Journal of Economic Psychology 31 (2010) 872–883

Contents lists available at ScienceDirect

Journal of Economic Psychology

journal homepage: www.elsevier.com/locate/joep

Inducing efficient conditional cooperation patterns in public goods games, an experimental investigation

Pablo Guillen [a,*], Enrique Fatas [b], Pablo Brañas-Garza [c]

[a] University of Sydney, Faculty of Economics and Business, Discipline of Economics, Room 340, Merewether Building (H04), Sydney NSW 2006, Australia
[b] LINEEX and University of Valencia, Facultad de Economía, Campus Tarongers, 46022 Valencia, Spain
[c] Facultad de Ciencias Económicas, Universidad de Granada, Campus Universitario de La Cartuja, E-18011 Granada, Spain

ARTICLE INFO

Article history:
Received 19 January 2009
Received in revised form 25 June 2010
Accepted 1 July 2010
Available online 8 July 2010

JEL Classification:
C9

PsycINFO Classification:
Group and interpersonal processes

Keywords:
Social dilemmas
Conditional cooperation

ABSTRACT

This study analyses the behavior in a repeated public goods game when subjects know about the possibility of existence of strict conditional cooperators. We employed a baseline treatment and a threat treatment in which subjects are informed about the possibility of being in a group together with automata playing a grim trigger strategy. We conjecture the resulting game allows for almost fully efficient outcomes. Contributions in the threat treatment increase by 40% before a surprise restart, and by 50% after the surprise restart. In line with the grim trigger strategy subjects contribute either all or nothing in the threat treatment.

1. Introduction

Since the article by Kelley and Stahelsky (1970) was published a stream of studies reports evidence of reciprocity or conditional cooperation in social dilemmas.[1] Cooperation is reported to decline over time in some social dilemmas like, for instance, repeated public goods games (see Fischbacher & Gächter, 2010; Neugebauer, Perote, Schmidt, & Loos, 2009).

In this study, we test whether the mere possibility of existence of a fraction of players credibly committed with a grim trigger strategy can suffice to avoid the decline of cooperation. This idea can be linked to the one in Kreps, Milgrom, Roberts, and Wilson (1982): if fully rational and egoistic individuals have the faintest suspicion they might be interacting with tit-for-tat players, there is room for cooperative equilibria. Note that both the grim trigger strategy and the tit-for-tat strategy are just particular forms of conditional cooperation.

Our experiment contrasts two scenarios. The first is a standard repeated public goods voluntary contribution game (*baseline* treatment). The second (*threat* treatment) is almost identical. The only difference is that in the latter some groups

* Corresponding author. Tel.: +61 2 9036 9188; fax: +61 2 9351 4341.
E-mail address: p.guillen@econ.usyd.edu.au (P. Guillen).
URL: http://www.econ.usyd.edu.au/16556.html (P. Guillen).

[1] See, for instance, Guttman (1986), Dawes and Thaler (1988), Andreoni (1995), Keser and van Winden (2000), Fischbacher, Gächter, and Fehr (2001), Brandts and Schram (2001), or Croson, Fatas, and Neugebauer (2005, 2006).

0167-4870/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.joep.2010.07.002

are composed by a combination of human subjects and computer controlled automata.[2] The automata in the threat treatment are noisy grim trigger strategy players. That is, as the experimental instructions explain, automata will contribute at least 90% of their endowment (45 units) to the public good as long as every other player in the group has always contributed at least 45 units in all previous periods, or contribute less than 10% (5 units) of their endowment until the last period otherwise.

This study focuses on the groups where no automata participated, as this provides a strong test for the grim trigger strategy threat and, at the same time, controls for the direct effect on cooperation due to the possible existence of automata.

Subjects in the threat treatment are aware of the possible existence of automata, but they are not given any objective probability. They face unmeasurable uncertainty because of the unknown behavior of others. This is not a common design choice. However standard laboratory public goods game, like our baseline treatment, entails unmeasurable uncertainty too. Indeed many articles studying finitely repeated public goods games find experimental subjects to behave in a very heterogeneous way, see for instance Fischbacher et al. (2001). According to their behavior subjects are usually classified as free riders, conditional cooperators and unconditional cooperators. Subjects are not informed about the objective probability of finding those behavioral types. Thus any given subject in a standard laboratory public goods game faces unmeasurable uncertainty regarding the composition of her group. In a way the threat treatment can be understood as not more, but actually less complicated than the very standard baseline treatment. Both self-regarding payoff maximizing players and conditional cooperators may respond in a similar way to the grim trigger strategy threat posed by the automata. That is imitating it to some extent. This way, human players become hard to distinguish from automata.

We do not observe beliefs but we rather directly manipulate subjects' beliefs in a way we expect to be mutually beneficial. This paper proposes the idea and analyses whether and to what extent inducing mutually beneficial beliefs may work in the context of the experimental laboratory. A similar idea is known as the Pascal Wager after the 17th century French philosopher Blaise Pascal. That is, the existence of God is not only uncertain but probabilities are unknown. Even though, a person should behave as if He exists, because living life accordingly has everything to gain and not much to lose. That of course may have beneficial effects to society as a whole if God, for instance, asks humans to be honest and trustworthy. A similar argument, taken from Islam, was made by the Imam Al-Juwayni about six centuries before Pascal. Al-Juwayni is known to have contributed most to Islamic canonical theology. More contemporarily Greif (2008) affirms that shared beliefs (about the behavior of others) are the engine of social rules.

Most papers pointing out the importance of beliefs, see for example Croson (2007), Neugebauer et al. (2009) or Fischbacher and Gächter (2010), do include belief elicitation. Typically experimental subjects are asked about others' contributions. We do not elicit beliefs in this study. We explicitly explain a strategy that could be played by a number of group members, thus making common knowledge of perfect rationality a completely unreasonable assumption. The right questions in that case would be not only about the subjective probability given by each subject to the existence of automata but also to whether the other subjects assign any probability to the existence of automata and so on. That would make the experimental design much more complicated and add little explanatory power.

