Supporting Online Material for

Lab Experiments for the Study of Social-Ecological Systems

Marco A. Janssen,* Robert Holahan, Allen Lee, Elinor Ostrom

*To whom correspondence should be addressed. E-mail: Marco.Janssen@asu.edu

Published 30 April 2010, *Science* **328**, 613 (2010)
DOI: 10.1126/science.1183532

This PDF file includes:

Materials and Methods
SOM Text
Figs. S1 to S10
Tables S1 to S7

Supporting Online Material

for

Lab Experiments for the Study of Social-Ecological Systems

Marco A. Janssen[1,*], Robert Holahan[2], Allen Lee[1], and Elinor Ostrom[1,2]

[1]School of Human Evolution and Social Change

Center for the Study of Institutional Diversity

Arizona State University

[2]Workshop in Political Theory and Policy Analysis

Indiana University

To whom correspondence should be addressed: Marco.Janssen@asu.edu

1. Material and Methods

1.1. Experiment Design

Our experiments were focused at understanding both the interactive and individual effects of communication and costly punishment in enhancing the likelihood of successful collective action of a social-ecological system. Unlike previous experiments that utilize static, one-shot, or repeated interactions to investigate these issues, we investigate a real-time dynamic resource-harvesting setting. The software used for this experiment is open-source and available at http://commons.asu.edu. Participants appropriate renewable tokens from a shared renewable resource environment.

In our experimental environment, groups of five participants share a 29 x 29 grid of cells. In the initial state, 25% of the grid space is filled with tokens, thus 210 tokens. The avatars are initially placed in the middle row of the screen with equal distances between the avatars. In order to collect a token a participant must position their avatar on the location of that token and explicitly press the space bar. Each token harvested is worth $0.02 USD. Participants have complete information on the spatial position of tokens and can watch the harvesting actions of other group members in real time.

Empty cells have each *second* the potential to generate new tokens. The probability that a given empty cell will generate a token is density-dependent on the number of adjacent cells with tokens. The probability p_t is linearly related to the number of neighbors: $p_t = p*n_t/N$ where n_t is the number of neighboring cells containing a green token, and N is the number of neighboring cells (N = 8, because we use a Moore neighborhood with a range of 1, and p = 0.01). If an empty cell is completely surrounded by eight tokens, it will generate a token at a higher probability than an empty cell that abuts only three tokens. At least one adjacent cell must contain a token for a new token generation to occur. Therefore, if participants appropriate all of the tokens on the screen, they have exhausted the resource and no additional token generation will occur. By designing the environment in this manner, we capture a key characteristic of many spatially dependent renewable resources. The optimum level of appropriation depends on the initial starting conditions, and probabilistic renewal of the empty cells. If we ignore the spatial variability, the optimal strategy is derived by waiting for two minutes to let the resource grow to 50% density, which leads to the highest growth rate (Fig. S1). The harvesting rate keeps the resource at a 50% density and all tokens are harvested during the last second of the experiment. This leads to an estimation of 664.

However, the spatial variability of token distributions and replenishments affect the number of tokens appearing on the screen. To calculate a distribution of the maximum number of tokens collected we performed 50,000 simulations of the harvesting of tokens using a simple strategy to maximize earnings. Each second, tokens with four tokens or more on the neighboring cells are collected, using a randomized non-sequential updating of the tokens. Since the software used in the experiments does not allow more than eight movements per second per person, the maximum number of tokens collected per second is 40 for a five person experiment. To the end of the experiment, tokens are collected such that not more than 40 * (seconds remaining) tokens remain on the screen. The average number of tokens collected is 665.65 with a standard deviation of 25.76.

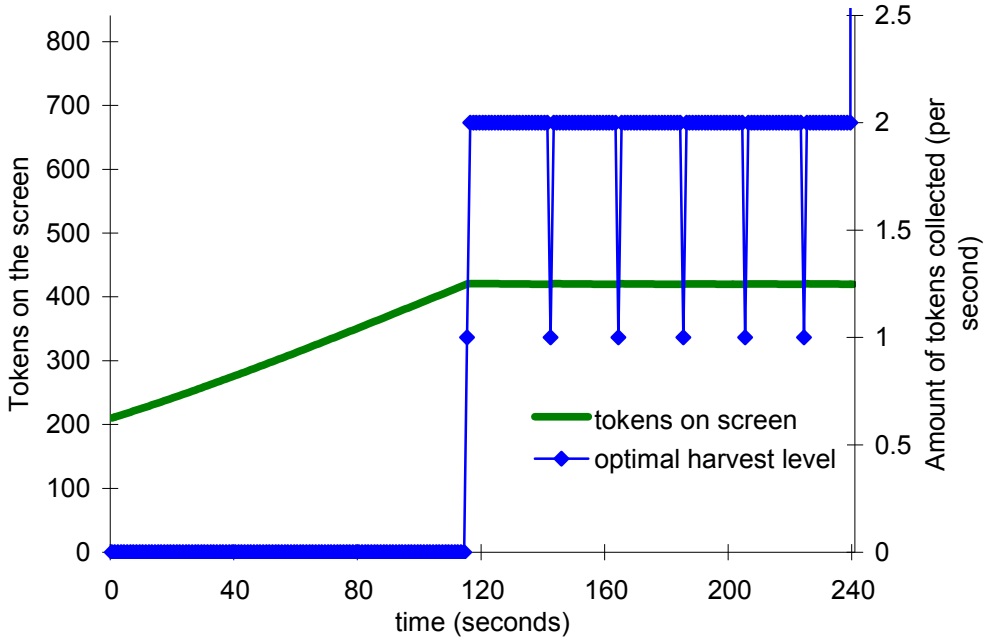


Fig. S1. Tokens collected and resource level in the optimal harvest strategy in an idealized spatial configuration of the tokens.

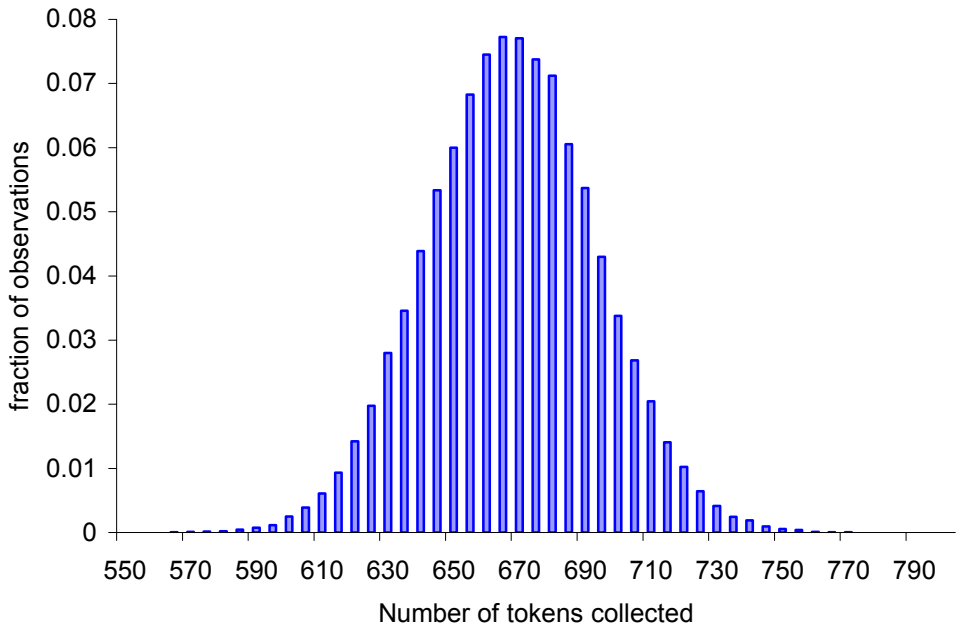


Fig. S2. Distribution of number of tokens collected using the maximum harvesting strategy.

We tested six treatments in an AB-BA, AC-CA, AD-DA format in which each treatment consists of three periods of neither communication nor costly punishment and three periods of costly punishment, text chat communication, or a combination of costly punishment and text chat communication (see Table S1). When participants have the opportunity to use costly punishment, they do so in real time, while appropriation occurs. This differs significantly from a variety of previous experiments that only examine ex post punishment. However, when participants have the opportunity to use text chat communication, they communicate for four minutes before a period begins. In this case, participants are identified by their randomly assigned identification numbers, which remain constant throughout the experiment. Our experiment allows participants in some treatments to ex ante discuss coordination strategies and in some treatments to use real-time costly punishment.

Table S1. Experimental design

Name	Number of groups (individuals)	Practice	Periods 1–3	Periods 4–6
NCP-CP	6 (30)	Individual resource	Neither communication nor punishment (NCP)	Communication plus costly punishment (CP)
CP-NCP	6 (30)	Individual resource	Communication plus costly punishment	Neither communication nor punishment (NCP)
NCP-P	5 (25)	Individual resource	Neither communication nor punishment (NCP)	Costly punishment (P)
P-NCP	6 (30)	Individual resource	Costly punishment	Neither communication nor punishment (NCP)
NCP-C	5 (25)	Individual resource	Neither communication nor punishment (NCP)	Communication (C)
C-NCP	5 (25)	Individual resource	Communication	Neither communication nor punishment (NCP)

Each experimental session consists of participants harvesting in six periods of four minutes each. Groups that quickly appropriate all of the tokens on the screen exhaust the resource and must wait for time to expire before continuing to the next period.

Participants can see the number of tokens earned during a period for each participant – indicated by number on the avatar – on top of the screen.

