ThE Papers 06/01

Departamento de Teoría e Historia Económica

Universidad de Granada

Inducing a Self-Fulfilling Prophecy in Public Goods Games[1]

Pablo Brañas-Garza, Universidad de Granada

Enrique Fatás, LINEEX and Universidad de Valencia

Pablo Guillén,[2] Harvard Business School

[1] We thank Al Roth for precious help.

[2] *Corresponding author:* Pablo Guillen, Baker Library B90, Harvard Business School, Soldiers Field, Boston MA 02163, pguillen@hbs.edu, Phone: +1 617 495 6753, Fax: +1 617 495 5287
.

Inducing a Self-Fulfilling Prophecy in Public Goods Games

Abstract

This study explores how a self-fulfilling prophecy can solve a social dilemma. We ran two experimental treatments, baseline and automata. Both consisted of a finitely repeated public goods game with a surprise restart. In the automata treatment it was announced that there might be automata playing a grim trigger strategy. This announcement became a self-fulfilling prophecy. That is, most participants actually followed a grim trigger strategy in the automata treatment resulting on an increase on the average contributions to the public good relative to the baseline treatment. Moreover, four out of nine groups managed to fully cooperate almost until the last period. Furthermore, after the surprise restart, when the automata threat is less credible, subjects' behavior was very close to that in the original game.

Keywords: *self-fulfilling prophecy, public goods game, grim trigger strategy, cooperation, automata, beliefs*

JEL Class.: C92, H41, C72

1. Introduction

A 'self-fulfilling prophecy' is a prediction that, in being made, actually causes itself to become true. The term was coined by the sociologist Robert K. Merton. We claim social dilemmas, like the voluntary contribution to a public good, may sometimes be solved by a self-fulfilling prophecy that induces cooperation. Many situations in real life can be modeled as a public goods game: keeping water wells in a good shape in Kenya as described by Ted Miguel and Mary Kay Gugerty, 2005; voluntary contributions to National Public Radio affiliates as studied by Jen Sheng and Rachel Croson, 2005; or even teamwork in organizations like described by Bengt Holmstrom, 1982. Sanctioning is a well known mechanism to maintain high contributions to a public good (see e. g. Ernst Fehr and Simon Gaechter, 2000). However, sanctioning can be only applied to some contexts, for instance to teamwork situations or to public goods in small communities. Sanctioning has at least two problems. First, it requires identifying the individual to be punished, which might not be obvious. Second, it can trigger a chain of retaliations and make cooperation to collapse (see Nikos Nikoforakis, 2004).

To design a self-fulfilling prophecy we looked at the experimental evidence in repeated public goods games in which it is a dominant strategy to contribute zero to the public good. However, in experiments, individual contributions start typically quite high, around 50% of the endowment on average, and they decline over time to less than 10% of the endowment. This behavior cannot be explained entirely as learning. Indeed, the experimental literature, e. g. James Andreoni (1988), shows how contributions jump up to 50% of the endowment after a surprise restart. A variety of studies[1] on public good games explain experimental results through reciprocity or conditional cooperation. These studies usually identify two types of subjects, cooperators and free riders. They observe that most of the time cooperators only keep giving high contributions to the public good if other people in the same group do the same; these are the so-called conditional cooperators. Free riders contribute zero most of the time making conditional cooperators contribute less. Thus, average contributions decline across time. Nevertheless, rational free riders should cooperate

[1] See e.g., Robert Sugden (1984), James Andreoni (1995), Thomas Palfrey and Jeffrey Prisbey (1997), Rachel Croson (1998), Urs Fischbacher et al. (2001), Claudia Keser and Frans van Winden (2000), Jordi Brandts and Arthur Schram (2001) or Rachel Croson et al. (2005).

if they know there are conditional cooperators in the group. That would mean sustained cooperation after a surprise restart. However, this is not found in the experimental literature. We find a couple of plausible explanations. The first is the lack of coordination of conditional cooperators in initial contributions. Indeed, most subjects contribute a significant amount of their endowment. However, a contributor of 80% of her endowment might feel exploited by somebody else contributing only 40% thus decreasing her contribution in the following period. On the other hand the reaction to very low contributions is usually not fast and strong giving more incentives to free ride after a restart.

Our candidate to work as a self-fulfilling prophecy is the grim trigger strategy. It can be formulated as "fully contribute to the public good while everybody does, contribute zero forever otherwise." Playing a grim trigger strategy would solve the coordination problem mentioned in the former paragraph. In addition, the response to low contributions would be the strongest. Even for a fully rational and self-regarding individual it makes sense to cooperate until the next to the last period if he thinks some of their partners are going to play the grim trigger strategy. This individual would contribute zero until the last period if they detect a deviation from cooperation. Hence, it makes sense for fully rational and self-regarding individuals to become grim trigger strategy players, except for the last period. Furthermore, a player who is sure there is no grim strategy player but thinks the other players believe there might be some grim strategy players has no interest in educating them. That is, a cooperative result not only depends critically on what are each player's beliefs about the existence of grim strategy players, but also in second, third or actually n^{th}-order beliefs. It is therefore unclear whether a grim trigger strategy threat can work as a self-fulfilling prophecy with real players in the experimental lab.