We conjecture *polarized* contributions in the threat treatment. That is, players contribute an amount either $\geqslant 45$ or 0. Additionally, cooperation is expected to unravel suddenly and sharply after one or more players contribute 0 at any point. Our results strongly support these conjectures. Each player in each of the groups in the threat treatment started contributing 45 or more of his endowment in period one. In line with the grim trigger strategy, contributions decreased sharply to nothing in some groups, after a player contributed 0 units. Nevertheless, four out of nine groups managed to fully cooperate until period 8 (out of 10), and three groups until period nine. Players tend to contribute either 45 or 0, and make mostly 0 contributions after anyone contributed less than 5. On average contributions are 40% higher in the threat treatment than in the baseline treatment.

We added a surprise restart[3] to the repeated public goods game, with the expectation that after the restart, players would have enough information to discard the existence of automata and therefore contribute less. This expectation turns out to be false as most players can not rule out the presence of at least one automata in their groups. Although the grim trigger strategy threat becomes weaker, it does not disappear at the beginning of the restart. Indeed, seven out of nine groups behave in the same way before and after the restart. For the remaining two groups the non-existence of automata becomes obvious during the restart. Then contributions decrease in a smooth way and polarization disappears. Altogether, contributions are even higher in the threat treatment than in baseline treatment after the restart. Thus, in contrast with the literature on confusion,[4] and in line with our conjecture players seem to somehow "best respond" to the environment. Most never learn the whole truth. Even if there are no conditional cooperators in a group belief manipulation can create a chance for sustained cooperation. Also it is probable that if conditional cooperators could commit themselves to a grim trigger strategy and announce it, similar mutually beneficial results may be created.

Achieving sustained cooperation in social dilemmas has been the focus of many studies. For instance, Fehr and Gächter (2000) observe dramatic gains in contributions when adding a costly punishment phase after each repetition of the public

[2] See the instructions for details. As it is clear from the next section, we were very careful to avoid any shadow of deception. The use of computerized players is not new, see Bardsley (2001) and Ferraro, Rondeau, and Poe (2003).

[3] Andreoni (1988), Croson (1996) or Cookson (2000) are traditional references of this relatively common technique.

[4] Andreoni (1995), Palfrey and Prisbey (1997), Houser and Kurzban (2002) and Ferraro and Vossler (2010).

goods game. Efficiency improvements are however modest because of the cost of sanctions. Moreover allowing costly punishment may not be free of problems. Firstly, as Bornstein and Weisel (2010) point out, it requires a proper identification of deviators. Secondly, it may actually reduce cooperation if free riders punish cooperators like in Herrmann, Thöni, and Gächter (2008). Thirdly, it may be ineffectual is counter-punishment is available.[5] On the other hand, Gächter, Renner, and Sefton (2008) show how punishment can act as a threat and induce cooperation with very little punishment actually happening, the longer the repeated interaction the better the prospects for punishment to work as a threat.

Other studies on conditional cooperation can be related to our paper. Fischbacher et al. (2001) elicit subjects' willingness to contribute given the average contribution level of group partners. The average conditional cooperator is willing to contribute less than the mean contribution of the other group members. In a similar vein, Fischbacher and Gächter (2010) provide a direct test of the role of preferences for cooperation in voluntary contributions and the way they decline over time. Preferences for cooperation are heterogeneous: individuals can be classified into free riders and conditional cooperators. The decline in cooperation is explained not only by the number of free riders in the group who contribute little or nothing in the first period, but also by the willingness of conditional cooperators to contribute less than group partners. Experiments were also conducted by Gunnthorsdottir, Houser, and McCabe (2007) and Gächter and Thöni (2005) in which after preferences for cooperation are elicited, homogeneous low contributor and high contributor groups are formed. In both studies groups formed exclusively by high contributors manage to maintain high levels of cooperation until it eventually declines. Gächter (2005) also observe that low contributors substantially contribute to the public good for a while when they are knowingly re-matched with other low contributors. One of the plausible explanations given by the authors is strategic cooperation, where low contributors who perhaps might be free riders who thought they are not matched with other free riders like themselves, but with conditional cooperators with pessimistic beliefs who gave low contributions in the past. This is in fact the reputation argument in Kreps et al. (1982).

The remainder of this paper consists of four sections. Section 2 describes the experimental design and procedures. Section 3 includes conjectures and hypothesis about the threat treatment. Section 4 summarizes the results while Section 5 concludes. Appendix shows the individual contributions organized in a table, summary statistics and a translation of the experimental instructions.

2. Experimental design

We designed an experiment in which subjects play 10 periods of a public goods game in groups of four players. In any given period, the players have to decide how much to allocate to a public account. These contributions are integers between 0 and 50. The sum of the contributions given by the four players is then multiplied by two. Afterwards, this amount is shared equally among the four members of the group. Therefore, the individual payoff of a group member i is:

$$\pi_i = (50 - g_i) + \frac{2 \cdot \sum_{j=1}^{4} g_j}{4},$$

where j stands for group members from 1 to 4 and g_i is his/her individual contribution.

This game is repeated 10× with a constant group composition, after which it is announced that the experiment is over and all subjects are invited to participate in a new one. Ten additional periods of the same are played in the so called surprise restart. Group composition is kept the same as it is in the original 10 periods.

Subjects receive information about their own and their partners' contributions at the end of each period. However, they only get the ranked vector of contributions. That is, in each period contributions are displayed without identifiers but only ranked from the highest to the lowest. Hence, there is no possibility of identifying a particular player's contribution across periods.

In the *baseline* treatment the game is exactly as described so far.[6] The *threat* treatment is exactly like the baseline treatment except that subjects are informed that: "In each group there might or might not be some computer simulated subjects. A number between zero (where there are no computer simulated players) and three (you are the only non-computer simulated player) has been determined by the computer. You will not be informed at any time about the characteristics of other group members, either simulated or human." No clues are provided about the number of automata present in any group.

