Supporting Material for

The Long-Run Benefits of Punishment

Simon Gächter*, Elke Renner & Martin Sefton

Centre for Decision Research and Experimental Economics, University of Nottingham, School of Economics, Sir Clive Granger Building, University Park, Nottingham NG7 2RD, United Kingdom.

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

13 September 2008

This file includes:

Materials and methods
Supporting analyses
Figure S1
Table S1
Notes and references
1. Materials and Methods

1.1 Experimental Design

*The basic decision situations.* The workhorse for our analysis is the public goods game with and without punishment (S1). The public goods game is a stylized model of situations that require cooperation to achieve socially beneficial outcomes in the presence of free-rider incentives. Examples from both our ancestral past (S2) and current life abound (S3): warfare, cooperative hunting, common pool resource management, contributing to community public goods, voting, paying taxes, fighting corruption, teamwork, work morale, neighborhood watch, recycling, tackling climate change, and so on. These are frequent n-person interactions with the common feature that cooperation leads to a group-beneficial outcome but is jeopardized by selfish incentives to free ride on others’ contributions.

To implement a cooperation game without and with punishment opportunities, we adapt a design developed by (S1). We conduct the public goods experiment with real monetary stakes and two conditions – a no-punishment condition (the N-experiment) and a punishment condition (the P-experiment). Groups of three members play the following public goods game in both conditions: Each member receives an endowment of 20 tokens. Participants decide how many tokens to keep for themselves and how many to contribute to a group project. Each of the three group members earns 0.5 money units for each token contributed to the project, regardless of his or her own contribution. Since the cost of contributing one token to the project is exactly one money unit while the private return on that token is only 0.5 money units, keeping all one’s own tokens is always in any subject's material self-interest – irrespective of how much the other three group members contribute. Yet, if all tokens are retained each group member earns 20 money units; on the other hand, each member earns $0.5 \times 60 = 30$ money units if each of them contributes their entire 20 token endowment.

The only and crucial difference between the P-experiment and the N-experiment is that participants in the P-experiment can punish each of the other group members after being informed of their contributions, while the N-experiment ends after participants are informed of the other group members' contributions. A punishment decision is
implemented by assigning the punished member between zero and five "deduction points". Each assigned deduction point reduces the punished member’s earnings by three money units, and costs the punishing member one money unit.

Participants were never 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. Punishment decisions were also implemented simultaneously and participants were only informed about the total punishment they received and not about who punished them.

**The time horizon.** In our experiment the basic decision situations described above are played repeatedly. To examine the implications of the time horizon for cooperation and punishment we conduct four treatments: the experiments either run for ten periods (N10 and P10 treatments) or for fifty periods (N50 and P50 treatments).

Most previous cooperation experiments ran for ten periods or less, so our N10 and P10 experiments replicate the most common length (S1, S4-13). Some studies lasted between twenty and thirty periods (S14-17). We are not aware of any cooperation experiment with punishment that lasted more than (an expected length of) thirty periods. Thus, our fifty-period experiments run for considerably more periods than the large majority of all punishment experiments. Moreover, no study looked at the role of the time horizons in experiments with and without punishment and thus isolates the interaction effects of time horizon with punishment opportunities.

**Further design considerations.** We conduct our experiments according to a fixed matching protocol (“Partners” (S18)) because we are interested in how punishment affects the welfare of groups with a stable membership. Thus, our statistical and econometric analysis is mainly conducted at the group level, which constitutes the independent unit of analysis. The supplementary analyses reported below complement this statistical analysis with regressions based on individual subject decisions.

We use n-person public good games rather than two-person prisoner’s dilemma games (S19). This is because many important cooperation problems in reality are multilateral in nature and give people more finely-tuned possibilities to cooperate than choosing between cooperate and defect. Furthermore, as has been shown theoretically (e.g., (S20-22)), cooperation is hard to sustain in groups larger than two. In a dyadic
relationship a player can also punish the opponent by defecting, whereas in \( n \)-person games such targeted punishment by defection is not possible – defecting as a reaction to a defector also punishes the other cooperators. Finally, sociological and anthropological studies on punishment suggest that punishment is typically implemented after a particular action (a norm violation) by the target of punishment has been observed (S23-25). For these reasons we move beyond dyadic relationships, and implement punishment as a separate stage in the game.