When participants can punish other participants, they can reduce the earnings of another participant by two tokens at a cost of one token. They do this by clicking key 1, 2, 3, 4, or 5, which is the number on the avatar of the other participant they wish to punish. Participants can punish only if they have net positive earnings in that period so they can pay the cost of the punishment, and they can punish only if the resource has not been fully depleted.

Communication among participants occurs during text-chat sessions. During four-minute sessions, participants can send public text messages to others in their group. Participants are identified by their avatar numbers, which remain the same throughout all periods, allowing individuals to associate the witnessed actions of other group members during a harvesting period with the discussions during the communication sessions.

1.2. Procedures

Participant recruitment took place using an existing voluntary database of undergraduate students maintained by the Interdisciplinary Experimental Laboratory (IELab) at Indiana University-Bloomington (http://www.indiana.edu/~ielab/). In total, 175 students participated in the experiments, but two groups of participants were dropped from the data analysis. In one group, an individual closed his browser during the experiment, rendering the data unusable, and in the other group a participant experienced a locked keyboard key during one period of the experiment, affecting the results of that period and potentially subsequent periods.

Two days before a session was conducted, a recruitment email was sent to members of the database informing them of an opportunity to participate in an experiment. Potential participants were informed that the experiment should last about 90 minutes, that they would be guaranteed a $10 show-up payment, and that they should expect to earn in the neighborhood of $15–$30 for their participation of about 60 minutes, but that the exact amount of earnings would depend on the results of the experiment. If a participant chose to sign up, (s)he clicked on a link in the email that took him/her to a secure website that listed all available experimental sessions. For each session we recruited 18 participants, three of whom became alternates.

Experiments were conducted throughout the 2008 calendar year at the IELab and included either 10 or 15 participants, depending on turnout. When participants arrived at the lab they were greeted by an experimenter and asked to read and sign an informed consent form while they waited for more participants to arrive. Our experiment required groups of five people and, in order to maintain some degree of anonymity, we only ran sessions if enough participants to fill either 2 or 3 groups showed up. Once a sufficient number of participants arrived, they were escorted into the computer room. Alternates were randomly selected through a lottery, paid $10, and asked to leave.

After everyone sat down at a computer terminal, the experimenter introduced himself and explained that they would be participating in a real-time group decision-making experiment with other participants in the room. At that point participants were asked to remain silent for the duration of the experiment, to turn off cell phones, and to put all other work away. The computer room at the IELab consists of individual computer terminals, separated by dividers to prevent participants from seeing one another's screens. When the experimental software was loaded, it randomly assigned each computer terminal to a group and gave each participant an identification number that remained constant throughout the experiment. In this way members of a group could gauge an individual's over-time behavior without actually knowing which participant in the room had been assigned which identification number. Once everyone was settled in and identification numbers had been assigned, the general instructions were read.

Our experiments use a real-time renewable resource environment. Participants see themselves on screen as a yellow avatar with the ability to collect green tokens by moving over the token and pressing the space bar on their keyboard (Fig. S3). The experimental environment requires some practice and we ask participants to answer two questions about the experimental environment after the general instructions are read to ensure they understand some key aspects of the regeneration of the resource. After that, participants participate in a four-minute practice period, in which they are individually placed in a 13 x 13 grid of cells to practice moving their avatars and collecting tokens. During the practice period, participants can reset the distribution of tokens by pressing the R key, so they can continue to practice even if they collect all the tokens on the screen before the four minutes expire.

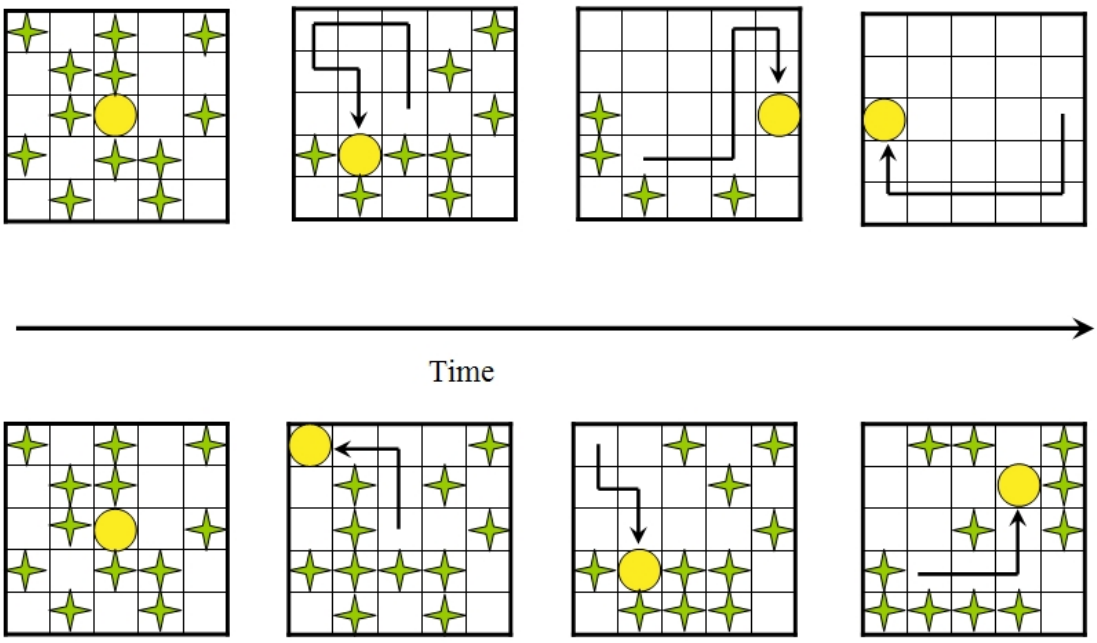


Fig. S3. Four snapshots of two harvesting strategies by two different types of subjects in a hypothetical situation of a 5 x 5 resource, where resource units are depicted by star-shaped objects. In the top row, the subject moves the avatar (circle) eight steps per time period. In the bottom row, the subject moves the avatar only four steps per time step.

At the completion of the practice period, participants were informed that the experimental environment would become five times larger and would be shared among the five members of the group. In addition, participants were told that they could no longer reset the distribution of tokens. Then the instructions for the first period were read. Since there are six treatments, the instructions for the first period varied slightly across treatments. After the first period was completed, participants were simply informed that the instructions for the next period were identical to the first period and the same was said before the start of period 3. Before the fourth period, however, new instructions were read to correspond with the experimental treatment. Participants were informed during the reading of the general instructions that the experiment would be six periods long, but they were not informed at any point that the experiment consisted of two stages of three periods each. Instead, they were only given instructions by period.

After the completion of the sixth period, participants were asked to fill out a brief survey that asked for basic demographic information and satisfaction with the experiment. While participants filled out the survey, one experimenter stayed in the computer lab, while another experimenter prepared participant payments in the reception area. Participants were then asked to leave one-by-one and sign for their cash payments. In this way, participants were not able to learn about the earnings of other participants.

1.3. Experiment Instructions

The software used to run the experiments is available at commons.asu.edu. We provide below the instructions for the no communication nor costly punishment treatment in the first three rounds[1] *and Communication Plus Costly Punishment in the last three rounds. The other treatments have different orders of instructions or deleted instructions when communication or costly punishment is not included in the treatment.*

Welcome. You have already earned 10 dollars by showing up at this experiment. You can earn more, up to a maximum of 40 dollars, by participating in this experiment, which will take about an hour to an hour and a half. The amount of money you earn depends on your decisions as well as the decisions of your group members during the six rounds of the experiment.

You appear on the screen as a yellow dot , and your other group members appear as blue dots . You move by pressing the four arrow keys to the right of your keyboard. You can move either up, down, left, or right. You have to press a key for every move of your yellow dot. In this experiment you can collect green diamond shaped tokens and you will earn two cents for each collected token. To collect a token, simply move your yellow dot over a green token and press the **space bar**. If you move over a token without pressing the **space bar** you will NOT collect that token.

The tokens that you collect have the potential to regenerate. After you have collected a green token, a new token can once again appear on that empty cell. However, the rate at which new tokens will appear depends on the number of adjacent cells that still have tokens. The more tokens in the 8 cells around an empty cell, the faster a new token will appear on that empty cell. Tokens generate new tokens. Thus the middle cell in Image 1 denoted with X will be regenerated at a faster rate than the middle cell in Image 2. When all neighboring cells are empty, there is no renewal.

If you have any questions at this time, please raise your hand and someone will come over to your station and answer it.

