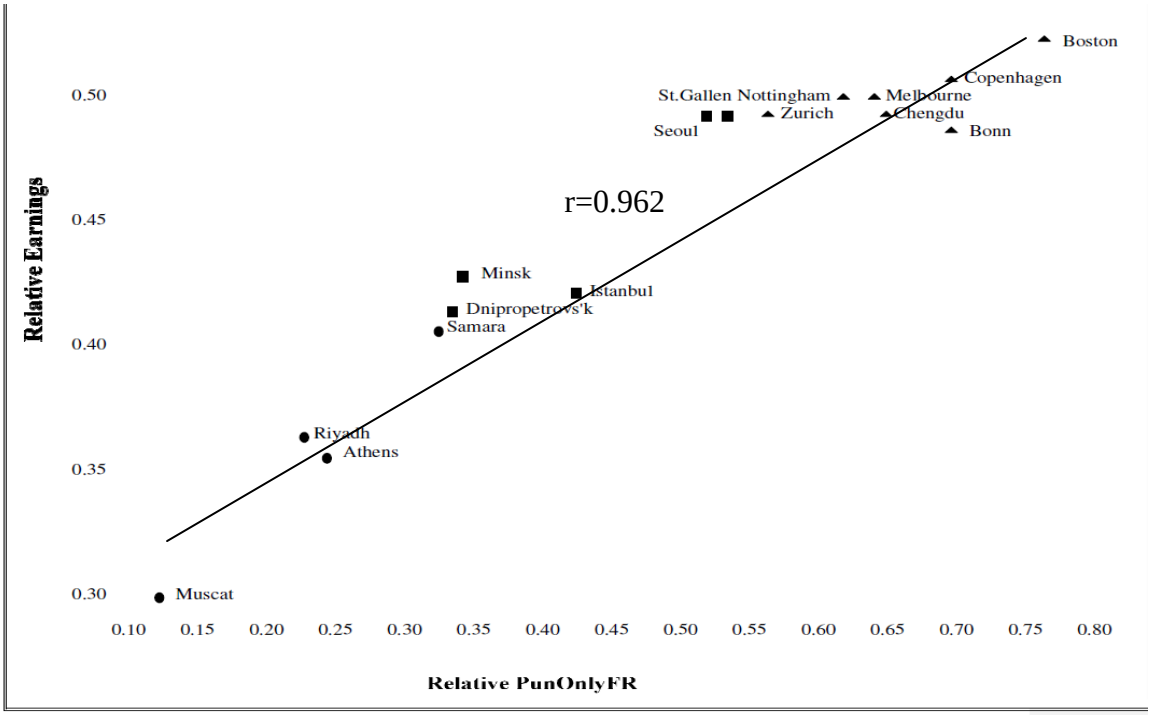


Figure 6b: Relationship between the relative earnings in sixteen participant pools from cities around the world, and the respective relative frequencies of strong reciprocators

We also correlated the ratio of strong reciprocators with the investigated countries' scores on various socio-economic and cultural factors, including the countries' norms of trust and civic cooperation (taken from the World Values Survey), the Rule of Law (taken from the World Bank), democracy and economic freedom (taken from the World Audit), the value orientations, survival versus self-expression and traditional versus secular-rational (Inglehart & Norris, 2003), and the country's GDP per capita (taken from the International Monetary Fund), in addition to Hofstede's four cultural dimensions: power distance, uncertainty avoidance (high score indicates more avoidance), individualism/collectivism (high score indicates more collectivism), and masculinity/femininity (Hofstede, 2005) (for the complete data and links to the data sources, see Appendix E). The highest correlations were with civic cooperation (r = 0.71, p=0.003), uncertainty avoidance (r = - 0.61, p= 0.02), power distance (r= -0.57, p =0.03), and democracy (r= - 0.54, p=0.03). Other variables that correlated moderately with the ratio of strong reciprocators were self-expression/survival (r=0.48, p=0.07) and economic freedom (r = 0.40, p= 0.07). R-square regressions performed on the ratio of strong reciprocators (as dependent variable), with various background variables (as independent variables), reveal that the three variables that best predict the strong reciprocators ratio are democracy, self-expression, and uncertainty avoidance (total r^2=0.70).

Summary and Concluding Remarks

In the present paper, we conducted an in-depth investigation of the role of punishment in the emergence and sustenance of cooperation. First, we reconsidered two widespread conventions concerning the functionality and dysfunctionality of punishment in providing public goods. One is the claim that punishment works as a deterrent to potential free riders, and the second is that antisocial punishment is revenge by low cooperators directed against high cooperators suspected of punishing them in previous rounds.

With regard to the conventional wisdom that punishment enhances cooperation, because it deters free riders, our findings show that despite its partial validity, this explanation is far from telling the complete story. Using data from a new experiment on participants from the city of Haifa, Israel, and from 16 experiments conducted by Hermann et al. (2008) on participant pools from cities around the world, we found that the effect of punishment on the emergence of cooperation is mainly due to *contributors* increasing their cooperation, rather than from free riders being deterred. Another unexpected phenomenon that our results suggest is that a sizable minority of contributors who get punished in a given round behave in the next round as strong reciprocators, and thus contribute their part in punishing free riders and in enhancing the overall group cooperation. Our findings also refute the conventional claim that punishment of high cooperators is an act of revenge by low cooperators. First, the data of our experiment reveal that punishment of cooperators was of the same magnitude in the strangers as in the 'partners' treatment. Second, in the two treatments in our experiment, the punishment points allocated in period t were not correlated with the punishment received in period t-1. Third, in our participant pool and in all participant pools reported in Hermann et al. (2008), the rates of punishers of only cooperators were quite low (see Tables 2, 3, & 1C). Fourth, in all participant pools, the norm keepers, who comprised the majority of punishers of high contributors, punished both high and low contributors. They also contributed about half of their endowments and thus are not low contributors, who might have acted vengefully against high contributors. They also invested as much in punishing free riders as in punishing cooperators (see Table 2).From a policy-making perspective, the revenge explanation leads to the conclusion that to shun revenge, the use of sanctions should be centralized in the hands of the state. In contrast to such a “Leviathan” solution (Hobbes, 1651), the proposed norm-enforcing explanation prescribes that a “mutual coercion” could be maintained (Hardin, 1968), provided that a common cooperation level, with lower and upper perimeters, could be agreed upon. If the people involved agree on the contribution level and no one exceeds it, there will be no excessive cooperation to be punished. The fourfold typology of punishment styles into

