Simulating the Effects of Sanctioning for the Emergence of Cooperation in a Public Goods Game

(Short Paper)

Ana L. C. Bazzan*
Inst. de Informática, UFRGS
Caixa Postal 15064
91501-970 P. Alegre, Brazil
bazzan@inf.ufrgs.br

Sílvio R. Dahmen
Inst. de Física, UFRGS
Caixa Postal 15051
90570-051 P. Alegre, Brazil
dahmen@if.ufrgs.br

Alexandre T. Baraviera
Inst. de Matemática, UFRGS
91509-900 P. Alegre, Brazil
baravi@mat.ufrgs.br

ABSTRACT

Several explanations have been proposed in order to explain why, in public goods games, cooperation does not collapse. In these games free-riders enjoy the benefits of other individuals who contribute in benefit of a community. In the present work we address a public goods game where individuals have the choice between contributing to a sanctioning institution and to a sanction-free one. In the former there is a possible sanction for those who do not contribute. Our results show that individuals who contribute to a sanctioning institution are better off after several repetitions of the game, despite the costs associated with sanctioning. This reproduces results found in experiments with human subjects, which point to advantages of sanctioning measures as a factor for the stabilization of cooperation.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence *Multiagent Systems, Coherence and Coordination, Intelligent Agents*

General Terms

Economics

Keywords

Multiagent Systems, Game Theory, Public Goods Game

1. INTRODUCTION

In AI in general and in multiagent systems in particular, decision-making by individuals (micro level) is highly affected by the group (macro level). This micro–macro effect is domain-dependent, has many facets, and is not well understood. In multiagent systems, not only the decision-making issue regarding individual agents is key to the performance of the system, but also the fundamental question that coordination among the various decision-makers is necessary. This issue arises in multiagent encounters because each agent faces the results of others' actions.

The agent-based approach has proven to be effective to analyze macro behavior arising from micro rules in classical scenarios of social sciences. Specifically, we are interested on agent-based simulation of public goods games. In these games, individuals incur a cost to create a benefit for a group. Just think about blood donation, recycling, and so on. They are problems because free-riders do enjoy the benefits created by the group without contributing themselves. Because free-riders are attracted by the benefits and proliferate, one may expect that eventually cooperation will collapse. However, human societies have somehow managed to solve this kind of problems. Therefore, there has been a great interest in public goods problems or dilemmas, and many researchers try to contribute to an understanding of the nature of these problems. The most popular explanations are based on signaling, reputation, and sanctions. We discuss the latter here; see [6] for an overview and further references.

Authors in [4] report laboratory experiments with *human* subjects playing the public goods game when two subjects can select among two institutions to contribute. Their work is the basis for ours and also supplies the experimental results we want to reproduce with a computational model based on learning agents. The main question the authors pose is the following: empirical evidence shows that the possibility of sanctions stabilize human cooperation while in their absence cooperation collapses. On the other hand the costs of punishment, for both those punished as well as punishers (for these incur into costs when punishing) might lead to the conclusion that people, if given a choice, would always opt for a sanction-free institution. The usual argument in favor of non punishing institutions is that these are the rational choice since payoffs are higher in a sanction-free institution as no extra costs of punishment are involved. However, recent experiments show that the proliferation of strong reciprocators, which are individuals willing to reward fair behavior and punish unfair ones even if they gain nothing with it, can be evolutionary stable [2].

What would happen however in a mobile environment, where newcomers from a noncooperative millieu are attracted by the higher payoffs? Will the number of strong reciprocators choosing sanctioning institutions grow enough to keep cooperation? This is investigated with the experiments, where people could move from one institution to the other.

Besides experimental results, there has been also studies on the public goods game which rely on simulation and/or analytical formulation. In [8], the authors describe the evolution of dynamics of the relationship among agents who are locally constrained, meaning that each agent has relationship with the two closest neighbors. The contribution by agents is modulated by a binary variable called

*Author partially supported by CNPq

Cite as: Simulating the Effects of Sanctioning for the Emergence of Cooperation in a Public Goods Game (Short Paper), Ana L. C. Bazzan, S. R. Dahmen and A. Baraviera, *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, Padgham, Parkes, Müller and Parsons (eds.), May, 12-16., 2008, Estoril, Portugal, pp. 1473-1476.

