Cooperation is not always so simple to learn…

M. Mailliard, F. Amblard, C. Sibertin-Blanc, P.Roggero

IRIT, LEREPS-CIRESS – Université de Toulouse 1
21, allées de Brienne, 31 042 Toulouse Cedex – France
mmaillia@univ-tlse1.fr, sibertin@univ-tlse1.fr, famblard@univ-tlse1.fr,
progger@univ-tlse1.fr

Abstract. In this paper, we propose to study the influence of different learning mechanisms of social behaviours on a given multi-agent model (Sibertin-Blanc et al. 2005). The studied model has been elaborated from a formalization of the organized action theory (Crozier and Friedberg 1977) and is based on the modelling of control and dependency relationships between resources and actors. The proposed learning mechanisms cover different possible implementations of the classifier systems on this model. In order to compare our results with existing ones in a classical framework, we restrain here the study to cases corresponding to the prisoner's dilemma framework. The obtained results exhibit variability about convergence times as well as emergent social behaviours depending on the implementation choices for the learning classifier systems (LCS) and on the LCS parameters. We conclude by analysing the sources of this variability and giving perspectives about the use of such a model in broader cases.

1. Introduction

The way social actions are coordinated by and among social actors has been a source of inspiration for many theories, as game theory, in very different domains such as economics, ecology (Dugatkin 1984), sociology or even in psychology. In a larger perspective, we choose as a research project to investigate the sociological theory of organized action proposed by (Crozier and Friedberg 1977), on the one hand to improve this discursive theory by proposing its formalization and on the other hand to apply it to model different socio-organizationnal phenomena. The work conducted on this project resulted in a proposed formalization of this theory (Sibertin-Blanc et al. 2005), a meta-model, that we briefly expose in the section 2.

Having this meta-model as a first milestone, we are then searching to improve it including social learning mechanisms enabling actors to act rationally. So we focus on LCS as a simple learning mechanism able to represent social learning. Different implementation choices being realistic, we decided to study those alternatives in order to make up our mind. We present those alternatives in section 3 as well as their sociological interpretations in the frame of the organized action theory.

Given those learning mechanisms, we are searching to understand in each case, which collective strategies could emerge and why. We then proceed to experimentations on each alternative in section 4. In order to make things understandable, we choose to make vary parameters corresponding to the sharing of resources in a perticular organization from which we generalize the famous prisoner's dilemma game.

The experimentations exhibit a variability in the emergent collective behaviours depending on the chosen parameters as well as a phase transition at the tipping-point corresponding to the individual transition from dependency on the needed resources to control on those resources. Results are given section 5. In section 6, we provide conclusions concerning the comparison of the different learning mechanisms. We conclude by giving the steps envisaged to follow up this work.

2. Formalization of the Organized Action sociological theory

A formalization of the Sociology of the Organized Action (SOA) leads to consider that constitutive elements of a social system are the *Actors*, the *Relations* and the *Resources*. In this paper, we will only focus on the Actors and the Resources (Fig. 1.).

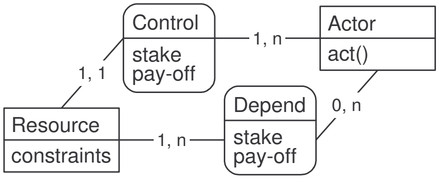


Fig. 1. Entity/Association Model of the structure of a social system in the frame of the SOA.

Actors could be defined as the entities of a social system. A Resource represents an object or a mean, needed by one or many Actors to achieve their action within the organization. Each Resource is linked to an Actor who controls it, and other associated ones being dependant on it. Each actor puts stakes on those Resources and receives in return a pay-off for each one of the resources he is linked to. The actor who controls a Resource decides of the distribution of the pay-offs among the Actors who depend on it and therefore influence their action capability. Every Actor controls one or more Resources and then possesses some freedom to act.

The pay-off corresponds to the quality of the Resource availability; more or better the usability of the Resource by an Actor, higher is its pay-off for this Resource.

The distribution of pay-offs and stakes on numerical scales enables, applying simple operations, to aggregate those values in synthetic and significant values. One can graduate the stakes on a scale between 0 and +10, and the pay-offs with the correspondence –10 to +10. As evidence, these numerical values just enable to perform comparison among them. To do so, we have to normalize the sum of the actors' stakes and then attribute the same amount of stakes to each actor for him to distribute on the relations he participates to. This normalization comes down to grant the same investment to each actor, i.e. the same possibility of personal implication in the social relations game.

