Trust in an N-Player Iterated Prisoner's Dilemma

Andreas Birk
Vrije Universiteit Brussel, Artificial Intelligence Laboratory,
Pleinlaan 2, 10G725, 1050 Brussels, Belgium
cyrano@arti.vub.ac.be

Abstract

The iterated Prisoner's Dilemma (iPD) is a standard problem for the evolution of cooperation. Published work so far dealt mainly with the 2-player case, which is neither very realistic nor very relevant for applications, as societies, in a general sense, usually are made of more than two agents. We present novel results where cooperation co-evolves with trust in an N-player iPD. In doing so, trust is expressed by group-formation, i.e., agents evolve preferences for other agents with certain markers.

1 Introduction

Since [Axe84], it is well-known that simple strategies like the famous Tit-For-Tat can be powerful means to establish cooperation among selfish agents in situations which can be modeled by the standard prisoner's dilemma [Col82]. But in respect to research on trust, the huge and partially very impressive body of work on the evolution of cooperation based on this framework has several drawbacks.

First, the prisoner's dilemma in its standard form is only a two-player game and it allows only binary decisions whether to cooperate or to defect. Both aspects, involving only two agents and the lack of continuity, are not very realistic and have been criticized many times before.

Second, building trust involves a non-rational component in the sense that decisions on how to deal with an other agent are not only based on previous interactions with exactly that agent, but also on other, presumingly subjective criteria. For example for humans, these criteria include the outer appearance, recommendations from others, and so on.

Both drawbacks are tackled here. We present a novel generalized version of the Prisoner's Dilemma as N-player game with continuous degrees of cooperation. In addition, we describe how trust can boost the evolution of cooperation in form of a co-evolution of group-formation and strategies.

2 A Continuous N-player Prisoner's Dilemma

Recently, Roberts and Sherratt published in "Nature" [RS98] results on the evolution of cooperation in an extension of the standard prisoner's dilemma (PD) to continuous degrees of cooperation. Partially inspired by this work and partially based on experiments with heterogeneous robots in an artificial ecosystem [BB98, Ste94], following further extension to an N-player case was developed.

In this artificial ecosystem, simple mobile robots, the so-called moles[1], can autonomously re-charge their batteries, thus staying operational over extended periods in time. As illustrated in figure 1, a so-called head can track mobile robots and it can perceive so-called pitfalls which are kind of inverse charging stations where the batteries of the moles are partially dis-charged via a resistor. When a mobile robot approaches a pitfall, which it cannot distinguish from a charging station, the head can warn the mobile robot. The mobile robot in exchange can share the benefit of the saved energy with the head.

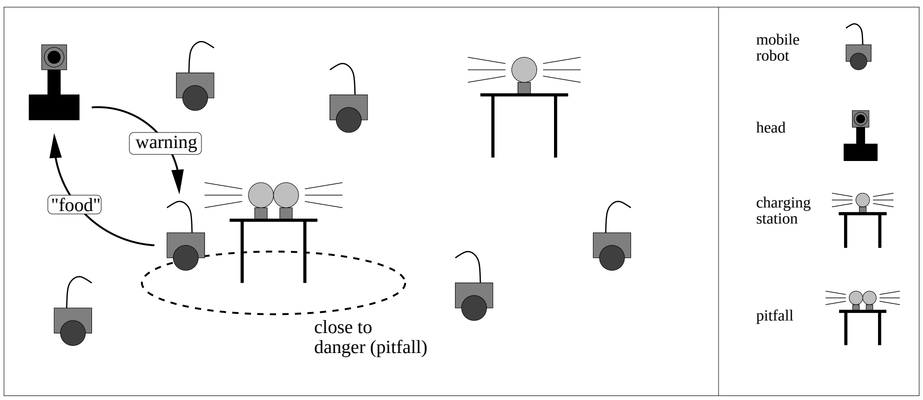

Figure 1: The extended artificial ecosystem of the VUB AI-lab, including a head and several moles. So-called pitfalls in the form of inverse charging-stations can suck energy out of a mole. Unlike moles, a head can distinguish pitfalls and charging-stations, and it can warn a mole when being close to a pitfall. The mole in return feeds a part of its benefit in form of energy to the head.

Let there be N moles and one head. Each mole $m_i (1 \leq i \leq N)$ has a gain G_i based on the avoidance of pitfalls due to warnings of the head. This gain only depends on the so-called headsight $hs \in [0, 1]$, i.e., the percentage with which the head perceives dangerous situations. Concretely, the gain is the headsight times one hundred energy-units (EU): $G_i = hs \cdot 100EU$.

