The Effect of Agreements in a Game with Multiple Strategies for Cooperation

Pedro Mariano and Luís Correia

{plm,lc}@di.fct.unl.pt, Informatics Department, New University of Lisbon
Quinta da Torre, 2829 - 516 CAPARICA, Portugal

Abstract

We present a model suited not only for the study of evolution of cooperation but also to study behaviours such as treason and exploitation. This game has multiple Pareto Optimal solutions, which causes shifts in the agent strategies that we can interpret as either treason or exploitation. This requires some form of coordination between agents to avoid penalising behaviours. We present results of our game with and without an agreement model. We show that our game provides rich evolutionary dynamics.

Introduction

The Iterated Prisoner's Dilemma has been used throughout the literature as a model of cooperation (Brembs 1996; Beaufils, Delahaye, & Mathieu 1997; Nowak, Bonhoeffer, & May 1994; Yamaguchi, Maruyama, & Hoshino 2000). In our view, treason, manipulation, and exploitation are not well modelled. Same authors have made extensions to the game in order to cope with player selection (Stanley, Ashlock, & Smucker 1995). Others have used emotions toward other players as a way to permit defection in the game to help partners (Bazzan, Bordini, & Campbell 1998). Manipulation is not well addressed in this game, while some authors engineer communication protocols for societies that clarify any information sent by lying agents (Sandholm 1999).

The game model we present has multiple Pareto Optimal strategies to choose from (where an agent cannot improve its outcome without impairing its opponent. This requires some coordination or a treaty (either explicitly or as a consolidated behaviour) among agents playing the game. The game also has the same characteristics of IPD, namely a set of play moves that are beneficial for both players (thus deemed cooperative) and another that is detrimental and used to penalise players. Multiple strategies with equal outcomes are well suited to study exploitation and treason. The coordination aspect was treated in (Mariano & Correia 2002). In this paper, we report the results related to treason.

Game Description

This game has been introduced in (Mariano & Correia 2002). Here we will only present the essential definitions to grasp the results we present. The game is inspired in resource sharing. In our case, there are two agents and one resource. Only one agent can have the resource in a single time unit. The agent with the resource can perform either the $none_g$ or *give* action. The agent without the resource can perform either the $none_t$ or *take* action. The *give* action has a bonus associated to it, b_g. The *take* action incurs losses in both the performer and the subject, c_{pt} and c_{st}. Possession or not of the resource is characterised by the parameters w_g and w_l.

In a single iteration, the wealth gain matrix for the player with resource is:

$$\begin{bmatrix} w_g & -w_l - c_{st} \\ -w_l + b_g & -w_l + b_g - c_{st} \end{bmatrix} \tag{1}$$

and for the player without the resource is:

$$\begin{bmatrix} -w_l & w_g - c_{pt} \\ w_g & w_g - c_{pt} \end{bmatrix} \tag{2}$$

When the game is played only once, the Nash equilibrium of the game is as follows: the agent with resource should play the *give* action; the agent without the resource should play the *take* action. This result was achieved using minmax analysis.

When the game is played several times (t time units), three cases are important in this game.

- The agent with the resource does not give it, and the other agent does not take it. Their wealth outcomes are the following:

$$t \times w_g \qquad \text{with resource} \tag{3}$$
$$-t \times w_l \qquad \text{without resource} \tag{4}$$

- When both agents give away the resource after t_{pg} time units (with $t \gg t_{pg}$), their wealth outcome is approximately:

$$\frac{t}{2}\left(w_g - w_l + \frac{b_g}{t_{pg}}\right) \tag{5}$$

- When both agents take the resource after t_{pt} time units without it (again $t \gg t_{pt}$), their wealth outcome is:

$$\frac{t}{2}\left(w_g - w_l - \frac{c_{pt}+c_{st}}{t_{pt}}\right) \tag{6}$$

When b_g is zero, expression 3 is greater than 5 which is greater than 6. These cases are extreme ones but we have presented them to show that: an agent is better of if it does not share the resource; however, if the opponent agent takes the resource, they get the worst result. Moreover, when b_g is zero there are multiple strategies with the same wealth outcome. Both agents can choose any t_{pg} value. We consider them as Pareto Optimal. This suggests some agreement between agents should be met. Consequently, it permits the study of treason and exploitation.