Image 1

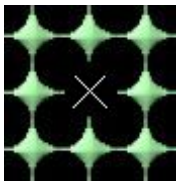
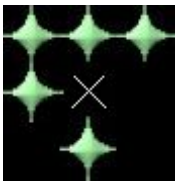


Image 2



---

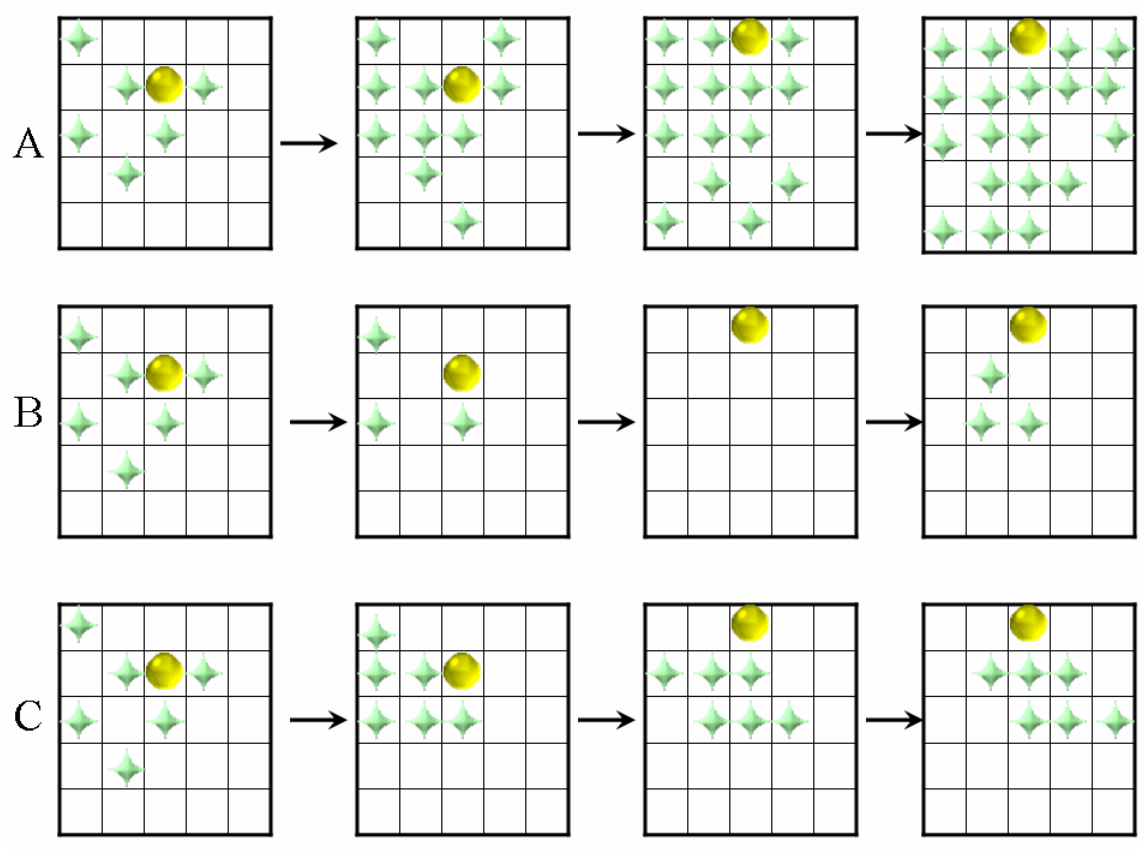
[1] We use the term rounds in the instructions of the experiment. Since the term round has a common use in experimental economics different from our use, we use the term period in the discussion of the results. Typically in a round participants make one decision before moving to the next round. In our experiment the round is a period of four minutes in which participants can make as many decisions as they want.

Before we go to the practice round, please answer the following questions to make sure you understand the instructions.

Which of the statements is incorrect:

- ***A*** ***Your decisions where to collect tokens affect the regeneration of tokens.***
- ***B*** ***When you have collected all tokens on the screen, no new tokens will appear.***
- ***C*** ***Tokens grow from the middle of the screen***
- ***D*** ***When you collect a token you need to press the [SPACEBAR] when your avatar is on a cell with a token***

Which sequence of situations is not possible?



- ☐ **A**
- ☐ **B**
- ☐ **C**

Practice Round Instructions

You will now have four minutes to practice with the experimental environment. The decisions you make in this round will NOT influence your earnings. At the beginning of the practice round half of the cells are occupied with green tokens. The environment is a 13 x 13 grid of cells.

During the practice round and only during the practice round, you will have the option to reset the tokens displayed on the screen. When you push the **R** key you will reset the distribution of the tokens to randomly occupying half of the cells with green tokens.

Please do not communicate with any other participant.

If you have any questions please raise your hand. **Do you have any questions so far?**

Round 1 Instructions

This is the first round of the experiment. The length of the round is 4 minutes. Like in the practice round you can collect green tokens. This time you earn **two cents** for each token collected. This time you **cannot** reset the distribution of green tokens.

In this round the renewable resource will become five times bigger. You will share this larger environment with four other random players in this room. Each of you has been randomly assigned to one of several equal-sized groups. Each member of the groups is collecting tokens from an identical, but separate resource.

Each of you has been randomly assigned a number from 1 to 5. These numbers will remain the same throughout the experiment, however, you will **not** be able to identify which person in the room has been assigned which number. Your anonymity is guaranteed. The other four players in your group will appear on the screen as blue dots  with a white number embedded in the dot. On the top right corner of the screen you can see how many tokens each player has collected. On the top left corner of the screen you can see a clock that displays the remaining time in the round.

If you have any questions please raise your hand. **Do you have any questions so far?**

Round 2 Instructions

Round 2 is the same as round 1.

If you have any questions please raise your hand. **Do you have any questions so far?**

Round 3 Instructions

Round 3 is the same as round 2.

If you have any questions please raise your hand. **Do you have any questions so far?**

Round 4 Instructions *[For Treatments with Communication and Punishing—Punishing-Only and Communication-Only Treatments Have the Same Instructions, but with Only the Relevant Information Listed]*

Round 4 is the same as the previous three rounds with two exceptions.

Before the next round starts you can anonymously communicate by text messages for four minutes with the other participants in your group. You can use this opportunity to discuss the experiment and coordinate your actions.

You may not promise anyone side-payments after the experiment is completed or make any threats. You are also not allowed to reveal your real identity. We are monitoring the chat traffic while you chat. Any violation of these two rules will disqualify you from further participation.

During the next round you have the option to reduce the earnings of another participant at a cost to your own earnings.

- If you press the numeric key 1-5 corresponding to another participant, you will reduce the number of tokens they have collected in this round by two tokens. This will also reduce your own token amount by one token. The decision whether or when to use this option is up to you.
- When you reduce the number of tokens of another participant, they will receive a message stating that you have reduced their tokens. Likewise, if another participant reduces your number of tokens, you will also receive a message. These messages will be displayed on the bottom of your screen.
- You may reduce the earnings of other participants as long as there are tokens remaining on the screen and while both you and the other participant have a positive number of tokens collected during the round. **Each time** you press the numeric key corresponding to another participant your token amount is reduced by **one**, and their token amount is reduced by **two**.

The length of this round is four minutes.

If you have any questions please raise your hand. **Do you have any questions so far?**

Round 5 Instructions

Round 5 is the same as round 4.

The length of this round is again four minutes.

If you have any questions please raise your hand. **Do you have any questions so far?**

Round 6 Instructions

Round 6 is the same as round 5.

The length of this round is again four minutes.

If you have any questions please raise your hand. **Do you have any questions so far?**

1.4. Survey

EXPERIMENT SURVEY

Computer number:____

Thank you for participating in this experiment.

While the experimenter is arranging for your payment, please take a few minutes to respond to the questions below. We would really appreciate learning your views on several questions and about the experiment you just finished.

1. Are you satisfied with your earnings in the experiment?

Not satisfied 1 2 3 4 5 very satisfied

2. Did the decisions from other players affect your own decisions?

Yes / No

3. When you made your decision, did you take into account if they affected the others in the group?

Yes/No

4. Did you understand the instructions of the experiments?

I did not understand anything 1 2 3 4 5 I did understand everything

If you **did reduce tokens** from other participants, answer question 5a:
5a Why did you reduce tokens from other participants? (Select *only* the most suitable statement)
○ Those other participants collected too many tokens.
○ Those other participants collected more than me.
○ Those other participants collected tokens too fast.
○ Those other participants did not follow our agreements.
○ I don't know.

If you **did *not* reduce tokens** from other participants, answer question 5b:
5b. Why did you not reduce tokens from other participants? (Select *only* the most suitable statement)
○ I did not want to give up some of my tokens.
○ I did not want to reduce earnings of others.
○ I was afraid of retaliation.
○ I did not see any reason for doing that.
○ I don't know.

6. Rank the five participants, including yourselves, in order of cooperativeness.

Least cooperative __ __ __ __ __ Most cooperative

SEE ALSO THE BACKSIDE

5c. Your group was able to exchange messages in between rounds. Please rank the others in your group as to who gave the most useful information in these discussions and who gave the least useful information.

Player_____ gave the most useful information

Player_____ gave the least useful information

5d. Did anyone in your group stand out in regard to their consistently taking actions which were "good exemplars" of the type of actions you discussed in your chat rounds?

Yes, Player_____ No_____

5e. Did anyone in your group stand out in regard to their consistently taking actions which were "bad exemplars" of the type of actions you discussed in your chat rounds?

Yes, Player_____ No_____

2. Supporting Analyses

2.1. Additional Statistics

The experiments were held in spring and fall semesters of 2008 at the IELab at Indiana University. The average age of the participants was 21 years, and 43% were male. The average response on the understanding of the instructions was 4.55 out of 5 (5 = complete understanding).

Table S2 presents the average gross number of collected tokens per group per period for blocks of three periods in which one of four conditions is applied: neither communication nor costly punishment, costly punishment, communication, or communication plus costly punishment. The differences of the gross and net earnings between periods 1–3 and periods 4–6 are tested by using the pair-wise two-tailed Mann-Whitney tests. The results are presented in Tables S3 and S4.

When the first three periods are neither communication nor costly punishment, allowing communication leads to a significant (p-value < 0.01) increase in both gross and net earnings. If costly punishment is allowed after three periods, there is no significant effect on the gross earnings, but there is a significant loss of net earnings due to penalties (p-value < 0.05). If both costly punishment and communication are allowed after periods 1–3, the gross and net earnings are significantly increased (p-value < 0.01).