Our use of three person groups has implications for the effectiveness of punishment. From previous studies we know that the severity of punishment a low contributor receives determines how strongly punishment can increase cooperation (S7, S11). The received punishment is a function of the number of potential punishers and the income reduction (“fine”) per punishment point received. We decided to keep the fine to cost-of-punishment ratio the same as in previous studies (1 punishment point cost 1 money unit and reduces the target’s income by three money units) (S1, S7, S11, S12). However, while previous experiments often used groups of four, so each group member could be punished by three other group members, in our setup each group member can be punished by only two other group members. This should make it harder for punishment to have a cooperation-enhancing effect in our experiment (S8).

1.2 Experimental procedures

We conducted our experiments in the CeDEx laboratory at the University of Nottingham (UK) (S26). We recruited our participants using the online recruiting software ORSEE (S27) which selects participants randomly from a large data base of volunteers interested in participating in experiments. We had 207 participants (69 groups of three members) in total. The participants were distributed across the treatments as follows – N10: \( n=60 \); N50: \( n=51 \); P10: \( n=45 \); P50: \( n=51 \). We had 12 sessions (three per treatment), with 4 to 7 groups in each session.

The participants were undergraduates from all fields of study at the University of Nottingham. The average age was 20.5 years and 52 percent were female. A post-experimental questionnaire revealed that, on average, a participant knew only 0.39 other
participants in a given session. Thus, the overwhelming majority of participants were strangers to one another.

The invitation to the experiment informed the participants that the experiments would last around 90 minutes. The reason why all sessions lasted 90 minutes was to ensure that there cannot be a selection effect of people into the experiment based on its duration. In the short experiments (N10 and P10), after they were finished with either N10 or P10, participants played an individual decision making task, which is entirely unrelated to the present experiments. In the long experiments (N50 and P50) participants only took part in the respective treatment.

Upon arrival, participants were randomly allocated to computer terminals which were separated by screens to ensure anonymity. After being seated participants were welcomed and received a set of written instructions (a copy is reproduced in the next section). The instructions were read out loud to ensure common information of procedures and payoffs. The instructions included a set of control questions that tested the participants’ understanding of payoff calculations. The experiment only started after all participants had solved all questions correctly.

All the interactions in the experiment were computer-mediated and took place anonymously. For conducting the experiments we used the experimental software z-Tree (S28). The program is available upon request.
1.3. Experimental Instructions

Here we document the experimental instructions we used in the long (50 periods) experiments. The short (ten periods) experiments were identical except for the number of periods.

Instructions N50

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

It is prohibited to communicate with the other participants during the experiment. If you have a question at any time raise your hand and the monitor will come to your desk to answer it.

During the experiment you can earn “points”. At the end of the experiment these points will be converted to cash at the following rate:

\[
1 \text{ point} = £0.01
\]

At the end of the experiment your entire earnings from the experiment will be paid to you in cash.

The experiment is divided into different periods. In all, the experiment consists of fifty periods. In each period the participants are divided into groups of three. You will therefore be in a group with two other participants. The composition of the groups will stay the same for all fifty periods. You are therefore with the same people in a group for all fifty periods.

In each period you will be endowed with tokens and have to decide how many tokens you would like to contribute to a project. Your earnings in the period will depend on the decisions made by you and the other two group members. The following pages describe the course of the experiment in detail:

Detailed Information on the Experiment

At the beginning of each period each participant receives an endowment of 20 tokens. Your task is 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 will appear:
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 tokens to contribute to the project, you have also decided how many tokens you keep for yourself: This is \((20 - \text{your contribution})\) tokens. After entering your contribution you must press the O.K. button. Once you have done this, your decision has been made and cannot be changed.

After all members of your group have made their decisions the following income screen will show you the total amount of tokens contributed by all three group members to the project (including your contribution). Also this screen shows you how many points you have earned in the period.
The Income Screen:

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

1. The tokens which you have kept for yourself (“Income from retained tokens”) whereby 1 token = 1 point.
2. The “income from the project”. This income is calculated as follows:

   Your income from the project = 0.5 times the total contributions to the project.

