

References and Notes

- J. D. Hays, W. C. I. Pitman, *Nature* **246**, 12 (1973).
- K. G. Miller *et al.*, *Science* **310**, 1293 (2005).
- A. Hallam, *Am. J. Sci.* **261**, 397 (1963).
- D. L. Turcotte, E. R. Oxburgh, *J. Fluid Mech.* **28**, 29 (1967).
- W. C. Pitman III, *Geol. Soc. Am. Bull.* **89**, 1389 (1978).
- D. B. Rowley, *Geol. Soc. Am. Bull.* **114**, 927 (2002).
- M. A. Kominz, in *Interregional Unconformities and Hydrocarbon Accumulation*, J. S. Schlee, Ed. (American Association of Petroleum Geologists, Tulsa, OK, 1984), vol. 36, pp. 109–127.
- X. Q. Xu, C. Lithgow-Bertelloni, C. P. Conrad, *Earth Planet. Sci. Lett.* **243**, 552 (2006).
- B. U. Haq, J. Hardenbol, P. R. Vail, *Science* **235**, 1156 (1987).
- C. G. A. Harrison, in *Sea-Level Change*, R. R. Revelle, Ed. (National Academy Press, Washington, DC, 1990), pp. 141–158.
- A. B. Watts, M. S. Steckler, in *Deep Sea Drilling Results in the Atlantic Ocean: Continental Margin and Paleoenvironment*, M. Talwani, Ed. (American Geophysical Union, Washington, DC, 1979), vol. 3, pp. 235–248.
- A. B. Watts, J. A. Thorne, *Mar. Petrol. Geol.* **1**, 319 (1984).
- G. M. Stampfli, G. D. Borel, *Earth Planet. Sci. Lett.* **196**, 17 (2002).
- A. Cox, B. R. Hart, *Plate Tectonics: How It Works* (Blackwell Science, Oxford, 1986).
- R. D. Müller, W. R. Roest, J.-Y. Royer, L. M. Gahagan, J. G. Sclater, *J. Geophys. Res.* **102**, 3211 (1997).
- R. D. Müller, M. Sdrolias, C. Gaina, W. R. Roest, *Geochem. Geophys. Geosyst.* 10.1029/2007GC001743 (2008); available at www.agu.org/journals/gc/.
- R. D. Müller, W. R. Roest, J.-Y. Royer, *Nature* **396**, 455 (1998).
- C. Heine, R. D. Müller, C. Gaina, in *Continent-Ocean Interactions Within East Asian Marginal Seas*, AGU *Monograph* **149**, P. D. Clift, D. E. Hayes, W. Kuhnt, P. Wang, Eds. (American Geophysical Union, Washington, DC, 2004), pp. 37–54.
- M. I. Ross, C. R. Scotese, *Tectonophysics* **155**, 139 (1988).
- D. C. Engebretson, A. Cox, R. G. Gordon, *Geol. Soc. Am. Spec. Pap.* **206**, 1 (1985).
- M. Nakanishi, K. Tamaki, K. Kobayashi, *Geophys. Res. Lett.* **19**, 693 (1992).
- T. Atwater, J. Severinghaus, in *The Eastern Pacific Ocean and Hawaii*, E. L. Winterer, D. M. Hussong, R. W. Decker, Eds. (Geological Society of America, Boulder, CO, 1990), vol. N, pp. 15–20.
- J. M. Whittaker *et al.*, *Science* **318**, 83 (2007).
- C. P. Conrad, C. Lithgow-Bertelloni, *Geology* **35**, 29 (2007).
- J. P. Cogne, E. Humler, V. Courtillot, *Earth Planet. Sci. Lett.* **245**, 115 (2006).
- C. A. Stein, S. Stein, *Nature* **359**, 123 (1992).
- A. G. Crosby, D. McKenzie, J. G. Sclater, *Geophys. J. Int.* **166**, 553 (2006).
- M. P. Doin, L. Fleitout, *Geophys. J. Int.* **143**, 582 (2000).
- G. Schubert, D. Sandwell, *Earth Planet. Sci. Lett.* **92**, 234 (1989).
- M. F. Coffin, O. Eldholm, in *Magmatism and the Causes of Continental Breakup*, B. C. Storey, T. Alabaster, R. J. Pankhurst, Eds. (Geological Society of London, 1992), pp. 21–34.
- T. H. Torsvik, M. A. Smethurst, K. Burke, B. Steinberger, *Geophys. J. Int.* **167**, 1447 (2006).
- D. L. Divins (National Geophysical Data Center, 2004), www.ngdc.noaa.gov/mgg/sedthick/sedthick.html.
- T. J. S. Sykes, *Mar. Geol.* **133**, 35 (1996).
- B. Parsons, *J. Geophys. Res.* **87**, 289 (1982).
- A sea-level rise Δs_l pushes the entire oceanic lithosphere down; thus, the resulting sea-level rise $\Delta s_l'$ is less than the changing mean ocean basement depth Δh , which is computed without this effect.
- P. L. Heller, C. L. Angevine, *Earth Planet. Sci. Lett.* **75**, 417 (1985).
- G. C. Bond, *Tectonophysics* **61**, 285 (1979).
- B. Steinberger, R. Sutherland, R. J. O'Connell, *Nature* **430**, 167 (2004).
- J. Ritsema, H.-J. van Heijst, J. H. Woodhouse, *J. Geophys. Res.* **109**, 10.1029/2003JB002610 (2004).
- T. W. Becker, L. Boschi, *Geochem. Geophys. Geosyst.* **3**, 48 (2002).
- S. P. Grand, *Philos. Trans. R. Soc. London Ser. A* **360**, 2475 (2002).
- B. Steinberger, *Phys. Earth Planet. Int.* **164**, 2 (2007).
- B. U. Haq, M. Al-Qahtani, *GeoArabia* **10**, 127 (2005).
- This paper benefited from reviews of early drafts by A. Dutkiewicz, J. Whittaker, and M. Gurnis, as well as from careful reviews by M. Kominz, R. O'Connell, and an anonymous reviewer, which improved the paper substantially. Ideas for New Jersey margin mantle-driven subsidence evolved during a sabbatical at the California Institute of Technology in discussions with M. Gurnis and S. Spasojevic. This work was funded by the Australian Research Council.

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5868/1357/DC1

Materials and Methods

Figs. S1 to S6

Tables S1 and S2

Database S1

9 October 2007; accepted 30 January 2008

10.1126/science.1151540

Antisocial Punishment Across Societies

Benedikt Herrmann,¹ Christian Thöni,² Simon Gächter^{1*}

We document the widespread existence of antisocial punishment, that is, the sanctioning of people who behave prosocially. Our evidence comes from public goods experiments that we conducted in 16 comparable participant pools around the world. However, there is a huge cross-societal variation. Some participant pools punished the high contributors as much as they punished the low contributors, whereas in others people only punished low contributors. In some participant pools, antisocial punishment was strong enough to remove the cooperation-enhancing effect of punishment. We also show that weak norms of civic cooperation and the weakness of the rule of law in a country are significant predictors of antisocial punishment. Our results show that punishment opportunities are socially beneficial only if complemented by strong social norms of cooperation.

Recent research has shown that altruistic punishment, that is, a person's propensity to incur a cost in order to punish free-loaders who fail to pull their weight in cooperative endeavors, can explain why genetically unrelated individuals are often able to maintain high levels of socially beneficial cooperation (1–4). This holds even when direct and indirect reciprocity (5, 6) or laws and regulations provide no incentives to behave cooperatively (7).

In this paper, we direct attention to a phenomenon that [with a few exceptions (8–10)] has been largely neglected: People might punish not only free-loaders, but cooperators too. For example, participants who had been punished in the past for contributing too little might retaliate against the cooperators because the cooperators are precisely those individuals most likely to punish the free-riding low contributors. Our experimental evidence from 16 participant pools with various cultural and economic backgrounds shows that antisocial punishment of prosocial cooperators is indeed widespread in many participant pools; interestingly, the participant pools in which most of the previous research on altruistic punishment has been conducted form the main exception.

Our observation of antisocial punishment grew out of our research goal to understand whether there are cross-societal differences in people's punishment and cooperation behavior. Previous large-scale cross-cultural evidence comes mainly from one-shot bargaining games conducted in small-scale societies around the world (11, 12). However, there is no systematic large-scale evidence on cooperation games. We therefore conducted cooperation experiments with and without punishment opportunities. Moreover, we ran our experiments as repeated games to see whether different cooperation levels emerge and remain stable across groups. Such a possibility is precluded in one-shot experiments.

Our research strategy was to conduct the experiments with comparable social groups from complex developed societies with the widest possible range of cultural and economic backgrounds (13) to maximize chances of observing cross-societal differences in punishment and cooperation. The societies represented in our participant pools diverge strongly according to several widely used criteria developed by social scientists in order to characterize societies (14–16). This variation, covering a large range of the worldwide available values of the respective criteria, provides us with a novel test for seeing whether societal differences between complex societies have any impact on experimentally observable disparities in cooperation and punishment behavior.

Experiments. The workhorse for our cross-societal analysis is the public goods game with and without punishment (1). The public goods game is a stylized model of situations that require

¹Centre for Decision Research and Experimental Economics, University of Nottingham, School of Economics, Sir Clive Granger Building, University Park, Nottingham NG7 2RD, UK. ²University of St. Gallen, FEW-HSG, Varnbuelstrasse 14, CH-9000 St. Gallen, Switzerland.

*To whom correspondence should be addressed. E-mail: simon.gaechter@nottingham.ac.uk

cooperation to achieve socially beneficial outcomes in the presence of free-rider incentives. Examples abound: warfare, cooperative hunting, voting, paying taxes, fighting corruption, contributing to public goods, teamwork, work morale, neighborhood watch, common pool resource management, recycling, tackling climate change, and so on. These are frequent situations with the common feature that cooperation leads to a group-beneficial outcome but is jeopardized by selfish incentives to ride free on others' contributions.

To implement a cooperation game with and without punishment opportunities, we adapted a design developed by (1). In each participant pool, we conducted the exact same public goods experiment with real monetary stakes and two treatment conditions: a no-punishment condition (the N experiment) and a punishment condition (the P experiment). Groups of four members played the following public goods game in both conditions: Each member received an endowment of 20 tokens. Participants had to decide how many tokens to keep for themselves and how many to contribute to a group project. Each of the four group members earned 0.4 tokens for each token invested in the project, regardless of whether he or she contributed any. Because the cost of contributing one token in the project was exactly one token whereas the return on that token was only 0.4 tokens, keeping all one's own tokens was always in any 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 tokens, there were no earnings to be

shared; on the other hand, each member would earn $0.4 \times 80 = 32$ tokens if each of them invested their entire 20-token endowment.

All the interactions in the experiment were computer-mediated (17) and took place anonymously. Participants were not informed about the identity of others in the group; they made their contribution decisions simultaneously, and, once the decisions were made, they were informed about the other group members' contributions.

The only and crucial difference between the P experiment and the N experiment was that participants in the P experiment could punish each of the other group members after they were informed about the others' investments, whereas the N experiment ended after participants were informed about the other group members' contributions. 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 tokens and cost the punishing member one token. All punishment decisions were made simultaneously. Participants were not informed about who punished them.

One of the goals of our experiment was to see whether and at what level punishment stabilized cooperation in the P experiment compared to the N experiment. To allow for the emergence of different cooperation levels, we therefore repeated the experiment 10 times under both conditions, keeping the group composition constant.

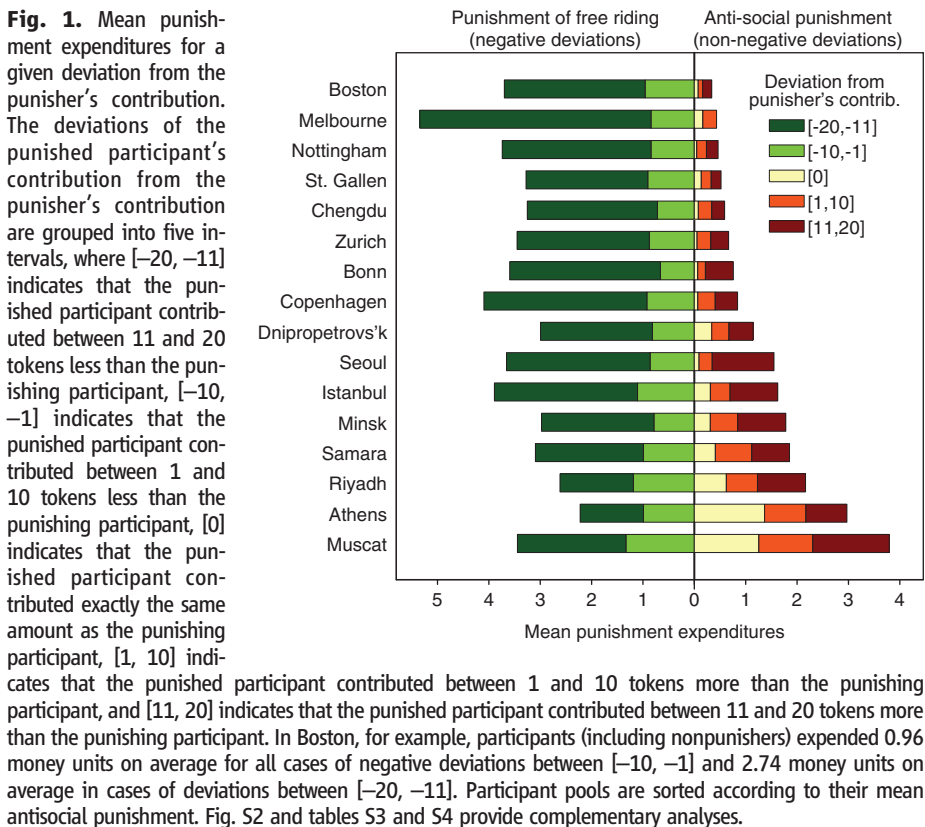
Because we were interested in whether people behave differently under the exact same cir-

cumstances, some methodological challenges arose. First, with regard to procedures, we followed the rules established in experimental economics (13). A second challenge was maximizing participant pool comparability to avoid confounds of participant pool differences with variations in sociodemographic composition. To minimize sociodemographic variability, we conducted all experiments with university undergraduates ($n = 1120$) who were similar in age, shared an (upper) middle class background, and usually did not know each other. We administered a postexperimental questionnaire to be able to control for further sociodemographic background characteristics (see table S2 for details).

Results. We first analyze people's punishment behavior across participant pools. Our perspective is how an individual who has contributed a certain amount to the public good punishes other group members who contributed either less, the same amount, or more than them. Figure 1 therefore displays punishment expenditures as a function of how much the punished individual's contribution deviated from the contribution of the punisher. We label the punishment of negative deviations punishment of free riding because the punished group member rode free on the punisher's contribution. Put differently, from the perspective of the punisher the target member behaved less prosocially than the punisher. In case the target member contributed the same amount or more, he or she behaved at least as prosocially as the punisher. We therefore call the punishment in these cases antisocial punishment.

Punishment behavior differed strongly across participant pools (Fig. 1). This holds in particular for antisocial punishment. A regression analysis of punishment behavior, which controls for the deviation, period effects, and sociodemographic composition, shows that antisocial punishment differed highly and significantly across participant pools [$\chi^2(14) = 64.9, P = 0.000$; tables S3 and S4]. Although there was very little antisocial punishment in some participant pools, in others people punished those who contributed the same or more than them as harshly as those who rode free on them. By contrast, punishment of free riding was only weakly significantly different across participant pools [$\chi^2(14) = 23.1, P = 0.059$; tables S3 and S4].

The punishment of free riding is likely triggered by negative emotions that arise from a violation of fairness norms and from feeling exploited (1, 2, 18). But what explains antisocial punishment? One plausible reason is that people might not accept punishment and therefore seek revenge (8–10). Revenge is a “human universal” (19) and part of a culture of honor in many societies. Our measure for vengeful punishment is the punishment people mete out as a function of received punishment in the previous period. Controlling for contributions of the punisher and the punished participant, we find a highly significant increase in antisocial punishment across all participant pools as a function of the amount of punishment received



in the previous period. Broken down by participant pools, the effect is highly significant (at $P < 0.01$) in seven participant pools, weakly significantly positive in two participant pools, insignificantly positive in six participant pools, and insignificantly negative in one (tables S3 and S4).

The presence of a punishment opportunity had dramatic consequences on the achieved cooperation levels (Fig. 2). Contributions were highly significantly different across participant pools [Kruskal-Wallis test with group averages over all 10 periods as independent observations, $\chi^2(15) = 113.1$; $P = 0.000$]. Cooperation was stabilized in all participant pools but at vastly different levels (Fig. 2A). Cooperation in about half of the participant pools remained at the initial level (period 1 of the P experiment), whereas contributions increased over time in the others (table S5). The most-cooperative participant pool (in which people contributed 90% of their endowment, on average) contributed 3.1 times as much as the least-cooperative participant pool (with an average contribution of 29% of the endowment). The differences in cooperation across participant pools are significantly negatively related to antisocial punishment: The higher antisocial punishment is in a participant pool, the lower is the average cooperation level in that participant pool (Fig. 2B).

As a consequence of the different patterns of punishment and cooperation, there were also substantial participant-pool differences in earnings in the P experiment. The average per-period earnings differed by more than 250 percentage points between the participant pool with the highest average earnings and that with the lowest average earnings (fig. S3 and table S6).

An important reason for the large participant pool differences in cooperation rates is the fact that participant pools reacted very differently to punishment received. Regression analyses (table S7) show that, in all but one participant pool, people who contributed less than the group average in period t and who were subsequently punished increased their contribution in period $t + 1$. The increase is only significant (at $P < 0.05$) in 11 participant pools, however, and the extent of the mean estimated increase per punishment point received varies considerably between participant pools. Thus, punishment did not have an equally strong disciplinary effect on free riders in all participant pools in the sense of steering low contributors toward higher contributions; in some participant pools, punishment had no cooperation-enhancing effect at all.