When communication is allowed during the first three periods, the absence of further communication in the last three periods does not lead to a significant change in the high levels of earnings. However, if the experiment starts with three periods of costly punishment, not allowing punishment in the last three periods leads to a significant loss of earnings in periods 4–6 in gross (p-value < 0.01) and net (p-value < 0.1) earnings. Finally, when communication and costly punishment are both allowed in the first three periods, turning to neither communication nor costly punishment leads to a significant (p-value < 0.01) reduction in gross and net earnings. The reason for the drops in the last two conditions can be explained by the modest—not significant—decline of earnings over the periods. Participants “learn” to collect all the tokens faster. The parametric test between the blocks of three periods does not include this learning effect. We also suspect that the condition of punishment options affects the behavior of participants. Without the option of using costly punishment, stopping communication after three periods does not lead to a significant drop. When costly punishment is possible credible threats could be used. Agreements in communication without credible threats may have more normative weight. When the enforcement mechanism is not possible after period 3 we see a reduction of the performance.

When we compare between treatments, we find that in the first three periods neither communication nor costly punishment is not different than only costly punishment (Tables S3B and S4B). Furthermore, in the first three periods, communication only does not lead to different earnings than communication plus costly punishment. Hence, costly punishment does not provide a significant effect on both gross and net earnings in the first three periods. In the last three periods, neither communication nor costly punishment is significant different than costly punishment only. This is caused by the higher earnings in the NCP condition if groups have communicated in the first three periods.

Finally, we compare the conditions NCP, C, P and CP if used in the first three periods, compared to the use in the last three periods, we see no significance difference for the C

condition, modest difference (p=0.075) for the NCP condition, and significance difference for the P and CP conditions ($p < 0.05$) (Tables S3C and S4C). If punishment only is used, the earnings are lower in the first three periods compared to the use of punishment in the last three periods. This is reversed for experiments when costly punishment is combined with communication.

Table S2. Experiment design and average gross number of collected tokens per group per period. We distinguish six treatments with different combinations of No communication not punishment (NCP), Communication (C), and Costly Punishment (P). The last two columns show the average amount of tokens collected per group per period, excluding the tokens lost by costly punishment.

Treatment		Experiment design		Tokens collected by group	
Name	Number of groups	Periods 1–3	Periods 4–6	Periods 1–3	Periods 4–6
NCP-CP	6	Neither communication nor punishment	Costly punishment plus text chat communication	292.89 (30.63)	451.39 (58.84)
CP-NCP	6	Costly punishment plus text chat communication	Neither communication nor punishment	402.22 (71.81)	331.22 (82.87)
NCP-P	5	Neither communication nor punishment	Costly punishment	265.00 (17.54)	247.67 (16.38)
P-NCP	6	Costly punishment	Neither communication nor punishment	280.39 (46.15)	256.72 (18.23)
NCP-C	5	Neither communication nor punishment	Text chat communication	274.07 (34.22)	411.60 (79.86)
C-NCP	5	Text chat communication	Neither communication nor punishment	441.33 (51.98)	415.47 (80.78)

Table S3A. Pair-wise two-tailed Mann-Whitney tests on equality of distributions of earnings in the first three rounds and the last three rounds. A cell (i,j) reports the relevant statistics. First, the direction of the relation is given where a positive (negative) sign mean that the average ***gross*** earnings in the last 3 periods are higher (lower) than the first three periods. The U statistic is given for the null hypothesis of equality in the distribution of ***gross*** group earnings between the first three periods and the last three periods, as well as the sample sizes. The corresponding p-value is reported in parentheses as last. Finally, we use *** = p-value < 0.01.

Treatment	Direction	U statistic	Sample sizes	P-value
NCP-C	(+)***	216	(15,15)	0.000
NCP-P	(-)	127.5	(15,15)	0.541
NCP-CP	(+)***	320	(18,18)	0.000
C-NCP	(-)	141.5	(15,15)	0.232
P-NCP	(-)***	262	(18,18)	0.001
CP-NCP	(-)***	258	(18,18)	0.002

Table S3B. Pair-wise two-tailed Mann-Whitney tests on equality of distributions of earnings between different treatments in the first three periods as well as the last three periods. A cell (i,j) reports the relevant statistics. The U statistic is given for the null hypothesis of equality in the distribution of ***gross*** group earnings between the two treatments, as well as the sample sizes and p-value.

Treatment comparison	First 3 periods	Last 3 periods
NCP vs C	U=717 (48,15), p-value = 0.000	U=591 (51,15), p-value = 0.001
NCP vs P	U=483 (48,18), p-value = 0.471	U=548 (51,15), p-value = 0.011
NCP vs CP	U=842 (48,18), p-value = 0.000	U=795 (51,18), p-value = 0.000
C vs P	U=267 (18,15), p-value = 0.000	U=224 (15,15), p-value = 0.000
C vs CP	U=176 (18,15), p-value = 0.145	U=183 (18,15), p-value = 0.086
P vs CP	U=301 (18,18), p-value = 0.000	U=170 (18,15), p-value = 0.000

Table S3C. Pair-wise two-tailed Mann-Whitney tests on equality of distributions of earnings of same type of periods (NCP, C, P or CP) in the first three rounds versus in the last three rounds. A cell (i,j) reports the relevant statistics. The U statistic is given for the null hypothesis of equality in the distribution of ***gross*** group earnings between the two treatments, as well as the sample sizes and p-value.

Treatment	Statistics
NCP	U= 1476 (51,48), p-value = 0.075
C	U= 138 (15,15), p-value = 0.305
P	U= 196 (18, 15), p-value = 0.028
CP	U= 225 (18,18), p-value = 0.047

Table S4A. Pair-wise two-tailed Mann-Whitney tests on equality of distributions in the first three periods compared to the last three periods. A cell (i,j) reports the relevant statistics. First, the direction of the relation is given where a positive (negative) sign means that the average ***net*** earnings in the last three periods are higher (lower) than the first three periods. In the net earnings, the number of tokens lost due to costly punishment is subtracted. The U statistic is given for the null hypothesis of equality in the distribution of ***net*** group earnings between the first three periods and the last three periods, as well as the sample sizes. The corresponding p-value is reported in parentheses as last. Finally, we use * = p-value < 0.1; ** = p-value < 0.05; *** = p-value < 0.01.

Treatment	Direction	U statistic	Sample sizes	P-value
NCP-C	(+)***	216	(15,15)	0.000
NCP-P	(-)**	163	(15,15)	0.037
NCP-CP	(+)***	317.5	(18,18)	0.000
C-NCP	(-)	141.5	(15,15)	0.232
P-NCP	(-)*	221.5	(18,18)	0.059
CP-NCP	(-)***	255	(18,18)	0.003

Table S4B. Pair-wise two-tailed Mann-Whitney tests on equality of distributions of earnings between different treatments in the first three periods as well as the last three periods. A cell (i,j) reports the relevant statistics. The U statistic is given for the null hypothesis of equality in the distribution of *net* group earnings between the two treatments, as well as the sample sizes and p-value.

Treatment comparison	First 3 periods	Last 3 periods
NCP vs C	U=717 (48,15), p-value = 0.000	U=591 (51,15), p-value = 0.001
NCP vs P	U=471 (48,18), p-value = 0.583	U=618 (51,15), p-value = 0.0002
NCP vs CP	U=834 (48,18), p-value = 0.000	U=785 (51,18), p-value = 0.000
C vs P	U=267 (18,15), p-value = 0.000	U=225 (15,15), p-value = 0.000
C vs CP	U=182 (18,15), p-value = 0.093	U=179 (18,15), p-value = 0.117
P vs CP	U=300 (18,18), p-value = 0.000	U=270 (18,15), p-value = 0.000

Table S3C. Pair-wise two-tailed Mann-Whitney tests on equality of distributions of earnings of same type of periods (NCP, C, P or CP) in the first three rounds versus in the last three rounds. A cell (i,j) reports the relevant statistics. The U statistic is given for the null hypothesis of equality in the distribution of *net* group earnings between the two treatments, as well as the sample sizes and p-value.

Treatment	Statistics
NCP	U= 1476 (51,48), p-value = 0.075
C	U= 138 (15,15), p-value = 0.305
P	U= 191 (18,15), p-value = 0.043
CP	U= 229 (18,18), p-value = 0.034

In Table S5 we present additional statistical analysis on the effects of punishment and communication on the amount of tokens collected. Like Table 1 of the article we use a multi-method fixed-effects linear regression (`mtmixed` using Stata 10.0). This technique command can be used to fit mixed models, models that contain both fixed and random effects.

We tested whether punishment events in the current period or previous periods had an impact on the amount of tokens collected (Table S5). We found that actual punishment has no effect on the amount of tokens collected, except when there was a combination of communication and costly punishment in the first three periods. There was a significant reduction of the number of tokens collected in NCP periods if costly punishment was used in the CP periods, in contrast to the groups who could use it, but did not do so.

We tested a possible learning effect of experiencing the same condition during multiple periods by including a dummy variable that indicates whether it is the first, second or third time in this condition. This dummy variable is zero when the period is not in that condition at all. We defined LearnNCP as the dummy for periods in the neither communication nor punishment condition. This variable is 1 for the first time the experiment has a NCP period, 2 for the second time, and 3 for the third time. LearnNCP is zero when it is not in the NCP condition. In the same way we define LearnCP, LearnC and LearnP.

Furthermore, we see that communication has a positive effect on the amount of tokens collected during a period. Every period of communication leads to an improvement, and when communication is not allowed anymore, the level of tokens collected remains at a high level (except when costly punishment was used).

Every period neither communication nor costly punishment was possible, the amount of tokens collected in a period was relatively reduced. Groups learn to overharvest the resource faster if they could neither communicate nor punish.

Table S5. Punishment, communication, and harvesting levels. A multilevel linear regression is performed with the gross number of tokens that groups collected for each period. The independent variables are a set of dummy variables: whether participants could communicate and/or punish during the period, and whether participants could have communicated and/or punished during the first three periods. Finally, learning effects are measured by including the period numbers and how many times the participants could use communication and/or punishment.

