Automata Simulation of N–Person Social Dilemma Games

Vladimir Akimov*
Instititute of the US&Canada Studies
of Russian Academy of Sciences
2/3 Khlebny St.,
121069 Moscow Russia

Mikhail Soutchanski†
Dep. of Computer Science,
Univ. Of Toronto,
M5S 1A4 Canada,
email: mes@cs.toronto.edu

15 May 1992. Revised: 7 August 1992; 19 April 1993

Abstract

Collective behavior of N players in a social dilemma game is simulated by automata exhibiting asymptotically cooperative behavior. In his automata models of simple biological systems M.Tsetlin assumed minimum of information available to the "players." Our automata were somewhat more sophisticated, using Markov strategies in their interactions. We investigated relationships between information received by the automata and the emergence of cooperation in a simulated evolution process. In some ways our approach is similar to that of R.Axelrod. It differs in that instead of determining the "most successful" strategy, we seek surviving strategies in a social dilemma environment. Previous results showed that cooperation could be established asymptotically under partially centralized control. In our model there is no such control. Our main result is that more sophisticated behavior of "self-seeking" automata compensates for the absence of such control. Moreover cooperation is established more rapidly when more information is available to the automata. [1]

*Lab. of Structural Analysis and Modeling of Decision Making

†10 King's College Road SF3303

[1] **Acknowledgment:** We would especially like to thank Prof. Anatol Rapoport for his kind help, in his role as editor, to get this paper out.

1 Dilemma game background

The history of attempts to describe international security and crisis bargaining by social dilemma games (SD) is long [9, 1]. Perhaps the most extensive empirically oriented treatment of SD in situation of international tension is [12], where 18 different international crises were typologized and studied. After [6] mathematically oriented sociologist and politologist encountered different generalizations of simple 2-persons, 2-actions dilemma and those generalizations have received increased interest. Papers [11, 5, 3] discussed the N-player case, and [8] studied a disarmament process in which each of two players has six options, each representing the number of warheads he has following each round.

A type of social dilemma can be conceived as an N–person game ($N \geq 2$), in which each player has a dominating option and in which the choice of dominating option by every player results in a Pareto-deficient equilibrium. We will suppose that every player has the same payoff structure and chooses between two options (to simplify definitions) designated respectively by C (representing *"cooperation"*) and D (representing *"defection"*). The payoff to each player associated with the choice of D will be designated by $D(m)$, where m represents the number of players choosing C ($0 \leq m \leq N-1$). The payoff associated with the choice of C will be designated by $C(m)$, where, as before, m is the number of players choosing C ($1 \leq m \leq N$). Our social dilemma game will be defined by the following conditions.

1. $D(m) > C(m+1)$, that is no matter what others do, each participant is better off choosing D than choosing C and becoming the (m+1)st cooperator, i.e. D is a dominating action (option);

2. $C(N) > D(0)$, the resulting equilibrium is deficient, i.e. universal noncooperation benefits everyone less than universal cooperation;

3. $D(m+1) > D(m)$, $C(m+1) > C(m)$, these conditions stipulate that stability increased for each individual as the number of cooperating players increases;

4. $(m+1)C(m+1) + (N-m-1)D(m+1) > mC(m) + (N-m)D(m)$, in addition to the previous condition is assumed that the society is more stable as a whole the more players who cooperate.

Since m is number of players that choose cooperation, *C(0)* and *D(N)* are impossible payoffs, hence m ranges in value between 1 and N-2. Boundary inequalities are obvious. For the sake of simplicity, we will assume that *C(m)* and *D(m)* are linear in m, hence each is determined by two points, for example, the minima ($C_{min} = C(1)$, $D_{min} = D(0)$) and the maxima ($C_{max} = C(N)$, $D_{max} = D(N-1)$).