The subjects are carefully informed about the strategy played by automata. This is a noisy grim trigger strategy. That is, the automata would cooperate until any group member defects by contributing less than 45 units. If there is any defection the automata will then defect until the end of the game. The strategy is "noisy" in the sense that automata choose an integer between 45 and 50 when cooperating and between 0 and 5 when defecting. There is the same probability of picking a particular number in each interval.

We obtained six and nine independent observations for the baseline and the threat treatments respectively. All the experiments were run in the Lineex laboratory at the University of Valencia (Spain) using Fischbacher's (2007) z-Tree toolbox. The average payoff (including a 5 EUR show-up fee) was 19.71 EUR. The sessions lasted about 75 min.

[5] Nikiforakis (2008) allows further rounds of sanctions in an experiment very similar to that of Fehr and Gächter (2000). He shows that in the presence of counter-punishment opportunities cooperators are less willing to punish free riders. As a result cooperation breaks down.

[6] Data for our baseline treatment is also used in the Croson et al. (2005) study on conditional cooperation.

3. Behavioral conjectures and hypotheses

The study is aimed at determining whether knowledge about the possibility of existence of strict conditional cooperators[7] results in higher contributions, higher efficiency and higher payoffs. It is important to note that a fully rational, self-regarding player (RSP) should realize that the possibility of an automaton playing the grim trigger strategy presents an opportunity for rational cooperation by eliminating the assumption of common knowledge of perfect rationality.

Conjecture 1. In the circumstances explained above a rational conditional cooperator adopts the grim trigger strategy. Hence we are not including human conditional cooperators in our analysis.

A precise analysis of the game played in the lab is hard. The game is certainly complicated as we did not provide subjects with probabilities and beliefs are unknown. Neither SPNE nor PBE can be used as solution concepts as they require perfect information or incomplete information based on known objective probabilities respectively. However, a RSP might qualitatively reasons as follows.

We can affirm RSPs contribute 0 in period 10 in any case. Moving back to period 9 RSPs would contribute 0 if until this point anyone has contributed less than 45 because the automata would have stopped cooperating after that. With the automata defecting the game becomes equivalent to a standard finitely repeated public goods game in which RSPs should contribute 0 units. If nobody has contributed less than 45 units before period 9 RSPs could contribute amounts either $\geqslant 45$ units or 0 units; by contributing amounts $\geqslant 45$ units in period 9 RSPs plan to exploit automata in period 10. By contributing 0 in period 9, RSPs would not postpone free riding knowing that everyone will defect in period 10. The same logic can be used going backwards to period 1. Therefore we could state the following two conjectures:

Conjecture 2. RSPs never contribute strictly more than 0 or strictly less than 45.

Conjecture 3. RSPs never contribute strictly more than 0 after anyone contributed strictly less than 45.

We can formulate testable hypotheses based on conjectures 1, 2, and 3:

Hypothesis 1. Polarization: Under the automata threat, contributions to the public good follow a step function, either $\geqslant 45$ or zero.

Hypothesis 2. Problems associated with coordination can cause declining contributions. In other words, due to lack of coordination, any given player can contribute zero at any point between period 1 and period 9 causing fellow group members to follow the grim trigger strategy and contribute zero in the following rounds.

4. Results

4.1. Periods 1 to 10

Every individual in every group in the threat treatment started contributing at or above the grim trigger strategy threshold. While data coming from the baseline are widely spread along the whole [0, 50] interval, threat treatment data are strongly *polarized*: subjects tend to contribute either zero or a value within the [45, 50] interval.

Fig. 1 shows the whole set of contributions for both the baseline (left panel) and the threat treatment (right panel). Periods are represented on the horizontal axis and contributions on the vertical axis. The diameter of each bubble represents how many subjects, in the baseline or the threat treatment, contribute a particular amount in a given period. Striped bubbles represent contributions greater than or equal to 45, solid grey bubbles represent contributions higher than 5 and smaller than 45, and white bubbles represent contributions greater than or equal to zero and smaller than 6. As seen in the figure, subjects in the threat treatment tend to contribute either 45 or 0 units, while those in the baseline treatment contribute in a much more scattered way.

From periods 1 to 10 in the threat treatment, subjects contributed 45 units 43 percent of the time, in contrast to only 1.3 percent of the time in the baseline treatment. Subjects contributed zero units 30.5 percent of the time in the threat treatment and 14.6 percent in the baseline.

Result 1. In line with *Hypothesis* 1 the grim trigger strategy threat causes polarization: subjects contribute following a step function, either $\geqslant 45$ or zero.

Fig. 2 shows the average contribution to the public good in both treatments. The average contribution in the threat treatment is always above the average contribution in the baseline treatment. Average contributions in the threat treatment start at more than 90 percent of the endowment whereas in the baseline, subjects start contributing around half of their endowment. Contributions decay across time in both treatments. The 0s happen at different times in the threat treatment so the average contributions decline smoothly.

[7] That is, the possible existence of players fully committed to a grim trigger strategy.

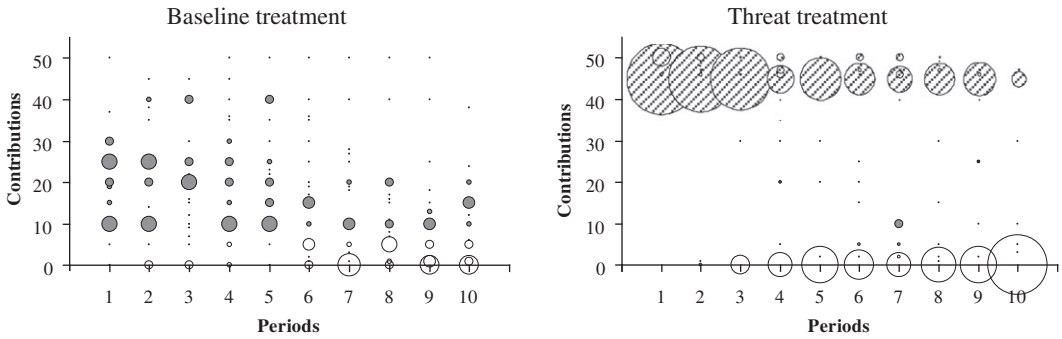

Fig. 1. Individual contributions under different treatments (periods 1 to 10).