We designed an experiment in which participants play ten periods of a public goods game in groups of four people. After period ten there was a surprise restart and the same game was played for another ten periods. Group composition never changed across periods. Players received information about their own and other group members' contributions. However, it was not possible to identify the actions of a certain player. Contributions were shown each period sorted from highest to lowest. In the baseline treatment the game was exactly the one described so far. In the automata treatment we told subjects they are in a group together with 0, 1, 2 or 3 automata that play a grim trigger strategy. The automaton would start by contributing a random amount between 90% and 100% of their endowment and they would keep doing it while everybody else in the group contributes at least 90%. Otherwise, automata would switch to contribute between 0% and 10% of their endowment. We did not provide any clue about what the number of automata in a certain group might be. Actually, there were 0 automata in all groups. Note that, if participants start fully contributing to the public good in the first period it will be hard to disprove the existence of grim strategy players. Indeed, since it is not possible to know how a certain player behaves across periods, the existence of automata can be only completely disprove if all four players do something not in line with the grim trigger strategy in the same period. This would be the case if subjects cooperate until period 9 and all contribute zero in period 10. Since automata are supposed to do some randomization there might be a 'back-door' way to conclude there are no automata by looking at contributions which do not seem to be random. However, a particular subject might think he understands contributions do not look random enough, but he can also think the others are not able to detect that lack of randomness. In this case it would be in his interest to keep playing the grim trigger strategy himself until the next to the last period. The interest of our surprise restart is to stress-test the power of the self-fulfilling prophecy. A fully cooperative restart will indicate that the prophecy is still causing itself to become true.

In the experiments, every player in every group in the automata treatment started contributing 90% of his endowment or more in period one. Cooperation unraveled in some groups because some player contributed zero. Nevertheless, four out of nine groups managed to fully cooperate until period eight, three until period nine. The self-fulfilling prophecy itself did not unravel. In period one of a surprise restart 51 out of 54 individuals playing the automata treatment contributed 90% of their endowments or more. In both the original ten periods and the restart average contributions were always higher than in the baseline. In line with the grim strategy, results are very polarized in the automata treatment, meaning that most of the time participants contributed 90% of their endowments (or more) or zero. Moreover, conditional cooperation pattern is altered in the automata treatment: 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. Exactly the opposite is true for the automata treatment.

Those results indicate that our design managed to generate a self-fulfilling prophecy which resulted in higher average contributions. In the context of our simple experimental environment, automata treatment can be thought of as an attempt to change the "corporate culture" that the subjects face. Of course, in a more complex and stochastic environment, the grim trigger strategy itself would be too rigid as a rule to use to sustain cooperation in the long term. But our results suggest that the amount of cooperation may be susceptible to influence through attempts to change the expected behavior of participants.

Section 2 explains the experimental design and procedures. Section 3 contains a discussion, Section 4 explains the results and Section 5 concludes. Finally, there is an Appendix including individual contributions organized in a table, summary statistics and a translation of the experimental instructions.

2. Experimental design and procedures

In our experiment subjects played a ten periods repeated public goods game. In any period players had to choose a contribution to a public fund. This contribution was an integer between 0 and 50. Four players formed each group. The sum of the contributions given by the four players was multiplied by two. Then, this amount was equally split among the members of the group. So, 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 individual contribution.

This game was repeated ten times by the same group. Then, there was a surprise restart and ten additional periods were played by the same group. Players received information about their own and partners' contributions. However, there was no possibility of identifying a particular player's contribution across periods. In each period contributions were displayed without identifiers ordered from highest to lowest.

We compare two treatments. In the *baseline treatment*[2] human subjects played the game described so far. In the *automata treatment* participants were told: "In each group there might or might not be some computer simulated subjects. A number between 0 (there

[2] We use the same baseline as in Rachel Croson et al. (2005).

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." Nevertheless, in the sessions run there were always 0 automata.

Participants were informed about the strategy automata played. This was a "noisy grim trigger strategy." That is, automata would cooperate until any group member defects. If there is any defection the automata would then defect until the end of the game. The strategy was noisy in the sense that automata would pick an integer from 45 to 50 when cooperating and an integer from 0 to 5 when defecting. There was the same probability of picking a particular number in each interval.

We ran two sessions of each treatment. There were three groups per session in the baseline. We had four groups in one of the automata treatment sessions and five in the other. So, finally we gathered six and nine independent observations for the baseline and the automata treatment s, respectively. All the experiments were run at the Lineex lab in University of Valencia (Spain) using Urs Fischbacher's (1999) z-Tree toolbox. Average payoff including a 5 EUR show-up fee was 19.71 EUR.