2 Automata with "expedient" behavior.

M.Tsetlin introduced several families of automata exhibiting "expedient" behavior, that is, behavior in some sense adapted to a *stationary random medium* [15]. Earlier, but in terms of *many-arms bandit problem*, one kind of automata was introduced by [10]. There is a long tradition within Tsetlin's scientific school to investigate individual and collective behavior in stochastic environment. To be successful in SD-games each player must be capable to adapt its behavior quickly to the dynamic behavior of the other participants. The best way to simulate this type of behavior is by means of automata models.

E.T. Gurvich developed an analytical method for investigating asymptotic behavior of game-playing automata [4]. M.E.Soutchanski used this method to propose a pattern of interactions among automata that guarantees asymptotically stable cooperation in SD games [14]. The main drawback of this proposal is that it presupposed synchronized switching to new options (these new options may coincide with the old ones), that is, some sort of partial centralization.

M. Tsetlin studied the behavior of automata characterized by a finite repertoire of actions (options) f^i $(i = 1, ..., k)$ and a finite set of states φ^i_j $(j = 1, ..., \nu)$. A deterministic automaton of this sort is described by the following canonical equations :

$$\varphi(t+1) = \Phi(\varphi(t), s(t+1)) \tag{1}$$

$$f(t) = F(\varphi(t)) \tag{2}$$

Here $t = 1, 2, ...$ is quantified time. Equation (1) indicates that the state of the automaton at time $t+1$ is determined by the state at time t and by the input $s(t+1)$ at time $t+1$. Equation (2) indicates that the automaton's action at time t is determined by its state φ at that time. The automaton chooses the same action f^i while its state belongs to the subset φ^i_j $(1 \leq j \leq \nu)$. For this reason, we will say that those states are *the memory* associated with the action f^i. When the automaton begin use an action f^i, it is in the deepest state φ^i_ν.

After the action has been chosen the automaton receives random response from the environment, probability of which is not known to the automaton. It is assumed that the player can distinguish only between a bad course of events or a good course of events, that is it can have as the input either 1 or 0, symbolizing accordingly, "reward" or "punishment", and each automaton has access only to its own input s. In contrast to classical game theory, no automaton is aware of other players, the actions chosen or any of the environment reward probabilities. An adaptation of automata behavior depends on feedback signals from the environment. Expediency of automata behavior is achieved by moving to the more deeper memory state if reward was received and, to the more shallower memory state if punishment was received.

The rules governing transitions from one state to another constitute the automaton's *tactic*. A number of tactics have been used in experiments on collective behavior of automata. All tactics have some similar features. Whenever being in state φ^i_1 an automaton receives a punishment, it switches to state φ^ℓ_ν $(i \neq \ell;\ i, \ell = 1, ..., k)$. It continues to use the same action f^i as long as it is in any state φ^i_j, $j > 1$. If it receives a punishment in the memory state with depth 1, it switches to the other action.

A *linear tactic* prescribes the transition from the memory state φ^i_j to the next deeper memory state φ^i_{j+1} $(j = 1, ..., \nu - 1)$ whenever the automaton receives a reward; and from state φ^i_j to the next shallower φ^i_{j-1} $(j = 2, ..., \nu)$ whenever it receives a punishment. When the automaton, being in state φ^i_ν, receives a reward, it remains in that state.

A *trusting tactic*, introduced by H.A. Robbins [10] differs from the linear tactic in that the automaton passes to state of maximum depth φ^i_ν whenever it receives a reward after the action f^i has been chosen.

A *quasi-linear tactic*, introduced by V.G.Sragovich [13] is like a linear tactic when the automaton receives a reward. When it receives a punishment, however, the automaton passes from state φ^i_j to either φ^i_{j-1} or to φ^i_{j+1} with equal probabilities. A reason for introducing the quasi-linear tactic was a result proved by [15] that expedient behavior cannot be "learned" by an automaton using a linear tactic in a stochastic environment when the probability of punishment exceeds the probability of reward. [13] showed that an automaton using quasi-linear or trusting tactic can demonstrate expedient behavior in a much broader range of environments.