cooperators (who cooperate but do not punish), strong reciprocators (who cooperate and punish free-riders), antisocial (who punish only cooperators), and norm keepers (who punish low and high cooperators) yielded several important findings, which we summarize as follows:

(1) ***Reaction to punishment***: Contrary to the common wisdom that punishment affects free riders, causing them to increase their contributions, which in turn enhances groups' cooperation, we found that punishment enhances cooperation, mainly by increasing the contributions of strong reciprocators and cooperators who do not punish (Figure 4), and that it causes a fraction of cooperators who do not punish to act as strong reciprocators (see Figures 5a and 5b).

(2) ***Punishment anticipatory effect***: Interestingly, we found that the anticipation of being punished was more effective than the actual application of punishment, particularly with regard to high contributors. Worth noting is that the "anticipation effect" of punishment was as strong in all previous public goods games in which rounds with no punishment preceded the punishment phase. In all reported results, a dramatic overshoot in cooperation occurred in the first round of the punishment phase, before any punishment had been applied, but this effect was somehow obscured from investigators' eyes. From a policy-making perspective, the differential effect of an anticipated punishment on participants playing different punishment strategies (see Figure 4) suggests that institutions that plan to impose sanctions on rules' violators (e.g., driving above a speed limit) might fare better by not appealing primarily to frequent rule breakers. Our findings suggest that normative individuals, who usually cooperate with and abide by the law, would enhance their cooperation more than the rule breakers, who do seem to be sufficiently deterred before they experience actual sanctions for breaking the law.

(3) ***Heterogeneity***: Our findings lend strong support to the view that individuals vary in their use of sanctions against others to achieve different objectives. Whereas some might use sanctions to promote cooperation, others might use sanctions to enforce and sustain what they consider desirable cooperation norms. Recent findings indicating that high contributors are perceived as norm breakers who should be excluded from the group (Parks & Stone, 2010) and that participants in a reward treatment primarily rewarded group members whose contributions were equal to theirs (Sutter et al., 2010) support this view.

No less interesting than the stark difference between cooperative and less cooperate participant pools is the relative distribution of punisher types. As Tables 3 and 1C show clearly, the high-contributing participant pools consisted mainly of cooperators who did not punish and strong reciprocators who cooperated and punished non-cooperators, whereas low-contributing participant pools consisted mainly of norm keepers who punished high and low contributors alike. An interesting question, which experimental studies like ours cannot answer satisfactorily, is if the different mixes of punishment strategies are transitory or are in a relatively stable equilibrium. The second possibility is intriguing, because it suggests that cultural differences in producing public goods reflect wider cultural differences in the evolution of social contracts and norms. According to such a perspective, the evolved norm in countries where we detected high percentages of cooperators and strong reciprocators favors high cooperation and disfavors low cooperation, whereas the evolved norm in countries where we detected high percentages of punishers of both high and low contributors is a mediocrity norm that favors normative behaviors and disfavors deviations from the norm. The possibility that different mixes of punishment styles could be in equilibrium is supported by agent-based simulations, as well as by experiments, showing that instead of evolving into homogeneous populations, the population dynamics often evolve into heterogeneous populations comprising multiple strategy types that coexist at equilibrium (Dugatkin & Wilson, 1991; Lomborg, 1996; Aktipis, 2004; Bowles & Gintis, 2004; Kurzban & Houser, 2005; Rand et al., 2010).

(4) ***Strong reciprocity***: Many studies on the role of punishment in the evolution of cooperation have emphasized the crucial role of strong reciprocity in promoting and sustaining cooperation (e.g., Gintis, 2000; Gintis et al. 2003; Bowles & Gintis, 2004; Fehr & Gintis, 2007). Gintis et al. (2003, p. 154) argued that "the evolutionary success of our species and the moral sentiments that have led people to value freedom, equality, and representative government, are predicated upon strong reciprocity and related motivations". Our study lends strong support to this view by showing that the ratio of strong reciprocators in a group is a potent predictor of the group's level of cooperation ($r \approx 0.86$) and its success in providing the public goods ($r \approx 0.96$!). In addition to its predictive potency, the ratio of strong reciprocators was found to be strongly associated with prominent, well-documented, socio-economic and cultural factors, primarily the countries' levels of civic cooperation, economic freedom, and the cultural dimensions of power distance and uncertainty avoidance (Hofstede, 2005). It follows that cooperation and efficiency in the production of public goods in different societies may be predicted *ex ante*, by knowing the cultures' scores on the aforementioned macro-level factors, or alternatively, by

measuring the rates of strong reciprocation in experiments on representative participant pools from the investigated cultures.

References

Aktipis, C. A. Know when to walk away: contingent movement and the evolution of cooperation . *Journal of Theoretical Biology* 231, 249–260, 2004.

Bowles, S., and Gintis, H. The evolution of strong reciprocity: cooperation in heterogeneous populations, Theoretical Population Biology 65 (1), 17–28, 2004.