The disciplinary effectiveness of punishment for below-average contributions is associated with the extent of antisocial punishment in a participant pool. There is a strong negative correlation between the mean antisocial punishment in a participant pool and the regression coefficient that measures the mean increase per punishment point received for a below-average contribution (Spearman's $\rho = -0.87$, $P = 0.000$, $n = 16$). One explanation is that the prospect of getting punished for at- or above-average contributions in some participant pools limits the low

contributors' incentives to increase their contributions. Another explanation has to do with how people perceive the moral message behind punishment because there is evidence that even nonmonetary sanctions (which signal social disapproval) can induce low contributors to increase their contributions (20). Participant pools might have differed in the extent to which people feel ashamed when being punished for low contributions.

A regression analysis (Table 1) summarizes our findings on the impact of punishment on cooperation. To also account for variations of punishment in different groups within participant pools, we use the group average contributions as independent observations.

The results show that groups that started at high levels in period 1 of the P experiment also had high group average contributions over the

remaining periods 2 to 10; groups that started at low levels in period 1 of the P experiment had low group average contributions over the remaining periods. Group average punishment of free riding relative to the punishers' own contributions is positively correlated with this group's average contribution, *ceteris paribus*. The opposite conclusion holds for antisocial punishment.

We also found significant participant pool differences in the N experiment, which serves as a benchmark for the P experiment [Kruskal-Wallis test with group averages over all 10 periods as independent observations, $\chi^2(15) = 46.5$, $P = 0.000$]. Mean contributions varied between 4.9 and 11.5 tokens of the least- and most-cooperative participant pool, respectively (Fig. 3). The span of 6.6 tokens between the least- and most-cooperative participant pool was thus substantially lower than the

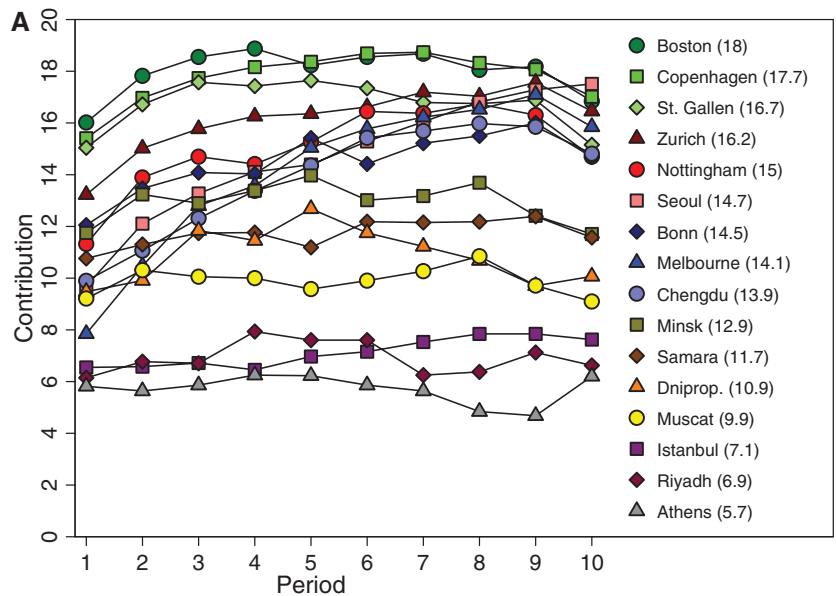
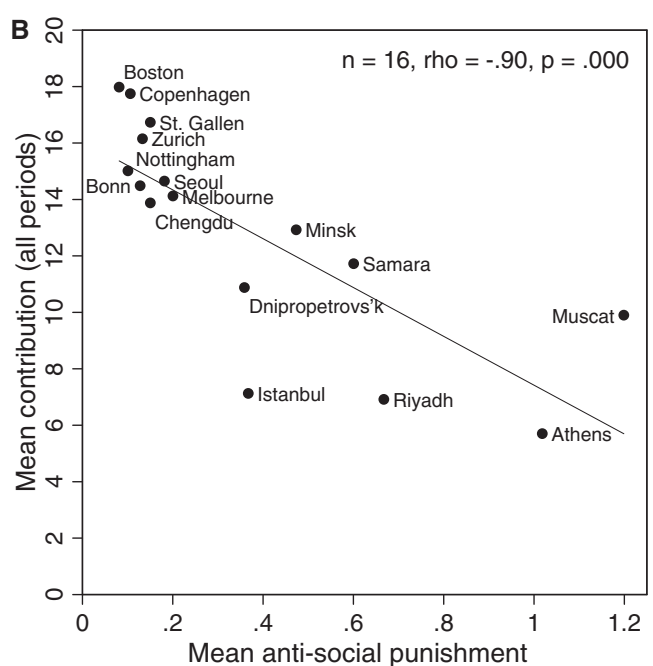


Fig. 2. (A) Mean contributions to the public good over the 10 periods of the P experiment. Each line corresponds to the average contribution of a particular participant pool. The numbers in parentheses indicate the mean contribution (out of 20) in a particular participant pool. **(B)** Mean antisocial punishment and mean contribution (across all periods) per participant pool. Rho indicates Spearman rank order correlation between participant pool averages.



span of 12.3 tokens in the P experiment (Fig. 2A). Moreover, in contrast to the P experiment, where contributions were stabilized at vastly different levels, contributions in the N experiment dwindled to lower levels almost everywhere (table S8 and fig. S4).

Compared with the N experiment, the presence of a punishment option had at least a weakly significant cooperation-enhancing effect in 11

participant pools (Wilcoxon signed ranks test with independent group average contribution rates across all periods as observations, fig. S4 and table S9); the change in cooperation between the N and the P experiment was not significant in the other five participant pools. Thus, the cooperation-enhancing effect of a punishment opportunity cannot be taken for granted. This

finding stands in contrast to previous results from experiments conducted in the United States and Western Europe, where punishment always increased cooperation in experiments with comparable fixed-group designs and parameters (8, 10, 20–22).

The reason for this result is related to antisocial punishment: the higher antisocial punishment was in a participant pool, the lower was the rate of increase in cooperation in the P experiment relative to the N experiment (Spearman's $\rho = -0.76$, $P = 0.001$, $n = 16$). Furthermore, participant pools' average cooperation levels in period 1 of the P experiment (where participants had not yet acquired any experience with punishment) were significantly negatively correlated with their subsequent mean expenditures on antisocial punishment: The more a participant pool expended on antisocial punishment in the later stages of the experiment, the lower was its initial cooperation level (Spearman's $\rho = -0.78$, $P = 0.000$, $n = 16$).

What explains the large participant pool differences in antisocial punishment and hence cooperation levels? Punishment may be related to social norms of cooperation. Social norms exist at a macrosocial level and refer to widely shared views about acceptable behaviors and the deviations subject to possible punishment (23, 24). Thus, if participant pools held different social norms with regard to cooperation and free riding, they actually might have punished differently. An interesting set of relevant social norms are norms of civic cooperation (14) as they are expressed in people's attitudes to tax evasion, abuse of the welfare state, or dodging fares on public transport. These are all situations that can be modeled as public goods problems. The stronger norms of civic cooperation are in a society, the more free riding might be viewed as unacceptable and the more it might be punished in consequence. The flip side of the argument is that cooperators, who behave in the normatively desirable way, should not get punished; strong norms of civic cooperation might act as a constraint on antisocial punishment.

The strength of the rule of law in a society might also have an impact on antisocial punishment. If the rule of law is strong, people trust the law enforcement institutions, which are perceived as being effective, fair, impartial, and bound by the law (25). Revenge is shunned. If the rule of law is weak, the opposite holds. Thus, the rule of law reflects how norms are commonly enforced in a society.

We construct the variable norms of civic cooperation from data taken from the World Values Survey (13) (fig. S1A). The variable is derived from answers of a large number of selected representative residents of a country to questions on how justified (on a 10-point scale; 1 is fully justified; 10 is never justified) people think tax evasion, benefit fraud, or dodging fares on public transport are. The more reproachful these behaviors are in the eyes of the average citizen, the stronger are a society's norms of civic cooperation.

Table 1. Punishment and cooperation levels. Ordinary least squares regressions with the group average contributions of all groups, which show any variation in contributions as independent observation ($n = 273$). The group average contributions over periods 2 to 10 are the dependent variables. The independent variables are the group average contributions in period 1, the group averages of punishment points assigned to group members who contributed less than the punishing participant (group average punishment of free riding) and to group members who are equally or more cooperative than the punishing participant (group average antisocial punishment). Model 1 does not control for the mean cooperation level in a participant pool, whereas model 2 controls for it by adding participant pool dummies. The adjusted r^2 increases by only 7% and the results remain robust, although the coefficient for antisocial punishment is reduced in size. $*P < 0.1$, $**P < 0.05$, $***P < 0.01$. Numbers in parentheses indicate robust standard errors.

	Dependent variable: group average contributions in period 2 to 10	
	1	2
Group average contributions in period 1	0.779*** (0.052)	0.720*** (0.065)
Group average punishment of free riding	0.521** (0.201)	0.480** (0.200)
Group average antisocial punishment	-2.247*** (0.350)	-1.256*** (0.325)
Constant	5.057*** (0.688)	5.899*** (1.221)
Participant pool dummies	No	Yes
Adjusted r^2	0.60	0.67
F test	136.9	31.3
P value	0.000	0.000
N	273	273

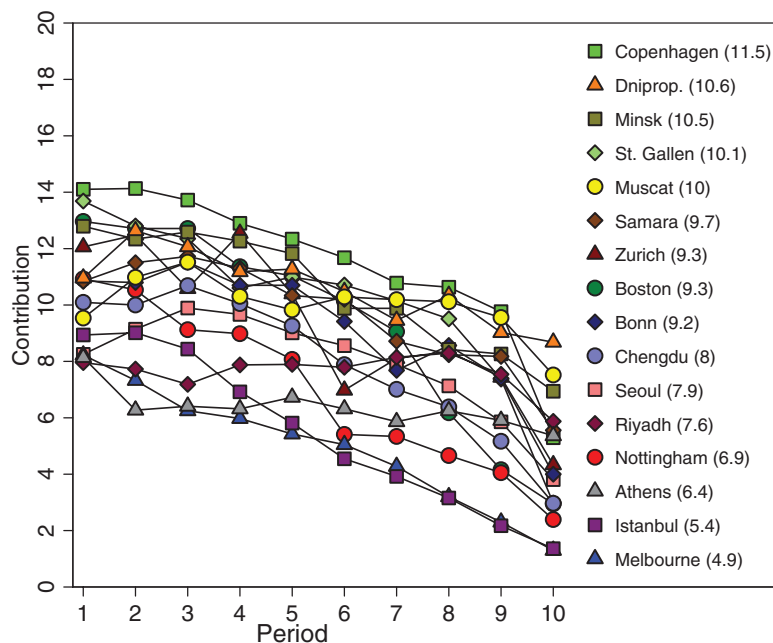


Fig. 3. Mean contributions to the public good over the 10 periods of the N experiment. Each line corresponds to the average contribution of a particular participant pool. The numbers in parentheses indicate the mean contribution (out of 20) in a particular participant pool.

tion (14). The country scores of our 16 participant pools vary between 6.91 and 9.79 (the mean is 8.53); the available world sample range ($n = 81$ countries; mean = 8.64) lies between 6.75 and 9.81. Thus, the societies of our participant pools cover almost the whole available worldwide range of the distribution of norms of civic cooperation.

The rule of law indicator (13) (fig. S1B) is based on a host of different variables that measure “the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, the police, and the courts, as well as the likelihood of crime and violence” (25). The theoretical range is -2.5 (very weak rule of law) to 2.5 (very strong rule of law). The empirically observed range of the 211 countries for which this indicator is available is -2.20 to 1.99 . The rule of law indicator varies between -1.23 and 1.96 in the countries of our participant pools.

Because both indicators reflect the views of the average citizen in a given society, it is likely that our participants, through various forms of cultural transmission (26), have been exposed to the prevalent social norms and have perceptions of the quality of the rule of law in their respective societies. Moreover, previous research, conducted in small-scale societies, suggests that experimentally observed behavior reflects socioeconomic conditions and daily experiences (11). Thus, there are good reasons to expect that the experimentally observable punishment behavior might be correlated with our indicators.

We investigated the link between punishment and the two indicators econometrically by running regressions of punishment expenditures on the variables norms of civic cooperation and rule of law. We distinguished between punishment of

free riding and antisocial punishment, and we also controlled for the punisher’s contribution, the contribution of the punished participant, the contribution of other group members, period effects, and the individual socioeconomic characteristics (to control for differences in participant-pool composition). The estimation method is Tobit (with robust standard errors clustered on the independent group).

The estimation results (Table 2) show that the stronger norms of civic cooperation are in the society, the harder people in the respective participant pool punish those who contributed less than them ($P < 0.01$). Rule of law has an insignificantly positive impact on the punishment of free riding. With respect to antisocial punishment, we found that both norms of civic cooperation and rule of law are significantly negatively correlated with punishment (at $P < 0.05$). In other words, antisocial punishment is harsher in participant pools from societies with weak norms of civic cooperation and a weak rule of law. Additional analyses (table S10) show that antisocial punishment also varies highly significantly with a variety of indicators developed by social scientists in order to characterize societies (table S1). Thus, the extent of antisocial punishment is most likely affected by the wider societal background.

Discussion. Evidence from economics, sociology, political science, and anthropology suggests that human social groups differ strongly in how successfully they solve cooperation problems (14, 27–29). In reality, many exogenous factors, institutional and environmental conditions as well as population characteristics, can explain varying degrees of cooperative success. Our contribution is to show experimentally that (antisocial) punishment can lead to very strong differences

in cooperation levels among comparable social groups acting in identical environments.

Antisocial punishment of cooperators existed in all our participant pools, but its importance and detrimental consequences varied strongly across them. Revenge is a likely explanation for antisocial punishment in most participant pools, but other (population-specific) motives might be relevant as well. Some antisocial punishment may be efficiency-enhancing in intent to induce the punished individual to increase his or her contributions. The fact that in most participant pools antisocial punishment was lower the higher the punished participant’s contribution was is consistent with this explanation (table S7A). Because punishment in our experiment was cheaper for the punisher than for the punished participant, people with a strong taste for dominance (30), a competitive personality (31), or a desire to maximize relative payoffs (32) might not only punish freeloaders but also cooperators, even including those who contributed the same amount as the punisher. Low contributors might also view high contributors as do-gooders who have shown them up. Punishment may therefore be an act of “do-gooder derogation” (33). Similarly, as observed in some bargaining experiments (12, 34, 35) in which people reject hyperfair offers, people for various reasons might be suspicious of others who appear too generous. Normative conformity, a desire and expectation to behave as all others do, is part of human psychology (36) and may lead to the punishment of all deviators, cooperators, and free riders alike. Punishment may be also related to in-group–out-group distinctions (37) because people might retaliate if punished by an out-group member (38). Societies also differ in the extent to which their social structures are governed by in-group–out-group distinctions. For instance, according to some cross-cultural psychologists (15, 39) in “collectivist” societies many interactions are confined to close-knit social networks, whereas in “individualistic” societies interactions are more permeable across social groups. Because in our experiment all participants were strangers to one another, people in collectivist societies might be more inclined than people in individualistic societies to perceive other participants as out-group members. Therefore, antisocial punishment might be stronger in collectivist than in individualistic societies. Our evidence is consistent with this possibility because in regressions similar to those of Table 2 antisocial punishment is highly significantly correlated with a widely used societal-level measure of individualism–collectivism (15) (table S10).

Our finding that social norms of cooperation and punishment are linked is of relevance for the debate about social capital (14) and in particular a literature that argues that informal sanctions often substitute for formal enforcement mechanisms if these are lacking or not working well (7, 27, 40–42). The fact that antisocial punishment is negatively correlated

Table 2. Punishment, norms of civic cooperation, and the rule of law. The dependent variable is assigned punishment points to participants who contributed less than the punishing participants (models 1 to 3) or to participants who contributed the same or more than the punishing participant (models 4 to 6). The independent variables are the country scores of norms of civic cooperation and rule of law. Controls include the participants’ own contribution, the punished participant’s contribution, the average contribution of the remaining two participants, the period, a dummy for the final period, and individual socioeconomic characteristics. We show the coefficients of Tobit estimates (43). Robust standard errors are calculated by using the group as the independent cluster. Table S10 contains further analyses.

	Punishment of free riding (negative deviations)			Antisocial punishment (nonnegative deviations)		
	1	2	3	4	5	6
Norms of civic cooperation	0.410*** (0.154)		0.371** (0.155)	−1.093*** (0.326)		−0.740** (0.371)
Rule of law		0.164 (0.111)	0.067 (0.110)		−0.641*** (0.221)	−0.618** (0.254)
Constant	−5.047*** (1.400)	−1.843*** (0.469)	−4.708*** (1.398)	5.622* (2.900)	−3.479*** (0.719)	2.422 (3.360)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
σ	3.323	3.457	3.322	5.583	5.665	5.566
<i>P</i> value	0.000	0.000	0.000	0.000	0.000	0.000
log pseudolikelihood	−12203	−13299	−12202	−10574	−11989	−10539
<i>N</i>	8350	8950	8350	19850	20660	19850

with the strength of the rule of law and also with cooperation levels suggests that the quality of the formal law enforcement institutions and informal sanctions are complements (rather than substitutes). Informal sanctions might be more effective in sustaining voluntary cooperation when the formal law enforcement institutions operate more effectively because antisocial punishment is lower in these societies. The detrimental effects of antisocial punishment on cooperation (and efficiency) also provide a further rationale why modern societies shun revenge and centralize punishment in the hands of the state.