“motivation" which is based on the actions of their nearest neighbors. This was later extended in [7] where the authors have studied the changes in persistence when agents are no longer locally constrained; rather, they interact in a small-world scenario. Finally [1] is a step towards an investigation about what happens public goods games when agents are spatially distributed in grid-like structure and can select between *two* goods.

Here, we depart from the idea that evolution of the cooperation and the dynamics of the game are based on the motivation variable as proposed in [8]. Rather, we allow agents to learn whether or not to contribute, how much to contribute, and especially, to select one ”community“ or group to join, namely either the sanctioning or the sanction-free one. In what follows we use the term “institution” (instead of community) to follow the terminology used in [4].

The goal is to formulate a learning-based approach that is able to simulate human subjects playing the public goods game in a setting similar to [4]. This is important for two reasons. First, it is not obvious how and whether all human beings use learning-based processes while participating at such experiments. Some people do experimentations that are not grounded on any rational behavior, not even bounded rational. Thus, it is an interesting question whether an explicit learning-based approach can replicate the data observed in [4]. Second, if some theory can be formed on how to replicate those results, other closer settings can be simulated without performing the actual experiments, which are both time consuming and expensive. The idea is that once we have a well calibrated basic model to at least partially replicate the experiments, we can use it in more interesting settings (as discussed in the conclusion).

In the next subsection, we present the public goods game and our learning-based model. Section 4 discusses the scenarios and details of the simulation settings, as well as our preliminary results. Section 5 reports some preliminary conclusions.

2. BACKGROUND

2.1 Overview of the Public Goods Game

In its original formulation, this game deals with public spending on some work for the comunity: roads, libraries, etc. Players are offered the opportunity to contribute to a common pool; benefits (obtained from tolls, membership fees) are equally distributed among all participants irrespective of their contributions. Clearly it would be “fair” for people to pay the same quantity for those items. However individuals are different, as they have different social and economic conditions and different stances which means that some contribute less than others. This being common-knowledge, if one assumes each player as rational s/he would default and contribute nothing. However this is not what occurs in reality.

2.2 Experimental Economics

A classical concern in AI is the use of rational behavior and its relation to the prediction of patterns of behavior appropriate to goal achievement. A rational behavior emerges when the preferences of the participants of a system (regarding the several combinations of actions) can be described via an utility function, and when each of these participants can analyse the outcome of every possible action available and select the one which maximises its expected utility. Due to cognitive limitations of individuals, the actual action selection and the rational model do not match each other.

A more coherent explanation about how players select an action is to assume that they are able to extrapolate from what they have observed in past interactions. In general players can learn to select the best action. This can be done by analysing the payoff got from each rule used to select strategies in the past. According to [5], in a set of actions $A_{i,t} = (a_1, ..., a_m)$, a learning rule is a rule that specifies the probabilities $P_{i,t} = (p_1, ..., p_m)$ as a function of the payoffs obtained by playing those strategies in the past. This means that, in the future, each strategy is selected according to a probability based on the reward.

2.3 Experiments on Public Goods Game

As mentioned before, authors in [4] wanted to investigate what would happen in a mobile environment where newcomers from a given institution are attracted by the potentially higher payoffs of another.

In order to do this, authors have performed experiments with 84 individuals who played 30 rounds of the game. The game was composed of three stages: S0, S1, and S2. In the first one participants were told to choose between a sanctioning and a sanction-free institution. In stage S1, they were given 20 monetary units (MU) or tokens and could contribute between 0 and 20 to the common public good. The total amount contributed was collected, multiplied by a positive rate r ($r > 1$), and distributed equally among all participants, no matter how much each individual contributed. Those tokens not put in the public fund remained in the personal account of the individual.

In the third and final stage participants are informed about the contribution of each member of his institution. If they belong to a sanctioning institution they may then reward fair behavior or punish bad behavior. In the first case each rewarded MU goes to the rewardee and costs the payer that same amount. Those who punish spend q MU’s, but those punished have to pay 3 MU’s for each MU from the punisher. At the end people are informed about the performance of both institutions. Their results indicate that people prefer punishing institutions, even if their payoffs are not as high as in a non-sanctioning one.