A particularly significant value is, for each actor, the sum on the whole set of Resources he is involved in, of a combination between his stake and the resulting pay-off he receive. We name this value the actor’s *satisfaction* (rather than utility because it is more linked to a context of bounded-rationality). It expresses the possibility for an actor to access the resources he needs in order to achieve his objectives, and then the means available to him considering these objectives. A linear version consists in considering the sum, on every relation he is involved in, of the stake by the pay-off:

$$Satisfaction(a) = \sum_{r/\ a\ participates\ to\ r} stake(a, r) * pay\text{-}off(a, r) \qquad \mathbf{(1)}$$

To attain or preserve a high level for this satisfaction is a meta-objective for every actor, as this level determines his possibility to achieve his concrete objectives. The strategic characteristic of an actor’s behaviour leads him, by definition, to aim and achieve his objectives and then to obtain an acceptable value (if not the optimum) for his satisfaction, that becomes the criterion for learning mechanisms we expose in the following section.

3. The learning mechanisms implemented

After the claim by (Conte and Paolucci 2001) among others for integrating so-called « intelligent » social processes in the agents when doing agent-based social simulations, several papers (Conte and Paolucci 2001, Flache and Macy 2002, Takadama et al. 2003) proposed some learning algorithms to be used by agents.

In this paper, we explore two models for social learning using the learning classifier systems, LCS (Holland et al. 2000), for the action selection. LCS are based on the learning of behavioural rules using a test-errors approach, reinforcing the rules depending on the results they produce in a given context. Recent works about reinforcement learning models exhibit that a reduced set of parameters and hypothesis may cover important hidden theoretical assumptions (Macy and Flache 2002). In order to advance prudently we decide to make their use explicit and to discuss the way they can be implemented.

Each model is a LCS without genetic algorithms nor bucket brigad retribution process. One can refer to (Sibertin et al. 2005) to access to the detail of the used algorithm. The main processes involved, namely retribution, oblivion and matching of the learned rules are respectively governed by three parameters:

- *reward* is a positive or negative reinforcement of the rules depending on Δ*satisfaction*;
- *oblivion* is a factor of reward mainly used to weaken the strength of each rules;
- *dmin* enables an agent to match a perceived situation with yet learned situation-action rules.

Finally, the election process performs the selection of the matching rule with the highest strength. If there is none, the covering process generates a random one.

As exprimed by (Molm 1991), many social scientists denote that satisfaction may be specific or global. We thus propose to study a specific and a global satisfaction model: *Specific Satisfaction CS* (SSCS) and *Global Satisfaction CS* (GSCS).

Our first definition of satisfaction as exprimed in equation (1) is in the scope of Molm’s definition of satisfaction: “*cognitive evaluations in which actors compare* *actual to expected outcomes*". We use this for the GSCS model. For the SSCS each actor *a* associates to each resource *r* a *satisfaction* expressed as:

$$\text{Satisfaction}_{a,r} = \text{stake}_{a,r} * \text{pay-off}_{a,\, r} \qquad \textbf{(2)}$$

The main difference between GSCS and SSCS results in a local or global feedback of actors' actions.

4. Experimentations conducted

In order to validate our model we propose to use a cross-validation, as proposed by (Takadama 2003), in a prisonner's dilemma (PD) game. Althougth Takadama's work has been a rich influence leading our work, we use another experimental protocol.

A general difference is to consider the simulator level analysis versus the social system level one. Firstly, this lead us to an exhaustive exploration of *oblivion/reward* and d_{min}, parameters on one side, and *stakes* parameters on the other one. This exploration goes beyond the constraints of the prisoner's dilemma and enables to situate the results in a wider area. Secondly, at the simulator level, we base the validation on a homogeneous behaviour whatever are the social system parameters, and at the social system level: we are searching for parameter values under which mutual cooperation in a PD game is the most frequent social behaviour, as in everyday life relationships. The last difference with Takadama work is that agents' representations are the same for both compared models

4.1. The Prisoner's Dilemma

The prisoner's dilemma was proposed by two mathematicians Merrill Flood and Melvin Dresher in 1950. It is exposed as a game where two players, the prisoners, have the choice to cooperate, *c*, or to defect, *d*. Players earn pay-offs depending on the choices of the both players, as shown in Table 1. If the two players cooperate they will receive the reward for the cooperation (R); if both defect they will be punished for the defection (P); and if one cooperates whilst the other defects, he is the sucker (S) and the other will earn the retribution of his temptation (T).