Independent variables	Dependent variable: tokens harvested by group (std. error)
Constant	293.875*** (13.077)
Communication in current period (0 = no, 1 = yes)	92.389*** (22.730)
Punishment possible and used in current period (0 = no, 1 = yes)	1.989 (24.719)
Punishment possible but not used in current period (0 = no, 1 = yes)	12.903 (25.552)
Communication and Punishment in current period (and used) (0 = no, 1 = yes)	16.979 (27.940)
Communication and Punishment in current period (and not used) (0 = no, 1 = yes)	-9.696 (30.206)
Communication in first three periods (0 = no, 1 = yes)	123.641*** (16.382)
Punishment in first three periods (and used) (0 = no, 1 = yes)	-13.007 (12.925)
Punishment in first three periods (and not used) (0 = no, 1 = yes)	-11.007 (16.052)
Communication and Punishment in first three periods (and used) (0 = no, 1 = yes)	-59.783*** (25.942)
Communication and Punishment in first three periods (and not used) (0 = no, 1 = yes)	11.121 (30.562)
LearnNCP	-9.576** (4.709)
LearnC	17.936** (8.475)
LearnP	-5.645 (10.819)
LearnCP	4.429 (3.402)
- Log likelihood	1034.724
Number of periods	198
Variance contributions	
Group	38.394 (5.651)
Individual	38.255 (2.116)
χ^2	68.73 (p-value < 0.001)

* P < 0.1, ** P< 0.05, *** P < 0.01

In Figure S4 the harvesting rates are given for periods of five seconds. A value of 25 means that during a five second interval every participant collected one token per second. In such a situation each person has to press at least two times per second a key on the keyboard to move and to collect. When there is no communication we see high levels of harvesting in the first minute of the experiment, which falls down to zero quickly because of the depletion of the resource. If communication is possible, we see a lower harvesting rate in the beginning of the experiment leading to a longer high level of harvesting, and higher returns.

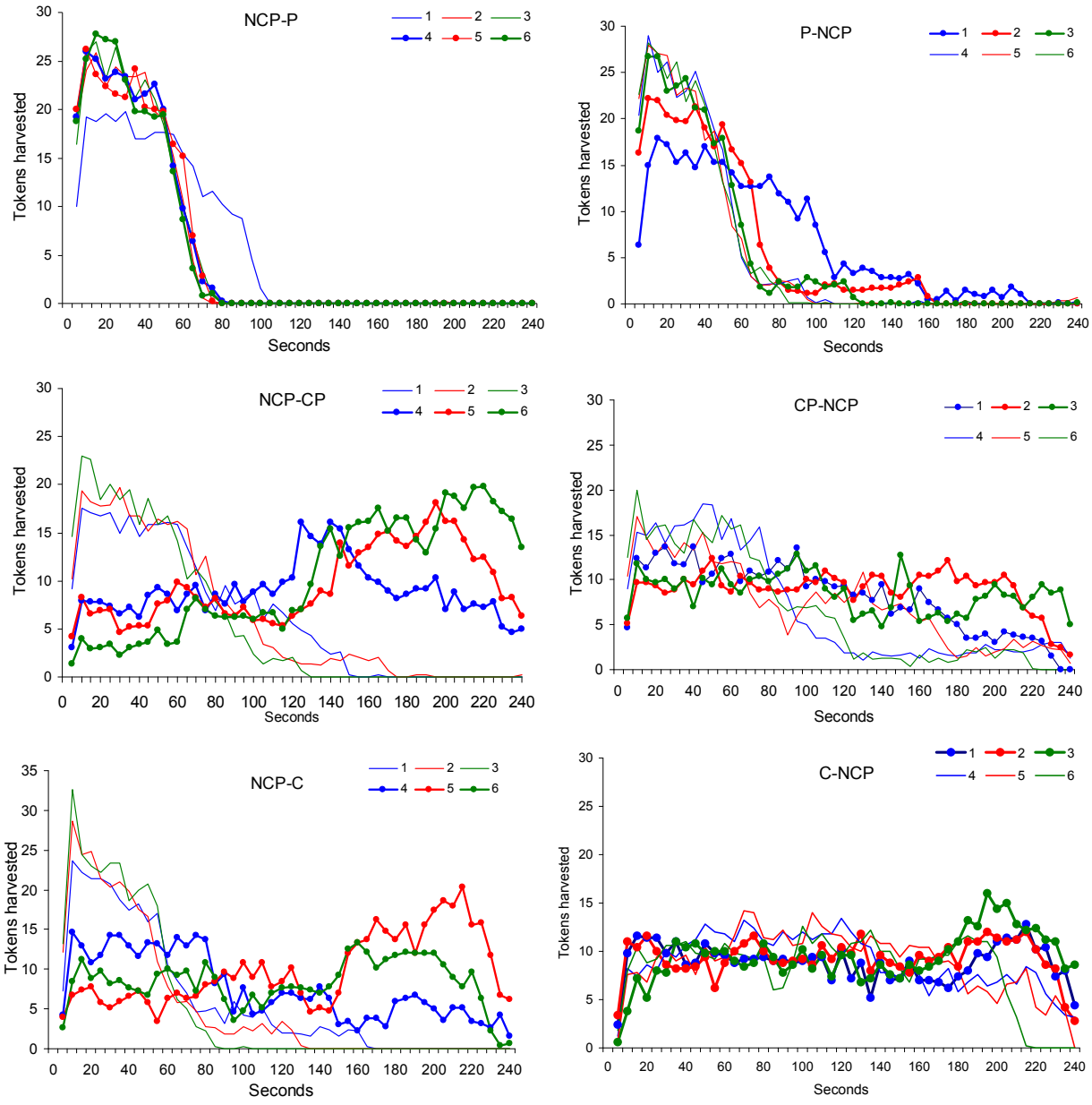


Fig. S4. Harvesting rates at Given Times. The diagrams show the average amount of tokens harvested at each period of five seconds for the five or six groups of each treatment. Each diagram shows a treatment condition and each line represents a particular period. The treatment is a combination of two sets of three periods of a specific condition. The names for these conditions are noted in the upper left of each display: (NCP for No Communication not Punishment, C for Communication, P for Costly Punishment, CP for Communication and Costly Punishment) is noted in the upper left of each display. The color and shape related to the data of each period is noted in the upper right.

Fig. S5 shows that punishment events are happening more frequently in the first period when costly punishment is possible, but this difference is not statistically significant, except a modest effect for treatment NCP-CP. We use pair-wise two-tailed Mann-Whitney tests and fine for treatment CP-NCP (U statistic for the null hypothesis of equality in the distribution: U=43 (sample sizes 12 and 6), p-value = 0.553), for treatment NCP-CP (U=56 (12,6), p-value = 0.083), for treatment P-NCP (U=46.5 (12,6), p-value = 0.335), and for treatment NCP-CP (U=28 (10, 5), p-value = 0.768).

We also do not find statistically significant differences in punishment distributions among treatments (using pair-wise two-tailed Mann-Whitney tests):

- CP-NCP vs NCP-CP: U=166 (18,18), p-value = 0.913
- CP-NCP vs P-NCP: U=194 (18,18), p-value = 0.324
- CP-NCP vs NCP-P: U=157.5 (18,15), p-value = 0.420
- NCP-CP vs P-NCP: U=200.5 (18,18), p-value = 0.227
- NCP-CP vs NCP-P: U=158 (18,15), p-value = 0.420
- P-NCP vs NCP-P: U=136 (18,15), p-value = 0.987

Fig. S6 shows that many periods do not experience any punishment events, while a few periods involve many events, probably including retaliation events. With communication there are more periods without punishment events. Using a pair-wise two-tailed Mann-Whitney test we find that U=710 (36, 33), p=0.165.

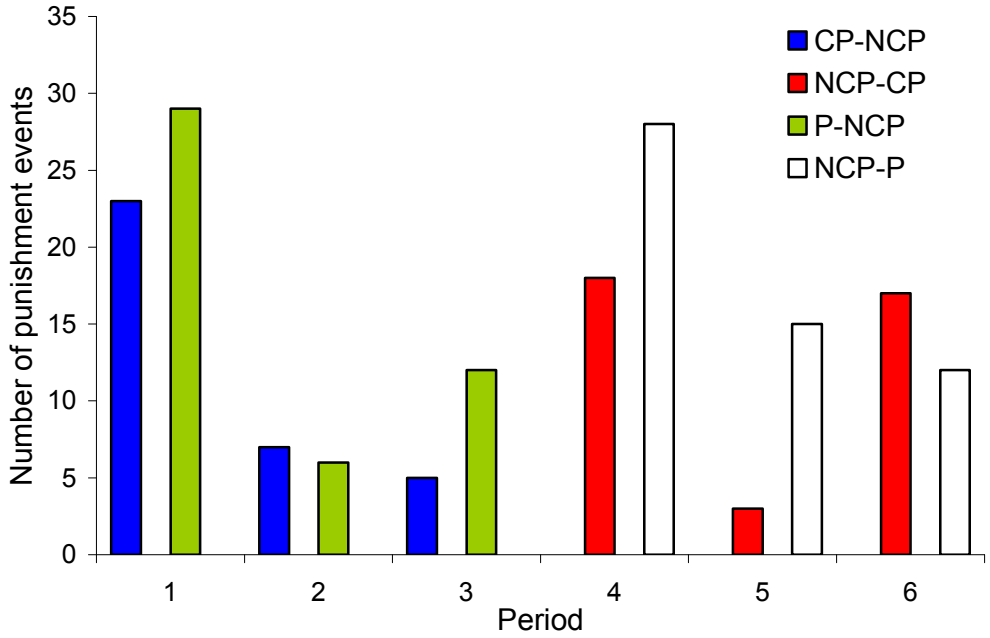


Fig. S5. Total number of punishment events for each period for each condition.