3. Discussion

The aim of our study was to determine empirically whether a self-fulfilling prophecy can be set-up in a public goods game resulting in higher contributions and higher payoffs to participants. This is an empirical question and we look for an answer in the experimental lab. A fully rational and self-regarding player in this experimental environment must realize that the possibility of an automaton playing the grim strategy presents an opportunity for rational cooperation by eliminating the assumption of common knowledge of rationality as explained by David Kreps et al., 1982.

A self-regarding and fully rational agent who thinks groups can be only composed of other rational players and automata will cooperate (C), meaning play between 45 and 50, to avoid automata retaliations and/or to look like an automata himself, or defect (D), meaning play zero. For the same player, a sequence of actions can never contain cooperation after defection. That rules out, for instance, a sequence like CCDCCCCCCD, where C means to contribute 45 of more and D means to contribute 0. This is so because automata will never cooperate after observing somebody contributes less than 45. Moreover, in equilibrium the same rational player must also start playing D immediately after observing somebody contributes less than 45. Indeed, it makes no sense to keep cooperating because the automata would now contribute from 0 to 5 and cooperation cannot be sustained with other rational and self regarding players.

Our design opens the door for beliefs such as "there are three grim trigger automata players in my group" which makes a best response to cooperate until the next to the last round.

Less demanding beliefs also allow cooperative outcomes. For instance, "there is one grim trigger automaton in my group and the rest of us are fully rational" is enough to make cooperation plausible in period nine. Consider the backwards induction process. In period ten the rational players should defect. In period nine, if cooperation has been maintained so far each rational player plays the game shown in table 1.

Table 1. Player R Payoffs at Period 9

	Number of other R players cooperating		
	0	***1***	***2***
C	872.5	895	940
D	895	917.5	940

Conditional on these beliefs there are two equilibria in period 9, everybody plays C or everybody plays D. In this example, backwards induction does not generate a unique non-cooperative equilibrium.

Furthermore, it is also possible to think of circumstances in which a rational player has incentives to behave as an automaton even though he is sure there are no automata in his group. Other group members might have wrong beliefs, but a rational player has no interest in showing them the truth.

One of the most interesting features of our automata treatment is it is actually very hard to obtain valuable information. If all four group members start cooperating the only way to disprove the existence of grim trigger strategy players is if all four group members violate the grim trigger strategy at the same time. Otherwise, there will always be the possibility there is at least one automaton.

Moreover, for non-fully rational individuals the grim trigger strategy is relatively safe to play. In the worst case scenario a given player fully cooperates in the first period and everybody else contributes zero. Hence, that player is going to make 25 units instead of 50 in the first period and 50 from period two to ten, it represents only a 5% loss over contributing always zero. If just one other player fully contributes in the first period there is no loss at all, and every other possible scenario offers higher profits than contributing zero all the time.

Finally, it is plausible that many people who participate in economic experiments are not able to put together a contingent action plan, that is, a strategy. They learn the grim trigger strategy from our experimental instructions and just use it. It is also plausible that most participants in economic experiments want to be cooperative to some extent, but they are afraid to be exploited. They might start contributing an amount bigger than zero in the first period to signal that they would actually like to cooperate. If these initial contributions are different, the ones who contribute more can think they are actually being exploited and decrease their contributions in the following rounds. This lack of coordination in the first period makes cooperation unravel. The grim trigger strategy offers a coordination mechanism, starting with a 'suggested' first period contribution.