References and Notes

- E. Fehr, S. Gächter, *Nature* **415**, 137 (2002).
- D. J. F. de Quervain *et al.*, *Science* **305**, 1254 (2004).
- Ö. Gülerk, B. Irlenbusch, B. Rockenbach, *Science* **312**, 108 (2006).
- B. Rockenbach, M. Milinski, *Nature* **444**, 718 (2006).
- R. Axelrod, W. Hamilton, *Science* **211**, 1390 (1981).
- M. A. Nowak, K. Sigmund, *Nature* **437**, 1291 (2005).
- E. Ostrom, J. M. Walker, R. Gardner, *Am. Polit. Sci. Rev.* **86**, 404 (1992).
- N. Nikiforakis, *J. Public Econ.* **92**, 91 (2008).
- M. Cinyabuguma, T. Page, L. Putterman, *Exp. Econ.* **9**, 265 (2006).
- L. Denant-Boemont, D. Masclet, C. N. Noussair, *Econ. Theory* **33**, 145 (2007).
- J. Henrich *et al.*, *Behav. Brain Sci.* **28**, 795 (2005).
- J. Henrich *et al.*, *Science* **312**, 1767 (2006).
- Materials and methods are available on Science Online.
- S. Knack, P. Keefer, *Q. J. Econ.* **112**, 1251 (1997).
- G. Hofstede, *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations* (Sage, Thousand Oaks, CA, 2001).
- R. Inglehart, W. E. Baker, *Am. Sociol. Rev.* **65**, 19 (2000).
- U. Fischbacher, *Exp. Econ.* **10**, 171 (2007).
- A. Falk, E. Fehr, U. Fischbacher, *Econometrica* **73**, 2017 (2005).
- J. Elster, *Ethics* **100**, 862 (1990).
- D. Masclet, C. Noussair, S. Tucker, M. C. Villeval, *Am. Econ. Rev.* **93**, 366 (2003).
- E. Fehr, S. Gächter, *Am. Econ. Rev.* **90**, 980 (2000).
- T. Page, L. Putterman, B. Unel, *Econ. J.* **115**, 1032 (2005).
- J. Coleman, *Foundations of Social Theory* (Belknap, Cambridge, MA, 1990).
- J. Henrich, N. Henrich, *Why Humans Cooperate: A Cultural and Evolutionary Explanation*, Evolution and Cognition Series (Oxford Univ. Press, Oxford, 2007).
- D. Kaufmann, A. Kraay, M. Mastruzzi, World Bank Policy Research Working Paper No. 4280, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=999979
- R. Boyd, P. J. Richerson, *Culture and the Evolutionary Process* (Univ. of Chicago Press, Chicago, 1985).
- E. Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action, the Political Economy of Institutions and Decisions* (Cambridge Univ. Press, Cambridge, 1990).
- R. J. Sampson, S. W. Raudenbush, F. Earls, *Science* **277**, 918 (1997).
- R. B. Edgerton, *Sick Societies: Challenging the Myth of Primitive Harmony* (Free Press, New York, 1992).
- T. H. Clutton-Brock, G. A. Parker, *Nature* **373**, 209 (1995).
- W. B. G. Liebrand, R. W. T. L. Jansen, V. M. Rijken, C. J. M. Suhre, *J. Exp. Soc. Psychol.* **22**, 203 (1986).
- K. Fließbach *et al.*, *Science* **318**, 1305 (2007).
- B. Monin, *Int. Rev. Soc. Psychol.* **20**, 53 (2007).
- D. L. Bahry, R. K. Wilson, *J. Econ. Behav. Organ.* **60**, 37 (2006).
- H. Hennig-Schmidt, Z.-Y. Li, C. Yang, *J. Econ. Behav. Organ.* **65**, 373 (2008).
- J. Henrich, *J. Econ. Behav. Organ.* **53**, 3 (2004).
- H. Bernhard, U. Fischbacher, E. Fehr, *Nature* **442**, 912 (2006).
- J.-K. Choi, S. Bowles, *Science* **318**, 636 (2007).
- H. C. Triandis, *Individualism and Collectivism*, R. E. Nisbett, Ed., *New Directions in Social Psychology* (Westview, Boulder, CO, 1995).
- A. Dixit, *Lawlessness and Economics: Alternative Modes of Economic Governance* (Princeton Univ. Press, Princeton, NJ, 2004).
- E. A. Posner, *Law and Social Norms* (Harvard Univ. Press, Cambridge, MA, 2000).
- R. C. Ellickson, *Order Without Law: How Neighbors Settle Disputes* (Harvard Univ. Press, Cambridge, MA, 1991).
- Alternative estimation methods, like Probit and Poisson, yield very similar results, both in terms of signs and statistical significance.
- All authors contributed equally to this work. The authors thank various workshop audiences, in particular the Arts and Humanities Research Council workshops Culture and the Mind in Sheffield, and I. Bohnet, R. Boyd, S. Burks, E. Fehr, U. Fischbacher, D. Gambetta, H. Gintis, G. Grimalda, J. Henrich, P. Richerson, B. Rockenbach, R. Sapolsky, and R. Zeckhauser for helpful discussions. We are grateful for financial support from the University of Nottingham, the Grundlagenforschungsfonds at the University of St. Gallen, the Latsis Foundation (Geneva), and the EU-TMR Research Network ENDEAR (FMRX-CT98-0238). This paper is part of the MacArthur Foundation Network on Economic Environments and the Evolution of Individual Preferences and Social Norms.

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5868/1362/DC1

Materials and Methods

SOM Text

Figs. S1 to S4

Tables S1 to S10

References and Notes

5 December 2007; accepted 28 January 2008

10.1126/science.1153808

REPORTS

Fiber-Optical Analog of the Event Horizon

Thomas G. Philbin,^{1,2} Chris Kuklewicz,¹ Scott Robertson,¹ Stephen Hill,¹ Friedrich König,¹ Ulf Leonhardt^{1*}

The physics at the event horizon resembles the behavior of waves in moving media. Horizons are formed where the local speed of the medium exceeds the wave velocity. We used ultrashort pulses in microstructured optical fibers to demonstrate the formation of an artificial event horizon in optics. We observed a classical optical effect: the blue-shifting of light at a white-hole horizon. We also showed by theoretical calculations that such a system is capable of probing the quantum effects of horizons, in particular Hawking radiation.

Laboratory analogs of black holes (1–3) are inspired by a simple and intuitive idea (4): The space-time geometry of a black hole resembles a river (5, 6)—a moving medium flowing toward a waterfall, the singularity. Imagine

that the river carries waves propagating against the current with speed c' . The waves play the role of light, where c' represents c , the speed of light in vacuum. Suppose that the closer the river gets to the waterfall, the faster it flows, and that at some point the speed of the river exceeds c' . Clearly, beyond this point waves can no longer propagate upstream. The point of no return corresponds to the horizon of the black hole. Imagine another situation: a fast river flowing out into the sea, getting slower. Waves cannot enter the river beyond the point where the flow

speed exceeds the wave velocity; the river resembles an object that nothing can enter: a white hole.

Nothing, not even light, can escape from a gravitational black hole. Yet according to quantum physics, the black hole is not entirely black but emits waves that are in thermal equilibrium (7–9). The waves consist of correlated pairs of quanta; one originates from inside and the other from outside the horizon. Seen from one side of the horizon, the gravitational black hole acts as a thermal black-body radiator sending out Hawking radiation (7–9). The effective temperature depends on the surface gravity (7–9), which, in the analog model, corresponds to the flow-velocity gradient at the horizon (1–5).

The Hawking temperature of typical black holes lies far below the temperature of the cosmic microwave background, so an observation of Hawking radiation in astrophysics seems unlikely. However, laboratory demonstrations of analogs of Hawking radiation could be feasible. One type of recent proposal (10–12) suggests the use of ultracold quantum gases such as alkali Bose-Einstein condensates or ultracold alkali fermions (12). When a condensate in a wave guide is pushed over a potential barrier, it may exceed the speed of sound (typically a few millimeters per second)

¹School of Physics and Astronomy, University of St Andrews, North Haugh, St Andrews, Fife, KY16 9SS, UK. ²Max Planck Research Group of Optics, Information and Photonics, Günther-Scharowsky-Strasse 1, Bau 24, D-91058 Erlangen, Germany.

*To whom correspondence should be addressed. E-mail: ulf@st-andrews.ac.uk

will be challenging if they are unstable.

New functions for poly(U) are emerging. The tail can be added to mRNAs to trigger their decay. mRNAs that direct the synthesis of histones—the major protein constituents of chromatin—during the cell division cycle are rapidly degraded once DNA replication is completed or blocked. Intermediate mRNAs in this decay process often have poly(U) tails of 8 to 10 residues (10). Small interfering RNAs directed against two candidate PUP enzymes blocked this degradation, implying that poly(U) addition is essential for their decay. Also, in *S. pombe*, a PUP adds poly(U) to actin mRNA, though its effect on turnover is unknown (9).

A poly(U) tail may enhance degradation by stimulating removal of the mRNA's 5' cap structure, a key step in mRNA turnover. Poly(U) tails enhance “decapping” in a cell-free system (11). Likely, the tails bind the Lsm protein complex, which associates with decap-

ping factors (11, 12). Indeed, depletion of Lsm1 inhibits histone mRNA turnover (10).

Addition of uridines probably has diverse consequences, including RNA stabilization (6); yet this modification often occurs on an RNA's road to ruin (see the figure). Aberrantly unmethylated microRNAs in the plant *Arabidopsis thaliana* are modified with oligo(U) and destroyed (2). Fragmentation of mRNA by microRNAs is accompanied by the addition of oligo(U) to the pieces before they disappear (13). The most common mRNA decay pathway involves association of the Lsm complex to the mRNA after poly(A) removal. Even this route may rely on evanescent, short oligo(U) because the Lsm complex preferentially binds 3'-terminal uridine tails.

The discovery of poly(U) tails on mRNAs opens unexplored territory in the RNA world. Dual-personality enzymes could switch an mRNA's fate from life to death simply by a

change in the nucleotide they accept. Others may well wait in the wings, along with proteins that target specific RNAs, or remove the tails. Count on new roles for poly(U), an expanding list of RNAs that receive it, and more startling enzymes that put it on and take it off.

References and Notes

1. R. Aphasizhev, *Cell. Mol. Life Sci.* **62**, 2194 (2005).
2. J. Li, Z. Yang *et al.*, *Curr. Biol.* **15**, 1501 (2005).
3. E. Lund, J. E. Dahlberg, *Science* **255**, 327 (1992).
4. G. Martin, W. Keller, *RNA* **13**, 1834 (2007).
5. J. E. Kwak, M. Wickens, *RNA* **13**, 860 (2007).
6. R. Trippe *et al.*, *RNA* **12**, 1494 (2006).
7. D. L. Mellman *et al.*, *Nature* **451**, 1013 (2008).
8. M. R. Macbeth *et al.*, *Science* **309**, 1534 (2005).
9. O. S. Rissland, A. Mikulasova, C. J. Norbury, *Mol. Cell. Biol.* **27**, 3612 (2007).
10. T. E. Mullen, W. F. Marzluff, *Genes Dev.* **22**, 50 (2008).
11. M. G. Song, M. Kiledjian, *RNA* **13**, 2356 (2007).
12. S. Tharun *et al.*, *Nature* **404**, 515 (2000).
13. B. Shen, H. M. Goodman, *Science* **306**, 997 (2004).
14. We thank R. Parker, J. Kimble, and the Wickens lab for comments. Work in the Wickens lab is supported by the NIH.

10.1126/science.1154946

BEHAVIOR

Punishment and Cooperation

Herbert Gintis

Even champions of modern society agree that it involves a loss of community (based on family and ethnic ties) and an expansion of civil society, with emphasis on the more impersonal interactions among individuals with minimal social ties. For two centuries, this dichotomy has anchored our understanding of modern Western society, applauded by its defenders as the fount of freedom (1), yet identified as the source of inequality (2), the decline of community (3), the destruction of the environment (4), and the impotence of grassroots political action (5). On page 1362 of this issue, Herrmann *et al.* (6) report their discovery that university students in democratic societies with advanced market economies show different social behavior from that exhibited by students in more traditional societies based on authoritarian and parochial social institutions. Their results suggest that the success of democratic market societies may depend critically on moral virtues as well as material interests, so the depiction of civil society as the sphere of “naked self-interest” is radically incorrect.

The standard view holds that human nature has a private side in which we interact morally with a small circle of intimates and a public

side in which we behave as selfish maximizers. Herrmann *et al.* suggest that most individuals have a deep reservoir of behaviors and mores that can be exhibited in the most impersonal interactions with unrelated others. This reservoir of moral predispositions is based on an innate prosociality that is a product of our evolution as a species, as well as the uniquely human capacity to internalize norms of social behavior. Both forces predispose individuals to behave morally even when this conflicts with their material interests.

These results are the latest to document a principle of reciprocity according to which people are more willing to sacrifice private gain for the public good as the cost of the sacrifice decreases and as expectations of the extent that others will sacrifice grows. In addition, individuals embrace such character virtues as honesty, trustworthiness, consideration, and loyalty (7). Of course, these moral predispositions moderate rather than eliminate considerations of self-interest and loyalties to kith and kin.

Suggestive evidence for the principle of reciprocity comes from daily life. For instance, political democracy has frequently been attained through popular collective action. Voting in elections is widespread despite its being personally time consuming, and the

Data from economic games show that the effectiveness of punishment in fostering cooperation varies greatly from society to society.

benefits are purely public (a single vote can change an electoral outcome only with infinitesimal probability). Moreover, citizens in democratic societies often vote to give substantial sums to charity, and to approve of poverty relief, although these measures increase the tax burden for the average voter.



Experimental evidence for reciprocity comes from behavioral game theory, which uses economic games in which subjects make choices under varied social conditions. For instance, Herrmann *et al.* employ a public goods game in which each of four anonymous subjects is initially given 20 tokens, and each is told he can place any number of these tokens in a public account. The tokens in the account are multiplied by 1.6 and the result divided evenly among the four. At the end of the experiment, the tokens are exchanged for real money.

In this game, each individual helps the group most by placing his 20 tokens in the

public account, and if all do so, each earns 32 tokens. However, if a single individual is selfish, he will place nothing in the public account, and his earnings will be $20 + 60(1.6)/4 = 44$ tokens. But, if all four are selfish, each earns only 20 tokens. Because the four subjects are strangers, the standard view of human nature suggests that there will be zero contributions. However, in the many times this game has been played in a variety of social settings, the older view is virtually never supported, and the average contribution is about half the initial endowment (8).

The public game indicates that individuals generally fall halfway between selfishness (keep all 20 tokens) and public-spiritedness (place all 20 tokens in the public account). However, mean contributions to the public account generally fall over many trials, reaching a very low level after 10 repetitions. By varying the rules of the game, researchers have concluded that the principle of reciprocity is responsible for the observed decay of cooperation: Subjects who contributed more than average on one round contribute less on the following round, showing their disapproval of the unfairness of their fellow players. Indeed, a single selfish individual in the group can lead contributions to spiral down to almost zero.

An innovation of Fehr and Gächter (9), used by Herrmann *et al.* as well, was to add punishment after each round of play. Each player A could specify that the player B associated with a particular contribution have three tokens deducted from his payoff, for each token deducted from A's payoff. Under these new conditions, the high contributors punished the low contributors who, in succeeding rounds, increased their contributions, so that in the 10th and final round, there was almost 100% cooperation. The behavioral propensity to cooperate with others at personal cost, and to punish non-cooperators even when this is personally costly in the long run, has been called strong reciprocity. The punishment meted out is considered altruistic because it increases the payoff of group members at a personal cost to the punisher.

The natural interpretation is that low contributors are selfish types who increase their contribution after punishment in order to avoid future punishment. However, many low contributors respond almost as much to symbolic as to monetary penalties, which indicates that many punishees are not self-interested but rather are motivated to increase their offers because they feel guilty for having violated a contribution norm (10). However, some researchers found a curious phenomenon. A few subjects, when punished, rather than contributing more,

suspected that it was the high contributors who punished them, and responded with antisocial punishment: They punished the high contributors in future rounds, leading the latter to reduce both their contribution and altruistic punishment (11).

Herrmann *et al.* collected data in 15 countries with widely varying levels of economic development. The subjects were university students in all societies. The authors found that antisocial punishment was rare in the most democratic societies and very common otherwise. Indexed to the World Democracy Audit (WDA) evaluation of countries' performance in political rights, civil liberties, press freedom, and corruption, the top six performers among the countries studied were also in the lowest seven for antisocial punishment. These were the United States, the United Kingdom, Germany, Denmark, Australia, and Switzerland. The seventh country in the low antisocial punishment group was China, currently among the fastest-growing market economies in the world. The countries with a high level of antisocial punishment and a low score on the WDA evaluation included Oman, Saudi Arabia, Greece, Russia, Turkey, and Belarus.

The most likely explanation is that in more traditional societies, the experimental setup represents a clash of cultures. On the one hand, high payoffs in the experiment require the modern

ethic of cooperation with unrelated strangers, so subjects who are reprimanded for low contribution are likely to respond with feelings of guilt and a resolve to be more cooperative in the future. In a more traditional society, many players may hold to the ethic of altruism and sacrifice on behalf of one's family and friends, with indifference toward unrelated strangers. When punished, such subjects are likely to respond with anger rather than guilt. Punishing the high contributors is thus a means of asserting one's personal values, which take precedence over maximizing one's payoff in the game.

References and Notes

1. M. Friedman, *Capitalism and Freedom* (Univ. of Chicago Press, Chicago, 1962).
2. M. D. Young, *The Rise of the Meritocracy* (Thames and Hudson, London, 1958).
3. W. C. McWilliams, *The Idea of Fraternity in America* (Univ. of California Press, Berkeley, CA, 1973).
4. G. Hardin, *Science* **162**, 1243 (1968).
5. M. Olson, *The Logic of Collective Action: Public Goods and the Theory of Groups* (Harvard Univ. Press, Cambridge, MA, 1965).
6. B. Herrmann *et al.*, *Science* **319**, 1362 (2008).
7. U. Gneezy, *Am. Econ. Rev.* **95**, 384 (2005).
8. J. O. Ledyard, in *The Handbook of Experimental Economics*, J. H. Kagel, A. E. Roth, Eds. (Princeton Univ. Press, Princeton, NJ, 1995), pp. 111–194.
9. E. Fehr, S. Gächter, *Nature* **415**, 137 (2002).
10. D. Masclet *et al.*, *Am. Econ. Rev.* **93**, 366 (2003).
11. O. Bochet *et al.*, *J. Econ. Behav. Organ.* **60**, 11 (2006).
12. I thank Samuel Bowles for advice in the preparation of this article.

10.1126/science.1155333

OCEANS

On Phytoplankton Trends

Victor Smetacek¹ and James E. Cloern²

How are phytoplankton at coastal sites around the world responding to ongoing global change?