Contribution We expect that at this point, the analysis performed could shed some light on the game that we are proposing. Although some resemblance with IPD and the lumberjack dilemma, this game requires coordination in order to both agents collecting equal wealth gains. While in IPD agents must play cooperate to get equal benefits, in our game agents need to coordinate their actions because there are multiple Pareto Optimal strategies. In IPD any play of defect other than cooperate leads to wealth differences. In our game, a deviation from a balanced combination of actions, can lead to punishment if agents do wish so, or to a new combination of actions that guarantees equal wealth gains (a new Pareto Optimal strategy).

As in the lumberjack dilemma, agents share a resource from where they collect wealth. In lumberjack dilemma and other tragedy of the commons games, the resource is a distinct object with a dynamic of its own (it grows or evolves, its value changes in response to agent actions). In our game, the resource is an object that agents must possess and, at a single time only one agent can collect wealth from it. In order to other agents being able to collect wealth, some action must be performed to change the owner of the resource. It can change due to an altruistic action (*give*) or through a punitive and costly action (*take*).

The simulation

We will now describe how we implemented the game introduced in the previous section. The topics are: what is the agent strategy, the agreement model used, the evolutionary framework used, the parameter values, and the lattice type environment used.

Since this game has multiple Pareto Optimal strategies, we introduce an agreement model that allows two agents playing the game to select one Pareto Optimal strategy. The goal of the simulations reported here is to compare the evolutionary dynamics between populations of agents that are able to establish an agreement from those that are not capable.

Agent strategy The agent strategy uses a probabilistic mechanism to perform the actions. For each action, the agent waits a number of ticks until it performs the action with some probability. Therefore, we can describe a single strategy by tuple (p_g, p_t, t_g, t_t), in which p_g is the probability of an agent doing the *give* action after it holds the resource t_g ticks, and conversely, p_t is the probability of an agent doing the *take* action after it stayed t_t ticks without the resource – two probabilities and two time intervals.

Agreement model Each agent has a preferred period to hold the resource, t_h^A. When two such agents meet to play the game, they must first negotiate the period they are going to hold the resource. There are two other parameters: one describes how much an agent will exploit the agreement, t_e^A, that is, after this time has elapsed it will definitely give away the resource; the other describes how much time an agent tolerates that it has not received the resource, t_t^A, that is, after this time has elapsed it will definitely take the resource. In the intervals $[t_h^A, t_h^A + t_e^A]$ and $[t_h^A, t_h^A + t_t^A]$ the agent actions are determined by its underlying strategy. This agreement model allows us to use the probabilistic strategy described previously, or another strategy such as an finite state automata.

When two agents meet to play the game, they have to decide the period to hold the resource. Two agents must negotiate between their parameters $t_{h\,1}^A$ and $t_{h\,2}^A$. In the results reported here, we use a simple rule: agents select the minimum value among them. We could use another rule, or some bargaining protocol. However, when the *give* bonus is zero, the multiple Pareto Optimal strategies are equivalent, since they all yield the same wealth outcome. Therefore, a simple rule suffices to reach an agreement.

Our model does not make any assumption about the truthfulness of the agent. An agent may announce some period to hold the resource, and during the course of the game use a completely different value. In this case, the agent is subject to its partner's tolerance time interval, t_t^A. The results we present are from simulations with truthful agents.

In our model agents always reach an agreement. We could for instance add another parameter that defines the probability of an agent making an agreement. If there were no agreement, agents would simply revert to their underlying strategy. Notice that, with agreement, this strategy is only applied in the exploitation and tolerance intervals.