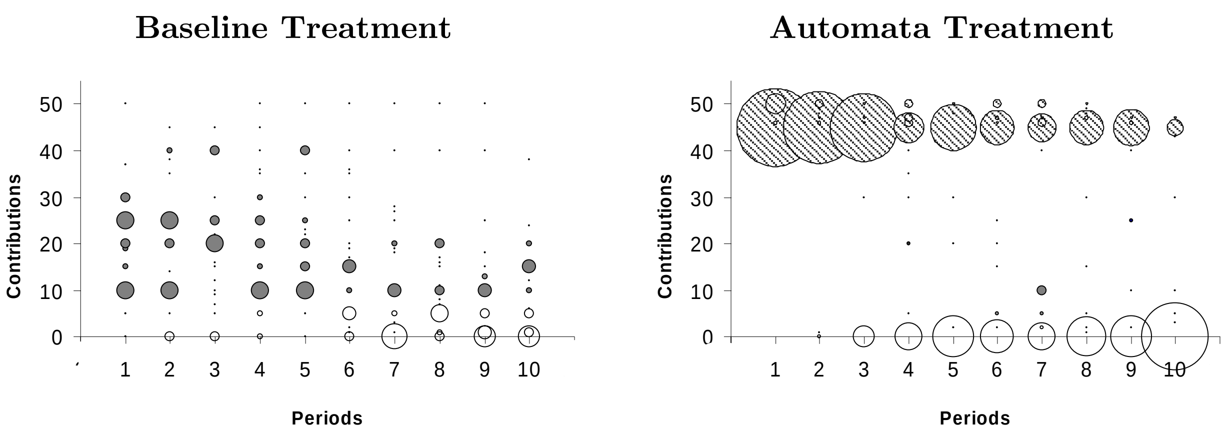
4. Results

4.1. Periods 1 to 10

A mere glance at the individual data illustrates how differently contributions are distributed. Every individual in every group in the automata treatment started contributing on or over the grim trigger strategy threshold. Whereas data coming from the baseline are widely spread along the whole [0, 50] interval, automata treatment data are strongly *polarized*: subjects tend to contribute either 0 or in the [45, 50] interval. The number of times somebody contributes any positive amount between 0 and 45 units is very small, around 10% on average. In fact it is: period 1 (0%), period 2 (3%), 3 (3%), 4 (17%), 5 (8%), 6 (17%), 7 (25%), 8 (14%), 9 (14%) and 10 (14%).

Figure 1 shows the whole set of contributions for both the baseline (left) and the automata treatment (right). Periods are represented on the horizontal axis and contributions on the vertical axis. The diameter of the bubbles represents how many subjects in the baseline or the automata treatment contribute a particular amount in a particular period. Striped bubbles represent contributions greater or equal to 45, solid grey bubbles represent contributions higher than 5 and smaller than 45, white bubbles represent contributions greater or equal to 0 and smaller than 6. Clearly, we observe that in the automata treatment subjects tend to contribute either 45 or 0, while in baseline contributions are much more scattered.

Figure 1: Individual contributions per treatment (periods 1 to 10)



In the automata treatment 156 out of 360 times (43%) participants played '45' while in the baseline they did it just 3 times out of 240 (1.3%). In contrast, 110 out of 360 (30.5%) subjects played '0' in the automata treatment while 35 out of 240 (14.6%) did it in the baseline.

Result 1: By inspection, the grim trigger strategy threat causes polarization: subjects tend to contribute 45 (or more), or 0.

Figure 2 shows the average contribution to the public goods in both treatments. The automata treatment average is always over the baseline. Note how average contributions in the automata treatment start over 90% of the endowment whereas in the baseline, on average, subjects start contributing around half of their endowment. Contributions decay across time in both treatments.

Figure 2: Average contributions per treatment (periods 1 to 10)

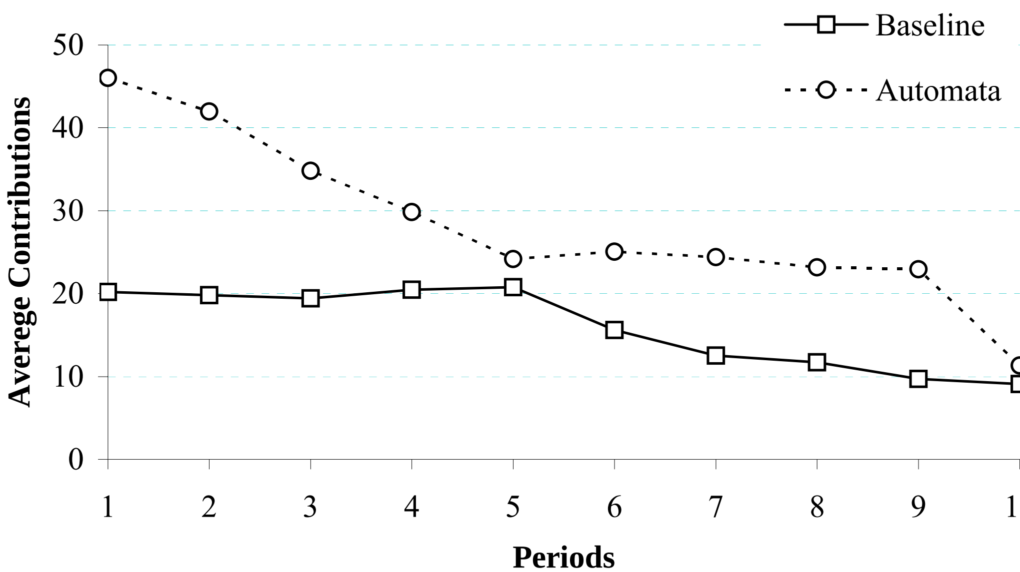


Table 2 shows estimations using random-effects regressions which include individual (subject) dummies. The first two rows contain information about the sample used and the periods considered. The explained variable is the individual contribution for all three models. Explanatory variables are Period, a dummy representing the automata treatment, the average contribution of partners in the previous period ($AvgCont_{t-1}$), the minimum contribution of partners in the previous period ($MinCont_{t-1}$).

Table 2: Panel data estimations (periods 1 to 10)

	[1]	[2]	[3]
Sample	All	Bas	Aut
Periods	1-10	1-10	1-10
Period	-2.47***	-0.78***	-0.81**
Automata	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

***α ≤ 1%; ** α ≤ 5%; * α ≤10%

Result 2: there is a significant and huge treatment effect. 'Automata' is significant at 1% in model [1]; contributions are 42% higher in the automata treatment than in the baseline treatment.