Phytoplankton—unicellular algae in the surface layer of lakes and oceans—fuel the lacustrine and marine food chains and play a key role in regulating atmospheric carbon dioxide concentrations. How will rising carbon dioxide concentrations in the air and surface ocean in turn affect phytoplankton? Answering this question is crucial for projecting future climate change. However, because phytoplankton species populations appear and disappear within weeks, assessing change requires high-resolution monitoring

of annual cycles over many years. Such long-term studies at coastal sites ranging from estuaries and harbors to open coastlines and islands are yielding bewildering variability, but also fundamental insights on the driving forces that underlie phytoplankton cycles (1).

An example of regularity is provided by a 45-year data set from weekly phytoplankton monitoring in Lake Windermere, England, which shows that the diatom species *Asterionella formosa* dominates phytoplankton biomass from autumn to spring but is virtually absent during summer; this species drives silicon cycling in the lake (2). In contrast, weekly data collected in Narragansett Bay in Rhode Island since 1959 reveal that the phytoplankton react with wide fluctuations in composition and timing of the annual biomass

¹Alfred Wegener Institute for Polar and Marine Research of the Helmholtz Foundation, 27570 Bremerhaven, and the University of Bremen, FB 2, 28334 Bremen, Germany. E-mail: Victor.Smetacek@awi.de ²U.S. Geological Survey, 345 Middlefield Road, Menlo Park, CA 94025, USA. E-mail: jecloern@usgs.gov

Supporting Material for

Anti-social punishment across societies

Benedikt Herrmann, Christian Thöni, Simon Gächter*

* To whom correspondence should be addressed. E-mail: simon.gächter@nottingham.ac.uk

12 February 2008

This file includes

1. Materials and methods

- 1.1 Research methodology: cross-cultural economic experiments
- 1.2 Further methodological issues
- 1.3 Subject pools and their societal/cultural background
- 1.4 Experimental instructions
- 1.5 Experimental procedures
- 1.6 Laboratories and software

2. Supporting analyses

- 2.1 Punishment behavior
- 2.2 Cooperation in the P-experiment
- 2.3 (Relative) payoffs in the N- and the P-experiment (efficiency)
- 2.4 Reactions to received punishment
- 2.5 Cooperation in the N-experiment and the change in contributions between the N- and P-experiment
- 2.6 Anti-social punishment and the economic and cultural background of societies

Tables S1 to S10

Figures S1 to S4

Notes and references

Supporting Material

1. Materials and methods

1.1 Research methodology: cross-cultural economic experiments

Our paper uses the methodology of experimental economics to study whether norms of cooperation are different across societies with various cultural and economic backgrounds. Our experimental data stems from 1120 participants in sixteen subject pools from fifteen countries with widely different socio-economic and cultural backgrounds: we have data from Asian subject pools (China and Korea), Arab societies (Oman and Saudi Arabia), English-speaking countries (Australia, USA and UK), Eastern Europe (Ukraine, Russia and Belarus), German-speaking Central Europe (Switzerland and Germany), Scandinavia (Denmark) and Southern Europe (Greece and Turkey). Our paper is therefore related to previous endeavors that try to understand cross-societal differences in controlled economic experiments. The methodology has been pioneered by Roth et al's cross-cultural bargaining study (*S1*). We will explain this methodology in detail below. First we relate our work to the previous literature and discuss our basic research methodology as well as the specific methods used for conducting our experiments.

Relation to previous literature

Our study differs from the existing studies in at least two major respects. First, the large majority of previous studies we are aware of used only between two and four subject pools when comparing societies (*S1-13*). Some recent exceptions are (*S14-16*) and in particular the studies led by anthropologist J. Henrich, whose research group conducted experiments in fifteen small-scale societies around the world (*S17*). A subsequent study also led by J. Henrich (*S18*) reports experiments with another thirteen small-scale societies as well as a rural Missouri and an Emory undergraduate subject pool. Our study investigates sixteen subject pools but from across fifteen different developed societies. Moreover, most studies compared Asian countries with USA, or Western European countries. Arab societies, for instance, are rarely studied (but see (*S14, S15*)).

A second distinguishing feature from previous cross-cultural economics experiments is that many of them test specific (proximate) hypotheses that are derived from the compared cultures (*S5, S6, S8-15*). Our approach is different since our goal is to understand a more fundamental issue – do we find evidence for social norm explanations of cooperation, a question which is partly motivated by evolutionary theories of cooperation (*S19*) rather than proximate mechanisms of cultural differences. In this regard our study is related to those of Henrich et al. (*S17, S18*).

Our research methodology

One research strategy would be to run experiments with sociologically different subject pools within a society (see for instance (*S7, S20, S21*)). Yet, the problem is then to ensure subject pool comparability across sixteen subject pools. Variability in the socio-demographic composition of subject pools could be confounded with the genuine subject pool differences we are interested in. Given that large-scale experiments with randomly selected representative subjects were not feasible for us, we ran the experiment with subject pools that are as comparable as possible. Undergraduates are such a subject pool. They are of a similar age, typically come from an urban and middleclass background and have a similar level of education. Comparing the same type of subject pool minimizes potential confounds with the socio-economic status of the subject pools.

At one stage of our analysis we are interested in how social norms of cooperation relate to punishment behavior. Social norms are standards of behavior that are based on widely shared beliefs about what constitutes acceptable behavior in a given situation (S22-25). Our approach to measure social norms is to take available societal-level variables as proxies for the prevalent values and norms in a given society, as social norms are a macrosocial phenomenon (S23). We will explain them in detail below. The data we use stem from large-scale representative and internationally comparable surveys (like the World Values Survey) or from other large-scale surveys (like Hofstede's cultural dimensions (S26)) and therefore arguably measure important aspects of norms and values that a majority of people hold in a given society (S26-30). It is therefore likely that the nationally prevalent social norms and values influence all members of a society to some degree through various forms of conformist transmission (through parents, teachers, peers and one's wider social networks) (S31, S32). A further advantage of our approach is that the norms and values are exogenous to the experiment in the sense that experimental play cannot influence the measurement of social norms.

Our methodology is similar to the approach by Henrich et al. (S33) who relate their experimental data (from ultimatum bargaining experiments) to society-level economic variables like market integration and the importance of cooperation for economic production. They find that ultimatum bargaining behavior can be explained by these variables. To our knowledge this is the first evidence that societal-level variables influence experimentally measured behavior. However, at least in comparison to the small-scale societies studied by Henrich et al., the societies of our subject pools are very similar in the dimensions of market integration and cooperation, since they are all based on extensive trading and division of labor among non-kin (S34). Yet, as we will describe below, our societies do differ in the norms and values they subscribe to and they also differ in their economic prosperity and the quality of their legal institutions. Our approach is to relate experimental behavior to those variables, that is, relevant social norms and variables that measure the quality of law enforcement (the "Rule of Law"). In this methodology our study is related to a meta-study of ultimatum bargaining behavior by Oosterbeek et al. (S35) who relate ultimatum bargaining behavior to variables derived from the World Values Survey and Hofstede's cultural dimensions. They find no significant influence of values on experimentally observed behavior. The exception is that a measure for traditional vs. secular values (see below) is correlated with lower offers in the ultimatum game. Finally, a research group led by G. Grimalda (S16) conducted public goods experiments in six societies around the globe and found that a macrolevel variable for the extent of a society's globalization has a positive impact on cooperation.

As we will show below, the societies of our subject pools span a very large range of the world-wide distribution of values, cultural dimensions, and economic and institutional conditions (among complex, more or less developed societies). Thus, if given this diversity we would not find any difference in experimentally observed behavior then we would conclude that (cooperation) experiments conducted in the developed world cannot pick up cultural/population differences (if they exist at all). It would indeed imply that cultural differences in economic decision making are only observable among small-scale societies (for which differences have been demonstrated convincingly). If we do detect differences then we know that cultural/population differences are not only confined to small-scale societies.

Methods in cross-cultural economic experiments

Language and experimenter effects

One potentially important issue in ensuring comparability of results in cross-cultural experiments is to minimize experimenter effects. We took several steps to minimize experimenter effects. A first step is to maximize social distance both between subjects and between subjects and the experimenter. We tried to achieve this by recruiting participants who were strangers to one

another and in particular to the experimenter and by conducting all experiments in computerized laboratories in which participants were separated by partitions that ensured maximal anonymity. Moreover, in a given session we had on average 21 participants, so that individuals could not be identified easily simply because of the sheer number of other participants. During the experiment subjects interacted anonymously by typing their decisions into a computer interface. The whole experiment was conducted in complete silence and there was no interaction with the experimenter after the experiment started (since participants were sitting behind their partitions the experimenter was invisible to the subjects during the experiment).

A second and very important instrument to minimize experimenter effects is to have written instructions that are the same everywhere, except for being written in the respective language. To ensure maximal comparability of instructions we first wrote them in German and English and then a native speaker of the respective language translated the instructions into the respective language. Another native speaker translated them back into English. We fine-tuned the translations until we could be sure that the rules of the experiment were explained as similarly as possible. All participants also had to solve the same set of control questions that ensured that subjects understood the rules of the experiment and payoff calculations (see Section 1.4 for a sample copy of the instructions and the control questions, and Section 1.5 for the procedures used).

However, there remain further sources of experimenter effects. Different experimenters may conduct the experiment slightly differently (for instance, by giving different explanations when asked in private). Our solution to this problem was that one of us (B. H.) was responsible for conducting the large majority of all sessions (S36). We conducted all sessions according to the same protocol (see Section 1.5). However, this does not solve the experimenter effect entirely. The reason is that subjects might behave differently towards a foreign experimenter than to a native one, and these reactions may even be subject-pool dependent. Our solution to this problem was that B. H. trained a local experimenter (and his helpers) according to a protocol we devised for all experiments. This training ensured that the local experimenter who read the script and gave explanations did so as similarly as possible in all subject pools. B.H. was present in all sessions he was responsible for and could therefore supervise both the preparation and the conduct of the experiment. Communication problems were eased by the fact that in addition to German and English B.H. speaks Russian and Arabic. For conducting the experiments in Chengdu and Seoul B.H. received support from bilingual speakers. In summary, we believe we did the utmost one can do to minimize experimenter effects.

Currency effects and stakes

A potentially relevant issue in subject pool comparability is the amount of money that is at stake in the experiments in the respective subject pool. Since the different subject pools come from countries with different currencies there is also the issue of currency effects. If we would calculate payoffs during the experiment in the denomination of the local currencies, subject pools would have faced very different nominal payoffs. To avoid any bias in the perception of numbers we calculated all incomes in the experiment in “Guilders” which we exchanged into the local currency at the end of the experiment.

Although the available evidence on stake effects in public goods experiments or related cooperation games suggests that stake size does not matter (S37-39) we nevertheless tried to ensure comparable stake levels. We therefore collected some information before the experiment about the likely average disposable income of a student in the respective subject pool. We also acquired information about the hourly wages of a typical student job. Given this information and the maximally possible earnings in the experiment, we chose the exchange rate of “Guilders” earned in the experiment into the respective local currency such that stakes were approximately similar across subject pools.

1.2 Further methodological issues

We conducted the majority of our experiments (45 out of 53 sessions) in the N-P sequence. We chose this sequence because we had only limited access to participants in most subject pools and therefore wanted to maximize the number of observations in the sequence we are primarily interested in. We chose the N-P sequence because our main interest is in how subjects react when a punishment option is introduced after subjects had experienced free riding in the N-experiment (an outcome which we expected given previous evidence from public good experiments (S40, S41)). In three locations (Samara, Minsk, and St. Gallen) we had access to large subject pools and could therefore also run the P-N sequence to test for order effects.

We do not find any evidence for order effects. For example, in St. Gallen the average contribution in the N-experiment in the N-P sequence was 9.7 tokens. In the N-experiment of the P-N sequence the mean contribution was 10.6 tokens. The difference is not significant according to a two-sided Wilcoxon test with group average contributions as the independent observations ($p=0.817$). Similarly, subjects in St. Gallen contributed 17.0 tokens on average in the P-experiment of the N-P sequence and 16.4 tokens in the P-N sequence. Again, a two-sided Wilcoxon test does not reject the null hypothesis of no difference in contributions ($p=0.751$). Mean punishment is also not significantly different between sequences. Very similar conclusions hold for the subject pools in Minsk and Samara since all two-sided Wilcoxon tests return $p>0.27$ (S42). Thus, since we have no evidence for the existence of sequence effects, we pool the data of the N-P and P-N sequences in the subject pools where we observe both sequences.

We address briefly another issue that is of potential relevance given that at one stage in our analysis we will relate subject pool behavior to country-level variables. The issue concerns potential differences between comparable subject pools within a country or culture.

We look at this issue in four ways: First we compare behavior in our two Swiss subject pools – St. Gallen and Zurich. We find neither a statistically significant difference in the N-experiment, nor in the P-experiment (group average contributions as independent observations; p -values > 0.409 ; two-sided Mann-Whitney tests). Second, we can also compare our Nottingham data to the comparable data of another UK subject pool. Nikiforakis (S43)(S44) conducted ten-period N- and P-experiments ($n=12$ fixed groups of four in each treatment) using very similar instructions and the same software as we did. His subjects were undergraduates of the University of London at Royal Holloway. In the N-experiment the Nottingham (Royal Holloway) subjects contributed 6.95 (6.1) tokens; in the P-experiment the Nottingham (Royal Holloway) subjects contributed 15.0 (15.3) tokens. The temporal developments of contributions in the Nottingham and Royal Holloway subject pools were also very similar in both experiments. Contributions were not significantly different in both treatments (group averages as independent observations, Mann-Whitney tests, p -values > 0.71). In a new set of experiments Nikiforakis and Normann (S45) got very similar results again. Third, Herrmann and Thöni (S46) conducted a one-shot public good experiment in four different comparable subject pools in urban and rural Russia (using the strategy method), with no detectable differences in responses across all four locations. Finally, we can compare our data from the Russian, Ukrainian and Belarusian subject pools. Although these are separate countries now they belonged to the former Soviet-Union and have for centuries been part of the Russian-orthodox culture. All of the pair-wise comparisons in both the N- and the P-experiment return p -values > 0.23 (two-sided Mann-Whitney tests with group averages as independent observations).

1.3 Subject pools and their societal/cultural background

Cultural and societal background of our subject pools

Table S1 summarizes the main cultural and economic background data of the societies of our subject pools. Our main interest is on “Norms of civic cooperation” and the “Rule of Law”, as there are theoretical reasons (see below) to believe that they shape people’s norms and expectations about how others will behave as well as their punishment behavior. The other variables mainly illustrate that the societies of our subject pools span a very large range of societal differences as stressed by economists, political scientists, sociologists, and cultural anthropologists with an interest to quantify cultural differences.

Subject pool	Country	Social capital variables		Economic prosperity	Law enforcement & Democracy		Cultural dimensions				Value orientations	
		Share of people who can be trusted	Norms of civic cooperation	GDP per capita in \$,000, PPP	Rule of Law	Democracy	Power distance	Individualism	Masculinity	Uncertainty avoidance	Survival (-) vs. Self-expression (+)	Traditional (-) vs. Secular rational (+)
Boston	USA	0.36	8.65	43.4	1.54	13	40	91	62	46	1.64	-0.53
Nottingham	UK	0.29	8.65	35.1	1.72	10	35	89	66	35	1.37	0.26
Copenhagen	Denmark	0.67	9.27	36.5	1.94	2	18	74	16	23	1.96	1.11
Bonn	Germany	0.38	8.89	31.1	1.73	11	35	67	66	65	1.08	1.13
Zurich St. Gallen	Switzerland	0.37	8.58	37.4	1.96	5	26	69	72	56	1.45	0.77
Minsk	Belarus	0.42	6.91	8.9	-1.23	137	n.a.	n.a.	n.a.	n.a.	-1.20	0.89
Dnipropetrovsk	Ukraine	0.27	7.61	7.6	-0.74	129	n.a.	n.a.	n.a.	n.a.	-1.68	0.90
Samara	Russia	0.24	8.05	12.1	-0.88	119	93	39	36	95	-1.86	1.08
Athens	Greece	0.24	7.46	26.0	0.71	34	60	35	57	112	0.62	0.73
Istanbul	Turkey	0.16	9.79	9.1	0.02	69	66	37	45	85	-0.35	-0.83
Riyadh	Saudi Arabia	0.53	8.32	16.7	0.22	129	80	38	53	68	0.12	-1.35
Muscat	Oman	n.a.	n.a.	18.8	0.75	99	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Seoul	South Korea	0.27	8.83	23.9	0.73	33	60	18	39	85	-0.43	1.08
Chengdu	China	0.55	9.34	7.6	-0.41	129	80	20	66	30	-0.61	1.16
Melbourne	Australia	0.40	9.02	32.9	1.79	8	36	90	61	51	2.00	-0.20
	Sample average	0.37	8.53	23.1	0.66	61.8	52.4	55.6	53.3	62.6	0.29	0.44
	N	83	81	180	211	150	71	71	71	71	83	83
Available world sample range	World minimum	0.03	6.75	0.7	-2.20	1	11	6	5	8	-1.86	-2.06
	World maximum	0.67	9.81	80.5	1.99	150	104	91	110	112	2.22	1.84
	World average	0.28	8.64	11.8	0.00		59.9	42.8	49.8	67.2	0.06	-0.18

Table S1. Economic and cultural background of our subject pools. Data are country-level averages. The social capital variables are taken from representative surveys as reported in the World Values Survey (S47). For the variable “Norms of civic cooperation” higher values indicate stronger norms. The GDP per capita data are taken from the International Monetary Fund (S48). The data on the strength of the Rule of Law range from -2.5 (weakest) to 2.5 (strongest); values are averages over the years 2002-2006; data taken from the World Bank (S49, S50). Data for Democracy are taken from World Audit (S51, S52); values are the ranks of the 150 countries in the sample. Lower ranks indicate more democracy. The cultural dimensions data are taken from (S53). The respective indicator increases in the score. The value orientations data are due to (S28).