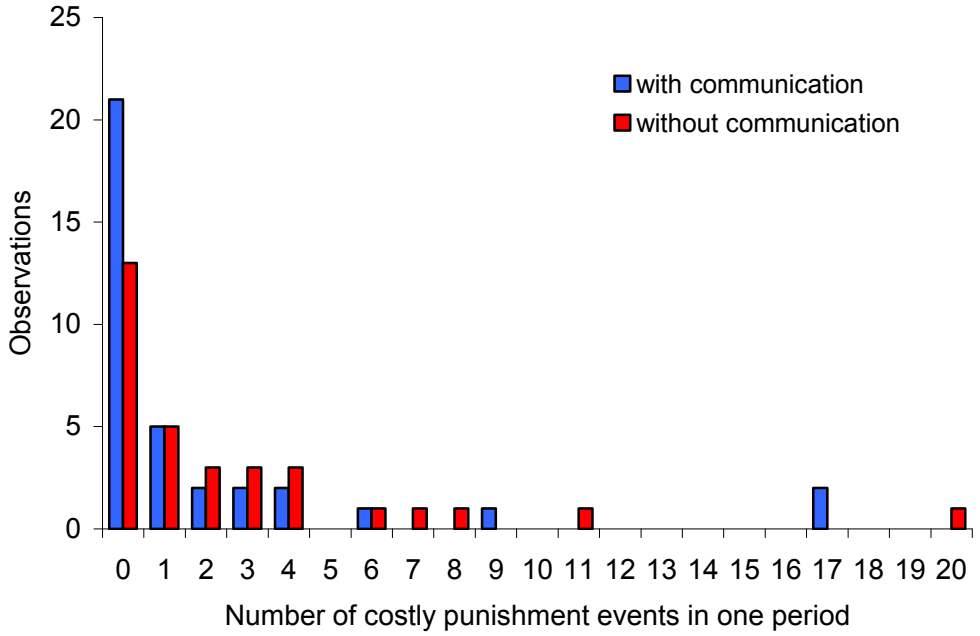
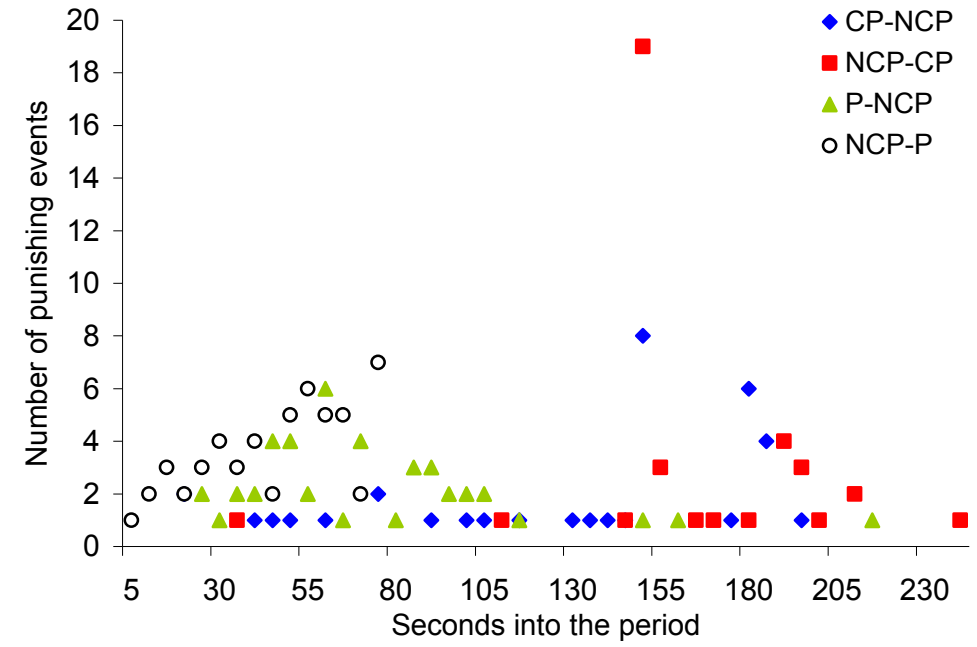
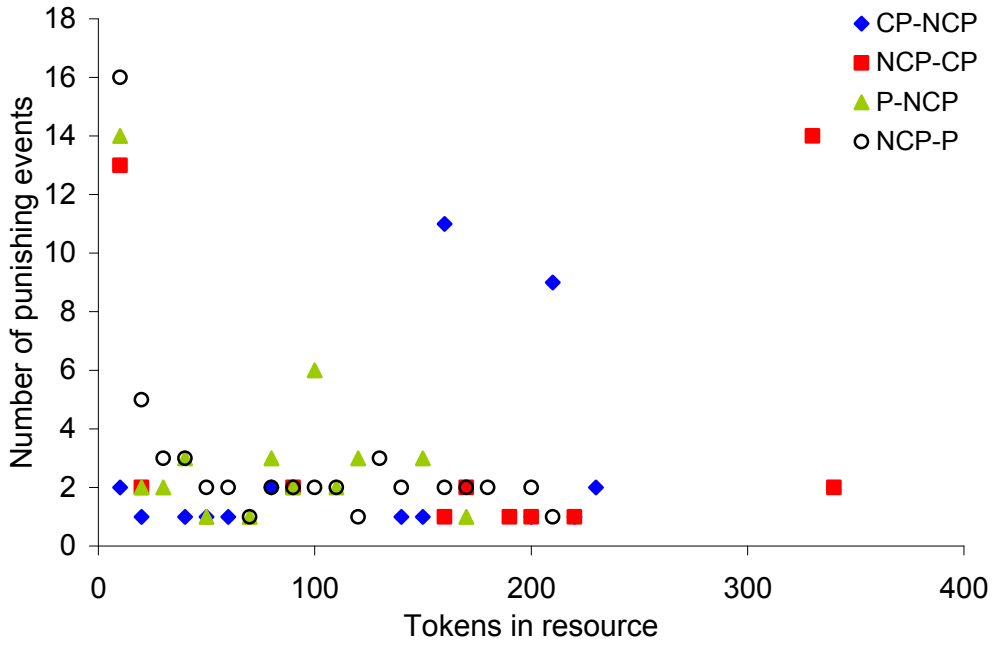


Fig. S6. Frequency of costly punishment events per period.

In Figures S7 and S8 we depict information when participants during a period punish each other. We see that they use punishment events during the whole experiment. However, when communication is not possible there is a tendency of punishment when not many tokens are on the screen anymore. We see sometimes spikes in the amount of punishment events, which are mainly retaliation wars (one participant start punishing and get punished back), or individuals who punish many times in a few seconds. For example, in one experiment of the NCP-CP case one individual had 16 punishment events in 4 seconds to the same person when there were around 300 tokens on the screen. This punishment event happened is triggered by one person starting to harvest earlier than agreed on.



Fig, S7. The time during the four minute periods when punishment events happen. We added the number of events per interval of five seconds. We distinguish different treatments but aggregated the three periods for each treatment.



Fig, S8. The resource level at the moment punishment events happen. We added the number of events per interval of 10 tokens. We distinguish different treatments but aggregated the three periods for each treatment.

To determine the effect of a punishment event on the actions of the participant being punished we calculated for all punishment events the number of tokens harvested in the 5 seconds before and after the punishment event for the participant who punished, the participant who is punished, and the three other participants (Table S6). We test whether there is a significant change in the harvesting rate of the participants before and after the punishment event, and how this relates to the other types of participants.

Using pair-wise two-tailed Mann-Whitney tests, we find that the punisher has significantly the lowest harvesting rate. For the 5 seconds before a punishment event in a CP period we have U=4271 (75, 75), p-value < 0.0001 and U=10017 (225, 75), p-value = 0.014. For the 5 seconds after a punishment event in a CP period we have U=3349 (75, 75), p-value = 0.041 and U=10441 (225, 75), p-value = 0.0019. For the 5 seconds before a punishment event in a P period we have U=6509 (102, 102), p-value = 0.0017 and U=19282 (306, 102), p-value = 0.0004. For the 5 seconds after a punishment event in a P period we have U=5908 (102, 102), p-value = 0.09, and U = 18832 (306, 102), p-value = 0.0016.

In CP periods the punished participants have a significant higher harvesting rate compared to those who are not involved in punishment events (U=10488 (225, 75), p-value = 0.0014). After the punishment event, there is no statistical significance between punished and those who are not involved in punishment events (U=8722 (225, 75), p-value = 0.662).

In the P periods the punished participants do not differ in the harvesting rate of those who are not involved in punishment events (U=15875, (306, 102), p-value = 0.792. After the punishment event, there is again no statistical significance between punished and those who are not involved in punishment events (U=16528 (306, 102), p-value = 0.369).

In the CP periods, the punisher has no significant change in the harvesting rate (U=2916 (75, 75), p-value = 0.695), while the punished participants has a significant drop in the harvesting rate (U=3559 (75, 75), p-value = 0.0043). The participants not involved in the punishment events do not change their harvesting rates (U=25777 (225, 75), p-value = 0.737).

In the P periods all type of participants had a significant drop in the harvesting rate: The punisher (U=6021 (102, 102), p-value = 0.050), the punished participants (U=6360 (102, 102), p-value = 0.0052), and those who are not involved in the punishment events (U=54164 (306, 306), p-value = 0.00072). Since the resource is declining rapidly in P periods, the reduction of available tokens is the likely cause of the decline of harvesting rate of all type of participants.