Evolutionary Algorithm The agent chromosome is composed of the 4 strategy parameters and 3 agreement model parameters. The initial population parameters are taken from a random uniform distribution: probability parameters ranging from 0 to 1 and time parameters from 0 to 15. Only the mutation operator was used: gaussian noise addition. A population of size 100 is distributed in a square lattice. At every generation, agents choose N opponents from their 4-neighbour to play a match. Each match consists of 3 games. Game length varies uniformly between 100 and 130 time units. Since an agent chooses and may be chosen, every agent plays in average $N \times 3 \times 2$ games. Agent selection probability is proportional to agent's wealth (roulette wheel). Each simulation run took 1000 generations.

The choice of the game length values and of the time parameters values must comply with the fact that $t \gg t_{pg}$ and $t \gg t_{pt}$. The lowest value of the game length should be much greater than the highest value for the time parameter so that the approximations of expressions 5 and 6 are valid and there is sufficient time for the resource to be shared equally between the agents.

Parameter Values Some of the game parameters and evolutionary parameters were varied. For each parameter set, 10 simulation runs were performed. Table 1 shows the different parameters values.

parameter	value
w_g	5
w_l	0, 1, 2
c_{pg}	0
c_{sg}	0
c_{pt}	2, 4, 6, 8
c_{st}	10, 20, 30
N	1, 2, 3, 4
p_{mut}	0.05

Table 1: Parameters values used in the simulations.

Results

Two main setups were prepared: one with the agreement model and the other without it. The goal is to study the effect of the agreement on what kind of evolutionary dynamics we obtain. The results without agreement model were taken from a previous work (Mariano & Correia 2002). In order to compare results, we measured the number of actions performed and the wealth accumulated for each main setup.

Agreement

In the simulations without the agreement model we obtain different strategies. There are populations where the *give* action is performed very often, and others where the *take* action is not. Typically, the higher the *take* action costs, the higher is the average number of *give* actions performed. The agreement model favours the decrease of *take* action performed. There are two cases for $c_{st} = 30$, with $c_{pt} = 2$ and $c_{pt} = 4$ where instead an increase was observed, as can be seen in table 2.

c_{pt}	c_{st} 10	20	30
2	20.9	4.5	2.6
4	19.1	5.5	2.0
6	16.7	6.1	2.1
8	14.8	6.8	2.6

(a) Without the agreement model.

c_{pt}	c_{st} 10	20	30
2	4.8	4.5	4.5
4	4.8	3.0	2.1
6	5.2	2.9	1.6
8	5.4	2.8	1.3

(b) With the agreement model.

Table 2: Average *take* action percentage obtained in simulations.

c_{pt}	c_{st} 10	20	30
2	−4.9	9.8	11.2
4	−6.5	6.5	12.8
6	−6.4	3.9	12.2
8	−6.5	1.0	9.9

(a) Without the agreement model.

c_{pt}	c_{st} 10	20	30
2	13.8	9.7	5.5
4	12.8	12.4	12.4
6	11.3	12.0	13.8
8	9.9	11.8	14.7

(b) With the agreement model.

Table 3: Average wealth obtained in simulations.

Average wealth outcome improves with the agreement model. Without it, when c_{st} is low, average wealth is negative. A high value favours the appearance of agent strategies based on the *give* action, and as a consequence, the average wealth increases (see table 3). The agreement model improves the average wealth specially when the c_{st} is lower. There are two cases for $c_{st} = 30$, with $c_{pt} = 2$ and $c_{pt} = 4$ where we obtained a poorer result. We also noticed that results are quite independent of parameter variations as can be seen in table 3.

Evolutionary dynamics

The simulations obtained without the agreement model provided more complex evolutionary dynamics. Different probability strategies appear. When the c_{pt} parameter is high, the population is composed of an agent strategy with the following characteristics: high probability to give and take the resource; the time to perform the *give* action is lower than the one to perform the *take* action. This strategy is cooperative as the agents give the resource frequently. In addition, it prevents exploitation since there is a small time window before an agent takes the resource. When the *take* action costs are lower, the agent strategy most common is characterised by: low probability to give the resource; long time interval to give the resource; short time interval before taking the resource.