The dilemma is constrained by the fact that temptation is more profitable than mutual cooperation (*cc*), that pays more than punishment, that is more valuable than to be the sucker: T > R > P > S; and that the best collective strategy is *cc*, 2 R > T + S.

The classical PD game is of minor interest compared to its iterated version where each player can potentially apply different actions over time and where the pay-offs are summed up. The iterated version of the PD has been widely explored and exposed (Hoffman 2000, Delahaye 1992, Macy and Flache 2002) since 1984 Axelrod's tournament.

The expressivity of the SOA formalization does not directly match the PD game. So we will present here how we make a projection from our model to enter in a PD game context. This projection leads us to a generalized PD game.

Lets define a *2-actors, 2-resources Organization* within SOA formalization. Let be two actors, *1* and *2*, and two resources, r_1 and r_2, such that each actor *i* controls the resource r_i. We normalize the stakes for each actor so that their sum is 10. Let be $s_{i,\, rj}$ and $p_{i,\, rj} \in [0;10]$ respectively the stake and the pay-offs of an actor *i* for a resource r_j. Let be $give_i$ and $take_i$ the possible actions each actor *i* can exert on its controlled resource r_i. Lets now define the effect of an action *action* applied by the controler *c* of a relation *r* as $effect_r(action)= \{\Delta p_{c,\, r}, \Delta p_{d,\, r}\}$ such that $\Delta p_{c,\, r}$ and $\Delta p_{d,\, r}$ are respectively the pay-off increments of the controler actor *c* and the dependant actor *d* of the relation *r*. Let be $effect(give)= effect^{-1}(take)=\{-1,1\}$, so that we are in a zero-sum game.

	$give_1$	$take_1$
$give_2$	$sat_1 =$ $s_{1,\, r2}$ - $s_{1,\, r1}$ $sat_2 =$ $s_{2,\, r1}$ - $s_{2,r\, 2}$	$sat_1 =$ $s_{1,\, r1}$ + $s_{1,\, r2}$ $sat_2 =$ -$s_{2,\, r1}$ - $s_{2,\, r2}$
$take_2$	$sat_1 =$ -$s_{1,\, r1}$ - $s_{1,\, r2}$ $sat_2 =$ $s_{2,\, r2}$ + $s_{2,\, r1}$	$sat_1 =$ $s_{1,\, r1}$ - $s_{1,\, r2}$ $sat_2 =$ $s_{2,\, r2}$ - $s_{2,\, r1}$

→

	c_1	d_1
c_2	R_1 R_2	T_1 S_2
d_2	S_1 T_2	P_1 P_2

→

	c	d
c	R R	T S
d	S T	P P

2-actors, 2-resources Organizations* ⊃ *Sociology of Organized Action Model ⊂ ***Generalized PD games*** ⊂ ***PD game***

Table 1. SOA Model enables to define 2-actors, 2-resources Organizations which include all Generalized PD games which include the famous PD game.

Lets now define the *Generalized PD Games*. In a *2-actors, 2-resources Organization,* lets consider that satisfaction sat_i of each actor *i* is determined by the played couple of action $\{action_1, action_2\}$, such that there are four potential different satisfaction values for each actor. Lets assume a syntactic equivalence between $\{sat_1(give_1, give_2), sat_2(give_1, give_2), \ldots\}$and $\{R_1, R_2,\ldots\}$ and another one between $\{give_i, take_i\}$ and $\{c_i, d_i\}$ as shown in the two first tables of Table 1.. Let be the constraints: $R_1 > S_1 > T_1 > P_1$, $R_2 > S_2 > T_2 > P_2$, $R_1+R_2 > T_1+S_2$ and $R_1+R_2 > T_2+S_1$.

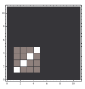


Fig. 2. *2-actors, 2-resources Organizations Matrix*. The X-axis and Y- axis respectively represents the relative autonomy[1] of *1* ($s_{1,\, r1}$) and *2* ($s_{2,\, r2}$). Generalized PD Games are ploted in grey or white, while PD Games are in white.

Lets define the classical *PD Game* as a *Generalized PD Game* respecting the following constraints: $R_1 = R_2$, $S_1 = S_2$, $T_1 = T_2$ and $P_1 = P_2$.