Boyd, R., and Richerson, P.J. Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology* 13:171–195, 1992.

Dugatkin, L. A., and Wilson, D. Rover: a strategy for exploiting cooperators in a patchy environment. *The American Naturalist* 138, 687–701, 1991.

Fehr, E., and Gächter, S. Altruistic punishment in humans. *Nature* 415: 137-140, 2002.

Fehr, E., and Fischbacher, U. Social norms and human cooperation. *Trends in Cognitive Sciences* 8: 185-190, 2004.

Fehr, E., and Gintis, H. Human motivation and social cooperation: experimental and analytical foundations, *Annual Review of Sociology* 33, 43-64, 2007.

Fehr, E., Fischbacher, U., and Gächter, S. Strong reciprocity, human cooperation, and the enforcement of social norms, *Human Nature* 13 (1):1-25, 2002.

Fischbacher, U. z-Tree: Zurich toolbox for ready-made economic experiments, *Experimental Economics* 10: 171- 178, 2007.

Gächter, S., and Herrmann, B. Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364: 791-806, 2009.

Gächter, S., Renner, E., and Sefton, M. The Long-Run Benefits of Punishment. *Science* 322:1510.

Gintis, H. Strong reciprocity and human sociality. *Journal of Theoretical Biology* 206:169-179, 2000.

Gintis, H. Punishment and cooperation. *Science* 319: 1345-1346, 2008.

Gintis, H., Bowles, S., Boyd, R., and Fehr E. Explaining altruistic behavior in humans. *Evolution and Human Behavior* 24 (3): 153-172, 2003.

Hardin, G. The tragedy of the commons, *Science* 162: 1243- 1248.

Hermann, B., Thöni, C, and Gächter, S. Antisocial punishment across societies. *Science* 319: 1362-1367, 2008.

Hobbes, T. *Leviathan*. Dent, London, 1651, 1947.

Hofstede, G., and Hofstede, G. J. *Cultures and organizations - software of the mind*, McGraw-Hill, New York, 2005.

Kurzban, R., and Houser, D. Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations. Proceedings of the National Academy of Sciences, USA 102 (5): 1803–1807, 2005

Liang, K.Y., and Zeger, S. L. Longitudinal data analysis using generalized linear models, *Biometrika* 73 (1): 13- 22, 1986.

Lomborg, B. Nucleus and shield: the evolution of social structure in the iterated prisoner's dilemma. *American Sociological Review*. 61, 278–307, 1996.

O'Gorman, R., Henrich, J., and Van Vugt, M. Constraining free riding in public goods games: designated solitary punishers can sustain human cooperation. *Proceedings of the Royal Society B* 276: 323-329, 2009.

Parks, C., and Stone, A. B. The desire to expel unselfish members from the group. *Journal of Personality and Social Psychology* 99 (2): 303–310, 2010.

Rand, D.G., Armao IV, J.J., Nakamaru, M., and Ohtsuki, H. Anti-social punishment can prevent the co-evolution of punishment and cooperation *Journal of Theoretical Biology* 265: 624 – 632, 2010.

Rotnitzky, A., and Jewell, N. P. Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data, *Biometrika* 77: 485- 497, 1990.

Schnedler, W. Likelihood estimation for censored random vectors, *Econometric Reviews* 24 (2), 195-217, 2005.

Sutter, M., Haigner, S., and Kocher, M. G. Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations. *Review of Economic Studies* 77: 1540-1566, 2010.

Tobin, J. Estimation of relationships for limited dependent variables, *Econometrica* 26: 24- 36, 1958.

Appendixes

Appendix A

Experiment's Instructions

In the following, we detail the complete instructions for the partners treatment. The instructions for the strangers treatment were identical. The only difference was that the participants were instructed that in each period of the experiment, they would be divided into groups of four members each and that the composition of the group would change from one period to another. We informed each participant that the composition of his or her group would change and that in each period, he or she would interact with different group members.

A.1. Instructions for the punishment phase in the partners treatment

"This is an experiment on decision-making. If you read the following instructions carefully, you can, depending on your decisions, earn a considerable amount of money. It is therefore very important that you read these instructions carefully.

It is prohibited to communicate with the other participants during the experiment.

If you have a question at any time raise your hand and the monitor will come to your desk to answer it.
During the experiment you can earn "money units". At the end of the experiment these money units will be converted to cash at the following rate:

1 Money Unit = 12 Agura[1]

At the end of the experiment your entire earnings from the experiment will be paid to you in cash. In addition to his or her earnings, each participant will receive 10 NIS for his or her participation in the experiment.

[1] Every 100 Agura equals 1 NIS (about $0.26)

The experiment consists of two parts. The instructions detailed below are for the first part. When this part ends, you will receive the instructions for the second part.
This part of the experiment is divided into **six** periods. In each period the participants are divided into groups of four. You will therefore be in a group with three other participants. The composition of the groups will stay the same for all six periods. **You are therefore with the same people in a group for all six periods.**