In sum, with communication punishment leads to a correction of the harvesting behavior of the punished participant, while there is no measurable difference in the behavior of participants who are punished in the treatment without communication.

Table S6. Number of tokens harvest on average by different types of participants in the 5 seconds before and after a punishment event.

		punisher	punished	Others
Without communication	5 seconds before	2.38	3.28	3.15
	5 seconds after	1.90	2.49	2.62
With communication	5 seconds before	1.81	3.91	2.95
	5 seconds after	1.76	2.85	3.02

We coded each punishment event. We checked the earnings and location of the participants at the moment of punishment. We could identify the following four situations:

- highest earning participant at that moment was punished.
- A participant who just was punished, punish the punisher (retaliation)
- A person who is not the highest earner was punished, but this person was the closest by the punisher.
- A person who is not the closest by and derives not the highest, but has higher earnings than the punisher, is punished.

We distinguish experiments with and without communication, and do not experience important differences between those two treatments. About two out of five times the highest earning participant is punished. At least 3 out of 4 times a person is punished who derived higher earnings up to that point. Using a pair-wise two-tailed Mann-Whitney test we find no statistical difference between both distributions ($U = 13.5$ (5, 5), $p=0.841$).

Table S7 presents the responses to survey questions on why participants use costly punishment or not. Without communication, participants argue that they punish others if they collect too many tokens or collect too fast. When communication is possible, participants' main reason for punishing others is because they did not follow the agreements. The main explanation for not punishing in both treatments is that participants did not see a reason for using punishment. Without communication, a second-ranked reason was being afraid of retaliation. This explanation is not common when communication is possible. However, not giving up earnings for themselves or others became important explanations when communication is possible.

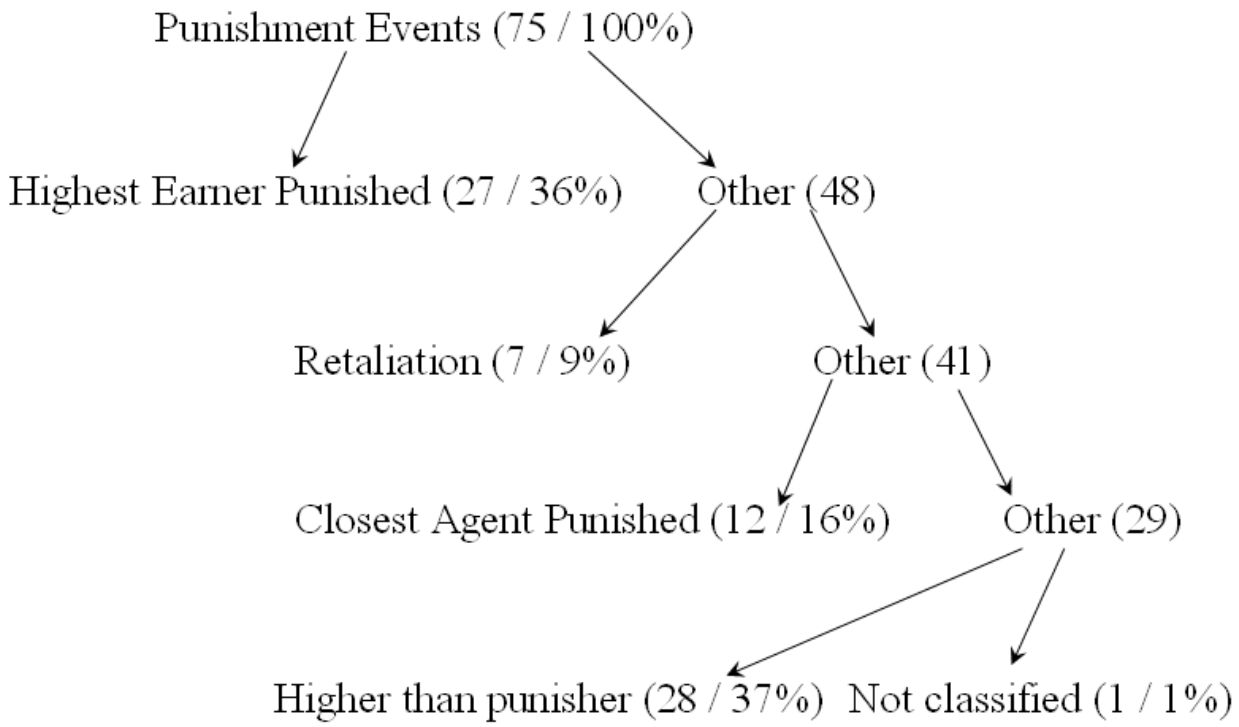


Figure **S9**: Scheme of coded punishment events in experiments where communication is possible.

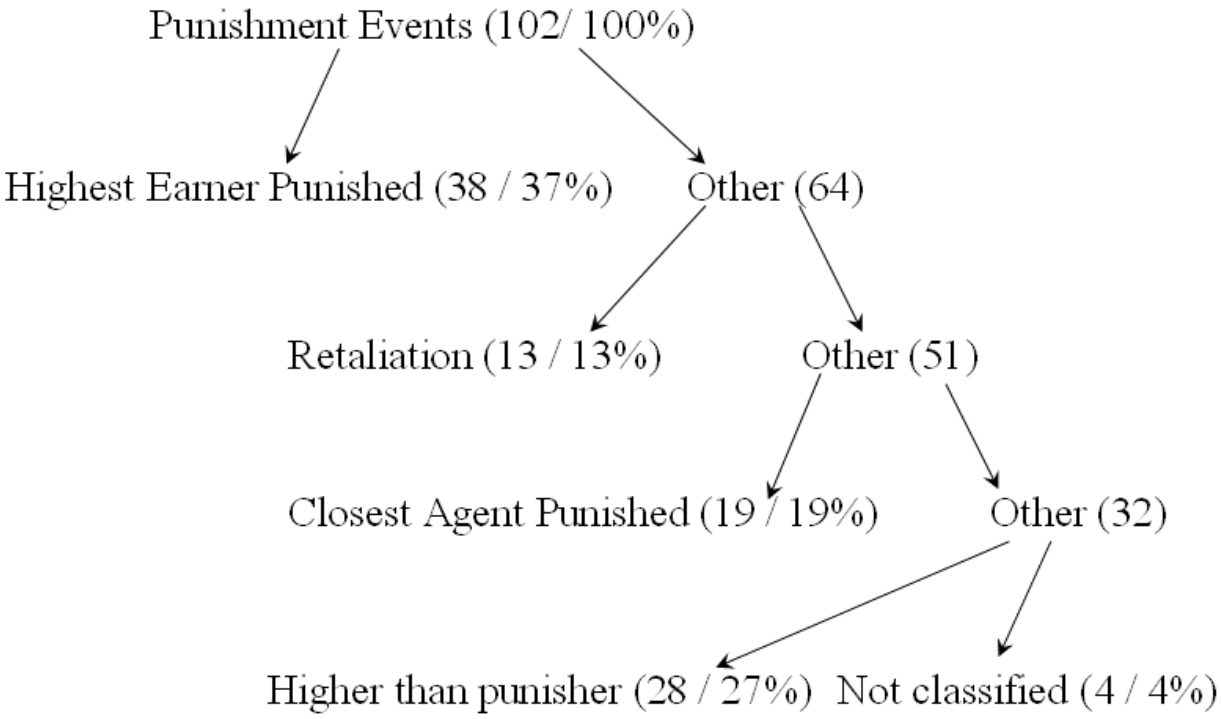


Figure **S10**: Scheme of coded punishment events in experiments where communication is not possible.

Table S7. Frequency of response to survey questions on why participants did or did not use the option to engage in costly punishment

Reason to punish	Punishment	Punishment and communication
Those other participants collected too many tokens	6	4
Those other participants collected more than me	5	0
Those other participants collected tokens too fast	6	2
Those other participants did not follow our agreements	1	11
I don't know	3	0

Reason not to punish	Punishment	Punishment and communication
I did not want to give up some of my tokens	4	9
I did not want to reduce earnings of others	0	7
I was afraid of retaliation	11	1
I did not see any reason for doing that	18	25
I don't know	1	1

2.2. Text Analysis

Some illustrative examples of chat between participants that show the use of informal punishment and the creation of informal institutional arrangements are show below.

The first set of examples shows the types of arrangements participants make during the communication periods.

Player 5: well

Player 2: If we dont go for all of them all at once, we can let them regenerate, correct?

Player 3: hey guys my opinion is that this is an experiment about greed and trust. we have to somehow trust eachother to ration out eating the dots so that we aren't stuckl for 3 minutes sitting with nothing to do

Player 1: do you guys want to all wait until the screen is filled

Player 4: well we need to set it up so that it is isnt just anarchy when we start, we need to let the things come regenerate

Player 3: I like number ones idea lets wait till the screen is filled and then start

Player 5: so maybe eat for 10 seconds rest for 10 seconds?

Player 4: well we just need to make sure we use the entire time limit too so don't try to just eat up as many as possible

Player 2: yeah

Player 4: well look for the spots where they will regenerate the fastest

Player 3: okay

Player 5: and then just have a clusterfest at that spot at the end?

Player 1: we could wait until the screen fills up entirely. Then we'll each eat 10, then wait again for the screen to fill up. Then just eat all of them...