Mann-Whitney tests confirms result 2, contributions are significantly different between treatments for periods 1***, 2***, 3***, 4* and 10* (see table A3). Result 3 supports the existence of significant differences in later rounds.

Result 3: there is a significant and similar decline in contributions in both treatments. 'Period' is significant at 1% in models [1], [2] and [3].

Models [2] and [3] detect a change in the conditional cooperation pattern. The lagged average contribution of group partners ($AvgCont_{t-1}$) is significant in the baseline[3], but the minimum contribution among group partners ($MinCont_{t-1}$) is not. However, the opposite is true for the automata treatment. This is not surprising, for a grim trigger strategy player the only thing that matters is whether the minimum contribution is under 45.

Result 4: the conditional cooperation pattern is altered by the treatment effect: $AvgCont_{t-1}$ is a good predictor in the baseline and $MinCont_{t-1}$ is not significant. Exactly the opposite is true for the automata treatment.

4.2. Updating beliefs

We can trace for each subject the maximum believable number of automata (MBNA). That is, the maximum number of automata compatible with the information a certain subject has at a particular time. Our automata play a noise grim trigger strategy. First of all, if the observed behavior is not noisy enough it cannot come from automata. We ran a battery of tests to check this aspect (see table A2 in the Appendix). If we consider that a player is able to remember the 30 contributions made by the other three members of his group during periods 1-10 our Kolmogorov-Smirnov (KS) test rejects "enough noise" in 35 out of 36

[3] This result is close to the one observed in Rachel Croson et al. (2005), where different conditional cooperation patterns are observed depending on the strategic structure of the game.

cases. We obtain the same result if we suppose that a player can remember only 20 contributions. If we allow an individual just to remember 10 contributions the test rejects "enough noise" 23 out of 36 times. Moreover, if we look at the series of individual contributions in order to 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 ($Z=0.949$, $p=0.329$). We can conclude that, although for a perfectly rational individual there is not enough noise compatible with the existence of automata there is indeed enough noise for some individuals if they cannot remember more than 10 contributions. Furthermore, it is always possible some players think they can deduce that there are no automata but other players cannot, so lack of common knowledge of rationality may persist.

Let us suppose hereafter players are not able to perceive whether or not subjects' behavior is noisy enough to come from a randomization process. Suppose they are still able to trace behavior compatible with a grim strategy. Table A1 in the Appendix shows the MBNA for each subject at every period. Notice that, at the time of the restart, MBNA for all individuals in groups 7 and 9 is bigger or equal than two. Some players in groups 3, 5, 6 and 8 cannot 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 his group. There is no group in which everybody can be sure there are no automata, therefore, the prophecy that there is a grim trigger strategy player can still make itself true.

4.3. Periods 11 to 20

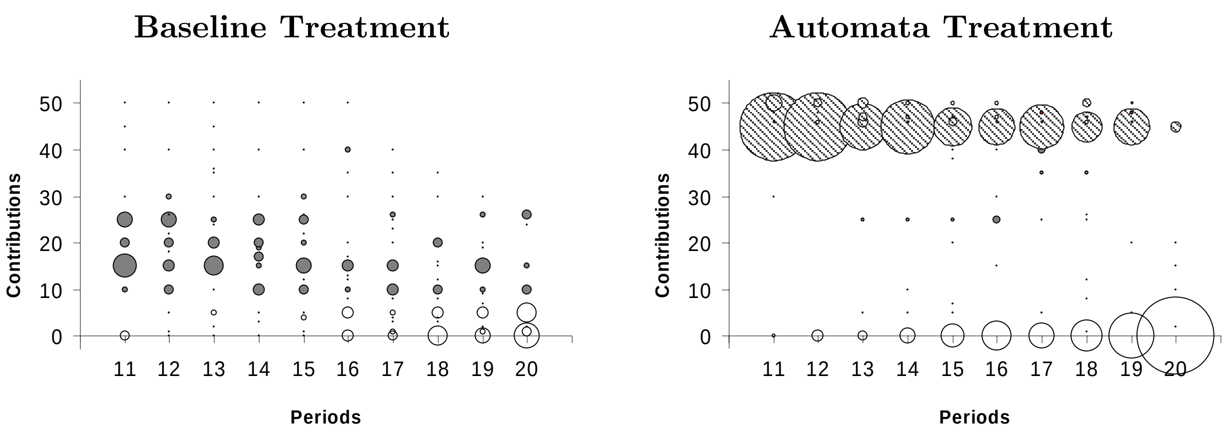
At the end of period 10 most players, 35 out of 36, can detect that there are no automata if they are sophisticated, meaning they can detect the insufficient randomization. If they are not sophisticated, i.e. if they can only detect strict departures from the grim trigger strategy, there are only three players who must know there are no automata at all. All but three subjects (S11, S51 and S53) started contributing 45 or more in period 11. Furthermore, those three who have to be sure there are no automata (S13, S22 and S41) also contributed 45 or more.