Your income in points from the period is therefore:

\[
(20 - \text{your contribution to the project}) + 0.5 \times (\text{total contributions to the project})
\]

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: 0.5 \times 60 = 30 points. If the total contribution to the project is 9 tokens, then you and all other group members receive an income of 0.5 \times 9 = 4.5 points from the project.

For each token you keep for yourself you earn an income of 1 point. Supposing 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 $0.5 \times 1 = 0.5$ points. However the income of the other group members would also rise by 0.5 points each, so that the total income of the group from the project would rise by 1.5 points. Your contribution to the project therefore also raises the income of the other group members. Similarly, you earn an income for each token contributed by the other members to the project. For each token contributed by any member you earn $0.5 \times 1 = 0.5$ points.

After you have viewed the income screen you will see an information screen showing much each group member contributed to the project:

Information Screen:

<table>
<thead>
<tr>
<th>Period</th>
<th>Contribution to the project</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>30 30</td>
</tr>
</tbody>
</table>

**Your contribution** is displayed in blue in the first column, while the contributions of the other group members of this period are shown in the remaining two columns. Note that the order in which others’ contributions are displayed will be determined at random in every period. The contribution in the second column, for example, could represent a different group member in different periods. The same holds true for the third column.

After you have viewed the income screen the period is over and the next period commences.
Now please complete the questions on the next sheet. If you have any questions please raise your hand. When everyone has completed the questions correctly we will begin the decision-making part of the experiment.

Questions

Please answer the following questions. They serve as a test of 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 two 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 14 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 6 tokens to the project?.......... 
   b) What is your income if the other group members together contribute a total of 22 tokens to the project?..........
Instructions P50

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

**It is prohibited to communicate with the other participants during the experiment.** If you have a question at any time raise your hand and the monitor will come to your desk to answer it.

During the experiment you can earn “points”. At the end of the experiment these points will be converted to cash at the following rate:

\[ 1 \text{ point} = £0.01 \]

Each participant receives a lump sum payment of **50 points** at the beginning of the experiment. This one-off 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 experiment your entire earnings from the experiment plus the lump sum payment will be paid to you **in cash.**

The experiment is divided into different periods. In all, the experiment consists of **fifty periods.** In each period the participants are divided into groups of three. You will therefore be in a group with two other participants. The composition of the groups will stay the same for all fifty periods. **You are therefore with the same people in a group for all fifty periods.**

Each period consists of **two stages.** In the first stage you will be endowed with tokens and have to decide how many 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.

**Detailed Information on the Experiment**

**The First Stage**

At the beginning of each period each participant receives an endowment of **20 tokens.** Your task is 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:
The First Stage Input Screen:

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 tokens to contribute to the project, you have also decided how many tokens you keep for yourself: This is \((20 - \text{your contribution})\) tokens. After entering your contribution you must press the O.K. button. Once you have done this, your decision has been made and cannot be changed.

After all members of your group have made their decisions the following income screen will show you the total amount of tokens contributed by all three group members to the project (including your contribution). Also this screen shows you how many points you have earned in the first stage.
The First Stage Income Screen:

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

(1) The tokens which you have kept for yourself ("Income from retained tokens") whereby 1 token = 1 point.
(2) The “income from the project”. This income is calculated as follows:

\[ \text{Your income from the project} = 0.5 \times \text{total contributions to the project} \]

Your income in points from the first stage of a period is therefore:

\[ (20 - \text{your contribution to the project}) + 0.5 \times (\text{total contributions to the project}) \]

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: 0.5 \times 60 = 30 points. If the total contribution to the project is 9 tokens, then you and all other group members receive an income of 0.5 \times 9 = 4.5 points from the project.

For each token you keep for yourself you earn an income of 1 point. Supposing 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 0.5 \times 1 = 0.5 points. However the income of the other group members would also rise by 0.5 points each, so that the total income of the group from the project would rise by 1.5 points. Your contribution to the project therefore also raises the income of the other group members.
Similarly, you earn an income for each token contributed by the other members to the project. For each token contributed by any member you earn $0.5 \times 1 = 0.5$ points.