Player 3: yeah that sounds good

Player 4: but we dont want them all off the screen

Player 4: we want to take up as much time as possible then wait till the end when we can eat them all

Player 3: no hes saying eat all of them when time is gonna run out

Player 4: yeah

Player 4: ok

Player 3: but until then everyone eats 10 at a time

Player 5: ok to summarize:

Player 3: that sounds like a good idea cuz we can monior how many have been eaten by each person

Player 5: we each eat 10

Player 5: then

Player 5: wait for the screen to fill up

Player 5: then go for it?

Player 1: we could continue to eat 10 at a time and waiting. then about 30 seconds left, everyone just eat as many as they can

Another example:

Player 1: if you want to generate more income

Player 4: whats up?

Player 1: you have to let the tokens multiply

Player 2: #4 that was real wreong of you

Player 4: what was wrong 2

Player 1: agreed?

Player 1: since we have so much time

Player 1: at the end

Player 5: oh

Player 2: i say we break up into quadrants like from geometry

Player 5: go after only ones in a bunch and then let it chill so they grow back

Player 5: that is a great idea

Player 2: you stole my token

Player 1: ok

Player 4: we need to make sure we end up for longer time4

Player 2: i was trying to see if it my multiply

Player 1: ya

Player 1: you need to let them multiply

Player 5: how do they multiply again?

Player 1: we always have so much time again

Player 2: if we each take a quadrant and move sporadically, they will re multiply

Player 1: let them stay connected

Player 1: to another token

Player 1: they multiply faster

Player 5: oh

Player 3: they multiply by groups of them

Player 1: go after the single ones first

Player 3: so if theres a bunch it will multiply faster

Player 1: ya

Player 5: nice

Player 1: leave them in a bunch
Player 1: they will mulitply faster
Player 2: yea thats why we need to each have our own section
Player 1: no need to rush so much time!
Player 1: no
Player 5: k
Player 4: so let us decide on sides of screen
Player 1: why
Player 1: what if i run out
Player 1: i wanna get more
Player 2: you won't
Player 5: haha
Player 1: ok
Player 3: we just have to slow down
Player 1: once the screen pops up
Player 2: i will take upper right
Player 4: so i will go after bottom right
Player 1: i will take the bottom left
Player 3: we all go crazy when they pop up
Player 5: i can get the bottom left
Player 3: i'll wonder
Player 1: no i said that first 5
Player 1: whatever it dosen't matter
Player 1: just scurry to any location
Player 4: lets try to work as a group
Player 1: i won't remember
Player 5: there are four quads and 5 players
Player 4: so again i am bottom right
Player 2: someone have center
Player 4: dont try to get in my space
Player 1: i'm bottom left
Player 4: well everyone has part of center
Player 2: when there is 30 sec left, free for all
Player 1: ok
Player 1: sounds good
Player 3: deal
Player 1: times almost up

Player 4: that would be too much

Player 2: once again im uper right

Player 1: ahhh get ready

Player 1: botoom left

Player 4: 'as we will end up before thaty

Player 3: wondering

Player 4: good luck to all

The next examples come from the treatment of neither communication nor costly punishment in the first three periods, and only communication in the second set of three periods. The examples show that participants are commenting on one another's behavior in period 5.

Player 2: What happened to the plan!?

Player 1: hey no one waited

Player 4: ok that was stupid

Player 5: wwe need to takr up more time

Player 4: no one waited?

Player 4: it was 4 and 2 that acctually did wait

Player 1: no one should eat until the whole thing fills up... but if one person starts eating it will ruin it

Another group, period 5:

Player 5: so who is the odd person

Player 3: we did better we just need to wait londer till we go nuts

Player 2: WHAT THE HECK FOUR

Player 4: i think we need to rethink on this

Player 1: 4 you cheated

Player 5: well

Player 5: yeah

Player 5: we need to keep them generating coins

Player 5: yeah agreed

Player 2: we could have lasted so much longer

Player 2: of course you think we should rethink

Player 1: so what's the plan

Player 1: ya

Player 5: that didn't go so well

Player 1: cuz you cheated

Player 1: new plan

Player 1: we need a new plan

Player 1: yes

Player 4: so i had to go to other side

Player 2: you had double everyone else

Player 1: we need to wait longer

Player 4: ok

Player 1: 4 you need to stop going crazy

Player 2: 4 shouldn't be selfish

Player 5: maybe everyone can look at the coins and keep even

Player 5: with everyone else's coin intake

Player 4: so lets try to switch sides

Player 2: i doubt it

Player 1: no we aren't doing quads anymore

Player 5: k

Player 2: what do you suggest?

Player 1: i suggest we wait for the whole grid to be filled

Player 5: i think we should really pace ourselves

Player 1: then go crazy

Player 4: if u guys dont agree with having sides we will all lose

Player 2: SMART lets do a time limit

Player 2: shut up four

Player 1: you totally came to my side

Player 2: mine to

Player 1: lets say that the last 60 seconds

Player 4: well what do u want

Player 5: there are not an equal amount of grids

Player 3: if we can wait to get the whole thing filled then we can all just go crazy

Player 1: is when we go crazy

Player 1: ya

Player 2: last 120 sec

Player 1: lets wait till its all filled up

Player 1: 100 seconds

Player 2: 110

Player 1: ok

Player 1: deal

Player 2: anyone else agree?

Player 3: so we're waiting till 110 to start ?

Player 1: you can still take one or two to help the multiplication alog

Player 4: well i think we should try to work our areas and then get crazy or whatever

Player 2: yup

Player 1: ok

Player 4: ok...

Player 2: once again, SHUT UP 4

Player 1: if you come into my area 4

Player 2: or mine

Player 1: we will all bomabard you

After period 5, the discussion continues as follows:

Player 2: FOUR

Player 1: ya

Player 1: no more quads ok

Player 1: we just wait

Player 1: cuz they waited

Player 1: ya 4 here dosen't want to wait

Player 5: we need to conserve the coins

Player 4: 1 got too crazy

Player 4: i am bored

Player 2: the other group all maximized

Player 2: but you couldn't wait

Player 2: so everyone else started

Player 1: and then when it hits 80 seconds

Player 1: we go at it

Player 1: ya

Player 1: so new plan: wait for max

Player 2: if four would wait and stick to the plan

Player 1: then at 80 seconds go at it

Player 1: deal?

Player 5: k

Player 4: y dont u guys just do

Player 5: 80 seconds

Player 2: sounds good

Player 2: i don't trust four tho

Player 3: 80 seconds we can all go crazy

Player 4: try to get the alone once
Player 2: i say we surround him
Player 1: ya me too
Player 1: i don't trust him
Player 4: as they dont have anything to do with multiplying
Player 2: and then all take off
Player 1: ok
Player 1: go for the alone ones
Player 2: shut up
Player 1: don't get the groups
Player 1: at 80 seconds we all get at them
Player 1: deal?
Player 4: ok i am the leader
Player 2: whatever guys
Player 4: as u can see from my scores
Player 1: no you;re not
Player 2: no
Player 1: thats b/c you cheat
Player 2: selfish
Player 4: so just follow what i say
Player 1: that's why you have those scores
Player 1: you do know btw
Player 3: u make 2 cents a thing ... good leadership reward
Player 1: that we earn money as a grou
Player 1: *group
Player 4: i am not liking one that much
Player 1: not individauls
Player 2: if we follow you no one will make money
Player 1: so if you wanna max your earnings
Player 5: true
Player 1: max the group earning
Player 2: exactly
Player 3: crazy at 80
Player 1: we all get the same amount of money
Player 1: so 4
Player 2: whats it gonna be
Player 1: you getting that score really dosen't mean much to you

Player 5: sounds great

Player 1: it means to all of us

Player 1: and if u waited we would have a good score

Player 4: well so do we stay in our area

Player 1: no area

Player 5: no areas

Player 1: !!!!

Player 2: no just WAIT

Player 1: guys surround four

Player 4: as 1 did rtry to get in mine

Player 2: YUP

Player 1: if he gets out of hand

Player 1: no i didn't

Player 2: gotcha

Player 1: i stayed in mine until you came to my area

Player 4: well try me guys

Player 1: surround ffour

In the second set of examples, the treatment is communication combined with costly punishment. We see comments like the following:

“i reduced #5 twice because he/she kept coming to my corner”

And discussions like:

Player 2: it went really well until everyone decided to take over my corner, thanks

Player 5: #2,why did you subtract me?..since we others've done our part

Player 4: me too

Player 1: If this is teamwork why is it that people are substracting from each other?

Player 5: we all moved to your corner..

Player 2: didn't we agree that eveyone will stay in their own corner until the end of the game

Player 3: because early in their game some people are moving into other people's corners to let their own regenerate

Player 5: where did it regenerate?

Player 2: if you stay within your own corner, they will regenerate quickly enough

Player 1: Obviously they are not regenerating so stop substracting from others!!! OK!!!!!!!!!!!!!!!!!

Player 3: well, they do regenerate if you don't pick all of them up #5

Player 3: 60 seconds in the middle is empty and #5 is romping around in my corner

Player 5: randomly?

Player 4: haha

Player 1: # 4, you need to worry about staying in your corner

Player 3: yeah, you need to worry about staying out of my corner

Player 1: If the shoe fits, you where it

Player 2: ok, how about if we rotated through each others squares but didn't collect all of the coins just enought to keep them regenerating

Player 5: should we change the corner?

Player 3: you're lucky this game's anonymous, you have NO CLUE how to play

Player 4: #2 that might get confusing

Player 2: ok true

Player 3: and you spelled *wear wrong

Player 2: stay in the same corner

A typical comment on the use of costly punishment:

Player 4: I dont think that stealing money from each other wil help anyone.

Player 1: agree

Player 4: we should agree not to do it

Player 2: I am fine with that

Player 3: i agree with noth stealing