Furthermore, in the beginning of each time step t, each mole m_i invests up to

[1] The names for this robot-"species" should not be taken literally. The name "moles" attributes to the fact, that these robots have no on-board vision. We use these names for reasons of convenience only.

seventy-five energy units to feed the head. This investment I_i is proportional to the continuous cooperation level $co_i \in [0, 1]$ of m_i: $I_i = co_i \cdot 75EU$.

The headsight hs depends on the amount of food the head receives from the moles, i.e., the head is completely fed when it receives the 75 energy units from every mole. Concretely, we define the headsight hs as the averaged sum of cooperation levels in time step t: $hs = \sum_{1 \leq i \leq N} co_i/N$.

The pay-off po_i for a mole m_i is the difference between gain and investment:

$$po_i = G_i - I_i = \sum_{1 \leq j \leq N} co_j/N \cdot 100EU - co_i \cdot 75EU \qquad (1)$$

So, a dilemma for the moles arises. On the one hand, it is in the interest of each mole that the head is well fed. On the other hand, there is the temptation to leave the task of actual feeding to others, as the head does not react to the behavior of a single mole, i.e., it does not punish a mole when it does not donate energy.

Note, that the pay-off for a mole depends on its own cooperation level and all of the cooperation levels of all other moles. Let $\bar{co}$ denote the average cooperation level of the group, i.e., $\bar{co} = \sum_{1 \leq i \leq N} co_i/N$. The pay-off for a mole m_i can directly be computed from co_i and $\bar{co}$. Namely, the pay-off function $f_p : [0, 1] \times [0, 1] \to I\!R$ is

$$f_p(co_i, \bar{co}) = co_i \cdot -75EU + \bar{co} \cdot 100EU \qquad (2)$$

Based on this, we can extend the terminology for pay-off values in the standard prisoner's dilemma, with pay-off types for cooperation (C), punishment (P), temptation (T), and sucking (S), as follows:

- Full cooperation as all fully invest: $C_{all} = f_p(1.0, 1.0) = 25.0$
- All punished as nobody invests: $P_{all} = f_p(0.0, 0.0) = 0.0$
- Maximum temptation: $T_{max} = f_p(0.0, \frac{N-1}{N}) \geq 50.0$
- Maximum sucking: $S_{max} = f_p(0.0, \frac{1}{N}) \leq -25.0$

For $co, \bar{co} \neq 0.0, 1.0$, we get the following additional types of pay-offs, the so-called partial temptation, the weak cooperation, the single punishment, and the partial sucking. They are not constants (for a fixed N) like the previous ones, but actual functions in $(co, \bar{co})$. Concretely, they are sub-functions of $f_p(co, \bar{co})$, operating on sub-spaces defined by relations of co in respect to $\bar{co}$ (table 1).

Note that for a fixed average cooperation level $\bar{co}$ and two individual cooperation levels $co' > co''$, it always holds that $f_p(co', \bar{co}) < f_p(co'', \bar{co})$. Therefore it holds for an individual player in a single game that:

- The partial temptation pays always better than weak cooperation.

Table 1: Additional pay-off types in the CN-PD

$co < \bar{co}$	$\bar{co} \leq co < 4/3 \cdot \bar{co}$	$co = 4/3 \cdot \bar{co}$	$co > 4/3 \cdot \bar{co}$
$T_{partial}(co, \bar{co})$	$C_{weak}(co, \bar{co})$	$P_{single}(co, \bar{co})$	$S_{partial}(co, \bar{co})$
$\in]0, 100[$	$\in]0, 25[$	$= 0$	$\in] - 100, 0[$
partial temptation	weak cooperation	single punish	partial sucking

- The partial temptation increases with decreasing individual cooperation.
- The absolute value of partial sucking increases with increasing individual cooperation.

This can also be stated as:

$$\begin{aligned} T_{max} > T_{partial}(.) > C_{all} > C_{weak}(.) > 0.0 \\ P_{single}(.) = P_{all} = 0.0 \\ S_{max} < S_{partial}(.) < 0.0 \end{aligned} \tag{3}$$

The equation 3 illustrates the motivation for the names of the different types of pay-off. The attribute *max* for temptation T and sucking S indicates that these are the maximum absolute values. The *partial* accordingly indicates that these values are only partially reached through the related T or S functions. The attribute *weak* for the cooperation function C relates to the fact that though the player receives a positive pay-off, it is less than in the maximum cooperation case where *all* players fully cooperate. When no player invests, *all* are punished with a Zero pay-off. Whereas in the *single* case, at least the individual player we are looking at gets punished with a Zero pay-off, other players can receive all types of pay-off.

3 Evolution of Cooperation, Boosted by Trust

For N players, no positive results for cooperation in the iterated PD are published up to our knowledge. In [Bir99], where also the Continuous N-player Prisoner's Dilemma (CN-PD) is described in more detail, we present a novel strategy, the so-called Justified-Snobism (JS). JS cooperates slightly more than the average cooperation level of the group of N players if a non-negative pay-off was achieved in the previous iteration, and it cooperates exactly at the previous average cooperation level of the group otherwise. So, JS tries to be slightly more cooperative than the average. This leads to the name for this strategy as the snobbish belief to be "better" (in terms of altruism) than the average of the group is somehow justified for players which use this strategy.