Figure 4 shows individual contributions after the restart for both the baseline and the automata treatment. The figures look very similar to those corresponding to periods 1-10. Now, in the automata treatment there are 158 out of 360 '45s' (43%) and 104 '0s' (29%).

In the baseline treatment there were just 2 out of 240 '45s' (0.83%) and 33 '0s' (13.8%). Everybody 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 less than 45.

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

Figure 4: Individual contributions per treatment (periods 11 to 20)



Result 5: By inspection, after a surprise restart the grim trigger strategy threat still causes polarization: subjects tend to contribute 45 (or more), or 0.

Figure 5 shows average behavior in periods 11 to 20. Average contributions in the automata treatment are again over average contributions in the baseline.

Figure 5: Average contributions per treatment (periods 11 to 20)

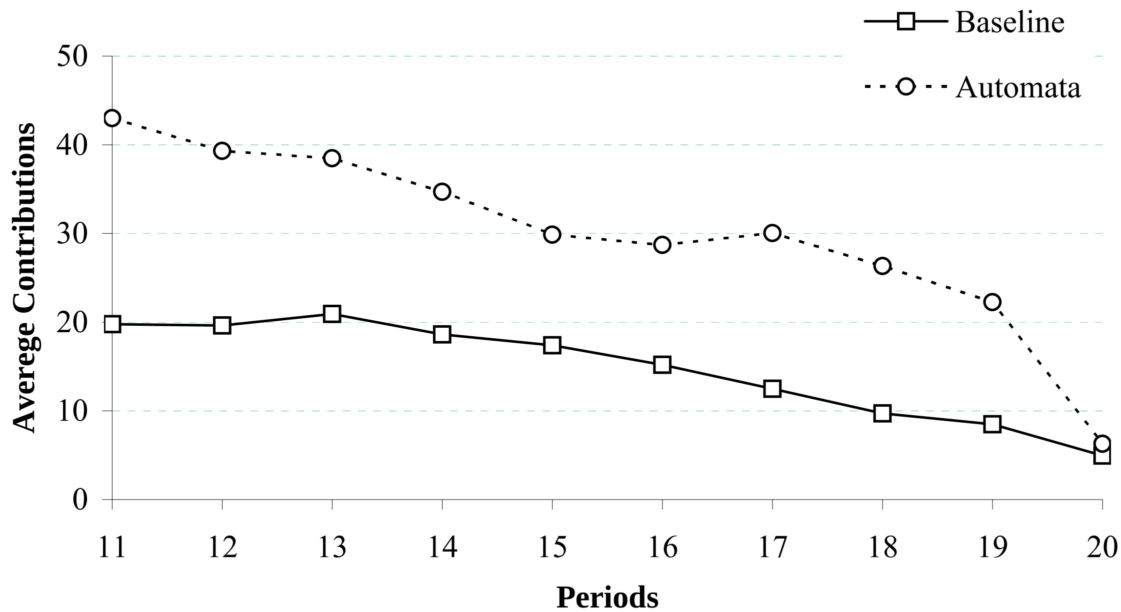


Table 3 shows estimations using random-effects regressions which include individual (subject) dummies. The only difference with respect to table 2 is that now data comes from the restart, periods 11 to 20.

Table 3: Panel data estimations (periods 11 to 20)

	[4]	[5]	[6]
Sample	All	Bas	Aut
Periods	11-20	11-20	11-20
Period	-2.60***	-1.36***	-1.49***
Treatment	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

***$\alpha \leq 1\%$; ** $\alpha \leq 5\%$; * $\alpha \leq 10\%$

Models [4], [5] and [6] give even stronger results than [1], [2] and [3] (included in Table 2).

Result 6: there is a significant and huge treatment effect. 'Automata' is significant at 1% in model [1]; contributions are 51% higher in the automata treatment than in the baseline treatment.

Again, Mann-Whitney tests confirm result 6; contributions are period by period significantly different between treatments except for period 19 (see table A3).

Result 7: there is a significant decline in contributions. 'Period' is significant at 1% in models [4], [5] and [6].

Models [5] and [6] detect a change in the conditional cooperation pattern. 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. However, the opposite is true for the Automata Treatment. This is not surprising, for a grim trigger strategy player the only thing that matters is whether the minimum contribution is under 45.

Result 8: the conditional cooperation pattern is altered by the treatment effect: $AvgCont_{t-1}$ is a good predictor in the baseline and $MinCont_{t-1}$ is not significant. Exactly the opposite is true for the automata treatment.

5. Conclusions

This study analyses how subjects behave in a repeated public goods game under a grim strategy threat. The purpose of doing that was to see whether a self-fulfilling prophecy can help to solve social dilemmas like public goods games. We ran 10+10 period public goods games with 4-player groups to explore how this threat affects individual contributions. We use a baseline treatment and an automata treatment in which subjects are informed about the possibility of being in a group together with 0, 1, 2 or 3 automata playing a grim strategy. There were zero automata in every group.