The existence of strong reciprocators, which remain always small in number, allied to a conformist behavior of the majority lead to a high level of cooperation which tends to stabilize the system. People conform to the established norms even when these do not lead to maximum payoffs.

3. MODEL

In experimental economics, players have to make repeated choices. This indicates that a kind of learning or at least adaptation is used. In [4] it is stated that humans have an ability for social learning and this supports the competitive advantage of sanctioning institutions. This points to a learning process going on at least at the collective level.

The aim in our experiments with an artificial population of agents is to test whether a naïve learning approach can lead to a similar behavior. It is not clear how human subjects decide. One theory is that the decision is not purely rational (otherwise to default would be the outcome). Rather, people have an idea of fairness that is brought to the interactions. It is not easy to model this notion of fairness, especially because it changes from individual to individual. Also, in these experiments, participants may be bounded rational, obey social norms etc. By performing the experiments with human subjects, one can analyse real decision-making, and possibly compare this to what comes out from the theory of rational decision-making.

The public goods game has N individuals. As mentioned, in our case, instead of using an individual motivation to decide whether or not to contribute as in [8], agents keep a history of their past selections and decide what to do. This decision has three main steps: select which institution to join; decide how much to contribute (in $[0, 20]$); and if in the sanctioning institution, whether and how much to punish.

In order to give this model a realistic taste, we let agents interact and contribute a random quantity that depends on the agent's type or tags. After all contributions are turned public (but anonymous), each agent in a sanctioning institution can choose to punish somebody who has contributed an amount less than a threshold that the agent thinks is "compulsory". All these issues may affect how players evaluate the selection of actions the next time it should contribute. As in standard public goods games, the return per agent is a function of the average contribution. The profit has a deterministic part (*e.g.* a fixed interest rate r) plus a fluctuating contribution that comes from the willingness (or not) of agents to contribute. Thus, the amount which is contributed by each agent is implicitly based on the actions of other agents.

In our model we try to keep the scenario as close as possible to the one described in [4]. There are two choices of institutions to join in step S0: the sanctioning institution and the sanction-free institution. Then, in S1, our players have to select a quantity to contribute given that they have received 20 MU. If a high number of agents opt to contribute the total quantity of 20 tokens, then all receive a high return as well. On the other hand, many players can opt to contribute a low quantity. Their decisions probably depend on the institution they are in. A free rider would probably join a sanction-free institution in the hope to exploit high contributors there. However, these tend to migrate to a sanctioning institution where they can exercise their punishment power.

We have $N = 81$ players and not 84 because we use a square grid of 9x9. At the end of each round, every player gets a reward that is computed based on the total contributed. A negative feedback is possible and occurs when somebody is punished. Assuming that player i punishes player j with q tokens, because the punishment is 3 times q, it may happen that j's balance gets negative.

In what follows we describe our learning-based model. Whenever adequate we use the value of the parameters of the experimental setting described in [4]. The model is detailed in Algorithm 1. For sake of example the actual value for the parameters as used in the experiment are mentioned in the text below. However all quantities can be used.

Initially, agents are created with a tag or type (line 5). We use the two types explicitly mentioned in [4], namely free riders (FR) and high contributors (HC), as well as another two tags that they have not explicitly denominated: we call these T1 and T2. FR's who contribute between 0 and 5 tokens; HC's contribute 15 to 20. T1 and T2 hence fill the gap: T1 are those who contribute between 6 and 9, whereas T2 are those contributing 10 to 14 tokens. These types are used to decide which institution agents join in S0. If an agent is a FR then its probability to select the sanction-free institution is around 90% (varies from agent to agent) because it is expected that a FR has a low tendency to join a sanctioning institution. This probability decreases for T1, T2, and HC. Note that a HC has a non zero probability of selecting the sanction-free institution. Conversely the probability to select the sanctioning institution is high for HC and low for FR. In the beginning of the simulation, on average, 50% of the agents select each institution.