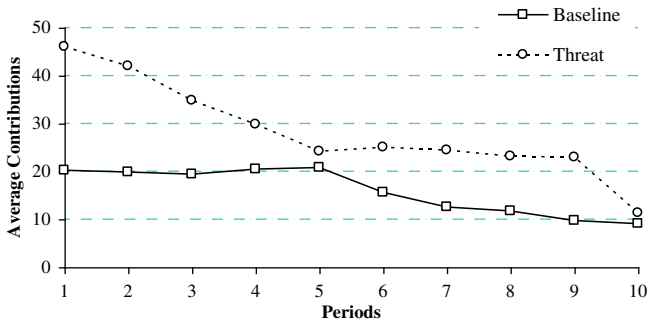

Fig. 2. Average contributions per treatment (periods 1 to 10).

Table 1 shows results from a panel data analysis.[8] The first two rows contain information about the sample used and the periods considered. The dependent variable is the individual contribution for all three models. The explanatory variables are the Period, a dummy representing the threat treatment (Threat), the average contribution of fellow group members in the previous period ($AvgCont_{t-1}$), and the minimum contribution of fellow group members in the previous period ($MinCont_{t-1}$).

Result 2. There is a significant and large treatment effect. "Threat" is significant at the 1% level in model [1] and contributions are estimated to be 42 percent higher in the threat treatment compared to the baseline treatment.[9]

Result 3. There is a significant and similar decline in contributions in both treatments. "Period" is significant at the 1% level in models [1], [2] and [3].

In models [2] and [3], a change in the conditional cooperation pattern arises. In line with Croson et al. (2005, 2006), the lagged average contribution of group partners ($AvgCont_{t-1}$) is significant in the baseline treatment, but the minimum contribution among group partners ($MinCont_{t-1}$) is not. The opposite is true for the threat treatment. This is not surprising as for a grim trigger strategy player the only thing that matters is whether the minimum contribution is lower than 45.

Result 4. The conditional cooperation pattern changes with the treatment. The treatment effect is as follows: $AvgCont_{t-1}$ is a good predictor in the baseline treatment and $MinCont_{t-1}$ is not significant. The opposite is true for the threat treatment.

Results 3 and 4 coupled with an inspection of individual contributions in Table A1 validate *Hypothesis* 2. Contributions decline sharply in the threat treatment after someone contributes <45 units.[10]

4.2. The maximum believable number of automata

The experimental design makes the exact number of automata in each group difficult to know. After each period, participants are informed of the individual contributions of the other three fellow group members in increasing order of contributions.

[8] The three models in Table 1 are estimated using random effects. Reported standard errors are corrected for robustness by clustering observations by group. This technique follows the approach designed by Liang and Zeger (1986) and implemented in Stata as a standard option for *xtreg* panel data regressions. The models in both Table 1 and Table 2 were also estimated using a *Tobit* approach censored at 0 and 50. Results are qualitatively the same.

[9] Similar results are obtained when using more conservative non-parametric techniques. Samples are compared across treatments period by period using a Mann–Whitney. Results are included in Table A3 in Appendix.

[10] Five out of nine times contributions decline sharply at some point because one individual contributes 0. In one case it happens after one individual contributes 1. In the remaining three cases cooperation is sustained until someone contributes 0 in the last period.

Table 1
Panel data estimations (periods 1 to 10).

	[1]	[2]	[3]
Sample	All	Baseline	Threat
Periods	1–10	1–10	1–10
Period	−2.47***	−0.80***	−0.81***
Threat	12.44**	–	–
$AvgCont_{t-1}$	–	0.49***	0.18
$MinCont_{t-1}$	–	0.16	0.53***
Const	29.52***	10.65***	12.64***
R^2-overall	0.23	0.43	0.59
N	600	216	324[a]

[a] Observations in the last two models do not add up to 600 due to the lagged nature of the variables (first period is missing so they sum 540).

* $\alpha \leqslant 10\%$.
** $\alpha \leqslant 5\%$.
*** $\alpha \leqslant 1\%$.

Remember that individual contributions are not identified with their contributor; hence tracing back individual contributions is not possible. Therefore, the only way of completely ruling out the existence of automata is when all the four group members' actions do not reflect the grim trigger strategy *at the same time*. For instance, in a group composed of four egoistic rational players who manage to cooperate until period 9, all would contribute 0 in period 10. After a surprise restart everyone knows there are no automata in the group.

In a very intuitive way, we can trace the maximum believable number of automata (MBNA) for each subject. That is, the maximum number of automata compatible with the information a certain subject has at a particular time. Table A1 in Appendix shows the MBNA for each subject in every period. The MBNA follows from the very same definition of grim trigger strategy and the fact that subjects only receive feedback about the ranked contributions of their fellow group members. From Table A1 we observe that, at the time of the restart, MBNA for all individuals in groups 7 and 9 is still $\geqslant 2$. A number of players in groups 3, 5, 6 and 8 cannot rationally believe in the existence of more than one automaton. Players S13 in group 1, S22 in group 2 and S41 in group 4 must be sure there are not automata at all in their group. There is no group in which every subject can be sure there are no automata.

An alternative approach to learn about the number of automata in a group is to measure how "random" the observed behavior is. Automata would randomize contributions between 45 and 50 when cooperating and between 0 and 5 when defecting. If the observed behavior is not noisy enough it cannot come from automata. We ran a battery of non-parametric tests to check the randomness of the observed behavior (see Table A2 in Appendix). Considering the extreme case where a player is able to perfectly recall all 30 contributions made by the other three members of his group during periods 1–10, a simple Kolmogorov–Smirnov (KS) test rejects "enough noise" for the data to come from three automata in 35 out of 36 cases. We obtain the same result if we suppose that a player can remember only 20 contributions, hence no rational subject should consider the data to be randomly generated by two automata.[11] In a similar vein, if an individual is assumed to remember just 10 contributions, there exists a set of 10 observations for 13 individuals where noise is big enough to come from an automaton. This test still rejects "enough noise" compatible with at least one automaton in 23 out of 36 times. Most of the time (in 73.5 percent of the occurrences) subjects choose to contribute either 0 or 45 units in the threat treatment. The KS tests suggest that it is extremely difficult to believe that the data observed by subjects is generated by more than one noisy automaton. In five out of nine groups, no subject ever contributed between 45 and 50.