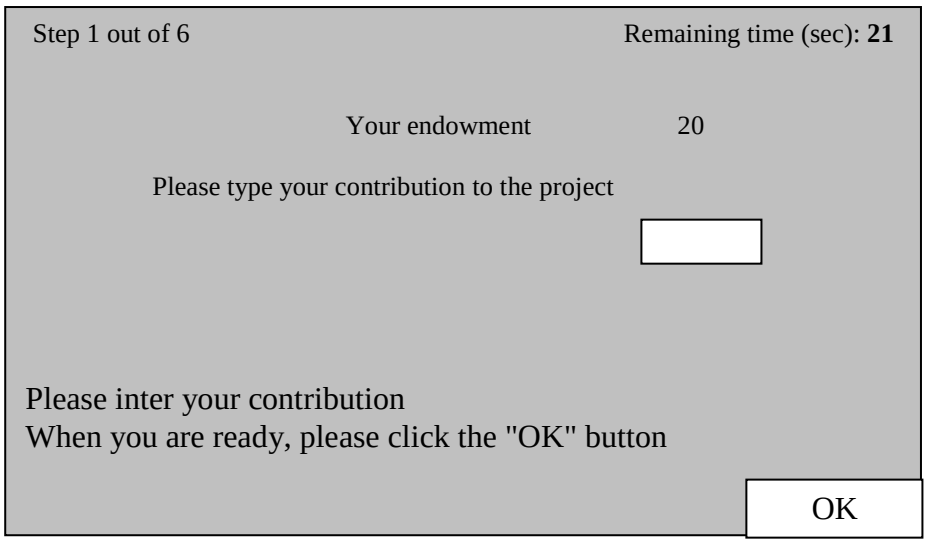
Each period consists of **two stages**. In the first stage you will be endowed with money units have to decide how many money units you would like to contribute to a project. In the second stage you will be informed of the contributions to the project by the three other group members. You will then decide whether or how much to reduce their earnings from the first stage by distributing points to them. The following pages describe the course of the experiment in detail:

Detailed Information on the Experiment

The First Stage

At the beginning of each period each participant receives an endowment of **20 money units**. Your task is to decide how many of the 20 money units you want to contribute to a **project** and how many of them to keep for yourself. The consequences of your decision are explained in detail below.
At the beginning of each period the following input-screen will appear:

The Input Screen:

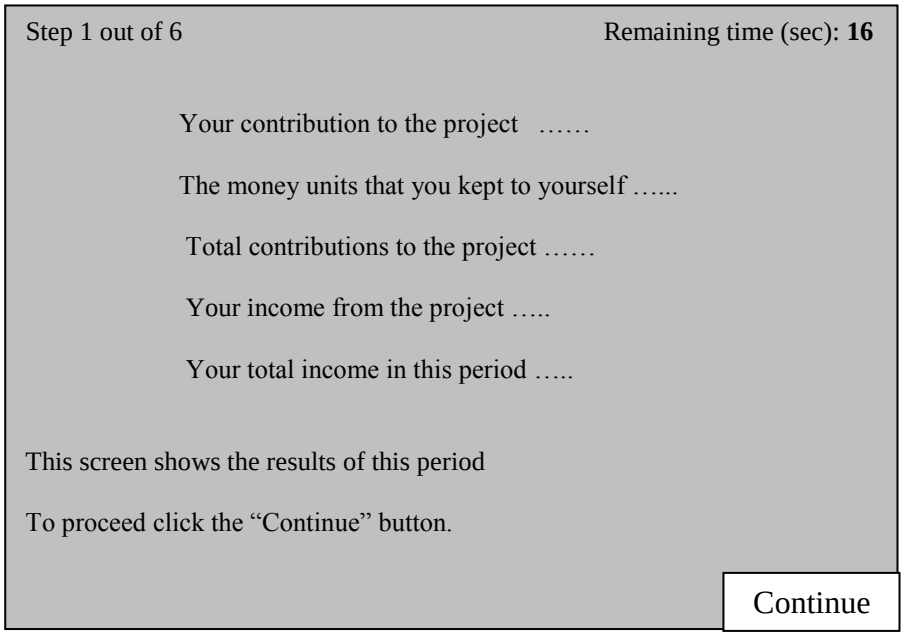


The number of the period appears in the upper left side of the screen. In the upper left side you can see how many seconds are left for you to make your contribution decision. You must make your decision before this time has passed.

Your **endowment in each period is 20 money units**. You have to decide how many money units you want to contribute to the project by typing a number between 0 and 20 in the input field. This field can be reached by clicking it with the mouse. As soon as you have decided how many money units to contribute to the project, you have also decided how many money units you keep for yourself: This is (**20 money units - your contribution**). After entering your contribution you must press the O.K. button. Once you have done this, your decision has been made and cannot be changed.

After all members of your group have made their decisions the following income screen will show you the total amount of money units contributed by all four group members to the project (including your contribution). Also this screen shows you how many money units you have earned in the period.

The Income Screen:



As you can see, your **income** consists of two parts:

(1) The money units which you have kept for yourself,

(2) Your "income from the project". This income is calculated as follows:

Your income from the project = 0.4 *times* the sum of contributions to the project.

Your **total income in the period,** in money units, is equal:

(20 – your contribution to the project) + 0.4 × (total contributions to the project)

The income of each group member from the project is calculated in the same way, i.e., each group member receives the same income from the project. Assume, for example, that the sum of the contributions of all group members is 60 money units. In this case each member of the group receives an income from the project of: 0.4 × 60 = 24 money units. If the total contribution to the project is 9 money units, then you and all other group members receive an income of 0.4 × 9 = 3.6 money units from the project.