Our experimental results suggest that:

i) The grim strategy threat induces subjects to contribute significantly more in repeated public goods games.

ii) The grim strategy threat causes subjects produces very polarized results. Subjects either contribute 45 (or more) or 0, very much in line with a grim trigger strategy. The prediction “there might be somebody playing a grim trigger strategy” causes itself to be true.

iii) The conditional cooperation pattern is altered by the treatment effect: 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. However, the opposite is true for the Automata Treatment.

iv) The self-fulfilling prophecy survived a surprise restart. That is, i), ii) and iii) still hold in a second batch of ten periods even when there is very limited support for a rational expectation about the existence of automata in most groups.

These results raise other questions. For instance, it is not clear whether similar effects can be attained by just announcing the grim trigger strategy. It should be also interesting to allow subjects to commit to play the grim trigger strategy. Furthermore, it would also make sense to examine different strategies.

Our results suggest that cooperation may be influenced by the social culture of participants, i.e. their expectations about what strategies others employ. Changes in the way this culture is transmitted and shared by individuals seem to have a mayor impact in their behavior and their expectations about future events. Whether or not these results hold for more complex and stochastic environments, where the grim trigger strategy itself would be too rigid as a rule to use to sustain cooperation remains an open question, is the target of ongoing research.

References

Andreoni, James. “Why Free Ride? Strategies and Learning in Public Goods Experiments.” *Journal of Public Economics,* 1988, 37, 291-304.

Andreoni, James. “Cooperation in public goods experiments: kindness or confusion?” *American Economic Review,* 1995, 85 (4), 891–904.

Brandts, Jordi and Schram, Arthur. “Cooperation or noise in public goods experiments: applying the contribution function approach. *Journal of Public Economics,* 2001, 79 (2), 399–427.

Camerer, Colin. “Behavioral Game Theory.” Princeton: Russell Sage Foundation/Princeton University Press, 2003.

Croson, R.T.A., 1998. Theories of altruism and reciprocity: evidence from linear public goods games. Working Paper, The Wharton School.

Croson, Rachel; Fatas, Enrique and Neugebauer, Tibor. “Reciprocity, matching and conditional cooperation in two public goods games.” *Economics Letters*, 2005, 87(1), 95-101.

Erev, Ido and Roth, Alvin. “Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria.” *American Economic Review*, 1988, vol. 88, no. 4, 848-81.

Fischbacher, Urs, Gaechter, Simon. and Fehr, Ernst. „Are people conditionally cooperative? Evidence from a public goods experiment.” *Economics Letters*, 2001, 71, 397–404.

Fischbacher, Urs. ”Z-tree. Zurich Toolbox for Readymade Economics Experiments - Experimenter’s Manual”, Working Paper No. 21, Insitute for Empirical Research in Economics, University of Zurich, 1999.

Holmstrom, B. "Moral Hazard in Teams." *Bell Journal of Economics*, 1982, 13(2): 324-40.

Healy, Paul. "Learning Dynamics for Mechanism Design: An Experimental Comparison of Public Good Mechanisms." Submitted to the *Journal of Economic Theory*, 2004.

Isaac, Mark; Walker, James and Williams, Arlington. "Group Size and the Voluntary Provision of Public Goods: Experimental Evidence Utilizing Large Groups." *Journal of Public Economics*, 1994, 54(1), 1-36.

Keser, Claudia and van Winden, Frans. "Conditional Cooperation and Voluntary Contributions to Public Goods." *Scandinavian Journal of Economics*, 2000, *102*(1), pp. 23-39.

Kreps, David; Milgrom, Paul; Roberts, John and Wilson, Robert. "Rational Cooperation in the Finitely Repeated Prisoners Dilemma." *Journal of Economic Theory*, 1982, 27:245-52.

Ledyard, John. "Public Goods: A Survey of Experimental Research," in John H. Kagel and Alvin E. Roth, eds., *Handbook of Experimental Economics*. Princeton: Princeton University Press, 1995, pp. 111-194.

McKelvey, Richard D. and Palfrey Thomas R. "An Experimental Study of the Centipede Game." *Econometrica,* 1992, *60*(4), 803-836.

Nikoforakis, Nikos. "Punishment and Counter-Punishment in Public Goods Games: Can we still govern ourselves." Economics Working Paper Archive at WUSTL.

Palfrey, Thomas R. and Prisbey, Jeffrey E. Anomalous behavior in public goods experiments: how much and why. *American Economic Review,* 1997, 87 (5), 829–846.

Shang, Jen and Croson, Rachel. "Field Experiments in Charitable Contribution: The Impact of Social Influence on the Voluntary Provision of Public Goods." 2005, *Mimeo*, Wharton School, University of Pennsylvania.

Sugden, Robert. "Reciprocity: the supply of public goods through voluntary contributions." *Economic Journal*, 1984 94, 772–787.