After you have viewed the income screen the first stage is over and the second stage commences.

**The Second Stage**

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

**The Second Stage Input Screen:**

Your contribution is displayed in **blue in the first column**, while the contributions of the other group members of this period are shown in the remaining two columns. Note that the order in which others’ contributions are displayed will be determined at random in every period. The contribution in the second column, for example, could represent a different group member in different periods. The same holds true for the third column.

You will have to decide how many deduction points to assign to each of the other two group members. You must enter a number for each of them. If you do not wish to change the income of a specific group member then you must enter 0. You can **assign up to 5 points to each group member**.
You will incur costs from assigning deduction points. Every deduction point you assign costs you 1 point. For example, if you assign 2 deduction points to one member, this costs you 2 points; if, in addition, you assign 4 deduction points to the other member this costs you an additional 4 points. In total you will have assigned 6 points and your total costs therefore amount to 6 points.

After you have assigned points to each of the other two group members you must click the button “calculation” (see the second stage input screen). On the screen you will then see the total costs of your assigned points. As long as you have not yet clicked the OK-button, you can still change your decision. To recalculate the costs after a change of your assigned points, simply press the “calculation” button again.

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

\[(\text{Costs of received deduction points}) = 3 \times \text{Sum of received deduction points}.\]

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

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

Your total income from the two stages is therefore calculated as follows:

\[
\text{If income from the first stage is greater than or equal to the cost of received deduction points:}
\]

\[
\text{Total income (in points) at the end of the second stage} = \text{period income} =
\]
\[
= \text{Income from the first stage}
\]
\[
- 3 \times (\text{sum of received deduction points})
\]
\[
- \text{sum of deduction points you have assigned}
\]

\[
\text{Or}
\]

\[
\text{If income from the first stage is less than the cost of received deduction points:}
\]

\[
\text{Total income (in points) at the end of the second stage} = \text{period income} =
\]
\[
= 0 - \text{sum of deduction points you have assigned}
\]
Please note that your income in points 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**

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

Now please complete the questions on the next sheet. If you have any questions please raise your hand. When everyone has completed the questions correctly we will begin the decision-making part of the experiment.
Questions

Please answer the following questions. They serve as a test of 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 first stage income? ...........
   The first stage 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 first stage income? ...........
   The first stage income of the other group members? ...........

3. Each group member has an endowment of 20 tokens. Suppose the other two group members contribute a total of 30 tokens to the project.
   a) What is your first stage income if you contribute 0 tokens to the project? ...........
   b) What is your first stage income if you contribute 14 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 first stage income if the other group members together contribute a total of 6 tokens to the project? ...........
   b) What is your first stage income if the other group members together contribute a total of 22 tokens to the project? ...........

5. Suppose at the second stage you assign the following deduction points to your two other group members: 5, 0. What are the total costs of your assigned deduction points? ...........

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

7. By how many points 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 points 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 points will your income from the first stage be changed if you receive a total of 10 deduction points from the other group members? ...........
2. Supporting Analyses

Figs. S1A and S1B complement the analysis in the main text by documenting the temporal development of the average contributions levels and the average expenditures on punishment, respectively. As Fig. S1A shows, contributions in P50 were substantially higher than contributions in P10 (with the exception of period 1). Contributions were also higher in the first ten periods of P50 than in P10. OLS regressions of the change in contributions between period t and period t-1 on the number of received punishment points for a below-average contribution showed that participants in the P50-experiments increased their contributions in the first ten periods by more than in the ten periods of the P10-experiments (S29).

Fig. S1B illustrates the development of the average punishment expenditures over time. Per period expenditures on punishing in P50 were less than thirty percent of those in P10 (p=0.0062, two-sided Mann-Whitney tests with group averages as independent observations). Expenditures in P10 were stable whereas in P50 they decreased steadily over time. The exception in P50 was the last period where the harshest punishment of the whole experiment occurred. In P50 the group-average costs of punishment were negatively correlated with the group average cooperation level (Spearman’s ρ=-0.811, p=0.0001, n=17), whereas in P10 this correlation was insignificant (ρ=-0.342, p=0.2122, n=15).