4.2. Experimental design

The simulations were conducted with the same complete experimental design for the both models (SSCS and GSCS).

A first set of parameters concerns the sociological model, as it is the *stake* of each actor for the relation he controls. Because of the normalization, we do not have to

[1] We can directly deduced the actors' relative dependency ($s_{1,\, r1}$ and $s_{2,\, r1}$) from their relative autonomy because of the stake normalization.

make vary the *stake* of the actor for the other relation. Moreover, the game being symmetric we only explore one half of the parameter space.

A second set of parameters concerns the LCS. We explore the d_{min} and the *oblivion/reward* ratio as follow. The d_{min} is a distance which is used to compare situation part of a rule to the perceived situation. The greater d_{min} the less thick is the exploration. The possible values we have chosen to explore for d_{min} are graduated on a logarithmic scale from 0 (maximum of one applicable rule) , 2^0,… to 2^5 (all rules are applicable). The *oblivion/reward* ratio is also essential because it permits to renew the rules whose situation part is matching the current situation. A high ratio value will conduct to a quick renewal of the rules, whilst at the opposite a low ratio will slow down the adaptation of the agent. The ratio value belongs to [0;1] and is incremented by step of 0.2. The reward is fixed at 5.

We have produced 50 runs within a maximum of 5000 steps for each parameter quadruplet {stake_of_r1_controler, stake_of_r2_controler, dmin, oblivion/reward}. For each model and for each parameter quadruplet we have observed many values (mean and standard deviation for the convergence time…) needed for the validation.

5. Results

How to read the five dimensions representation ?

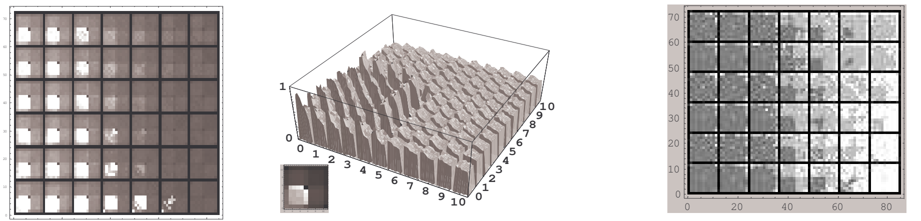
In the following of this document we use a combinaison of two bidimensional matrices to present some observations in a single bidimensional matrix. The construction principle is to affect a matrix to an element of the other one. This enable us to represent a value from the social system point of view in function of the LCS parameters or the opposite. If we analyse a value under the social system point of view, the main matrix will be the *stake matrix*, else it will be the *LCS matrix* (X-axis represents *dmin*, Y-axis represents oblivion/reward). 3-dimentional matrices are the result of applying the mean of each sub-matrix of the five dimensions representation.

5.1. Results for the Specific Satisfaction Classifier System

The frequency of occurence of each action pair (*cc*, *cd* or *dd*) at convergence point is 1 for each Generalized PD Game whatever are the simulator parameter values. As an evidence of violation of our second validation criterium, we do not need more investigation with this model.

5.2. Results for the Global Satisfaction Classifier System

The GSCS seems to give a largest variety of results than the SSCS. As we can observe on Fig. 3.a the larger d_{min} the lower the convergence time. The *oblivion/reward* ratio also speeds up the convergence as it increases. We observe a phase transition for a value of *dmin* between 2 and 4. Another observation presents in the mean convergence matrix (Fig. 3.b.) is a distinct transition which appears in the generalized PD game area. Homogeneity validation[2] is true for dmin > 8 (Fig. 3.c).



a. Mean convergence *LCS view*

b. Mean convergence *Social system view*

c. Behavior homogeneity (dark gray: invalidated)

Fig. 3. Convergence and LCS homogeneity. A transition appears in the generalized PD area (**b.**)

The results of the analysis of the frequency of occurence of each action pair at convergence give us that the *cc* pair emerges as more frequently as the *dmin* value is higher than 16 or lower than 2, and as the oblivion/reward rate increases; the *dd* pair emerges more frequently as *dmin* increases and as the oblivion/reward rate decreases; the *cd* pair emerges more frequently as the *dmin* value is upper 1 and lower 16.

We now can verify our second criterium, that is predominance of mutual cooperation. The criterium is true if *cc* are 20% more frequent than each other pair, that is when dmin is lower than 2 or equal to 32 and oblivion/reward is upper 0.4.