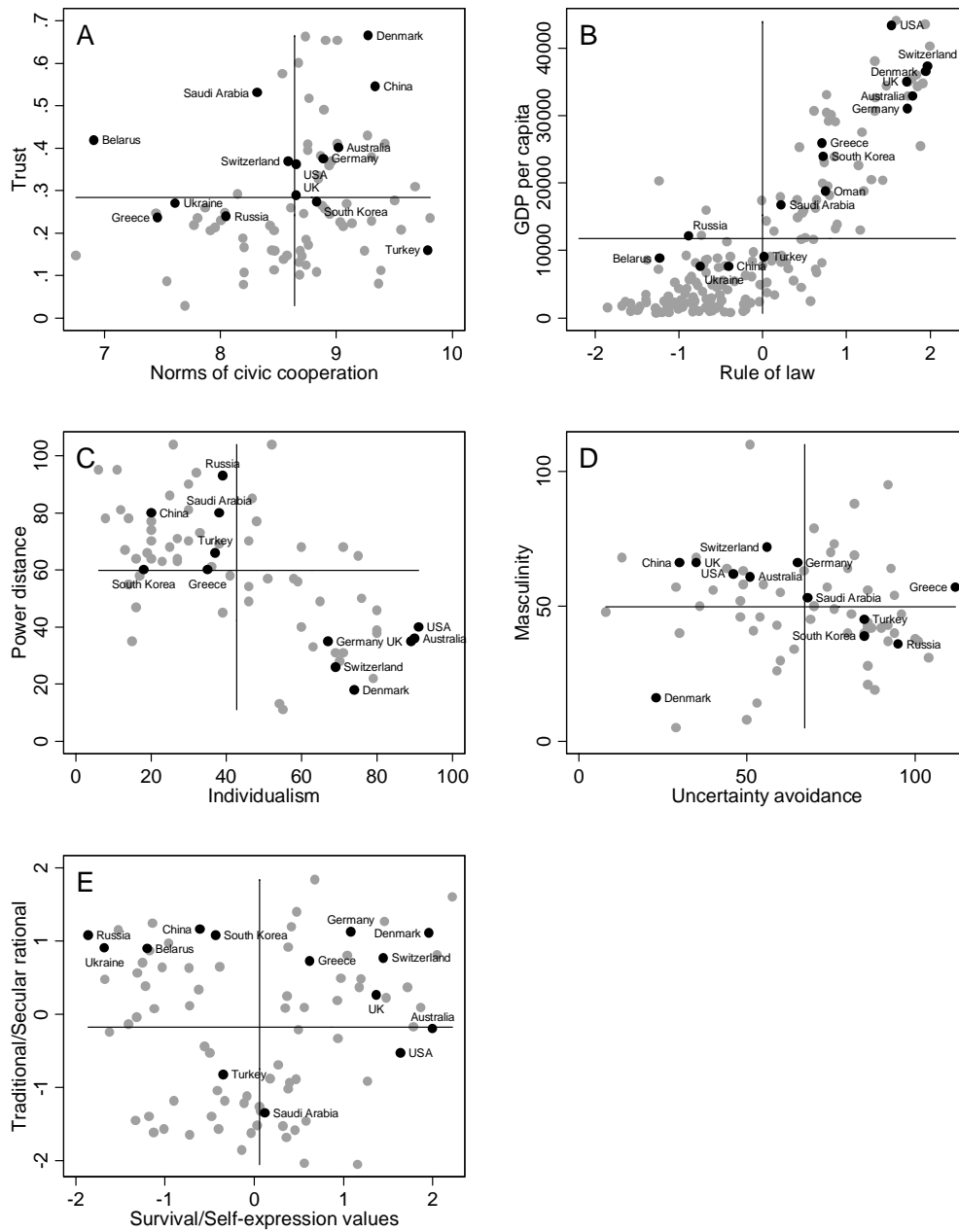


Figure S1: Distribution of norms, socio-economic conditions, values, and cultural dimensions in the countries where we conducted our experiments (black dots with labels) and all other countries for which data are available in the respective data set (grey dots without labels). Lines indicate world averages of the respective variable. **(A)** Distribution of share of people who say others can be trusted, and Norms of Civic Cooperation (1=very weak norms of civic cooperation; 10=very strong norms of civic cooperation); data taken from the World Values Survey (*S54*). **(B)** GDP per capita, in US-\$ (PPP); data taken from the International Monetary Fund (*S48*); Strength of the Rule of Law; (-2.20=weakest rule of law; 1.99=strongest rule of law); averages over the years 2002-2006; data taken from the governance indicators of the World Bank (*S49*, *S50*). Panel B does not contain the data from the outlier Luxembourg (GDP per cap. 80'471, Rule of Law 1.99). **(C and D)** Cultural dimensions according to Hofstede (*S26*, *S53*). Panel C: Power distance (11=lowest; 104=highest) and Individualism (6=least individualist; 104=most individualist); Panel D: Masculinity (5=least masculine; 110=most masculine) and Uncertainty Avoidance (8=most uncertainty tolerant; 112=most uncertainty avoidant). Data taken from Tables 2.1, 3.1, 4.1 and 5.1, respectively, in (*S53*). **(E)** Distribution of value orientations according to (*S27-29*); data taken from (*S28*). Survival vs. Self-expression values (-1.86=strongest emphasis on survival values; 2.22=strongest emphasis on self-expression values); Traditional vs. secular-rational values (-2.06=strongest emphasis on traditional values; 1.84=strongest emphasis on secular-rational values).

Figure S1 illustrates the values of the respective indicator for the societies of our subject pools (black dots with labels) relative to the respective indicator in the world wide available data set (grey dots without labels). Figure S1 supports the claim made in the main text that the societies of our subject pools cover a very large range of the world's distribution of trust and norms of civic cooperation (panel A), the GDP per capita and the strength of the Rule of Law (panel B), the cultural dimensions according to Hofstede (panels C and D), and value orientations (panel E). In the following we describe these society-level indicators in more detail.

Norms of Civic Cooperation

“Norms of civic cooperation” is one of the two main variables of interest to us. Norms of civic cooperation are highly relevant in our context, because they measure the strength of social norms in several domains that can be modeled as cooperation problems with free rider incentives. Norms of civic cooperation are social norms that might constrain people's narrow self-interest and help in the provision of public goods. Examples include norms against littering, welfare fraud, tax evasion, and traveling without a ticket on public transport. In general, the stronger civic norms of cooperation in a society are the more efficiently might collective action problems be solved (from a societal point of view). In general, cooperative norms are part of a society's “social capital” and might increase allocative efficiency by mitigating monitoring costs and contract enforcement problems (S55).

To measure norms of civic cooperation we follow (S55) and take data from the World Values Survey (WVS) (S47). The WVS is a representative survey conducted with more than 1000 representatively selected respondents in the respective country. Thus, the WVS data should measure the representative average values of norms of civic cooperation.

Specifically, respondents are asked whether a particular behavior can be justified or not. The statements are (i) “Claiming government benefits to which you are not entitled”, (ii) “Avoiding a fare on public transport”, and (iii) “Cheating on taxes if you have a chance” (in the WVS these are questions f114, f115 and f116, respectively) (S56). The subjects answered on a ten point scale between “Never justifiable (=1) and “Always justifiable (=10). We calculated the average of the three items and rescaled it such that a value of one means very weak civic norms and a value of ten denotes very strong civic norms. The strength of norms of civic cooperation in the WVS ranges from 6.75 to 9.81 and the world average is 8.64. The values in the societies of our subject pools range from 6.91 to 9.79 and the average is 8.49.

In addition to norms of civic cooperation another frequently used “social capital” variable from the WVS is “trust” (S23, S55, S57-59). It has received a lot of attention in cross-country comparisons. In a widely used question people are asked whether they think that “most people can be trusted” or that “you can't be too careful when dealing with people” (question a165 in the WVS). We document the share of people who say that others can be trusted for the countries of our subject pools in Table S1 and in Figure S1A.

Figure S1A shows that the countries of our subject pools cover a large range of the world-wide variation in both norms of cooperation and trust. With respect to trust, our experimental data come from subject pools in countries that are far above the world average (Saudi Arabia, Denmark, China) as well as below the world average (Russia, Greece, and Turkey). Likewise, our subject pools also come from countries at the opposite ends of the spectrum with regard to the strength of norms of civic cooperation in their respective countries. Denmark, China and Turkey are among the countries in the world with the strongest norms of civic cooperation, and Ukraine, Greece and Belarus are among the countries with the weakest norms of civic cooperation.

The Rule of Law

The second main variable of interest is the Rule of Law. The Rule of Law indicator is a governance indicator developed by the World Bank (*S49*). Governance indicators measure traditions and institutions by which authority is exercised in a society. The Rule of Law indicator measures “the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, the police, and the courts, as well as the likelihood of crime and violence” (*S50*, p. 4). Specifically, it measures how well private and government contracts can be enforced in courts, whether the legal system is perceived as being fair, how important the black market and organized crime are, the quality of the police etc. The Rule of Law is one of six other indicators of governance (the others are “Voice and Accountability”, “Political Stability”, “Government Effectiveness”, “Regulatory Control” and “Control of Corruption”). These indicators are constructed from a host of other indicators and questionnaires (*S50*). The Rule of Law indicator, which is of main interest to us, is highly correlated with all other indicators – this holds for the countries of our subject pools (Spearman’s $\rho > 0.85$, $n = 15$, $p < 0.0001$) as well as in the whole available set of countries ($\rho > 0.83$, $n = 206$, $p < 0.0001$). Thus, it can be seen as an indicator how well formal institutions work in a country. The reason why this is interesting is that there are many arguments that even if contracts cannot be enforced well enough by courts they may, for instance, be self-enforced by social norms and punishment, networks, and long-term relationships (*S60-64*).

The Rule of Law indicator is also very highly correlated with the GDP per capita, a frequently used measure of economic prosperity (Spearman’s $\rho = 0.79$, $n = 175$, $p = 0.000$) (*S65*, *S66*). The correlation between GDP and the Rule of Law also holds in our subject pools: $\rho = 0.88$, $n = 15$, $p = 0.000$. Our subject pools come from societies that differ strongly in their economic prosperity. The “GDP per capita” ranges from \$7.6K in the poorest society of our sample to \$43.4K in the richest society in our sample. Compared to the world sample our subject pools stem from the 52nd percentile to the 99th percentile. Four of our subject pools come from societies below the world average, one society is close to the world average and the rest is substantially richer than the world average. Table S1 records the values and Figure S1 (panel B) plots them.

Further variables used for cross-country comparisons

Aside from the two variables of our main interest we report a series of other measures of societal norms, economic, or political conditions in Table S1 and Figure S1. The variable “Democracy” is taken from “World Audit” (*S51*), an organization which reports indicators for the status of democracy in a country. World Audit ranks a total of 150 countries with regard to the quality of democracy in these countries. World Audit combines four sub-indicators from different sources to calculate the Democracy variable. These are measures for political rights, civil liberties and press freedom reported by Freedom House (*S67*), and a measure for corruption reported by Transparency International (*S68*). The variable “Democracy” is scaled such that lower ranks mean more political rights, civil liberties and press freedom and less corruption. The numbers in Table S1 report the resulting rank of the societies represented in our subject pools. The societies of our subject pools cover almost the whole range of ranks: from rank 2 (= best overall democracy in our sample) to 137 (= worst overall democracy in our sample). (*S69*)

A set of variables that has gained considerable attention among scholars interested in cultural differences is the cultural dimensions of national cultures, as developed by G. Hofstede (*S26*, *S53*). We use the data as provided in (*S53*). Hofstede conducted his surveys among comparable employees of one big company (IBM) in 74 countries around the globe. Thus, the survey methodology behind the cultural dimensions is different from the WVS. Hofstede’s book (*S26*) provides extensive methodological details.

Hofstede argues that there are four distinct cultural dimensions that characterize different societies (*S70*): “Power distance” which measures how hierarchical/egalitarian a society is; “In-

dividualism” which measures how important the individual relative to the collective in a society is – cultures high on collectivism stress the importance of extended families and cohesive private networks where people care primarily for fellow in-group members, whereas in individualistic societies group boundaries are more permeable; “Masculinity”, which measures how strongly a society emphasizes gender differentiation and is dominated by males; and “Uncertainty avoidance”, which measures how tolerant a society is of uncertainty and ambiguity. We record the relevant indices in Table S1 and in Figure S1 (panels C and D) (S71). The societies of our subject pools differ quite strongly on all these dimensions and our sample is fairly representative of the overall available sample.

A second influential investigation of cultural differences is due to Inglehart and co-workers (S28-30), who argue that societies can be characterized by two dimensions: “traditional vs. secular-rational values” and “survival vs. self-expression values”. “Traditional vs. secular-rational values” refers to people’s attitudes on topics like abortion, national pride, obedience, and respect for authorities. “Survival vs. self-expression values” refers to attitudes on the importance of economic and physical security over self-expression and quality-of-life; homosexuality, happiness and trust. The data for these variables are available in the WVS (the variables are “tradrat5” and “survself”) (S72). We document the relevant values for our subject pools in Table S1 and Figure S1 (panel E). The countries of our subject pools span a fair range in both dimensions.

Subject pool details

Table S2 summarizes some key background figures of our 1120 participants. We aimed at recruiting subjects who were as homogeneous as possible across subject pools with respect to their socio-economic background. A second goal was to maximize the likelihood that subjects were strangers to one another. With these two goals in mind we recruited university students from various schools and universities in a given city as our subjects. Moreover, our choice of university students makes it likely that subjects share a similar (upper) middleclass background in their respective society. However, since participation in the experiment was voluntary and we could not “cherry-pick” our subjects we elicited important socio-economic background information in a post-experimental questionnaire. Our goal was to capture the most important variables (personal characteristics, family background, and social integration) that might influence cooperation and the readiness to punish. The main purpose of this information is to use it in our econometric analyses as control variables for subject pool composition effects.

Table S2 lists the cities and universities where we conducted our experiments, as well as summary statistics of the socio-economic background of our participants. The experimental sessions were conducted in computer laboratories of the universities indicated in Table S2. The fourth column shows the exchange rate of the experimental currency unit to the local currency. In the next column we report the total number of subjects who participated in the experiments at a given location. In most of the locations we first conducted 10 periods of the public goods experiment without the punishment option (the N-experiment) and then another 10 periods with the punishment option (the P-experiment). We refer to this sequence as the N-P sequence. In three subject pools (St. Gallen, Minsk, and Samara) we had access to large subject pools and therefore also conducted the reverse sequence (the P-N sequence) (S73). We show the number of subjects who participated in the P-N sequence in parentheses. Furthermore, we report the average age of the participants, and the percentage of female participants in each subject pool.

We measured the degree of between-subject anonymity by asking subjects in a post-experimental questionnaire how many other participants in the session they had known before. Most participants had not known anyone; the average participant had only known less than nine percent of other participants. Thus, on average, our subjects were indeed mostly strangers to one another. However, there are some differences between subject pools. In our statistical analysis we include the variable “percent known participants” in a given session to control for the degree of non-anonymity in a given session.

The dummy variable “Urban background” is based on a question on the size of the city in which a participant spent most of his or her life. This variable aims at capturing the degree of social anonymity subjects are possibly used to: social control is usually stronger in small vil-lages than big cities and this might matter for cooperation. The underlying variable contains four categories: (1) city size is up to 2’000 inhabitants; (2) between 2’000-10’000 inhabitants; (3) between 10’000-100’000 inhabitants and (4) more than 100’000 inhabitants. The variable “Ur-ban background” takes the value 1 if the subject spent most of his or her life in cities of at least 10’000 inhabitants (categories (3) and (4)). Almost two thirds of our subjects actually had an urban background.

We also asked participants about their personal judgment of whether their family income was “substantially below average”, “somewhat below average”, “average”, “somewhat above average”, or “far above average”. Family income is certainly positively correlated with both education and socio-economic status. As a proxy for socio-economic status we use the dummy variable “middle class” (1 if family income is at least average; 0 otherwise). Three-quarters of our participants have a middle-class background.

The dummy variable “Single child” aims at capturing the family background in socializa-tion. Participants who grew up with siblings might have been more strongly socialized into norms of cooperation and fair sharing than participants who grew up without siblings. Slightly less than fourteen percent of our participants grew up with no siblings.

The dummy variable “Membership” records whether a subject is a member of any civic voluntary association (political, interest groups, sports, culture, nonprofits, others). People who are members of a voluntary association might be more used to voluntary cooperation and its enforcement through social control than non-members. Moreover, scholars interested in social capital have argued for the relevance of this variable for a society’s social capital (S74, S75). Almost eighty percent of our participants report at least one membership.

City	University	Country	Exchange rate	Total # of subjects (# of subjects in P-N)	% known participants	Mean age	% Female	% Urban background	% Middle class	% Single child	% Membership
Boston	Harvard University	USA	USD .03	56 (0)	1.5	25.5	39.3	76.8	80.4	7.1	75.0
Nottingham	University of Nottingham	UK	GBP .015	56 (0)	0.3	20.1	51.8	55.4	87.5	7.1	98.2
Copenhagen	University of Copenhagen	Denmark	DKK .3	68 (0)	2.5	24.4	27.9	62.5	87.5	15.0	77.5
Bonn	University of Bonn	Germany	EUR .025	60 (0)	1.3	22.5	55.0	71.7	83.3	13.3	75.0
Zurich	University of Zurich	Switzerland	CHF .07	92 (44)†	1.3	21.7	34.8	33.7			
St. Gallen	University of St. Gallen	Switzerland	CHF .07	96 (48)	7.6	20.7	34.4	47.9	84.4	4.2	85.4
Minsk	Belarusian National Techni- cal University	Belarus	BYR 17	68 (36)	5.2	19.8	2.9	55.2	74.6	6.0	59.7
Dnipro- petrovs'k	Dnipropetrovs'k Regional Institute of Public Admini- stration	Ukraine	UAH .03	44 (0)	50.7	23.7	31.8	72.7	65.9	32.6	65.9
Samara	Samara State University	Russia	RUB .2	152 (72)	19.0	20.0	53.9	77.0	71.1	21.3	67.8
Athens	Panteion University	Greece	EUR .02	44 (0)	8.9	20.3	43.2	70.5	84.1	13.6	77.3
Istanbul	Bogazici University	Turkey	TRY .04	64 (0)	11.4	20.4	31.3	82.8	65.6	10.9	87.5
Riyadh	Imam Muhammad bin Saud University	Saudi Arabia	SAR .15	48 (0)	5.3	21.1	0.0	50.0	79.2	0.0	79.2
Muscat	Sultan Qaboos University	Oman	OMR .006	52 (0)	14.9	21.3	36.5	34.6	57.7	0.0	94.2
Seoul	Chung-Ang University	South Korea	KRW 20	84 (0)	1.7	23.8	46.4	81.0	81.0	8.3	85.7
Chengdu	Southwest Jiaotong Univer- sity	China	CNY .07	96 (0)	5.8	23.4	46.9	66.7	54.2	35.4	91.7
Melbourne	University of Melbourne	Australia	AUD .09	40 (0)	2.6	19.5	40.0	77.5	75.0	5.0	82.5
Total				1120 (200)	8.7	21.6	37.9	63.6	74.5	13.3	79.8

Table S2: Subject pool details. † In Zurich the 44 subjects in parentheses played only the P-experiment.