Agents then receive each 20 tokens (line 16) and decide how many tokens to contribute (line 17). This decision is also based on the type. As stated, FR contributes between 0 and 5, etc. The actual choice is random in those intervals. The total contributed is then multiplied by r ($r = 1.6$ in the experiments with human subjects and also here) and divided equally among the N participants. Thus the return per agent is this amount less the amount actually contributed (line 20) because it has been credited before (line 16).

Later agents receive other 20 tokens to decide to punish other agents. In the experiments with human subjects there was also the possibility of rewarding fair behavior by others in the sanctioning institution. However this was seldom used by the subjects thus we have not implemented it. The decision whether or not to punish is exercised only in the sanctioning institution of course. Here agents look at the internal threshold (to punish) they have and compare with their own contribution level. If the contribution was higher than the threshold, then it will punish somebody (line 27). Because the threshold to punish is high, normally only individuals labelled HC or T2 will punish. The actual quantity of sanction q is decided randomly (line 29). The agent to be punished is drawn probabilistically based on the inverse of its contribution level. Those punished have their balance decreased by $3 \times q$ (line 34). Agents then update the array of rewards they keep regarding each type. After they have experiences with all types, they start selecting their types probabilistically and play as described above.

Algorithm 1 Learning-based Behavior in the public goods game

1: INPUT: global variable t_{max} // max. time steps
2: **while** not t_{max} **do**
3:   **for** each agent i **do**
4:     **if** i has not experienced playing all types **then**
5:       set type randomly // FR, T1, T2, HC
6:     **else**
7:       set type probabilistically according to return of each type in the past (see Section 2.2 and [5])
8:     **end if**
9:   **end for**
10:   read global variable r // interest rate; $r = 1.6$ used
11:   **for** each agent i **do**
12:     $return \leftarrow 0$
13:     S0: choose institution according to type // probability of selecting SI: increases from FR to T1 to T2 to HC
14:   **end for**
15:   **for** each agent i **do**
16:     $return \leftarrow return + MU$ // receive MU to play; $MU = 20$ used
17:     S1: choose contribution amount c_i according to type and contribute
18:   **end for**
19:   **for** each agent i **do**
20:     $return \leftarrow return + \frac{(\sum_i c_i) * r}{N} - c_i$
21:   **end for**
22:   **for** each agent i **do**
23:     $return \leftarrow return + MU$ // receive MU to punish; $MU = 20$ used
24:   **end for**
25:   **for** each agent i **do**
26:     S2: if in SI, choose whether to punish somebody based on internal threshold:
27:     **if** $c_i \geq threshold\_to\_punish$ AND in SI **then**
28:       select whom to punish according to contribution amount of others in SI
29:       decide how much to punish (q_i) randomly $\in [0, MU]$
30:       $return \leftarrow return - q_i$
31:     **end if**
32:   **end for**
33:   **for** each agent i **do**
34:     if punished: $return \leftarrow return - 3 \times q_i$
35:     update reward array for given type
36:   **end for**
37: **end while**
38: END

4. SIMULATIONS

In order to evaluate how the simulations match the experiments, we measure the number of agents in both institutions, the main issue of the experiments reported in [4] (Figure 1). Since the approach is learning-based, the process takes more time than the 30 steps taken in the actual experiments. In our simulations it takes between hundreds and thousands of time steps in order to converge to a situation with a small number of agents in the sanction-free institution. However, the simulation cost is much lower than running the actual experiments, so that one can afford to run the simulation that long.

In the beginning agents keep experimenting playing the "roles" associated with the different types. There are shifts from one institution to the other (which also happen in experiments with human subjects) and on average 50% are in each one. This cannot be seen in Figure 1 because we plot one in each 10 actual choices. By the choice between 10 and 20, almost all agents have experience with all four types and start selecting their types probabilistically according to the return they had in the past. In Figure 1 we can see a clear, though slow, trend to select the sanctioning institution.

Hence the learning based model was able to reproduce the experimental data. It can be used as a starting point to test the effect of other configurations and other values for key parameters. For instance the conductor of the experiment could play with the model to try to predict what happens when r is higher or varies with time, sometimes being less than one. Or when there is a high rate of FR's. It is clear that such a model ceases to be valid if the conditions are too far from those used in the experiments with human subjects as for instance players do not remain anonymous, and so on.