The time horizon also strongly affected net earnings (Fig. 1 in the main text). We complement the analysis in the main text by comparing payoffs at the individual level. Eighty-four percent of the P50-participants earned more per period than the average N50 participant (who earned 23.8 money units per period). In P10 only 29 percent of participants earned more than the average in N10.
We complement the statistical and graphical analysis of Fig. S1 and Fig. 1 of the main text with regression analyses which model the time effects in more detail. Specifically, we include dummies for each block of five periods and use the first block of five periods (periods 1 to 5) as reference group. We also add a dummy for “Fifty-period experiment” if the experiment lasts for 50 periods. To control for endgame effects (S30) we add a dummy “Period 10 in N10 and P10” and “Period 50”. Since we control for time effects in periods > 5 by including period dummies, the dummy variable “Fifty-period
experiment” measures the difference in the respective variable to the ten-period experiment.

We have three dependent variables: “Contributions”, “Expenditures on Punishment” and “Net Earnings”. We distinguish between treatments and run the regressions separately for the N-experiment and the P-experiment. For the Contributions regressions we use Tobit, because contributions are censored at 0 and 20, and these were also very frequent choices. Similarly, we also use Tobit for the Expenditures on Punishment regressions because the expenditures on punishment are censored at 0 and 10, with 0 being very frequent. In the case of Net Earnings we use OLS because censoring is not a problem. Net Earnings can vary between -10 and +40, and we have very few observations at the extreme values. We record the results in Table S1.

The results show that contributions decreased during the N-experiment, whereas in the P-experiment contributions increased over time. Contributions were insignificantly higher in the first ten periods of N50 compared to N10. The result that the average contributions over all periods in N50 were even lower than in N10 (Fig. 1A) is surprising in view of theories of finitely repeated games which predict higher cooperation in N50 than N10 (S31). There was a significant drop in contributions in the last period of N10, but not in the last period of N50 (though note from Fig S1A that by period 49 contributions in N50 were already very low).

In P50 contributions were weakly significantly higher in the first ten periods compared to P10. Contributions in P50 were significantly higher than in P10. Finally, note that there was a significant drop in contributions in the last period of P50.
<table>
<thead>
<tr>
<th></th>
<th>Contributions</th>
<th>Expenditures on Punishment</th>
<th>Net Earnings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment: N-experiments</td>
<td>P-experiments</td>
<td>P-experiments</td>
</tr>
<tr>
<td>Fifty-period experiment</td>
<td>1.212</td>
<td>5.708</td>
<td>-2.006</td>
</tr>
<tr>
<td></td>
<td>(2.210)</td>
<td>(3.155)*</td>
<td>(1.012)**</td>
</tr>
<tr>
<td>Period 6-10</td>
<td>-3.769</td>
<td>2.745</td>
<td>-0.526</td>
</tr>
<tr>
<td></td>
<td>(1.170)***</td>
<td>(1.081)**</td>
<td>(0.441)</td>
</tr>
<tr>
<td>Period 11-15</td>
<td>-5.877</td>
<td>4.000</td>
<td>-1.962</td>
</tr>
<tr>
<td></td>
<td>(1.601)***</td>
<td>(1.684)**</td>
<td>(0.630)***</td>
</tr>
<tr>
<td>Period 16-20</td>
<td>-5.721</td>
<td>5.574</td>
<td>-1.710</td>
</tr>
<tr>
<td></td>
<td>(1.951)***</td>
<td>(2.585)**</td>
<td>(0.952)</td>
</tr>
<tr>
<td>Period 21-25</td>
<td>-5.713</td>
<td>6.670</td>
<td>-2.331</td>
</tr>
<tr>
<td></td>
<td>(2.106)***</td>
<td>(3.131)**</td>
<td>(0.872)***</td>
</tr>
<tr>
<td>Period 25-30</td>
<td>-5.172</td>
<td>9.216</td>
<td>-4.066</td>
</tr>
<tr>
<td></td>
<td>(2.031)**</td>
<td>(3.226)***</td>
<td>(1.264)***</td>
</tr>
<tr>
<td></td>
<td>(1.757)***</td>
<td>(3.475)***</td>
<td>(1.515)***</td>
</tr>
<tr>
<td>Period 36-40</td>
<td>-7.186</td>
<td>10.962</td>
<td>-4.006</td>
</tr>
<tr>
<td></td>
<td>(2.592)***</td>
<td>(3.629)***</td>
<td>(1.317)***</td>
</tr>
<tr>
<td>Period 41-45</td>
<td>-8.985</td>
<td>13.556</td>
<td>-5.138</td>
</tr>
<tr>
<td></td>
<td>(3.440)***</td>
<td>(3.774)***</td>
<td>(1.115)***</td>
</tr>
<tr>
<td></td>
<td>(3.440)***</td>
<td>(3.371)***</td>
<td>(1.242)***</td>
</tr>
<tr>
<td>Period 10 in P10 and N10</td>
<td>-8.142</td>
<td>-1.459</td>
<td>-0.164</td>
</tr>
<tr>
<td></td>
<td>(2.645)***</td>
<td>(2.019)</td>
<td>(0.953)</td>
</tr>
<tr>
<td>Period 50</td>
<td>-4.233</td>
<td>-8.235</td>
<td>4.779</td>
</tr>
<tr>
<td></td>
<td>(2.845)</td>
<td>(3.129)***</td>
<td>(1.247)***</td>
</tr>
<tr>
<td>Constant</td>
<td>9.656</td>
<td>12.551</td>
<td>-0.070</td>
</tr>
<tr>
<td></td>
<td>(1.461)***</td>
<td>(2.019)***</td>
<td>(0.777)</td>
</tr>
<tr>
<td>Observations</td>
<td>3150</td>
<td>3000</td>
<td>3000</td>
</tr>
<tr>
<td>Wald $\chi^2$(11)</td>
<td>34.53***</td>
<td>55.57***</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table S1: Regression analyses of Contributions, Expenditures on punishment, and Net earnings. All explanatory variables are dummies. The estimation method is Tobit for Contributions and Expenditures on punishment and OLS for Net earnings.