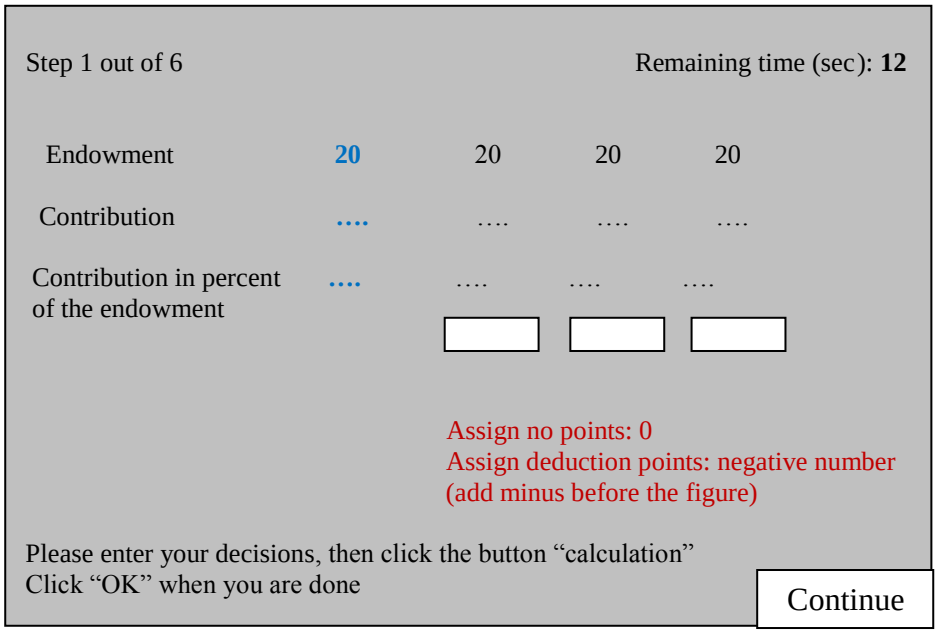
For each money unit you keep for yourself you earn 1 money unit. Supposing you contributed this money unit to the project instead, then the total contribution to the project would rise by one money unit. Your income from the project would rise by 0.4 × 1 = 0.4 money unit. However the income of each other group member would also rise by 0.4 money units each, so that the total income of the group from the project would rise by 1.6 money units. Your contribution to the project therefore also raises the income of the other group members.

Similarly, you earn an income for each money unit contributed by the other members to the project. For each money unit contributed by any member you earn 0.4 × 1 = 0.4 money units.
After you have viewed the income screen the first stage is over and the second stage commences.

The Second Stage

In the second stage you will see how much each group member contributed to the project.
Moreover, in this stage you can decide whether to **decrease** the income of each other group member by assigning **deduction points**. The other group members can also decrease your income if they wish to. This is apparent from the input screen at the second stage:

The Second Stage Input Screen:



Your contribution is displayed **in blue in the first column**, while the contributions of the other group members of this period are shown in the remaining three columns. Note that the order in which others' contributions are displayed will be determined at random in every period. The contribution in the second column, for example, could represent a different group member in different periods. The same holds true for the third and fourth columns. You will have to decide how many deduction points to assign to **each** of the other three group members. You must enter a number for each of them. If you do not wish to change the income of a specific group member then you must enter 0. You can **assign up to 10 points to each group member**.

You will incur costs from assigning deduction points. Every deduction point you assign costs you 1 money unit. For example, if you assign 2 deduction points to one member, this costs you 2 money units; if, in addition, you assign 9 deduction points to another member this costs you an additional 9 money units; if, in addition you assign 0 deduction points to the third group member this will cost you 0 money units. In total you would have assigned 11 points and your **total costs** therefore amount to 11 money units.

After you have assigned points to each of the other three group members you must click the button "**calculation**" (see the second stage input screen). On the screen you will then see the

total costs of your assigned points. As long as you have not yet clicked the **OK button**, you can still change your decision. To recalculate the costs after a change of your assigned points, simply press the "calculation" button again.

If you assign 0 deduction points to a particular group member (i.e., enter "0"), you will not alter his or her income. However, if you assign **one deduction point** to a group member you will **decrease** the income of this group member by **3 money units**. If you assign a group member **2 deduction points** you will **decrease** the group member's income by **6 money units**, and so on. Each deduction point that you assign to another group member will reduce his or her income by 3 money units. Similarly, each deduction point assigned to you by another group member will reduce your first stage income by three money units:

Cost, in money units, of received deduction points = 3 × Sum of received deduction points.

How much the income at the second stage is decreased depends on the sum of deduction points received. For instance, if somebody receives **a total** of **3 deduction points** (from all other group members in this period), his or her income would be decreased by **9 money units**. If somebody receives a total of **4 deduction points,** his or her income is reduced by **12 points**.

There is one exception to this rule. If the cost of received deduction points exceeds the group member's first stage income, his or her first stage income will be reduced to zero.
However, even in this case the group member must still incur the costs of any deduction points he or she assigned.

Your total income from the two stages is therefore calculated as follows:
If the income from the first stage is greater than or equal to the cost of received deduction points:

Total income, in money units, at the end of the second stage = period income =

= Income from the first stage
– 3 × (sum of received deduction points)
– sum of deduction points you have assigned.

OR

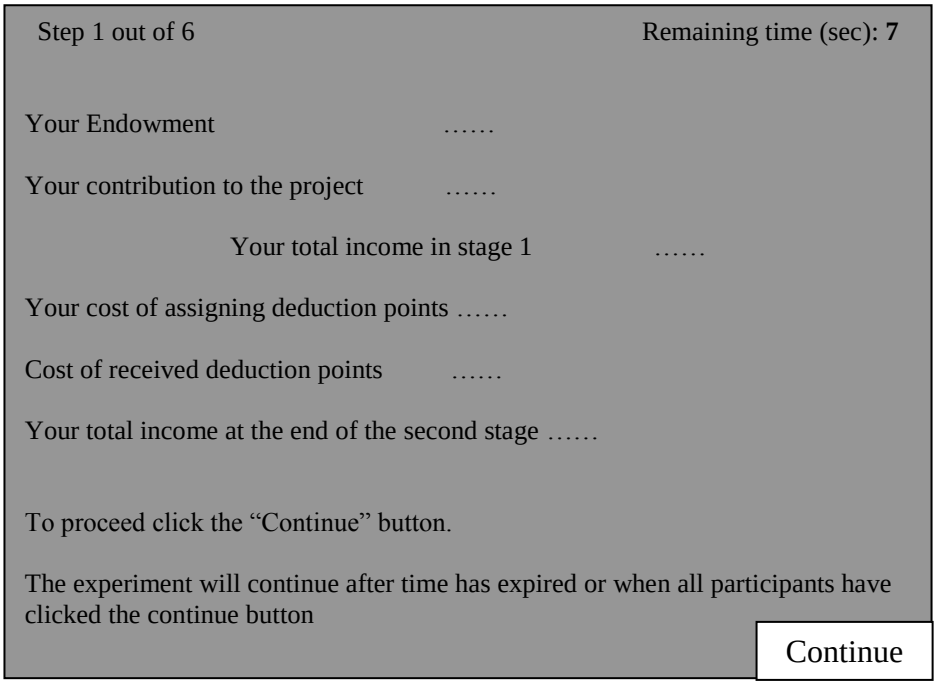
If income from the first stage is less than the cost of received deduction points:

Total income, in money units, at the end of the second stage = period income

= **0** – sum of deduction points you have assigned.

Please note that your income in money units at the end of the second stage can be negative if the costs of your assigned points exceed your income from the first stage minus the income reduction by the received deduction points. You can, however, avoid such losses with certainty through your own decisions!

After all participants have made their decision, your income from the period will be displayed on the following screen:



After you have viewed the income screen the period is over and the next period commences.

Now please complete the questions on the next sheet. They serve as a test of your understanding of payoff calculations. If you have any questions please raise your hand. When everyone has completed the questions correctly we will begin the decision-making part of the experiment".

A.2. Questions

1. Each group member has an endowment of 20 money units. Suppose nobody (including you) contributes any money units to the project. What is:

(a) Your first stage income? (20 - ___) + (0.4 X ___) = ___

(b) The first stage income of the other group members? (20 - ___) + (0.4 X ___) = ___

2. Each group member has an endowment of 20 money units. Suppose you contribute 20 money units to the project. All other group members contribute 20 money units each to the project. What is:

(a) Your first stage income? (20 - ___) + (0.4 X ___) = ___

(b) The first stage income of the other group members? (20 - ___) + (0.4 X ___) = ___

3. Each group member has an endowment of 20 money units. Suppose you contribute 10 money units to the project. All other members contribute a total of 30 money units to the project.

What is your first stage income? (20 - ___) + (0.4 X ___) = ___

4. Suppose at the second stage you assign the following deduction points to your three other group members: -9, -5 and 0.

What are the total costs of your assigned deduction points? _____

5. What are your costs if you assign a total of 0 deduction points? _____

6. By how many money units will the income of another player be reduced if you assign to him (-5) deduction points?

A.3. Instructions for the no-punishment phase in the 'partners' treatment

"This part of the experiment also contain six periods. It is identical in all aspects to the previous part, except that it does not include a possibility for assigning deduction points.

Like in the previous part, in each period your total income in this part is equal to:

(20 - your contribution to the project) + 0.4 × (total contributions to the project)

The income of each group member from the project is calculated in the same way.

When the six periods of this part of the experiment are completed, the experiment will end and you will receive the sum of your income in the two parts of the experiment.

Your total profit in the experiment will be equal to:

Your total income from the first six periods + your total income from the second six periods + 10 NIS [2] for your participation in the experiment

When the experiment ends, please wait patiently for the experimenter, who will call you and pay you your profits in the experiment.

When you have finished reading the instructions, please wait until the rest of the participants in the experiment finish reading. The second part of the experiment will begin after all the participants have finished reading the instructions".

Appendix B

Investment in punishment as function of the difference between the punished and the punishing group members

We performed Tobit regressions with punishment points as the dependent variable and the difference between the punished and the punishing group members as the independent variable. In the partners treatment, the reported robust standard errors are clustered on groups as the independent units of observations (S1), and in the strangers treatment, the reported standard errors are clustered on sessions as the independent units of observations.

The points assigned in the partners treatment for punishing cooperators are best explained by the punisher's contribution and the mean contribution of others. The less the punisher's contribution, the more points he or she assigned to punishing cooperators (regression coefficient of -0.334, $p<0.0001$). Correspondingly, participants assigned more points to punishing cooperators when the mean contribution of the two members in their group was higher (regression coefficient of 0.307, $p<0.0001$). This finding indicates that participants punish cooperators more when high contributions are more frequent. As could be expected, punishment of cooperators also correlated positively with the contribution of the punished group member (regression coefficient of 0.087, $p<0.05$). In the strangers treatment, the punishment of cooperators was negatively associated with the punisher's contribution (regression coefficient of - 0.083), but this association was only marginally significant ($p=0.054$).

For the punishment of free riders (negative deviations), the analysis revealed that for both treatments, the relationship between the assigned punishment points and the deviation between

[2] 10 NIS = $2.6

the punished and the punishing group members was positive and statistically significant (for the partners treatment, the regression coefficient = 0.078, z = 2.92, p = 0.003; for the strangers treatment, the regression coefficient = 0.180, z = 5.82, p< 0.0001). Conversely, we observe different patterns of punishing cooperators (non-negative deviations) in the two treatments. As the figure shows, in the partners treatment, the more the cooperators contributed relative to the punishing group members, the higher the points assigned to punishing them. We detected no similar relationship for the strangers treatment. Tobit regressions support these observations. In the partners treatment, the detected relationship between the points assigned for punishing cooperators and the deviation between the punished and the punishing group members was positive and statistically significant (regression coefficient = 0.277, z = 8.15, p < 0.0001). By contrast, in the strangers treatment, the regression analysis revealed that the relationship between the above-mentioned variables was not significant (regression coefficient = 0.029, z = 0.79, p = 0.43).