5. CONCLUDING REMARKS AND FUTURE WORK

Traditional methods of analysis in many-actor systems (social sciences, economics) are being replaced by approaches able to explicitly deal with decision-making modulated by the interaction among individuals. This is important in many areas of AI such as multiagent systems and Alife. However the gap between individual rules and macro behavior is not very well studied as this problem has many facets and is domain dependent. Here we explore this problem in a public goods game, a metaphor for many interactions among cooperative and non-cooperative agents.

Experimental results show that the existence of strong reciprocators or high contributors are evolutionary stable and players, given the choice between a sanction-free and a sanctioning institution opt for the latter. The goal of this paper was to reproduce these findings in a theoretical model based on learning, where agents can choose to join a sanction-free institution or a sanctioning institution according to their types.

Our results on the analysis of the number of agents in both institutions show that as the game evolves, agents choose a sanctioning over a sanction-free institution. This is in accordance with the findings of [4] and give it support within that particular context.

It would be interesting to study a situation where agents can also change their type in a more deliberative way, as opposed to a reactive, probabilistic behavior. This deliberation could then be grounded on reputation an agent builds in its community. This has been also studied in the context of public goods games (e.g. [3]). In the setting explored here a reputation-based behavior could not be implemented because agents remained at least partially anonymous. However if we relax some constraints that were kept in order to comply with the experiments, reputation combined with punishment could be explored. Also, in order to follow the experiments with human subjects, we have assumed that the information players receive is complete. We did not consider loss of information, or people who would simply play randomly. Finally, since we run experiments with agents in a grid, we want to explore the fact that agents may have information only about their close neighbors (e.g. reputation). Therefore, a future direction is to play with these questions in the simulations.

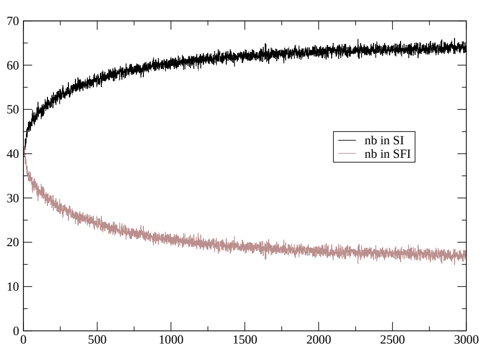

Figure 1: Number of Agents in Both Institutions

6. REFERENCES

[1] A. L. C. Bazzan, R. da Silva, S. R. Dahmen, and A. Baraviera. Emerging cooperation in a public goods game with competition. In *Proc. of the The 23rd Annual ACM Symposium on Applied Computing*, pages 8–12. ACM Press, March 2008.

[2] S. Bowles and H. Gintis. The evolution of strong reciprocity: cooperation in heterogeneous populations. *Theoretical Population Biology*, 65(1):17–28, February 2004.

[3] H. Brandt, C. Hauert, and K. Sigmund. Punishment and reputation in spatial public goods games. *Proc Biol Sci*, 270(1519):1099–1104, May 2003.

[4] O. Gürek, B. Irlenbusch, and B. Rockenbach. The competitive advantage of sanctioning institutions. *Science*, 312(5770):108–111, April 2006.

[5] C. B. Harley. Learning the evolutionary stable strategy. *Journal of Theoretical Biology*, 89:611–631, 1981.

[6] J. Henrich. Cooperation, punishment, and the evolution of human institutions. *Science*, 312(5770):60–61, April 2006.

[7] R. Silva, A. Baraviera, S. R. Dahmen, and A. L. C. Bazzan. Dynamics of a public investment game: from nearest-neighbor lattices to small-world networks. In C. Bruun, editor, *Advances in Artificial Economics, The Economy as a Complex Dynamic System*, number 584 in Lecture Notes in Economics and Mathematical Systems, chapter 16, pages 211–233. Springer, 2006.

[8] R. Silva, A. L. C. Bazzan, A. Baraviera, and S. R. Dahmen. Emerging collective behavior in a simple artificial financial market. In F. Dignum, V. Dignum, S. Koenig, S. Kraus, M. P. Singh, and M. Wooldridge, editors, *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, volume 1, pages 313–319. ACM, July 2005.