We can also test whether particular subjects tried to imitate the randomized play of automata in the [45, 50] interval in order to trick other group members into believing the existence of automata. By looking at the series of individual contributions we can check whether they are noisy enough to perfectly imitate automata behavior. A KS test only accepts it for players S32 ($Z = 0.949$, $p = 0.329$) and S73 ($Z = 0.949$, $p = 0.329$). Only two subjects out of 36 pretend to be automata, and thus the data seems to be generated by a deterministic process rather than a random process in the majority of groups.

In summary, after 10 periods, the threat of grim trigger automata becomes weaker. It is possible that an extremely faint threat suffices to support behavior in line with Hypotheses 1 and 2. Whether or not the threat is weak enough to make behavior less cooperative after a surprise restart remains an open question. The analysis of the second block of 10 rounds helps us to clarify this issue.

4.3. Periods 11 to 20

All but three subjects (S11, S51 and S53) started contributing 45 after the surprise restart or more in period 11. Even the three players (S13, S22 and S41) who have to be sure there are no automata also contributed 45 or more.

[11] Strictly speaking, there is no single combination of 20 observations noisy enough to come from two random automata.

Fig. 3 shows individual contributions after the restart in both the baseline treatment and the threat treatment. The figures look very similar to those corresponding to periods 1–10. Now in the threat treatment subjects contributed 45 units 43 percent of the time, while in the baseline only 2 percent of the time. In the threat treatment, 29 percent of the time subjects contributed 0 units, compared with 13.8 percent in the baseline. All the subjects in groups 2, 3, 4, 6, 7, 8, and 9 started contributing 45 or more. Only one individual in group 1 and two in group 5 started contributing <45 units.

It has to be pointed out that everybody in group 1 (after period 11) and in group 4 (after period 14) should know for sure about the inexistence of automata. Cooperation unravels in both groups. It would be interesting to see what would happen in these groups after a second restart.

Result 5. For the 10 periods after the surprise restart the grim trigger strategy threat continues to cause polarization: subjects tend to the step function and contribute 45 or 0 units. Hypothesis 1 continues to hold after the surprise restart.

Fig. 4 shows the average contributions of subjects during periods 11 to 20. Average contributions in the threat treatment are again greater than the average contributions in the baseline.

Table 2 shows results from a panel data analysis. Models in Table 2 are estimated with the same methods used for Table 1.

Result 6. There is a significant and large treatment effect after the surprise restart. The "Threat" variable is significant at the 1 percent level in model [4]; contributions are 51 percent higher in the threat treatment than in the baseline treatment.

Result 7. There is a significant decline in contributions after the surprise restart. "Period" is significant at the 1 percent level in models [4], [5] and [6].

In models [5] and [6] a change in the conditional cooperation pattern is detected. The lagged average contribution of group partners ($AvgCont_{t-1}$) is significant in the baseline, but the minimum contribution among group partners ($MinCont_{t-1}$) is not. Again, the opposite is true for the threat treatment. This is not surprising because for a grim trigger strategy player the only thing that matters is whether the minimum contribution is less than 45 units.

Result 8. The change in the conditional cooperation pattern persists after the surprise restart: $AvgCont_{t-1}$ is a good predictor of contributions in the baseline and $MinCont_{t-1}$ is not significant. The opposite is true for the threat treatment.

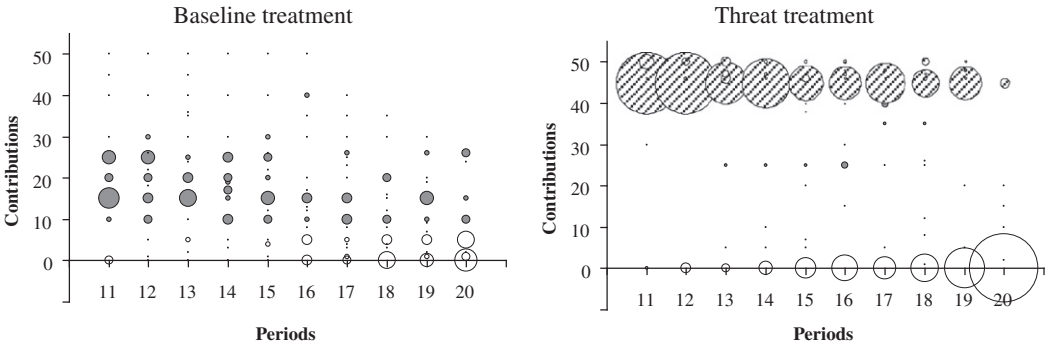

Fig. 3. Individual contributions per treatment (periods 11 to 20).

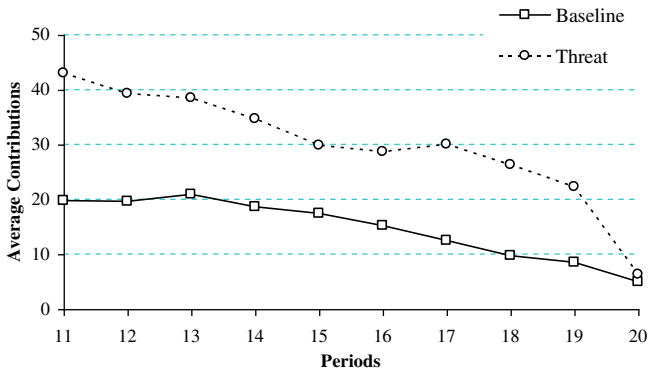

Fig. 4. Average contributions per treatment (periods 11 to 20).

Table 2
Panel data estimations (periods 11 to 20).