It can be shown that JS is a successful strategy for the CN-PD and especially that JS is evolutionary stable. Here, we show that the evolution of cooperation can be boosted by co-evolving strategies and trust. As mentioned in the introduction, trust is seen as a kind of subjective criterion guiding the interaction with others. More concretely, a strategy is based on more or less objective measures on the performance of other agents, namely their cooperation-level in previous iterations. Trust in contrast is based on secondary, derived measures, here the "outer appearance" of an agent in form of a marker.

The main idea is as follows. In the beginning, agents are randomly marked with labels from a finite set $L = \{l_1, ..., l_k\}$. Then, they evolve strategies in an iterated CN-PD, also based on a random initialization, i.e., there is no link between markers and strategies in the beginning. At the same time, agents evolve preferences to play CN-PD games with agents with certain markers, i.e., they start to trust agents with a certain label. Note, that this trust can not be justified by any rational means, at least in the beginning of the iterated games, as there is no link between a certain label and a certain strategy.

4 Results

In the experiment described here, the population *pop* of agents is very large, namely $N_{pop} = 1000$. Within this population, agents are grouped together to play an iterated CN-PD with 20 players per game. Trust is a basis for this group-formation as follows. The trust function $f_t : L \to [0, 1]$ of an agent a_i maps labels to a weight w, such that $f_t(l_j) = w$ represents a_i's preference to interact with an agent with label l_j. If w is high, i.e., close to One, a_i prefers to interact with agents with label l_j, or it simply trusts them. If w is low, i.e., close to Zero, a_i prefers not to interact with agents with label l_j, or it simply does not trust them.

Trust influences the group-formation as the players for each CN-PD are selected from the population with roulette-wheel selection with the weights as bias. This means that agents that trust the same types of other agents tend to be more likely to be put together in one group.

The co-evolution of trust and strategies proceeds in rounds. Within a round r, k iterations of the CN-PD are played, with $k = 50$ for the experiments reported here. The groups for each game are formed as described above at the beginning of each round and kept fixed within a round. Each agent a_i has one of the following strategies to determine its cooperation level co_i:

Follow-the-masses (FTM) : match the average cooperation level from the previous iteration, i.e., $co_i(t) = \bar{co}(t-1)$

Hide-in-the-masses (HIM) : subtract a small constant c from the average cooperation level, i.e., $co_i(t) = \bar{co}(t-1) - c$

Occasional-short-changed-JS (OSC-JS) : a slight variation of JS, where occasionally the small constant c is subtracted from the JS-investment

Occasional-cheating-JS (OC-JS) : an other slight variation of JS, where occasionally nothing is invested

Challenge-the-masses (CTM) : Zero cooperation when the previous average cooperation is below one's one cooperation level, a constant cooperation level c' otherwise, i.e.,

- $co_i(t-1) \geq \bar{co} : co_i(t) = c'$
- $co_i(t-1) < \bar{co} : co_i(t) = 0$

Non-altruism (NA) : always completely defect, i.e., $co_i(t) = 0$

Anything-will-do (AWD) : always cooperate at a fixed level, i.e., $co_i(t) = c'$

The fitness $f(a_i)$ of agent a_i is determined by the running average of its pay-offs. Reproduction of agents is proportional to their fitness as roulette-wheel selection keeps the population size fixed to 1000.

The trust of an agent a_i is updated based on the (very limited) experiences with other agents with a certain label. Concretely, the weight of trusting agents with label l_j is updated in each game proportionally to the pay-off and the number of agents with that label in the group. This means that when many agents with label l_j are in the group and the pay-off is high, then the agent a_i increases its trust in agents with that label l_j.

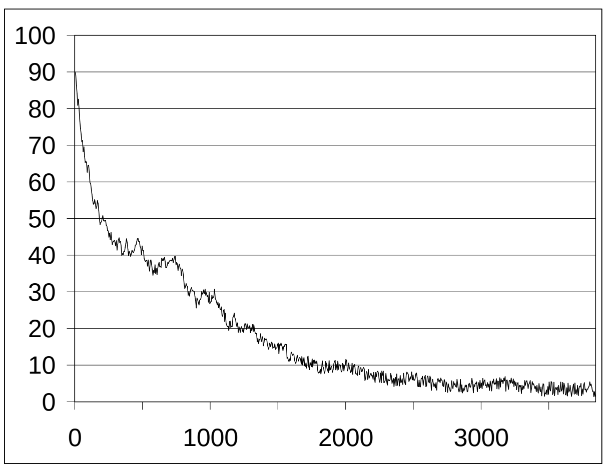

Figure 2: The percentage of agents in the population which can not "decide" which types of agents they should trust. In the beginning of the run, this percentage is high as most agents change their preference in every time step, more or less randomly guessing. After a while, fixed preferences evolve.