1.4 Experimental instructions

Here we provide a sample copy of the experimental instructions we used in our N-P experiments. The instructions for other sequences of treatments were adapted accordingly. The instructions were originally written in German (by S.G.) and translated into the respective language. We had them translated back to minimize translation-induced differences in meaning.

You are now taking part in an economic experiment financed by various foundations for research. 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 with care.

These instructions are solely for your private use. **It is prohibited to communicate with the other participants during the experiment.** Should you have any questions, please ask us. If you violate this rule, you will be dismissed from the experiment and forfeit all payments.

During the experiment we will not speak in terms of [national currency], but in Guilders. During the experiment your entire earnings will be calculated in Guilders. At the end of the experiment the total amount of Guilders you have earned will be converted to [national currency] at the following rate:

1 Guilder = [corresponding amount in national currency]

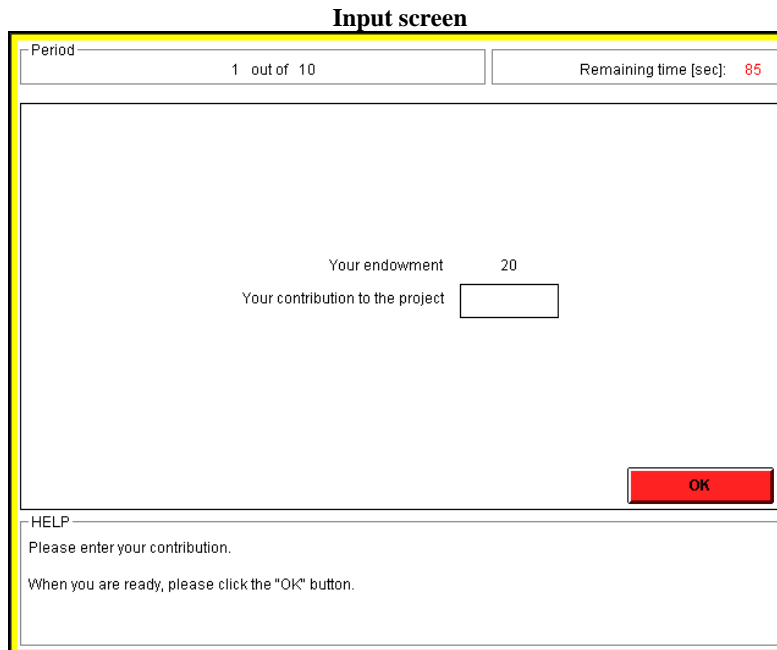
At the end of the experiment your entire earnings from the experiment plus the show-up fee will be paid to you **in cash**.

The experiment is divided into 10 separate periods. In each period the participants are divided into groups of four. You will therefore be in a group with 3 other participants. **The composition of the groups will stay the same for all ten periods.** In the following pages we describe the experiment in detail.

Detailed Information on the Experiment

At the beginning of each period each participant receives **20 tokens**. We call this his or her endowment. Your task is to decide how to use your **endowment**. You have to decide how many of the 20 tokens 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 for the first stage will appear:

Input screen



Period 1 out of 10 Remaining time [sec]: 85

Your endowment 20

Your contribution to the project

OK

HELP
Please enter your contribution.
When you are ready, please click the "OK" button.

The period number appears in the top left corner of the screen. In the top right corner you can see how many more **seconds** remain for you to decide on your contribution. You will have 90 seconds in the first two periods and 60 seconds in the remaining periods. Your decision must be made within the time limit.

Your **endowment in each period is 20 tokens**. You have to decide how many tokens 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 points to contribute to the project, you have also decided how many

points you keep for yourself. This is **(20 – your contribution)** tokens. After entering your contribution you must click the O.K. button. Once you have done this, your decision can no longer be revised.

After all members of your group have made their decision the following screen will show you the total amount of tokens contributed by all four group members to the project (including your contribution). This screen also shows you how many Guilders you have earned at the first stage.

Result screen

Period: 1 out of 10	Remaining time [sec]: 38
<p>Your contribution to the project</p> <p>Sum of contributions</p> <p>Income from retained tokens</p> <p>Income from the project</p> <p>Your total income in this period</p> <p>Your total income including this period</p>	
<input type="button" value="continue"/>	
<p>HELP</p> <p>This screen lists the results of this period.</p> <p>After time has expired or you have clicked the button, the experiment continues.</p>	

Your **income** consists of two parts:

(1) The tokens which you have kept for yourself (“Income from retained tokens”) whereby

1 token = 1 Guilder.

(2) The “income from the project”. This income is calculated as follows:

Your income from the project =
0.4 times the total contributions to the project.

Your **income in Guilders** of a period is therefore:

(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 tokens. In this case each member of the group receives an income from the project of: $0.4 \cdot 60 = 24$ Guilders. If the total contribution to the project is 9 tokens, then you and all other member of the group receive an income of $0.4 \cdot 9 = 3.6$ Guilders from the project.

For each token, which you keep for yourself you earn an income of 1 Guilder. Suppose you contributed this token to the project instead, then the total contribution to the project would rise by one token. Your income from the project would rise by $0.4 \cdot 1 = 0.4$ tokens. However the income of the other group members would also rise by 0.4 tokens each, so that the total income of the group from the project would rise by 1.6 tokens. Your contribution to the project therefore also raises the income of the other group members. On the other hand you earn an income for each token contributed by the other members to the project. For each token contributed by any member you earn $0.4 \cdot 1 = 0.4$ tokens.

In the first two periods you have 45 seconds and in the remaining periods 30 seconds to view this income screen. If you are finished before the time is up, please click the “continue”-button.

Next, the information screen appears, which reveals the contributions of the other group members.

Information screen

Period 1 out of 10	Remaining time [sec]: 42															
<table style="margin-left: auto; margin-right: auto;"> <tr> <td style="padding: 5px;">Endowment</td> <td style="padding: 5px;">....</td> <td style="padding: 5px;">....</td> <td style="padding: 5px;">....</td> <td style="padding: 5px;">....</td> </tr> <tr> <td style="padding: 5px;">Contribution to the project</td> <td style="padding: 5px;">....</td> <td style="padding: 5px;">....</td> <td style="padding: 5px;">....</td> <td style="padding: 5px;">....</td> </tr> <tr> <td style="padding: 5px;">Contribution in % of the endowment</td> <td style="padding: 5px;">....</td> <td style="padding: 5px;">....</td> <td style="padding: 5px;">....</td> <td style="padding: 5px;">....</td> </tr> </table>		Endowment	Contribution to the project	Contribution in % of the endowment
Endowment												
Contribution to the project												
Contribution in % of the endowment												
<input type="button" value="continue"/>																
<p>HELP Here you can see the other participants' decisions. When you are ready to continue, please click the button.</p>																

This screen shows how much each of the other group members contributed to the project. **Your contribution** is displayed **in blue in the first column**, while the contributions of the other group members are shown in the remaining three columns. Please note that the order in which contributions are displayed is changed randomly in each period. The contribution in the second column, for example, in general stems always from a different group member. The same holds for the contributions in the other columns. Besides the absolute contributions, the contributions as a percentage of the endowment are also displayed. Do you have any questions?

Control questions:

Please answer all control questions. They serve as a test for your understanding of payoff calculations.

1. Each group member has an endowment of 20 tokens. Suppose nobody (including you) contributes any tokens to the project. What is:
Your income?.....
The income of the other group members?.....
2. Each group member has an endowment of 20 tokens. Suppose you contribute 20 tokens to the project. All other group members each contribute 20 tokens to the project. What is:
Your income?.....
The income of the other group members?.....
3. Each group member has an endowment of 20 tokens. Suppose the other three group members contribute a total of 30 tokens to the project.
a) What is your income if you contribute 0 tokens to the project?.....
b) What is your income if you contribute 15 tokens to the project?.....
4. Each group member has an endowment of 20 tokens. Suppose you contribute 8 tokens to the project.
a) What is your income if the other group members together contribute a total of 7 tokens to the project?.....
b) What is your income if the other group members together contribute a total of 22 tokens to the project?.....

General explanation for the second experiment

We now repeat the experiment and introduce some changes. Each participant receives a lump-sum payment of **25 Guilders** at the beginning of the experiment (in addition to the show-up fee). This payment can be used to pay for eventual losses during the experiment. However, you can always avoid losses with certainty through your own decisions. At the end of the subsequent ten periods the whole experiment is finished and you receive:

Your income from the first 10 periods
+ your income from the second 10 periods (including the lump-sum payment of 25 Guilders)
= Total sum of Guilders
+ show-up fee.

This experiment consists of **two stages** in each period and altogether there are 10 periods. The first stage is identical to the previous experiment. At the first stage you have to decide how many tokens out of 20 you would like to contribute to a project (and hence you decide with it how many tokens you keep for yourself). Your income from the first stage will be calculated exactly in the same way as in the previous experiment.

For each token you keep for yourself, you earn an income of 1 Guilder. For any token you contribute to the project, you and all other group members will earn 0.4 Guilders. Therefore, each token that another group member contributes to the project will increase your income by 0.4 Guilders.

What is different in the new experiment?

Now there is a **second stage** introduced that follows the display of the income screen at the end of the first stage.

The second stage

At the second stage you see how many tokens each of the other group members contributed to the project. In addition, in this stage you can **decrease** the income of each other group member by assigning **deduction points** or by leaving the income **unchanged**. The other group members can also decrease **your** income if they wish to. This is apparent from the input screen at the second stage:

Input screen at the second stage

The screenshot shows a software interface for the second stage of the experiment. At the top, it displays 'Period 1 out of 10' and 'Remaining time [sec]: 178'. Below this, there is a table with four columns representing different group members' contributions. The rows are labeled 'Endowment', 'Contribution', and 'Contribution in percent of the endowment'. The first column (your own contribution) has a '-' sign next to it. Below the table, there are three input boxes for 'Your decision in stage 2'. Below these boxes, there are two lines of red text: 'Assign no points: 0' and 'Assign deduction-points: negative number'. At the bottom right, there are two red buttons: 'calculation' and 'OK'. At the bottom left, there is a 'HELP' section with the text: 'Please enter your decisions. Note the sign of your assigned points. Then click the button "calculation". Click "OK" when you are done.'

Besides the period and time display, the screen shows how much each group member contributed to the project at the first stage. Your contribution is displayed in blue in the first column, while the contributions made by the other group members are shown in the remaining three columns. Please note that the order in which contributions are displayed changes randomly in every period. The contribution in the second column, for example, generally represents a different group member each time. The same holds true for the contributions in the other columns. That way you are informed about the contributions but not about the identities of the other group members. In addition to the absolute contributions, the contribution as a percentage of the endowment is displayed.

You now have to decide whether, and if so, how many deduction points to assign to **each** of the other three group members. In any case 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. If you want to distribute deduction points, you must put a **negative sign** in front of the number (without spaces between them).

For this decision you have 180 seconds in the first two periods and 120 seconds in the remaining periods. You can move from one input field to the other by pressing the tab -key (\rightarrow |) or by using the mouse.

If you distribute deduction points, you have costs in Guilders that depend on the amount of deduction points you distribute. You can **assign between -10 and 0 points to each group member**. The larger the amount of deduction points that you assign, the larger your costs. The following formula indicates the relationship between the number of assigned points and the costs of assigning points:

$$\text{Costs of assigning deduction points} = \text{Sum of assigned deduction points.}$$

Every assigned deduction point costs you 1 Guilder. For example, if you assign 2 deduction points to one member, this costs you 2 Guilders; if, in addition, you assign 9 deduction points to another member this costs you 9 Guilders; and if you assign 0 points to the last group member this has no cost for you. In total you have assigned 11 points and your **total costs** therefore amount to 11 Guilders (2+9+0).

You can determine the total cost on the computer. To perform the calculation you have to click the button “**calculation**” (see the input screen at the second stage). You can do this after you have entered the deduction points. On the screen you will 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 (within the remaining time). To recalculate the costs after a change of your assigned points, simply press the “calculation” button again.

If you assign 0 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 (i.e., enter “-1”) you will **decrease** the income of this group member by 3 Guilders. If you assign a group member **2 deduction points** (i.e., enter “-2”), you will **decrease** the group member’s income by 6 Guilders, and so on. Each deduction point that you assign to another group member will reduce his or her income by 3 Guilders.

Whether or by how much the income at the second stage is decreased in total depends on the total of the received deduction points. If somebody, for instance, receives a **total of 3 deduction points** (from all other group members in this period), his or her income would be decreased by **9 Guilders**. If somebody receives a total of **4 deduction points**, his or her income is reduced by **12 Guilders**. Your total income from the two stages is therefore calculated as follows:

$$\begin{aligned} \text{Total income (in Guilders) at the end of the second stage} &= \text{period income} = \\ &= \text{income from the first stage} \quad (1) \\ &- 3 * (\text{sum of received deduction points}) \quad (2) \\ &- \text{costs of deduction points you have assigned} \\ &\quad \text{if } (1) + (2) \text{ is larger or equal to } 0; \\ &\text{OR} \\ &= 0 - \text{costs of deduction points you have assigned} \\ &\quad \text{if } (1) + (2) \text{ is less than } 0 \end{aligned}$$

Please note that your income in Guilders 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:

Income screen at the end of the second stage

Period	1 out of 10	Remaining time [sec]: 40
<p>Your income at the first stage </p> <p>Your costs of assigning deduction points </p> <p>Amount of received deduction-points </p> <p>Income reduction through deduction points </p> <p>Your income in this period </p> <p>Your total income including this period </p>		
<input type="button" value="continue"/>		
<p>Help</p> <p>This screen shows the results of the second stage. The experiment will continue after time has expired or when all participants have clicked the button.</p>		

Do you have any questions?

Control questions:

5. Suppose at the second stage you assign the following deduction points to your three other group members: -9,-5,0. What are the total costs of your assigned deduction points?.....
6. What are your costs if you assign a total of 0 points?.....
7. By how many Guilders will your income from the first stage be changed if you receive a total of 0 deduction points from the other group members?.....
8. By how many Guilders will your income from the first stage be changed if you receive a total of 4 deduction points from the other group members?.....
9. By how many Guilders will your income from the first stage be changed if you receive a total of 15 deduction points from the other group members?.....

1.5 Experimental procedures

We used these experimental procedures as guidelines for conducting the experiment and for ensuring that the procedures and verbal explanations were as similar as possible across subject pools. We document the experimental procedures as used in the N-P sequence in which we conducted the large majority of sessions. The procedures for the P-N sequence that we conducted in some subject pools were adapted accordingly.

1) Preparations before the experiment: Start up computers and z-Leafs, distribute instructions, prepare cards with participant IDs corresponding to the computer numbers in the lab, so that participants can be randomly allocated to computers. Also prepare X-cards for those who cannot participate in case of excess show-ups (the number of participants has to be divisible by four) and no volunteers for non-participation can be found.

2) Welcoming of participants outside the lab:

“Good afternoon, we are glad that you are participating in our experiment. You will learn in the lab what the experiment is about. In our experiment you will be divided into groups of four members. Therefore, we need a number of participants that is divisible by four. Now there are x people present.”

If the number of show-ups is divisible by four:

“We can now begin the experiment”.

If the number of show-ups is not divisible by four:

“The number of participants needs to be divisible by four. Currently, xx people are present. Is there anybody who would like to leave voluntarily?”... “In case there are no volunteers, we will choose the participants by a lottery: Everybody has to draw a card with a participation ID on it. This number corresponds to the computer number in the lab. Those who draw a card with an X on it can not participate and receive [the local show-up fee].”

After all participants have drawn a card they enter the lab.

3) Introduction in the lab:

“Thanks again for coming. You will learn from the instructions we have distributed to your computer place what the experiment is about. Before the experiment starts, I would like to point out that during the whole experiment communication is not allowed.”

“The data are only of scientific value, if we can be sure that you have made your decision without interferences of others. If you communicate with somebody, we can’t be sure and the experiment would lose its scientific value. In this case, we would not be able to pay you. Therefore, it should be in the common interest to follow this strict ban of communication. If you have any questions, please raise your hands. We will come to your place and answer the questions in private.”

“At the end of the instructions you will find control questions. These are not an exam. They serve only to ensure your understanding of how your earnings in this experiment will be calculated. You can start now reading the instructions.”

4) Participants read the instructions and solve the control questions. Questions are asked and answered in private and silently at the participant’s computer place. Questions like “What is the purpose of this experiment?”, “Why should I put money into the group project?”, etc. are not answered. It is best to say that we can only explain the rules of the experiment and how earnings in the experiment will be calculated. Once a participant has finished reading the instructions and answered the control questions, we check whether the answers are correct; in case of an incorrect answer the error is pointed out and the participant is asked to try again, until he or she has found the correct answer. The experiment cannot start before all participants have answered all control questions correctly.

5) When all have correctly answered all control questions, the experiment is summarized:

“All have answered the control questions correctly. Before we start the experiment, we summarize the experiment.”

“As you know you will do this experiment with three other group members. You will never learn who the other three group members are. This experiment lasts 10 periods. You are always in the same group. In each period you have to decide how many of the 20 tokens you contribute to a project and how many tokens you keep for yourself. Please be aware that you can not transfer tokens to the next period. In each period you

start with a new endowment of 20 tokens. You make your decision about the contribution to the project by entering the amount of your contribution in the Input screen.”

“When all four group members have made their decisions, an income screen will appear. The income screen lists your contribution to the project, the total sum of contributions of all four group members to the project, your income from the retained tokens, and your income from the project. “

“The income from retained tokens is the difference between 20 and your contribution to the project. Your income from the project is calculated as 0.4 times the total sum of all four group members’ contribution to the project.”