It is also remarkable that for dmin equals 32 and oblivion/reward ratio is upper 0.4 our first criterium about the homogeneity of the LCS behaviour is also validated, and thus the GSCS is validated for both criteria.

6. Discussion

We have proposed to compare the behaviour of two learning classifier systems in a generalized PD game in regard to a social system criterium, namely the predominance of the mutual cooperation, and to a simulation criterium, namely the LCS behavioural homogeneity at convergence point independently of the social system state.

As a matter of fact the SSCS model has no feedback on the whole system and thus is not suitable for agent adaptation in a mutual dependancy context. At the opposite, the GSCS enable values validating our criteria (oblivion/reward ≥ 0.6 and dmin=32).

As a first feedback of this study, and in regard to the transitions that we have exhibit, it would be interesting to relaxe our homogeneity criterium and to take into account that mutual cooperation could have a co-adaptation cost. The second feedback is that the frequency of occurrence for the *cd* action pairs is very high and seems to be due to an edge effect of the model that we have to study.

The explanation of the phase transition requires an interpretation of the behaviour of a *2-actors, 2-resources Organization* which will cover the classical interpretation of the PD Game. Such a work is out of the scope of this article.

[2] Standard deviation < 0.2 * mean (white on Fig. 4.c), or mean convergence < 50 (light grey).

7.Conclusion

As an evidence, the presented case is constraining considering the learning mechanisms that could be proposed. Following works by increasing the number of actors or by using different organisations of the relationships between actors and resources could lead to suggest other learning mechanisms. But in this case, as a result of our study, the GSCS is more adequate to our modelling purposes.

We have also to mention that the proposed model, as a formalization of the sociological theory of organized action, has a far broader spectrum of application than the one presented in the paper. Among others, we applied or derived this meta-model on the study of the emergence of territorial coalitions (Mailliard et al. 2005), to classical cases from the strategic analysis literature as the Trouville case (Mailliard et al. 2003).

Moreover, this interdisciplinary work even presented as the use of computer sciences as tools for sociological theories, benefits also to computer sciences as a source of inspiration in order to propose original coordination mechanisms among computational agents (Sibertin-Blanc et al. 2005).

References

Axelrod, R.: The evolution of cooperation. Basic Books, New York, (1984).

Conte, R., Paolucci, M.: Intelligent Social Learning. JASSS vol. 4, no. 1, (2001).

Crozier, M., Friedberg, E.: L’acteur et le système: les contraintes de l'action collective. Seuil (1977).

Delahaye, J. P., L'altruisme récompensé ? *Pour La Science*, 181:150-156, (1992).

Dugatkin, L.A.: Cooperation among Animals: An Evolutionary Perspective. Oxford University Press, (1997).

Flache A., Macy, M.W.: Stochastic colusion and the power law learning. Journal of Conflict Resolution, (2002).

Hoffmann, R.: Twenty Years on: The Evolution of Cooperation Revisited, Journal of Artificial Societies and Social Simulation (JASSS) vol. 3, no. 2, (2000).

Holland, J, All.: What Is a Learning Classifier System? LCS’99, LNAI 1813, 3-32 (2000).

Macy, M.W., Flache. A.. Learning Dynamics in Social Dilemmas. Proceedings of the National Academy of Sciences U.S.A. May 14;99(10):7229-36,(2002).

Mailliard,M.,Amblard, F., Sibertin-Blanc C.: Modélisation multi-agents pour la formalisation de théories sociologiques: Le cas de la sociologie de l’action organisée appliquée à l’étude de la dynamique du pays Quercy-Rouergue. In Proceedings of the SMAGET, France (2005).

Mailliard, M., Audras, S., Casula, M. : Multi Agents Systems based on Classifiers for the Simulation of Concrete Action Systems. In Proceedings of the 1st EUropean Workshop on Multi-Agent Systems (EUMAS), Oxford University, (2003).

Molm, L.: Affect and social exchange: satisfaction in power-dependence relations. American Sociological Review, vol. 56, (1991).

Sibertin-Blanc, C., Amblard, F., Mailliard, M.: A coordination framework based on the Sociology of the Organized Action, In Proceedings of the From Organization to Organization Oriented Programming in MAS, AAMAS, Utrecht University, (2005).

Simon, H.: The sciences of the artificial, MIT Press, 3rd edition (1996).

Takadama, K., and al.: Cross-Element Validation in Multiagent-based Simulation: Switching Learning Mechanisms in Agents. JASSS vol. 6, no. 4, (2003).