Note, that this trust is really subjective in some sense. First, it is based on very limited data, i.e., there are many agents with label l_j in the population, but the agent a_i builds up some belief by interacting with just a few of them. Second, within the group to which agent a_i belongs to at the moment, there are (most probably) many different agents in respect to labels. The update of trust does not distinguish between those labels, though different agents, and accordingly labels, do (most probably) contribute very differently to the pay-off that a_i receives.

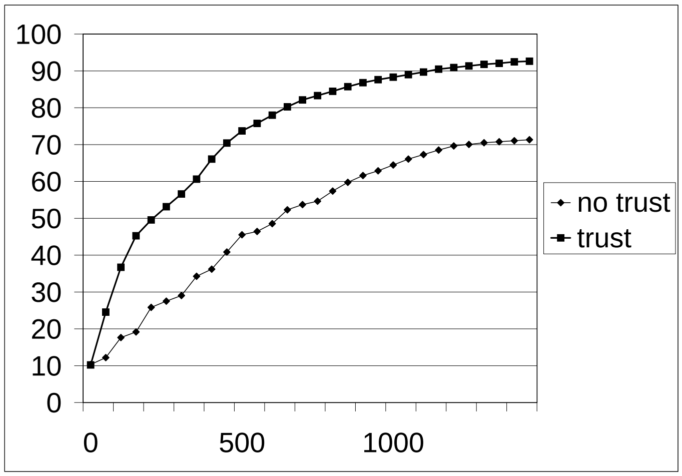

Figure 3: The general cooperation levels, averaged from respectively fifty runs with and without a co-evolution of trust. When trust is activated, a higher general cooperation level is reached much faster than without trust.

Nevertheless, a stable relation of trust emerges. This means, the agents evolve fixed preferences for interacting with agents with a certain label. Figure 2 shows the percentage of agents which can not "decide" which label they should trust. More precisely, the graph shows the percentage of agents where the preference of a particular agent for a certain label in the current step is different from its preference in the previous step. In the beginning of the run, the percentage of "undecided" agents is very high, i.e., the agents are more or less randomly guessing in each step which label to trust. After a while, this indecision is dropping to almost Zero, i.e., the agents evolve fixed preferences for certain labels.

Especially, the evolution of trust boosts the evolution of cooperation in these experiments. Figure 3 shows the development of the general cooperation level for both cases, namely respectively fifty averaged runs with and without a co-evolution of trust. When the co-evolution of trust is activated, a higher general level of cooperation is reached much faster than without an evolution of trust.

5 Conclusion

We presented an extension of the iterated prisoner's dilemma, a well-known framework for research on the evolution of cooperation, to the particular problems involved with trust.

First, a general drawback of the prisoner's dilemma is tackled, namely its limitation to two players and the limitation to the binary decision whether to cooperate or not. Concretely, an N-player prisoner's dilemma with continuous levels of cooperation is described.

Second, building trust involves a somewhat subjective criterion on whether to interact with an other agent or not. Therefore, we present the idea to base trust on (in the beginning) meaningless markers, i.e., the so-to-say outer appearance of an agent. We present experiments where trust, as preference to interact with agents with certain labels, gets updated through limited interactions with a limited number of agents. Nevertheless, stable relations of trust can emerge. Especially, the established trust can significantly boost the evolution of cooperation.

References

[Axe84] R. Axelrod. *The Evolution of Cooperation.* Basic Books, 1984.

[BB98] Andreas Birk and Tony Belpaeme. A multi-agent-system based on heterogeneous robots. In *Collective Robotics Workshop 98.* Springer LNAI, 1998.

[Bir99] Andreas Birk. Evolution of continuous degrees of cooperation in an n-player iterated prisoner's dilemma. Technical Report under review, Vrije Universiteit Brussel, AI-Laboratory, 1999.

[Col82] A.M. Colman. *Game Theory and Experimental Games.* Pergamon Press, 1982.

[RS98] Gilbert Roberts and Thomas N. Sherratt. Development of cooperative relationships through increasing investment. *Nature*, 394 (July):175–179, 1998.

[Ste94] Luc Steels. A case study in the behavior-oriented design of autonomous agents. In Dave Cliff, Philip Husbands, Jean-Arcady Meyer, and Stewart W. Wilson, editors, *From Animals to Animats 3. Proc. of the Third International Conference on Simulation of Adaptive Behavior.* The MIT Press/Bradford Books, Cambridge, 1994.