“After the income screen the so called “information screen” will appear. Here you will find a table. In the first column your contribution to the project (absolute and in %) is listed. In the other columns the other group members’ contribution are listed in a randomly chosen order. Do you have any questions?”

6) The N-experiment is started. Subjects take their decisions undisturbed and unobserved by the experimenters, who only observe the experiment by monitoring the software.

7) After the N-experiment has finished, the second experiment is announced:

“This experiment is now finished. Another experiment follows that will last 10 periods as well. After this, the entire experiment is finished. You will then have to answer a short questionnaire and will then get paid. We will now distribute the instructions for the new ten periods.”

8) Subjects read the new instructions and solve the control questions. Once all participants have solved the control question correctly, a short description of the procedure of the second experiment follows:

“The second experiment differs from the previous experiment. Now there is a second stage added. Please look at page 7 of the instructions. There you find the input screen of the second stage. This screen is similar to the information screen you know already from the first experiment.”

“New in this two-stage experiment is the possibility for you to assign deduction points (between 0 and -10) to the other group members. One deduction point costs you 1 guilder and reduces the income of the group member to whom you assign the deduction point by 3 guilders.”

“If you assign deduction points, you have to put a negative sign before the number. This two-stage experiment will be repeated 10 times with the same people in the group. Do you have any questions?”

9) Subjects play the P-experiment.

10) End: After the P-experiment has finished, a short questionnaire is announced and subjects are told that they will be paid in private after they had answered all questions.

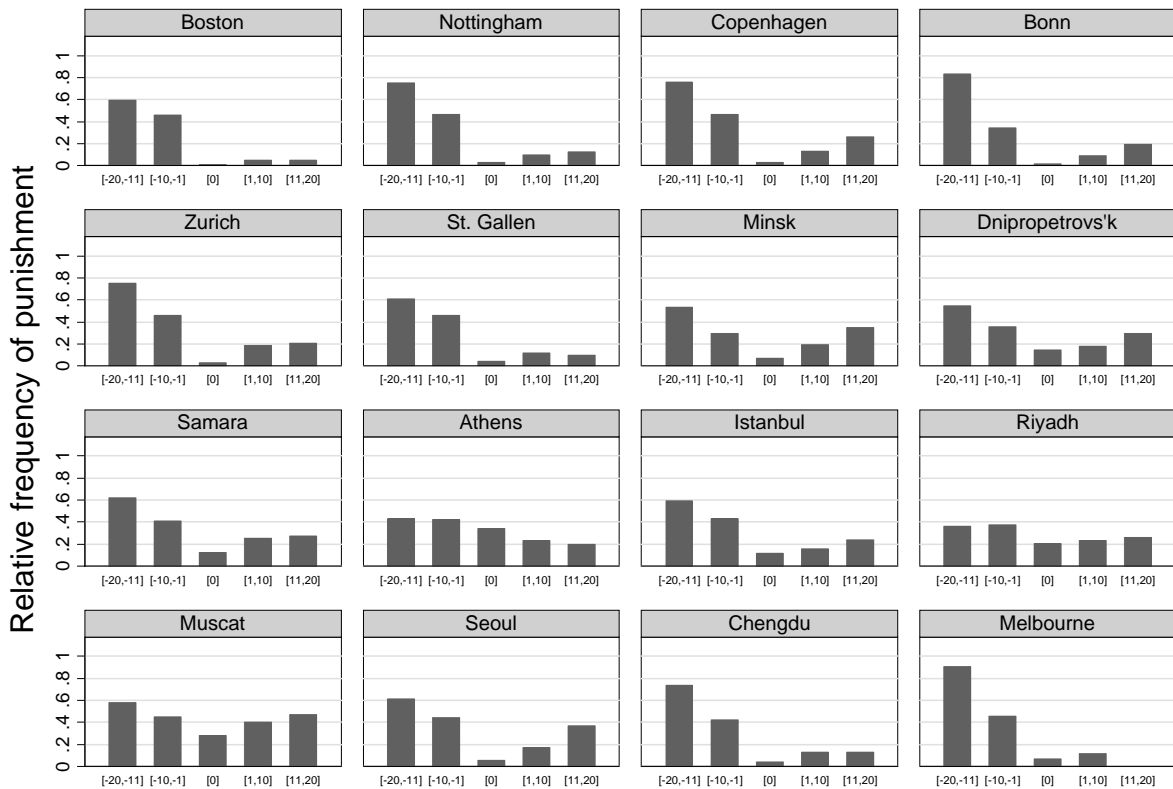
1.6 Laboratories and software

We conducted all experiments in networked computer laboratories, where participants in all sixteen laboratories were separated by partitions that ensured their anonymity. In all subject pools we used in the software “Zurich toolbox for ready-made economic experiments (z-Tree)”, developed by Urs Fischbacher (S76, S77). We had all texts that appeared on the computer screens translated into the respective language.

2. Supporting analyses

2.1 Punishment behavior

Figure 1 in the main text shows that punishment behavior was very different across subject pools, in particular for non-negative deviations (i.e., situations in which the punished subject contributed at least as much as the punisher). Figure 1 displays average punishment in case a given deviation from the punisher’s contribution has occurred. Figure S2 complements Figure 1 by depicting the relative frequency of punishment. Consistent with Figure 1A we find that the probabilities for punishing a free rider (bars at [-20,-11] and [-10,-1] respectively) are much more similar across subject pools than the probability of punishing anti-socially.



Graphs by Subject pool

Figure S2. Relative frequency of punishment for a given deviation from the punisher’s contribution. The deviations of the punished subject’s contribution from the punisher’s contribution are grouped into five intervals, where [-20,-11] ([-10,-1]) indicates that the punished subjects contributed between 11 and 20 tokens (1 and 10 tokens) less than the punishing subject; [0] indicates that the punished subject contributed exactly the same amount as the punishing subject and [1,10] ([11,20]) indicates that the punished subject contributed between 1 and 10 tokens (11 and 20 tokens) more than the punishing subject.

Our next step is to corroborate the graphical analyses by a regression analysis. We distinguish between negative deviations and non-negative deviations. We use a Tobit estimation procedure to account for the fact that our dependent variable “Assigned punishment points” is censored at 0 and 10 punishment points. We report robust standard errors clustered on groups as the independent units of observations (S78).

Our explanatory variables are (i) the “Punished subject’s contribution”, (ii) the “Punisher’s contribution”, (iii) the “Average contribution of others” (that is, the average contribution of the

two group members other than the punisher and the punished subject), (iv) the “Period” index (to capture time effects), and (v) a dummy “Final period” to capture last round effects in punishment. Table S3A and S3B report the results of our model for each subject pool. We distinguish between negative deviations and non-negative deviations.

Table S3A reports the estimation results for all situations in which the punished subject contributed less than the punishing subject. In all subject pools the punisher assigned the fewer punishment points the higher the punished subject’s contribution was (i.e., the coefficient of “Punished subject’s contribution” is negative). In fourteen subject pools the variable is highly significantly negative, whereas in two subject pools the coefficient is insignificantly negative. In other words, in these two subject pools the amount of assigned punishment was unrelated to the punished subject’s deviation from the punisher’s contribution; whereas in all other subject pools a subject got punished more the more he or she deviated from the punisher’s contribution.

The variable “Punisher’s contribution” measures to what extent the level of the punishing subject’s contribution influenced the assigned punishment points for a given negative deviation of the punished subject from the punisher’s contribution. The variable “Punisher’s contribution” is positive in all but one subject pools, which implies that punishment was harsher, *ceteris paribus*, the more the punisher contributed to the public good. This effect is (weakly) significant in eleven subject pools; in five subject pools the amount of punishment assigned was unrelated to the punisher’s own contribution level, *ceteris paribus*. “Average contribution others” is positive in all subject pools and significant (at $p < 0.05$) in twelve subject pools.

Our experiment was not designed to test for motives behind the punishment decision. However, we can use our data to investigate one specific motive for punishment, namely revenge. Punishment is motivated by revenge if – in addition to the contributions – the punishment in a given period t is also positively influenced by the punishment received in the previous period $t-1$. To capture the revenge motive we therefore add the number of received punishment points in the previous period as explanatory variable (in period one this variable is set to zero). If, *ceteris paribus*, received punishment in the previous round increases the probability of the use of punishment in the actual round we interpret this as evidence for revenge. The link between the punishment of free riding and having experienced punishment in the previous period is not clear. Positive and negative coefficients occur equally often in our subject pools.

The variable “Period” is mostly insignificant and weakly significantly negative in two subject pools. In other words, in these two subject pools there was some tendency of punishment to decrease over time, *ceteris paribus*. In the other subject pools the variable “Period” is mostly negative but insignificant. There was no last round effect in punishment of free riding because the dummy variable “Final period” is not significantly different from zero, with the exception of three subject pools, two positive and one negative.

Table S3B investigates the determinants of anti-social punishment. We find that, *ceteris paribus*, the level of the punished subject’s contribution was unrelated to punishment; the variable “Punished subject’s contribution” is significantly negative in only three subject pools and significantly positive in one subject pool. The variable “Punisher’s contribution” is (weakly) significantly negative in eleven subject pools. The average contribution level of the other two group members had a significantly positive effect on assigned punishment in eight subject pools and an insignificant effect in all other pools. The coefficient for “received punishment in $t-1$ ” is positive in all but one subject pools. In nine subject pools this effect is at least weakly significant. Thus, in the majority of subject pools revenge is one likely explanation for anti-social punishment. The “Period” variable is mostly negative and (weakly) significant in six subject pools. The variable “Final period” is significantly positive in five subject pools and insignificant in all others.

	Boston	Nottingham	Copenhagen	Bonn	Zurich	St. Gallen	Minsk	Dniprop.	Sumara	Athens	Istanbul	Riyadh	Muscet	Seoul	Chengdu	Melbourne
Punished subject's contribution	-0.229*** (0.058)	-0.280*** (0.082)	-0.279*** (0.048)	-0.332*** (0.071)	-0.229*** (0.032)	-0.172*** (0.045)	-0.333*** (0.068)	-0.475*** (0.127)	-0.306*** (0.056)	-0.474*** (0.103)	-0.422*** (0.073)	-0.102 (0.139)	-0.062 (0.105)	-0.416*** (0.071)	-0.295*** (0.044)	-0.505*** (0.051)
Punisher's contribution	0.273** (0.133)	0.169 (0.120)	-0.001 (0.094)	0.276*** (0.090)	0.155*** (0.043)	0.079* (0.045)	0.192*** (0.064)	0.152** (0.067)	0.147*** (0.032)	0.157** (0.074)	0.141 (0.087)	0.024 (0.098)	0.215* (0.119)	0.104 (0.088)	0.134** (0.059)	0.382*** (0.063)
Average contribution others	0.090 (0.071)	0.124*** (0.047)	0.346*** (0.084)	0.129 (0.126)	0.101*** (0.032)	0.202*** (0.044)	0.086 (0.066)	0.144** (0.067)	0.122*** (0.040)	0.258*** (0.042)	0.334*** (0.042)	0.172** (0.075)	0.134 (0.097)	0.223*** (0.040)	0.092*** (0.036)	0.150*** (0.044)
Received punishment in t-1	-0.275 (0.182)	-0.121 (0.092)	-0.231** (0.112)	-0.038 (0.092)	0.043 (0.081)	-0.056 (0.104)	0.263** (0.109)	0.208*** (0.067)	0.159*** (0.043)	0.121** (0.054)	-0.018 (0.052)	0.256*** (0.075)	-0.002 (0.089)	-0.026 (0.067)	0.113* (0.064)	0.032 (0.091)
Period	-0.167** (0.067)	-0.157 (0.121)	-0.060 (0.091)	-0.121 (0.082)	-0.030 (0.043)	-0.080 (0.083)	0.027 (0.140)	-0.123 (0.113)	-0.080* (0.042)	-0.107 (0.131)	-0.116 (0.085)	-0.059 (0.153)	0.032 (0.125)	0.098 (0.079)	0.050 (0.068)	-0.018 (0.062)
Final period	0.038 (1.220)	0.045 (1.244)	1.081 (0.771)	1.279* (0.754)	0.547 (0.431)	1.546** (0.763)	-0.973 (1.081)	-0.439 (1.108)	0.121 (0.328)	-0.467 (0.892)	-0.416 (0.691)	-0.106 (1.040)	0.028 (0.819)	-0.474 (0.589)	-0.155 (0.509)	-1.674** (0.788)
Constant	-3.108*** (0.950)	-0.885 (0.914)	-1.780** (0.889)	-2.785** (1.339)	-1.348 (0.830)	-2.307** (1.070)	-3.857*** (1.119)	-1.646 (2.008)	-1.655** (0.754)	-1.369 (0.936)	-1.555* (0.803)	-3.050*** (1.009)	-4.616** (2.352)	-0.939 (0.794)	-1.115* (0.660)	-2.020** (0.874)
sigma	3.489	2.727	2.607	2.746	2.365	2.935	4.362	4.042	3.373	3.052	3.525	4.951	4.520	2.690	2.323	2.036
p	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.667	0.000	0.000	0.000
ll	-290	-680	-359	-717	-1049	-910	-916	-694	-2612	-701	-959	-886	-1089	-984	-1108	-454
N	178	440	240	556	721	585	707	492	1761	505	654	619	645	709	819	347

Table S3A: Punishment of negative deviations from the punisher's contributions, per subject pool. Tobit estimates with robust standard errors. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	Boston	Nottingham	Copenhagen	Bonn	Zurich	St. Gallen	Minsk	Dniprop.	Samara	Athens	Istanbul	Riyadh	Muscet	Seoul	Chengdu	Melbourne
Punished subject's contribution	-0.732** (0.296)	0.003 (0.065)	-0.041 (0.076)	0.160 (0.173)	-0.129** (0.053)	-0.249** (0.103)	0.107 (0.072)	-0.023 (0.040)	-0.066 (0.071)	-0.241 (0.205)	0.004 (0.123)	0.042 (0.067)	0.148*** (0.048)	0.005 (0.052)	-0.088 (0.063)	-0.142 (0.180)
Punisher's contribution	-0.361* (0.187)	-0.127* (0.072)	-0.352*** (0.063)	-0.330*** (0.103)	-0.151*** (0.037)	-0.135 (0.102)	-0.374*** (0.084)	-0.105 (0.071)	-0.223*** (0.068)	-0.457 (0.298)	-0.251** (0.098)	0.029 (0.085)	-0.087** (0.045)	-0.296*** (0.076)	-0.194*** (0.057)	-0.100 (0.164)
Average contribution others	0.459* (0.268)	0.214*** (0.078)	0.005 (0.154)	-0.017 (0.090)	0.133** (0.058)	0.256* (0.134)	-0.073 (0.125)	0.141* (0.083)	0.067 (0.055)	-0.026 (0.205)	0.136 (0.103)	0.036 (0.072)	0.188*** (0.061)	0.205*** (0.065)	0.054 (0.076)	0.616* (0.347)
Received punishment in t-1	1.670*** (0.560)	0.357*** (0.131)	0.188 (0.120)	0.356*** (0.135)	0.335*** (0.099)	0.686*** (0.178)	0.171 (0.107)	0.077 (0.083)	0.314*** (0.068)	0.384* (0.199)	0.055 (0.110)	0.168 (0.150)	-0.214 (0.192)	0.240*** (0.072)	0.240 (0.149)	0.412* (0.231)
Period	-0.823 (1.043)	-0.162 (0.163)	-0.111 (0.175)	0.116 (0.096)	-0.075 (0.072)	-0.297** (0.138)	-0.491** (0.192)	-0.163** (0.078)	-0.425*** (0.084)	-0.337** (0.151)	-0.360** (0.166)	-0.021 (0.167)	0.023 (0.120)	0.044 (0.099)	0.051 (0.102)	-0.444 (0.298)
Final period	16.230*** (5.108)	-0.135 (1.996)	1.620 (1.612)	-0.597 (1.803)	1.568** (0.737)	5.734*** (1.873)	1.989 (1.723)	2.189*** (0.801)	2.114** (0.902)	-0.260 (1.023)	0.975 (1.431)	0.238 (0.860)	0.811 (0.938)	0.103 (0.594)	0.666 (0.989)	2.112 (3.012)
Constant	-26.357* (13.789)	-7.172*** (1.942)	-2.144 (2.631)	-7.228** (3.546)	-1.856 (1.178)	-7.788*** (2.964)	-1.718 (1.642)	-2.976** (1.200)	-1.094 (1.056)	-0.472 (3.187)	-3.809*** (1.242)	-4.958*** (1.867)	-4.236** (1.839)	-3.791*** (0.720)	-3.642*** (1.102)	-11.970** (5.463)
sigma	17.650	3.497	5.034	4.972	2.799	6.201	6.462	3.317	5.292	6.215	5.488	5.191	4.861	3.148	3.999	5.548
p	0.057	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.884	0.001	0.000	0.000	0.716
ll	-159	-338	-424	-376	-732	-718	-921	-615	-2429	-875	-844	-821	-1305	-815	-751	-353
N	1502	1240	1800	1244	2039	2295	1333	828	2799	815	1266	821	915	1811	2061	853

Table S3B: Punishment of non-negative deviations from the punisher's contribution, per subject pool. Tobit estimates with robust standard errors. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

In Table S4 we investigate subject pool differences in the whole data set. We add subject pool dummies to see whether there are significant *ceteris paribus* differences in mean punishment across subject pools. The estimation contains all subject pool dummies and no constant term. In addition to that we add individual socio-economic control variables elicited by a post-experimental questionnaire (see Table S2). Some of these controls were not elicited in Zurich. Data from Zurich are therefore omitted in the estimates reported in Table S4. These variables are important to disentangle subject pool effects from possible biases in the composition of the different subject pools. We record the estimation results in Table S4. We report estimates for free-rider and anti-social punishment with and without the parameter for received punishment in the previous period. The first set of variables confirms the insights gained from Tables S3A and S3B. Punishment of free riding increases in the punisher’s contribution and the average contribution of the other group members. It decreases in the punished subject’s contribution and has a negative time trend. Anti-social punishment decreases in the punisher’s contribution and is unrelated to the contribution of the punished subject.