Table 1c

Frequencies and relative frequencies of four punishment strategies across the sixteen participant pools investigated in Hermann et al. (2008)

The top figure in each cell, in first four rows, indicates the percentage of participants who used the corresponding strategy. The figure below indicates the participants' number, and the two figures below it indicate the mean contribution and net-profit of participants using the strategy, correspondingly.

City	Punishment Strategy				Total
	NoPun	PunOnlyFR	PunCoop	Pun2Way	
Boston (18)	23 41.07% 18.46 28.85	25 44.64% 18.37 27.72	2 3.57% 12.55 26.27	6 10.71% 16.35 25.81	56
Copenhagen (17.7)	22 32.35% 19.06 29.87	32 47.06% 17.99 27.38	5 7.35% 18.22 27.96	9 13.24% 13.42 23.14	68
St. Gallen (16.7)	20 20.83% 17.12 27.76	41 42.71% 18.00 26.54	7 7.29% 13.24 25.84	28 29.17% 15.48 22.28	96
Zurich (16.2)	16 17.39% 17.35 28.70	45 48.91% 16.68 25.27	6 6.52% 16.42 28.08	25 27.17% 14.36 21.86	92
Nottingham (15)	10 17.86% 15.76 28.11	28 50.00% 16.30 24.67	5 8.93% 9.56 19.30	13 23.21% 13.76 21.55	56
Seoul (14.7)	13 15.48% 17.22 26.83	38 45.24% 15.46 24.75	3 3.57% 9.27 18.99	30 35.71% 13.05 22.32	84
Bonn (14.5)	13 21.67% 12.51 25.49	31 51.67% 15.92 24.65	4 6.67% 13.73 26.61	12 20.00% 13.18 20.46	60
Melbourne (14.1)	7 17.50% 13.03	21 52.50% 14.11	1 2.50% 14.90	11 27.50% 14.78	40

City	Punishment Strategy				Total
	NoPun	PunOnlyFR	PunCoop	Pun2Way	
	23.83	23.56	19.52	22.03	
Chengdu (13.9)	9 9.38% 16.49 28.71	56 58.33% 15.06 24.42	5 5.21% 11.36 24.70	26 27.08% 10.90 20.89	96
Minsk (12.9)	16 23.53% 15.77 26.71	19 27.94% 14.36 18.46	5 7.35% 6.90 16.66	28 41.18% 11.40 17.87	68
Samara (11.7)	16 10.53% 13.38 21.84	46 30.26% 14.14 20.34	8 5.26% 9.23 18.62	82 53.95% 10.29 15.58	152
Dniprop. (10.9)	6 13.64% 10.15 19.89	13 29.55% 13.13 20.74	6 13.64% 8.03 18.81	19 43.18% 10.47 17.33	44
Muscat (9.9)	5 9.62% 3.98 19.14	6 11.54% 15.27 9.69	8 15.38% 4.39 12.27	33 63.46% 11.16 9.63	52
Istanbul (7.1)	7 10.94% 4.97 19.41	25 39.06% 9.62 18.99	8 12.50% 4.74 15.24	24 37.50% 5.995 14.79	64
Riyadh (6.9)	10 20.83% 6.82 17.48	9 18.75% 8.79 14.95	4 8.33% 2.86 16.25	25 52.08% 6.93 11.65	48
Athens (5.7)	5 11.36% 7.00 19.80	10 22.73% 8.03 17.96	5 11.36% 0 0.11	24 54.55% 5.65 12.57	44
Total	198	445	82	395	1120