	[4]	[5]	[6]
Sample	All	Baseline	Threat
Periods	11–20	11–20	11–20
Period	−2.60***	−1.36***	−1.49***
Threat	14.94***	–	–
$AvgCont_{t-1}$	–	0.50***	0.15
$MinCont_{t-1}$	–	−0.15	0.58***
Const	29.02***	15.55***	16.63***
R^2-overall	0.29	0.28	0.64
N	600	216	324

* $\alpha \leqslant 10\%$.
** $\alpha \leqslant 5\%$.
*** $\alpha \leqslant 1\%$.

It should be noted that models [5] to [8] give parallel results to what is obtained in models [1] to [4]: behavior in the restart is extremely similar to that in the 10 original periods.

Results 7 and 8 plus an inspection of individual contributions in Table A1 validate Hypothesis 2. Contributions decline sharply in the threat treatment after someone contributes less than 45.[12]

5. Conclusions

The introduction of a grim trigger strategy threat increases contributions with respect to a baseline by 40 percent before a surprise restart and by 50 percent after the restart (Results 2 and 6). However, it does not prevent contributions from decreasing over time. Trend coefficients are still negative and significant in the threat treatment (Results 3 and 7). Experimental results validate Hypothesis 1: subjects either contribute $\geqslant$45 or 0 units, which is very much in line with the grim trigger strategy. Decay in contributions becomes sudden and sharp in the threat treatment. This occurs as a result of the zero contribution of one subject which is followed by zero (or very low) contributions from her fellow group members in the subsequent periods. The conditional cooperation pattern is altered. Results 4 and 8 show how the lagged average contribution of group partners is a good predictor of behavior for the baseline, but the minimum contribution among group partners is not. The reverse is the case for the threat treatment. Symptomatically, only in the two groups in which all members learn about the non-existence of automata, do contributions decay in the usual smooth way.

Our results suggest that subjects seem to play in a rather illuminated way somehow "best responding" to a very complicated environment. Subjects hardly learn the truth about group composition, which goes in their favor. Conditional cooperators may not exist at all in a particular group and yet subjects might almost fully cooperate. People might be good responding to environments containing unmeasurable uncertainty just because they do it in a day to day basis so they are equipped with the tools to deal with it.

On the other hand people might be not so good when dealing with measurable uncertainty because of the computational burden it entails. For that reason an obvious extension of this paper would be to compare our threat treatment with another in which objective probabilities are provided. Or in general, to compare the reaction to environments characterized by unmeasurable uncertainty with other characterized by measurable certainty.

It is not clear whether the kind of belief manipulation that was achieved here is attainable by just explaining the grim trigger strategy only to subjects rather than introducing a threat, or allowing subjects to commit to play a given cooperative strategy. This is the goal of ongoing research.

Appendix Experimental instructions

See Tables A1-A3.

The purpose of this experiment is to study individual decision making. Instructions are simple and you follow them carefully you will receive an amount of cash confidentially at the end of the experiment. That is, nobody will be informed of the payoffs of other participants. You can ask any question to us by raising your hand. Any other communication between participants is forbidden and it will be punish with the immediate expulsion from the experiment.

(1) The experiment consists of 10 independent periods. You will be a member of the same 4 people group throughout all the 10 periods. The results of each group are completely independent from others.

[12] Four out of 9× contributions decline sharply at some point because one individual contributes 0. In two cases contributions decline in a way not in line with the grim trigger strategy when someone contributes 30 and 25 respectively. These two cases correspond to the groups in which subjects have sufficient information to disproof the existence of automata. In the remaining three cases cooperation is sustained until someone contributes 0 in the last period.

Table A1
Individual contributions in the Threat treatment and maximum believable number of automata (MBNA).[a]

	S11	S12	S13	S14	S21	S22	S23	S24	S31	S32	S33	S34	S41	S42	S43	S44	S51	S52	S53	S54	S61	S62	S63	S64	S71	S72	S73	S74	S81	S82	S83	S84	S91	S92	S93	S94
1	45	45	50	45	45	45	45	50	50	45	45	45	45	45	45	45	45	45	45	50	45	45	45	50	45	45	46	50	46	45	45	45	45	45	45	50
2	50	50	0	45	46	45	45	0	48	47	45	46	45	45	45	1	45	45	45	45	45	45	45	45	45	45	49	45	45	45	45	45	50	45	45	45
3	30	45	0	45	0	45	0	0	45	46	45	45	45	0	0	0	0	45	45	45	47	45	45	45	45	45	45	45	45	45	45	45	45	45	50	45
4	35	5	50	30	20	0	20	40	45	50	45	46	0	0	0	0	0	0	0	0	46	45	45	46	50	45	47	45	47	45	45	45	45	0	45	47
5	20	45	0	30	0	0	0	0	45	45	45	45	0	0	0	2	0	0	0	0	45	45	45	45	45	45	48	45	45	50	45	45	0	0	0	45
6	25	5	0	15	5	0	0	45	45	47	47	45	2	0	0	0	20	0	0	45	50	45	45	45	45	45	50	45	45	46	50	45	0	0	0	0
7	5	0	5	2	0	10	0	50	45	50	45	46	2	10	10	10	0	0	0	0	45	45	45	50	45	45	46	45	45	46	47	45	0	0	0	40
8	15	30	0	45	0	0	0	0	45	48	45	45	5	0	0	1	0	0	0	0	49	45	45	47	45	45	47	45	50	45	45	45	0	0	2	0
9	25	45	0	40	10	0	0	0	45	47	45	45	2	0	0	0	0	0	0	45	45	45	45	46	45	45	46	45	0	45	0	45	0	25	0	0
10	30	0	0	43	0	10	0	3	0	45	45	0	0	0	0	0	0	0	0	0	0	45	0	45	45	0	47	45	0	0	5	0	0	0	0	0
11	30	45	50	45	46	45	45	50	45	50	45	45	45	45	45	45	0	45	0	50	45	45	45	45	45	45	48	45	45	45	50	45	45	45	50	45
12	45	45	50	45	0	50	45	45	45	45	45	50	45	45	45	46	0	0	0	0	45	45	45	45	45	45	45	45	46	45	48	45	45	45	45	45
13	25	50	50	45	5	0	0	47	45	46	45	46	40	45	45	45	0	0	25	45	47	50	45	45	46	45	46	45	45	45	45	45	45	50	47	45
14	45	50	50	0	10	5	0	45	45	45	45	45	25	45	45	44	0	0	25	0	47	45	45	45	45	45	47	45	45	0	45	45	46	45	45	45
15	20	50	25	40	7	5	0	0	45	47	45	46	38	45	45	45	0	0	25	0	45	45	45	45	45	45	50	45	0	0	0	0	46	45	46	45
16	30	45	50	40	0	0	0	25	45	46	45	45	45	25	25	15	0	0	0	0	50	45	45	45	47	45	45	45	0	0	0	0	47	45	48	45
17	35	45	45	25	5	0	0	48	45	45	45	46	35	40	40	0	0	0	0	0	45	45	45	45	45	45	45	45	0	40	0	0	48	45	45	45
18	25	50	50	0	8	0	0	0	50	46	45	45	12	35	35	1	0	0	0	0	45	45	45	45	45	45	46	45	0	0	0	0	47	45	48	45
19	20	45	0	45	0	0	0	0	45	50	45	0	5	0	0	0	0	0	0	0	45	46	45	45	45	45	45	45	0	0	0	0	48	45	48	45
20	15	0	0	20	0	0	0	0	0	0	0	10	0	0	0	2	0	0	0	0	0	45	0	0	0	0	45	45	0	0	0	0	0	45	0	0
	50		3 automata																																	
	50		2 automata																																	
	50		1 automaton																																	
	50		0 automata																																	