The simulations with the agreement model showed some examples of treason behaviour. They are characterised by high values for the t_e^A parameter. The lower is the c_{st}, (cost of the subject of the *take* action), the higher is the tendency to treason. However, tolerance to this behaviour depends on the c_{pt} parameter. The costlier it is to penalise a traitor, the higher is the tolerance level (t_t^A parameter).

In both simulation setups, the population goes through phases of cooperation, switching between Pareto Optimal strategies, and phases characterised by negative wealth outcome.

In other simulation results (Mariano & Correia 2002) we varied the *give* action costs. In this case we verified that when there is a bonus to give the resource, there is a cooperative agent strategy with the previous characteristics. We observed that in this case the time to perform the *give* action is generally 1 to 2 time units.

Conclusions and future work

We have presented a game model that exhibits multiple Pareto Optimal strategies. In order to agents be able to obtain above average wealth outcome, they have to cooperate. However, multiple Pareto Optimal strategies require some coordination or an agreement between agents. Even in the absence of such agreement, we obtained such cooperative strategies. This is mainly due to the parameter values that favoured this outcome. When we added an agreement model, results improved, since the average wealth did not show a strong relationship to game parameters.

The agreement model favoured the appearance of cooperative strategies. However, treason behaviour and its counter measures were observed. Their rate depended on the *take* action costs. The possibility not to enter an agreement should be assessed, as we have interactions between agents that do not establish any agreement, and others that make agreements.

The existence of multiple Pareto Optimal solutions leads to changes in the agent population. These changes are caused by new strategies that exploit or break the agreement. The population can stabilise into a new Pareto strategy. On the other hand, it can go through a phase of penalising strategies. Despite the evolutionary dynamics presented here, further research into the population diversity should be conducted.

The agreement model negotiation may be kept simple. A simple solution suffices to establish the agreement terms.

References

Bazzan, A.; Bordini, R.; and Campbell, J. 1998. Moral sentiments in multi-agent systems. In Müller, J. P.; Singh, M. P.; and Rao, A. S., eds., *Intelligent Agents V: Agent Theories, Architectures, and Languages*, volume 1555 of *Lecture Notes in Computer Science*, 113–132. Springer-Verlag.

Beaufils, B.; Delahaye, J.-P.; and Mathieu, P. 1997. Our meeting with gradual: A good strategy for the iterated prisoner's dilemma. In Langton, C. G., and Shimohara, K., eds., *Artificial Live V: Proceedins of the Fifth International Workshop on the Synthesis and Simulation of Living Systems*, 202–209. MIT Press.

Bedau, M. A.; McCaskill, J. S.; Packard, N. H.; and Rasmussen, S., eds. 2000. *Artificial Live VII: Proceedings of the Seventh International Conference on Artificial Life*. MIT Press.

Brembs, B. 1996. Chaos, cheating and cooperation: potential solutions to the prisoner's dilemma. *Oikos* 76(1):14–24.

Mariano, P., and Correia, L. 2002. A game to study coordination and cooperation. In *5th Workshop on Deception, Fraud and Trust in Agent Societies*, 101–112.

Nowak, M. A.; Bonhoeffer, S.; and May, R. M. 1994. Spatial games and the maintenance of cooperation. *Proceedings of the National Academy of Sciences* 91:4877–4881.

Sandholm, T. W. 1999. Distributed rational decision making. In Weiss, G., ed., *Multiagent systems: a modern approach to distributed artificial intelligence*. MIT Press. chapter 5, 201–258.

Stanley, E. A.; Ashlock, D.; and Smucker, M. D. 1995. Iterated prisioner's dilemma with choice and refusal of partners: Evolutionary results. In Morán, F.; Moreno, A.; Merelo, J. J.; and Chacón, P., eds., *Advances in Artificial Life*, volume 929 of *Lecture Notes in Computer Science*, 490–502. Springer-Verlag.

Yamaguchi, Y.; Maruyama, T.; and Hoshino, T. 2000. A co-evolution model of scores and strategies in IPD games: toward the understanding of the emergence of the social morals. In Bedau et al. (2000), 362–366.