Appendix D

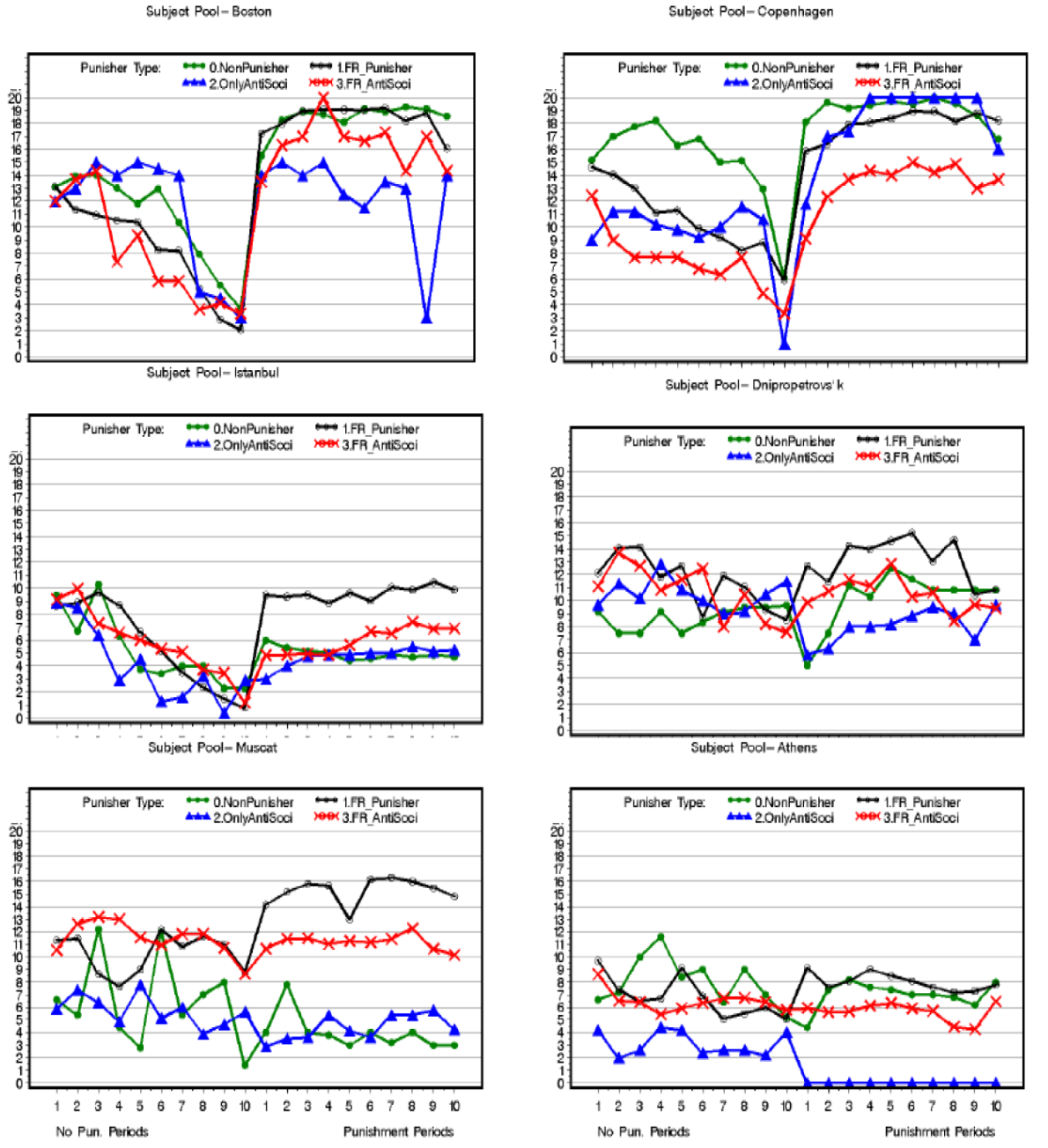


Figure 1d: Mean contributions (in MUs) of different punishers' types in the no-punishment and punishment phases for representative cities with high, moderate and low contributions

Appendix E

Table 1e

Socio-economic and cultural background variables

City	Country	Trust index	Level of trust	Civic cooperation	GDP per capita (in $1000)	Rule of law	Democracy	Self expression vs. survival	Secular vs. traditional	Economic freedom rank	Economic freedom score	Masculinity	Individualism	Uncertainly avoidance	Power distance
Melbourne	Australia	92.40	0.40	9.02	32.90	1.79	8	2.00	-0.20	4	82.5	61	90	51	36
Minsk	Belarus	85.20	0.42	6.91	8.90	- 1.23	137	-1.20	0.89	135	47.9				
Chengdu	China	120.90	0.55	9.34	7.60	-0.41	129	-0.61	1.16	135	52.0	66	20	30	80
Copenhagen	Denmark	131.90	0.67	9.27	36.50	1.94	2	1.96	1.11	8	78.6	16	74	23	18
Bonn	Germany	75.80	0.38	8.89	31.10	1.73	11	1.08	1.13	23	71.8	66	67	65	35
Athens	Greece	54.60	0.24	7.46	26.00	0.71	34	0.62	0.73	88	60.3	57	35	112	60
Muscat	Oman				18.80	0.75	99			34	69.7	53	38	68	80
Samara	Russia	55.40	0.24	8.05	12.10	-0.88	119	-1.86	1.08	143	50.5	36	39	95	93
Riyadh	Saudi Arabia	105.80	0.53	8.32	16.70	0.22	129	0.12	-1.35	54	66.2	53	38	68	80
Seoul	South Korea		0.27	8.83	23.90	0.73	33	-0.43	1.08	35	69.8	39	18	85	60
St.Gallen/ Zurich	Switzerland	107.40	0.37	8.58	37.40	1.96	5	1.45	0.77	5	81.9	70	68	58	34
Istanbul	Turkey	10.20	0.16	9.79	9.10	0.02	69	-0.35	-0.83	67	64.2	45	37	85	66
Nottingham	UK	61.70	0.29	8.65	35.10	1.72	10	1.37	0.26	16	74.5	66	89	35	35
Dniprop.	Ukraine	60.00	0.27	7.61	7.60	-0.74	129	-1.68	0.90	164	45.8				
Boston	USA	78.80	0.36	8.65	43.40	1.54	13	1.64	-0.53	9	77.8	62	91	46	40

Two 'social capital' variables (norms of trust and civic cooperation) are taken from the World Values Survey (1), the Rule of Law taken from the World Bank (2, 3), Democracy and Economic Freedom ranks are taken from the World Audit (4, 5), value orientations (survival vs. self-expression and traditional vs. secular-rational) (6), four cultural dimensions (power distance, uncertainty avoidance, individualism/collectivism and masculinity/femininity) (7) and the country's GDP per capita taken from the International Monetary Fund (8).

References

1. http://www.worldvaluessurvey.org
2. http://www.govindicators.org

3. D. Kaufmann, A. Kraay, M. Mastruzzi, *World Bank Policy Research Working Paper No.4280*, (2007); available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=999979.

4. http://www.worldaudit.org/

5. http://www.worldaudit.org/statsexcel.xls (accessed 11 February 2008).

6. R. Inglehart, P. Norris, *Rising Tide: Gender Equality and Cultural Change around the World* (Cambridge University Press, Cambridge, 2003).

7. G. Hofstede, G. J. Hofstede, *Cultures and Organizations. Software of the Mind* (McGraw-Hill, New York, 2005.

8. International Monetary Fund, *World Economic Outlook Database*, April 2007; available at: http://www.imf.org/external/pubs/ft/weo/2007/01/data/index.aspx.