[a] Different gray tones represent in the table represent the maximum number of automata a particular subject may believe on at a certain time (regardless of random behavior). For instance, in period 3 subject S13 have to be sure there are no automata on his group because none of his partners behaved consistently with the grim strategy in the former period. Subject S72 can believe there are three automata in his group until period 19.

Table A2
Randomization tests.

	"10"			"20"			"30"		
	Sample	*N*	K-S	Sample	*N*	K-S	Sample	*N*	K-S
vs11	10	10	1.58***	20	20	2.90***	30	24	3.47***
vs12	10	10	1.58***	20	18	2.82***	30	18	2.82***
vs13	10	10	1.58***	20	17	2.66***	30	17	2.66***
vs14	10	10	1.58***	20	20	2.68***	30	22	2.98***
vs21	10	10	2.21***	20	20	3.80***	30	26	4.51***
vs22	10	10	1.58***	20	20	3.69***	30	27	4.35***
vs23	10	10	1.58***	20	20	3.35***	30	25	4.00***
vs24	10	10	2.53***	20	20	4.02***	30	25	4.60***
vs31	10	10	0.94	20	20	2.01***	30	30	3.46***
vs32	10	10	1.58***	20	20	3.13***	30	30	4.38***
vs33	10	10	0.63	20	20	1.78***	30	30	3.28***
vs34	10	10	0.63	20	20	2.23***	30	30	3.65***
vs41	10	10	2.21***	20	20	3.80***	30	27	4.61***
vs42	10	10	1.58***	20	20	2.90***	30	28	3.96***
vs43	10	10	1.58***	20	20	2.90***	30	28	3.96***
vs44	10	10	1.89***	20	20	3.57***	30	28	4.53***
vs51	10	10	2.84***	20	20	4.24***	30	30	5.29***
vs52	10	10	2.84***	20	20	4.24***	30	29	5.19***
vs53	10	10	2.84***	20	20	4.24***	30	29	5.19***
vs54	10	10	–	20	20	–	30	29	–
vs61	10	10	1.58***	20	20	3.35***	30	30	4.56***
vs62	10	10	0.94	20	20	2.46***	30	30	3.83***
vs63	10	10	0.94	20	20	2.46***	30	30	3.83***
vs64	10	10	1.89***	20	20	3.57***	30	30	4.74***
vs71	10	10	0.63	20	20	2.23***	30	30	3.65***
vs72	10	10	0.63	20	20	2.01***	30	30	3.46***
vs73	10	10	2.53***	20	20	4.02***	30	30	5.11***
vs74	10	10	0.63	20	20	2.23***	30	30	3.65***
vs81	10	10	1.58***	20	20	3.35***	30	30	4.56***
vs82	10	10	1.58***	20	20	3.35***	30	30	4.56***
vs83	10	10	0.63	20	20	2.46***	30	30	3.83***
vs84	10	10	0.63	20	20	2.68***	30	30	4.01***
vs91	10	10	0.63	20	20	2.46***	30	30	3.83***
vs92	10	10	0.63	20	20	2.46***	30	30	3.83***
vs93	10	10	1.26*	20	20	3.13***	30	30	4.38***
vs94	10	10	1.26*	20	20	3.13***	30	30	4.38***
Summary									
# Rejections	23	63%		35	97%		35	97%	
# Non-rejections	13	36%		1	3%		1	3%	

** $\alpha \leqslant 5\%$.
* $\alpha \leqslant 10\%$.
*** $\alpha \leqslant 1\%$.

(2) In each group there might or might not be some computer simulated subjects. A number between 0 (there are no computer simulated players) and 3 (you are the only non-computer simulated player) has been determined by the computer. You will not be informed at any time about the characteristics of other group members, either simulated or human.
(3) At each period every participant receives an amount of 50 Eurocents (ECU). Your only decision is to decide how many do you want to assign to a Common Account. The remaining Eurocents will be assigned to your Private Account.
(4) Profits from your Private Account equal the amount you assigned to it and they are independent of other people decisions.
(5) Profits from your group's Common Account are determined according to the sum of money assigned to this account by everybody in your group (that is, what you decide to assign plus what the other three group members decide to deposit on the Common Account). This sum is multiplied by two and divided in four equal parts, one for each group member.
(6) Computer simulated subjects follow a simple rule. They will assign a randomly determined amount between 45 and 50 ECU to the Common Account in the first period and they will keep doing the same in following periods while everybody else in the group assigns at least 45 ECU to the Common Account. Otherwise, they will they will assign to the Common Account a randomly determined amount between 0 and 5 ECU.
(7) Summarizing, you profit in a determined period will be determined in the following way:

$$\text{Profit} = \text{Profit from Private Account} + \text{Profit from Common Account 50 Eurocent} - \text{what I assigned to the common account} + (2 \times \text{Common Account})/4.$$

(8) You will be paid your 10 period accumulated profits privately at the end of the experiment.