Expenditures on punishment were significantly lower in the first ten periods of P50 compared to P10 and in the remaining periods dropped significantly. There was a significant increase in punishment in period 50. This observation suggests that punishment is to a large degree non-strategic and emotions-driven (S32) because it
occurred despite no further interactions and after the experience of 49 rounds by which sufficient learning should have taken place.

The results from the Net earnings regressions mirror those of the Contributions regressions. In the N-experiments there was no significant difference between earnings in N10 and earnings in the first ten periods of N50, earnings are significantly lower in later periods, and there was a significant last period effect in N10 but not N50. In the P-experiments earnings are significantly higher in the first ten periods of P50 compared to P10, and earnings are significantly higher in later periods, with the exception of period 50 where a sharp drop in earnings is observed.

There are alternative ways to analyze data from experiments where sharp endgame effects occur. Our analysis is based on the view that behavior in the endgame is an integral part of how the fifty-period game is played, and so we report results based on all fifty periods of data. Some researchers view the endgame effect as an artificial aspect of the experiment, since interactions outside the laboratory rarely have fixed termination periods, and discard the data from later periods. The result that punishment is beneficial in long, but not short, duration interactions is strengthened slightly if we discard final period data: average net earnings in P50 are significantly higher than in N50 (by 3.02 money units per period, \( p = 0.0063 \), but average net earnings in P10 are significantly lower than in N10 (by 4.88 money units per period, \( p=0.0266 \)) (both tests: two-sided Mann-Whitney test with group average earnings as independent observations).

Notes and References

S29. P10-experiments: $\Delta c = 0.94(0.47) \times $ Punishment points received in t-1; P50-experiments: $\Delta c = 2.17(0.58) \times $ Punishment points received in t-1; $\Delta c$ denotes the change in contributions between period t and t-1 in case the participant in t-1 contributed less than the other two group members on average. Robust standard errors clustered on the independent groups in parentheses; * $p < 10\%$; ** $p < 5\%$; *** $p < 1\%$.
S33. We are grateful to Michalis Drouvelis for his help in conducting the experiments. We thank various workshop audiences and S. Bowles and E. Fehr for helpful discussions.