Appendix: Table A1. Maximum believable number of automata[4]

	S11	S12	S13	S14	S21	S22	S23	S24	S31	S32	S33	S34
1	45	45	50	45	45	45	45	50	50	45	45	45
2	50	50	0	45	46	45	45	0	48	47	45	46
3	30	45	0	45	0	45	0	0	45	46	45	45
4	35	5	50	30	20	0	20	40	45	50	45	46
5	20	45	0	30	0	0	0	0	45	45	45	45
6	25	5	0	15	5	0	0	45	45	47	47	45
7	5	0	5	2	0	10	0	50	45	50	45	46
8	15	30	0	45	0	0	0	0	45	48	45	45
9	25	45	0	40	10	0	0	0	45	47	45	45
10	30	0	0	43	0	10	0	3	0	45	45	0
11	30	45	50	45	46	45	45	50	45	50	45	45
12	45	45	50	45	0	50	45	45	45	45	45	50
13	25	50	50	45	5	0	0	47	45	46	45	46
14	45	50	50	0	10	5	0	45	45	45	45	45
15	20	50	25	40	7	5	0	0	45	47	45	46
16	30	45	50	40	0	0	0	25	45	46	45	45
17	35	45	45	25	5	0	0	48	45	45	45	46
18	25	50	50	0	8	0	0	0	50	46	45	45
19	20	45	0	45	0	0	0	0	45	50	45	0
20	15	0	0	20	0	0	0	0	0	0	0	10

	S41	S42	S43	S44	S51	S52	S53	S54	S61	S62	S63	S64
1	45	45	45	45	45	45	45	50	45	45	45	50
2	45	45	45	1	45	45	45	45	45	45	45	45
3	45	0	0	0	0	45	45	45	47	45	45	45
4	0	0	0	0	0	0	0	0	46	45	45	46
5	0	0	0	2	0	0	0	0	45	45	45	45
6	2	0	0	0	20	0	0	45	50	45	45	45
7	2	10	10	10	0	0	0	0	45	45	45	50
8	5	0	0	1	0	0	0	0	49	45	45	47
9	2	0	0	0	0	0	0	45	45	45	45	46
10	0	0	0	0	0	0	0	0	0	45	0	45
11	45	45	45	45	0	45	0	50	45	45	45	45
12	45	45	45	46	0	0	0	0	45	45	45	45
13	40	45	45	45	0	0	25	45	47	50	45	45
14	25	45	45	44	0	0	25	0	47	45	45	45
15	38	45	45	45	0	0	25	0	45	45	45	45
16	45	25	25	15	0	0	0	0	50	45	45	45
17	35	40	40	0	0	0	0	0	45	45	45	45
18	12	35	35	1	0	0	0	0	45	45	45	45
19	5	0	0	0	0	0	0	0	45	46	45	45
20	0	0	0	2	0	0	0	0	0	45	0	0

	S71	S72	S73	S74	S81	S82	S83	S84	S91	S92	S93	S94
1	45	45	46	50	46	45	45	45	45	45	45	50
2	45	45	49	45	45	45	45	45	50	45	45	45
3	45	45	45	45	45	45	45	45	45	45	50	45
4	50	45	47	45	47	45	45	45	45	0	45	47
5	45	45	48	45	45	50	45	45	0	0	0	45
6	45	45	50	45	45	46	50	45	0	0	0	0
7	45	45	46	45	45	46	47	45	0	0	0	40
8	45	45	47	45	50	45	45	45	0	0	2	0
9	45	45	46	45	0	45	0	45	0	25	0	0
10	45	0	47	45	0	0	5	0	0	0	0	0
11	45	45	48	45	45	45	50	45	45	45	50	45
12	45	45	45	45	46	45	48	45	45	45	45	45
13	46	45	46	45	45	45	45	45	45	50	47	45
14	45	45	47	45	45	0	45	45	46	45	45	45
15	45	45	50	45	0	0	0	0	46	45	46	45
16	47	45	45	45	0	0	0	0	47	45	48	45
17	45	45	45	45	0	40	0	0	48	45	45	45
18	45	45	46	45	0	0	0	0	47	45	48	45
19	45	45	45	45	0	0	0	0	48	45	48	45
20	0	0	45	45	0	0	0	0	0	45	0	0

50	3 automata
50	2 automata
50	1 automaton
50	0 automata

50	3 automata
50	2 automata
50	1 automaton
50	0 automata

[4] 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	35	35
	63%	97%	97%
# non-rejections	13	1	1
	36%	3%	3%

***α ≤ 1%; ** ≤ 5%; * ≤10%

Table A3. Period by period tests

Round	*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*

***$\alpha \leq 1\%$; ** $\leq 5\%$; * $\leq 10\%$

Instructions

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 ten independent rounds. You will be a member of the same 4 people group throughout all the ten periods. The results of each group are completely independent from others.

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 round will be determined in the following way:

Profit =	Profit from Private Account	+	Profit from Common Account
	50 Eurocent – what I assigned to the Common Account	+	(2 x Common Account)/4

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

9. 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.