Table A3
Mann–Whitney tests.

Period	*Z*-MW
1	−6.336***
2	−5.571***
3	−3.724***
4	−1.856*
5	−0.346
6	−1.107
7	−1.1481
8	−0.579
9	−1.109
10	1.800*
11	−5.381***
12	−4.287***
13	−4.006***
14	−3.303***
15	−2.107**
16	−2.226**
17	−2.735***
18	−1.888*
19	−0.794
20	2.525*

* $\alpha \leqslant 10\%$.
** $\alpha \leqslant 5\%$.
*** $\alpha \leqslant 1\%$.

You will see the amounts everybody in your group assigned to the Common Account sorted from the biggest to the smallest, but you will not know who assigned what. You will see also the sum of money assigned overall, that is, the Common Account. Moreover you will be informed of your profits, differentiating what comes from the Common Account and the Private Account.

References

Andreoni, J. (1988). Why free ride? Strategies and learning in public goods experiments. *Journal of Public Economics, 37*, 291–304.
Andreoni, J. (1995). Cooperation in public goods experiments: Kindness or confusion? *American Economic Review, 85*(4), 891–904.
Bardsley, N. (2001). Collective Reasoning. *Critical Review of International Social and Political Philosophy, 4*, 171–192.
Bornstein, G., & Weisel, O. (2010). Punishment, cooperation and cheater detection in noisy social exchanges. Hebrew University of Jerusalem Discussion paper 528.
Brandts, J., & Schram, A. (2001). Cooperation or noise in public goods experiments: applying the contribution function approach. *Journal of Public Economics, 79*(2), 399–427.
Cookson, R. (2000). Framing effects in public goods experiments. *Experimental Economics, 3*(1), 55–79.
Croson, R. (1996). Partners and strangers revisited. *Economics Letters, 53*(1), 25–32.
Croson, R. (2007). Theories of commitment, altruism, and reciprocity: Evidence from linear public goods games. *Economic Inquiry, 45*(2), 199–216.
Croson, R., Fatas, E., & Neugebauer, T. (2005). Reciprocity, matching and conditional cooperation in two public goods games. *Economics Letters, 87*(1), 95–101.
Croson, R., Fatas, E., & Neugebauer, T. (2006). *Excludability and contribution: A laboratory study in team production*. Working Paper, The Wharton School.
Dawes, R., & Thaler, R. (1988). Anomalies: Cooperation. *Journal of Economic Perspectives, 2*(3), 187–197.
Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review, 90*, 980–994.
Ferraro, P. J., & Vossler, C. A. (2010). The source and significance of confusion in public goods experiments. *The B.E. Journal of Economic Analysis & Policy*, *10*(1). Article 53. doi:10.2202/1935-1682.2006.
Ferraro, P., Rondeau, D., & Poe, G. L. (2003). Detecting other-regarding behavior with virtual players. *Journal of Economic Behavior and Organization, 51*(1), 99–109.
Fischbacher, U. (2007). z-Tree. Zurich toolbox for readymade economics experiments – Experimenter's manual. *Experimental Economics, 10*(2), 171–178.
Fischbacher, U., & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods. *American Economic Review, 100*(1), 541–556.
Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters, 71*, 397–404.
Gächter, S., & Thöni, C. (2005). Social learning and voluntary cooperation among like minded people. *Journal of the European Economic Association, 3*(2–3), 303–314.
Gächter, S., Renner & Sefton (2008). The long-run benefits of punishment. *Science, 322*(5907), 1510.
Greif, A. (2008). Self-enforcing institutions: Comparative and historical institutional analysis (Han Fuguo, Zhejiang University, Trans.). *Comparative Economic and Social Systems* (in Chinese).
Gunnthorsdottir, A., Houser, D., & McCabe, K. (2007). Disposition, history and contributions in public goods experiments. *Journal of Economic Behavior and Organization, 62*, 304–315.
Guttman, J. M. (1986). Matching behavior and collective action: Some experimental evidence. *Journal of Economic Behavior and Organization, 7*, 171–198.
Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science, 319*, 1362–1367.
Houser, D., & Kurzban, R. (2002). Revisiting kindness and confusion in public good experiments. *American Economic Review, 92*(4), 1062–1069.
Kelley, H., & Stahelsky, A. (1970). Social interaction basis of cooperators' and competitors' beliefs about others. *Journal of Personality and Social Psychology, 16*(1), 66–91.
Keser, C., & van Winden, F. (2000). Conditional cooperation and voluntary contributions to public goods. *Scandinavian Journal of Economics, 102*(1), 23–39.
Kreps, D., Milgrom, P., Roberts, J., & Wilson, R. (1982). Rational cooperation in the finitely repeated prisoners dilemma. *Journal of Economic Theory, 27*, 245–252.

Liang, K. Y., & Zeger, S. (1986). Longitudinal data analysis using gerealized linear models. *Biometrika, 73*, 13–22.
Neugebauer, T., Perote, J., Schmidt, U., & Loos, M. (2009). Selfish biased conditional cooperation: On the decline of cooperation in repeated public goods experiments. *Journal of Economic Psychology, 30*(1), 52–60.
Nikiforakis, N. (2008). Punishment and counter-punishment in public goods games: can we still govern ourselves? *Journal of Public Economics, 92*(1–2), 91–112.
Palfrey, T., & Prisbey, J. (1997). Anomalous behavior in public goods experiments: How much and why. *American Economic Review, 87*(5), 829–846.