The subject pool dummies measure whether there was a significant difference in mean punishment for a given value of the other explanatory variables. Almost all dummies are significant, which is *per se* not informative. The crucial question is whether they differ from each other. A test for equality of the subject pool dummies for the second model reported in Table S4 yields $\chi^2(14) = 22.8$, $p = 0.063$, which implies that there were only weakly significant *ceteris paribus* subject pool differences in mean punishment of free riding.

The third and fourth model in Table S4 estimate the impact of the exact same variables for all situations in which the target subject contributed at least as much as the punishing subject. We find that (i) punishment decreased significantly in the punisher’s contribution, (ii) was unrelated to the punished subject’s contribution, (iii) increased significantly with the average contribution level of the other group members, and (iv) decreased over time with the exception of the final period in which punishment of non-negative deviations was significantly higher than in all other periods. The test for equality of the subject pool dummies for the fourth model reported in Table S4 yields $\chi^2(14) = 59.7$, $p = 0.000$, which implies that there were highly significant *ceteris paribus* subject pool differences in the punishment of non-negative deviations.

2.2 Cooperation in the P-experiment

Figure 2A in the main text suggests that contributions in all subject pools are stable or even increase over time. Thus, one goal is to test whether there is a significant time trend or not. Table S5 documents the results of Tobit estimations explaining the individual contribution by the explanatory variables “Period” (i.e., the period number) and a dummy “Final period”. We chose a Tobit estimation procedure because contributions are constrained between 0 and 20 tokens by design, and in almost all subject pools we find that contributions at 0 and 20 tokens are the two most frequent contribution levels. We calculate robust standard errors clustered on groups.

On the basis of previous evidence from experiments with comparable parameters (S43, S45, S79-85) we predict that “Period” is non-negative, that is, contributions do not decline over time. The results reported in Table S5 support this prediction. We report separate estimates for each subject pool. The variable “Period” is significantly positive (at 10 percent or better) in nine subject pools and insignificantly different from zero in the other subject pools. Thus, across all periods punishment stabilized or increased average cooperation everywhere. Eleven subject pools showed a (weakly) significantly negative endgame effect (variable “Final period”), whereas one subject pool exhibited a significantly positive endgame effect.

	Dependent variable: Assigned punishment points			
	Punishment of free riding		Anti-social punishment	
Punished subject's contribution	-0.292*** (0.020)	-0.290*** (0.020)	-0.004 (0.032)	-0.027 (0.029)
Punisher's contribution	0.147*** (0.020)	0.148*** (0.020)	-0.259*** (0.029)	-0.209*** (0.026)
Average contribution others	0.138*** (0.017)	0.138*** (0.017)	0.112*** (0.028)	0.093*** (0.025)
Received punishment in t-1		0.101*** (0.021)		0.297*** (0.042)
Period	-0.069*** (0.025)	-0.076*** (0.025)	-0.168*** (0.038)	-0.200*** (0.037)
Final period	0.069 (0.200)	0.125 (0.201)	1.450*** (0.367)	1.703*** (0.374)
Boston	-2.072*** (0.694)	-2.167*** (0.684)	-8.117*** (1.254)	-7.972*** (1.184)
Nottingham	-1.524** (0.629)	-1.604*** (0.616)	-6.870*** (1.471)	-7.033*** (1.372)
Copenhagen	-2.035*** (0.568)	-2.071*** (0.561)	-8.927*** (1.423)	-8.697*** (1.363)
Bonn	-1.832*** (0.696)	-1.921*** (0.678)	-6.349*** (1.152)	-6.350*** (1.078)
St. Gallen	-1.870*** (0.669)	-1.938*** (0.658)	-5.876*** (1.201)	-5.936*** (1.108)
Minsk	-2.826*** (0.696)	-2.980*** (0.684)	-3.606*** (0.918)	-3.882*** (0.827)
Dnipropetrovsk	-2.910*** (0.601)	-2.973*** (0.598)	-4.302*** (0.991)	-4.407*** (0.965)
Samara	-1.804*** (0.605)	-2.001*** (0.583)	-3.055*** (0.999)	-3.385*** (0.870)
Athens	-1.595*** (0.572)	-1.799*** (0.558)	-2.380 (1.451)	-3.099** (1.210)
Istanbul	-1.253** (0.549)	-1.385** (0.541)	-4.682*** (0.894)	-4.879*** (0.865)
Riyadh	-2.200*** (0.654)	-2.407*** (0.632)	-3.273*** (0.880)	-3.609*** (0.827)
Muscat	-1.425** (0.669)	-1.769*** (0.665)	-0.486 (1.084)	-1.076 (0.953)
Seoul	-1.423** (0.588)	-1.545*** (0.582)	-4.634*** (0.982)	-4.716*** (0.918)
Chengdu	-1.705*** (0.605)	-1.811*** (0.599)	-6.004*** (1.033)	-5.981*** (0.985)
Melbourne	-1.230** (0.569)	-1.318** (0.554)	-5.161*** (1.400)	-5.269*** (1.301)
Socio-economic controls	yes	yes	yes	yes
sigma	3.440	3.429	5.450	5.356
p	0.000	0.000	0.000	0.000
ll	-13259	-13235	-11708	-11569
N	8950	8950	20660	20660

Table S4: Punishment behavior in the pooled data set. Tobit estimates with robust standard errors, clustered on groups; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. “Punisher’s contribution” is the contribution of a punishing subject in period t . “Punished subject’s contribution” is the contribution of the punished subject in period t . “Average contribution others” is the average contribution in period t of the two group members other than the punished and the punishing subject. “Period” is the period number and “Final period” is a dummy for period 10. The socio-economic control variables are those of Table S2.

Dependent variable: Contribution

	Boston	Nottingham	Copenhagen	Bonn	Zurich	St. Gallen	Minsk	Dniprop.	Samara	Athens	Istanbul	Riyadh	Muscat	Seoul	Chengdu	Melbourne
Period	1.573 (1.116)	1.581*** (0.569)	2.014** (1.015)	0.564*** (0.150)	1.054*** (0.288)	0.672 (0.455)	0.203 (0.216)	0.104 (0.236)	0.309** (0.148)	-0.191 (0.136)	0.368* (0.211)	-0.001 (0.299)	0.085 (0.137)	1.280*** (0.143)	1.318*** (0.298)	1.406*** (0.279)
Final period	-15.701* (8.631)	-7.686** (3.031)	-11.887* (6.650)	-2.596** (1.131)	-3.825** (1.903)	-7.233*** (2.329)	-2.850* (1.670)	-1.484 (1.490)	-1.616* (0.903)	1.743** (0.686)	-1.302 (1.051)	-0.578 (2.066)	-1.973 (1.316)	-1.209* (0.688)	-4.423*** (1.478)	-4.167*** (1.555)
Constant	37.592*** (9.665)	12.689*** (2.587)	24.926*** (5.172)	13.122*** (1.401)	15.330*** (1.770)	23.783*** (3.416)	14.335*** (1.688)	11.442*** (1.992)	11.344*** (0.945)	5.820*** (1.112)	4.651** (1.840)	6.030*** (1.358)	9.484*** (1.379)	10.091*** (1.505)	9.372*** (1.219)	8.051*** (0.710)
sigma	25.124	13.598	18.642	7.436	10.296	16.047	10.808	11.228	10.222	6.234	10.049	8.583	10.781	8.478	9.550	6.341
p	0.191	0.013	0.120	0.001	0.001	0.006	0.214	0.606	0.074	0.027	0.170	0.935	0.314	0.000	0.000	0.000
ll	-518	-1156	-784	-1532	-1863	-1545	-1779	-1200	-4203	-1201	-1779	-1344	-1466	-1991	-2356	-991
N	560	560	680	600	920	960	680	440	1520	440	640	480	520	840	960	400

Table S5: Contributions over time in the P-experiment. Tobit estimates with robust standard errors clustered on groups. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

2.3 (Relative) payoffs in the N- and the P-experiment (efficiency)

The fact that subject pools exhibited vastly different levels of cooperation and punishment also led to large differences in earnings. We document the average per-period earnings (in experimental money units) in Table S6. The final column of Table S6 reports the percentage change in earnings in the P-experiment relative to the earnings in the N-experiment.

In the P-experiments the average per-period earnings of subject pools varied from 11.0 to 27.9 money units, that is, by more than 250 percent. In the N-experiments differences in earnings were much smaller, ranging from 23.3 to 26.9 money units.

	Average earnings in		Percentage change relative to N-experiment
	N-experiment	P-experiment	
Boston	25.6	27.9	9.1%
Copenhagen	26.9	27.7	2.8%
Melbourne	23.0	23.1	0.5%
Zurich	25.6	25.1	-1.8%
Nottingham	24.2	24.1	-0.4%
St. Gallen	26.1	25.5	-2.2%
Seoul	24.8	24.0	-3.0%
Chengdu	24.8	23.9	-3.6%
Bonn	25.5	24.1	-5.4%
Minsk	26.3	20.0	-23.9%
Istanbul	23.3	17.0	-26.9%
Dnipropetrovs'k	26.4	18.9	-28.4%
Samara	25.8	17.8	-30.8%
Riyadh	24.6	13.9	-43.6%
Athens	23.8	13.2	-44.6%
Muscat	26.0	11.0	-57.9%

Table S6: Average per-period earnings in experimental currency units in the N- and the P-experiment. Subject pools are sorted by the percentage change in earnings relative to the N-experiment.

Figure S3 complements Table S6 by showing the development of relative earnings over time (that is, earnings in the P-experiment / Earnings in the N-experiment). This earnings ratio is a measure of the relative efficiency of treatments. An OLS regression of relative earnings on “Period” showed that the relative efficiency increased over time in eleven subject pools, and stayed constant in five subject pools (Minsk, Dnipropetrovs'k, Athens, Riyadh, and Muscat).

However, the speed at which the relative improvements occurred varied greatly between pools. During the ten periods of the experiment, eight subject pools were eventually able to achieve higher earnings in the punishment condition relative to the no-punishment condition. The points in time at which this relative improvement occurred ranged from period 2 in the “fastest” subject pool to period 7 in the “slowest” pool (Figure S3).

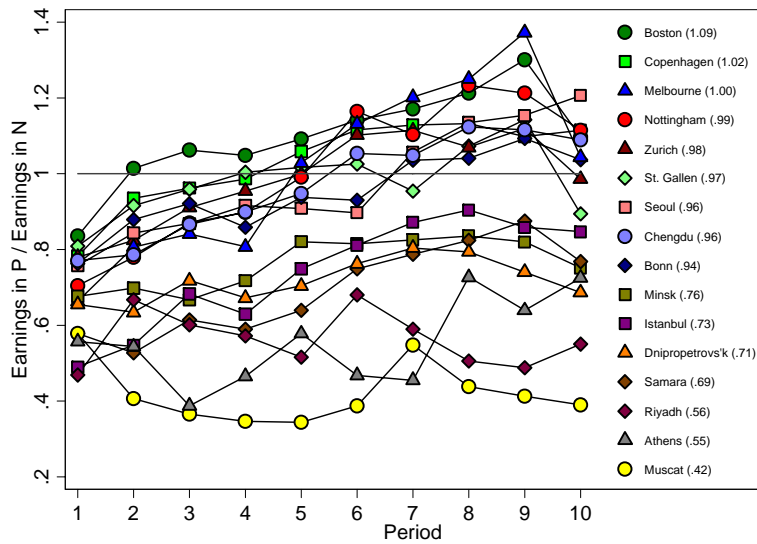


Figure S3: Relative earnings in the P- and the N-experiment over time. We calculated for each subject pool and every period the average earnings in the N-experiment and in the P-experiment. On the vertical axis we depict the ratio between the two earnings. The numbers in parentheses show the earnings ratio over all ten periods.

2.4 Reactions to received punishment

A clear message from Figure 2A in the main text is that subject pools contributed very differently in the P-experiment. However, Figure 2A only shows the aggregate consequence (in terms of the mean contribution level) when a punishment option is present. In the following analysis we use regression techniques to disentangle the average cooperation level as a function of received punishment.

There are two situations: a subject can either have contributed less than the group average or at least the group average. We look at how a subject reacted who got punished for a contribution above the group average (or equal) or for a contribution below the group average. We model this as follows: The dependent variable is the change in contributions between period t and $t+1$. The explanatory variables are the amount of punishment received, and variables measuring the time trend (“Period” and “Final period”). We estimate this model either for the situations in which a group member contributed less than the other group members or at least as much as the other group members, and for each subject pool separately. Table S7 reports the results of OLS estimations (robust standard errors clustered on the independent groups).

Table S7A documents the behavioral reactions in case a subject contributed less than the other group members on average. The estimated coefficient of “Received punishment” is positive in all but one subject pools, and significant (at five percent) in eleven subject pools. The size of the coefficient varies strongly between the subject pools and ranges from virtually zero to more than one effort unit. In the main article we document that this coefficient is negatively connected with anti-social punishment in the respective subject pool. We also estimated the reaction to punishment for the minimal contributor(s) in a group (given that they contribute less than 20). The results (not reported in the Table) are quite similar and there is also a strong correlation between subject pool averages in anti-social punishment and the reaction to punishment (Spearman’s $\rho=-0.83$, $p=0.000$, $n=16$).

Table S7B reports the analogous regressions for all situations in which a subject contributed at least as much as the group average. We find that in seven subject pools subjects lowered their contributions at least weakly significantly per punishment point received; in all but one other subject pools subjects did not change their contributions significantly.

Change in contribution if a subject contributed less than the other group members																
	Boston	Nottingham	Copenhagen	Bonn	Zurich	St. Gallen	Minsk	Dniprop.	Samara	Athens	Istanbul	Riyadh	Muscat	Seoul	Chengdu	Melbourne
Received punishment	1.152*** (0.216)	0.571** (0.253)	0.876*** (0.168)	0.772*** (0.159)	0.719*** (0.116)	0.749*** (0.181)	0.123 (0.110)	-0.019 (0.078)	0.247** (0.089)	0.049 (0.081)	0.365*** (0.099)	0.082 (0.072)	0.141 (0.097)	0.559*** (0.101)	0.425*** (0.071)	0.518*** (0.115)
Period	-0.319 (0.299)	-0.363* (0.176)	-0.298 (0.322)	-0.000 (0.183)	-0.017 (0.135)	0.178 (0.191)	-0.134 (0.118)	-0.427** (0.138)	-0.077 (0.084)	-0.124 (0.103)	0.003 (0.097)	-0.079 (0.145)	-0.268** (0.121)	-0.166** (0.062)	-0.204* (0.119)	-0.292* (0.151)
Final period	0.144 (3.835)	-0.139 (1.978)	1.590 (2.788)	-0.464 (1.001)	-0.430 (0.808)	-4.348** (1.880)	1.799 (1.279)	1.276 (1.159)	-0.168 (0.920)	3.298** (1.478)	-0.299 (0.792)	-0.022 (1.296)	0.975 (1.755)	-0.686 (1.065)	-2.115*** (0.730)	-3.518** (1.084)
Constant	1.781 (1.919)	2.635* (1.357)	2.392 (1.993)	0.904 (0.810)	1.045 (0.733)	0.298 (0.825)	2.996*** (0.886)	4.262*** (0.704)	2.187*** (0.563)	1.386* (0.745)	0.304 (0.531)	1.737** (0.733)	2.536*** (0.738)	2.180*** (0.313)	3.126*** (0.599)	2.525** (0.817)
r ²	0.374	0.210	0.320	0.276	0.193	0.199	0.018	0.040	0.039	0.053	0.146	0.006	0.019	0.238	0.146	0.331
N	51	133	89	192	232	186	213	173	610	201	244	234	215	284	302	134

Table S7A: OLS estimation with robust standard errors clustered on groups; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the change in the contribution ($g_{i,t+1} - g_{i,t}$) given the subject contributes less than the group average in t ($g_{i,t} - g_{mean,t} < 0$). "Received punishment" is the sum of the points received from the three other group members in period t . "Final period" is the period number. "Final period" is a dummy for the last period.

Change in contribution if a subject contributed more than the other group members or equal																
	Boston	Nottingham	Copenhagen	Bonn	Zurich	St. Gallen	Minsk	Dniprop.	Samara	Athens	Istanbul	Riyadh	Muscat	Seoul	Chengdu	Melbourne
Received punishment	0.005 (0.059)	0.370* (0.186)	-0.582* (0.312)	-0.681*** (0.069)	-0.054 (0.164)	0.058 (0.165)	-0.351*** (0.090)	-0.270 (0.318)	-0.265** (0.102)	0.001 (0.051)	-0.174** (0.071)	-0.071 (0.177)	-0.097* (0.052)	-0.332* (0.191)	-0.340 (0.297)	0.040 (0.059)
Period	-0.066 (0.064)	0.011 (0.044)	-0.017 (0.048)	-0.019 (0.057)	-0.045 (0.062)	-0.201** (0.088)	-0.245* (0.119)	-0.126 (0.177)	0.053 (0.083)	0.051 (0.125)	0.130 (0.076)	0.018 (0.165)	0.052 (0.139)	-0.031 (0.053)	-0.005 (0.052)	-0.025 (0.063)
Final period	-1.099 (0.913)	-1.654* (0.794)	-1.250 (1.012)	-1.995** (0.748)	-1.252 (0.742)	-0.329 (0.521)	-0.655 (0.858)	1.658 (1.130)	-1.033 (0.727)	1.259 (0.703)	-0.619 (0.525)	-1.573 (1.855)	-1.133 (1.252)	0.257 (0.294)	-0.823 (0.588)	-0.622 (0.597)
Constant	-0.095 (0.260)	-0.633 (0.427)	-0.174 (0.290)	-0.380 (0.288)	-0.292 (0.366)	0.167 (0.261)	0.198 (0.618)	-0.934 (0.767)	-1.747*** (0.532)	-1.679 (1.006)	-1.251** (0.498)	-1.575 (1.022)	-1.408 (0.847)	-0.021 (0.301)	-0.549 (0.357)	0.199 (0.392)
r ²	0.026	0.026	0.056	0.104	0.021	0.022	0.051	0.013	0.023	0.020	0.027	0.007	0.009	0.021	0.015	0.017
N	453	371	523	348	596	678	399	223	758	195	332	198	253	472	562	226

Table S7B: OLS estimation for the change in contribution given the subject contributed at least as much as the group average, i.e. ($g_{i,t} - g_{mean,t} \geq 0$). The explanatory variables are the same as in Table S7A.