The parameter ν is called the memory span of the automaton. It can be interpreted as a measure of the automaton's "patience" or, perhaps, "stubbornness". It represents the maximum number of time quanta during which the automaton will continue to choose the same action, even if it is punished every time it does so, before switching to another action.

Whenever automata play a game, they choose actions simultaneously, and, after that, they will concurrently receive random binary inputs from stationary random environment in accordance with the game payoff functions.

3 Our methodology and our purpose.

Using the automata approach the collective behavior based upon *the least interaction* principal was investigated by Tsetlin's scientific school. In accordance with this school paradigm automata do not "communicate" with each other, and their behavior depends only on media signals. The paradigm describes rather well some processes in biological systems.

However, it is difficult to imagine a collective behavior in human societies without communication and some kind of distributed knowledge about the best ways of behavior in a given collective.

Our approach suggests two innovations in automata paradigm. The main contribution is modeling each decision maker, or player, by an automaton using not pure actions but deterministic Markov decision strategies. We suppose each automaton has a finite set of Markov strategies defined on the set of possible actions (C and D). That means an automaton will not select the fixed action during its bias to some strategy, but its actions will be determined in accordance with its current strategy by one step prehistory of game playing. Thus automata use rules of behavior instead of actions. Moreover, we suppose automata can "communicate" and obtain some kind of knowledge from society experience. In this paper, it is a knowledge about the strategy of the champion (the most successful automaton). This knowledge will be useful when automaton looks for a new strategy. Additionally, we should suppose that players have a knowledge about the choice of the majority of other players at each game stage because without that knowledge a concept of Markov strategy is not applicable.

We would not like to confine ourselves to using a more or less appealing principle of rationality chosen in advance. Our approach does not follow the main stream of game theory but is to some extent closer to the evolutionary approach proposed in [2]. Unlike decision rules submitted to the tournament our automata can not use more sophisticated strategies than Markov ones. It is not surprising that we limit our consideration only by Markov strategies: the most prominent one is **"TIT-FOR-TAT"** - the winner of the computer tournament that was described by R.Axelrod. Other differences between our approaches will be described later.

Our purpose is to investigate the evolution of the strategies in a collective. We are interested in establishing the correlations between the conditions of information exchange and available knowledge on the one hand, and a type of asymptotic collective behavior as well as speed of establishing asymptotic behavior on the other hand. We would like to investigate which strategies will survive and which will perish under different conditions and what type of collective behavior this process will result in. It is important to note that contrary to [14] in the current model decision makers are fully decentralized and all participants are self-seeking individuals. The main result of our computer experiments is that the cooperation is really emerged between the players having sophisticated lines of behavior without any centralization.

4 Description of the model

Our approach is applicable under the following assumptions:

- Players choose their actions simultaneously.
- The outcomes are not fully determined, i.e. the game is played in a stochastic environment, hence the payoffs $C(m)$ and $D(m)$ should be regarded as expected values of random variables. This condition is the main difference between our orientation and R.Axelrod's one. It seems that such a setting is much more adequate for simulating real phenomena of social life where participants not only are unaware of exact values of the payoffs but moreover don't know what is the game they play. Let us note that payoffs of the players must be comparable and have to be measured on an absolute scale. Further, they certainly have to be symmetric: $C(m)$ and $D(m)$ payoffs are the same no matter which particular players choose cooperation. Those assumptions are common for N-person SD games.
- The information available to each player is not very much: each player knows only its random payoff (but not its mean value), its own current Markov strategy, its own action on the previous play and, maybe, some other simple facts.

- Each run of the game consisted of some hundreds of thousands of iterations.

The simplest SD game characterized by the above mentioned criteria is a Two-person prisoner's Dilemma represented by the following table 1 (with $T > R > P > S$):

	C	D
C	R , R	S , T
D	T , S	P , P

Table 1: Canonical matrix

	C	D
C	70 70	25 90
D	90 25	40 40

Table 2: Game matrix.

A generalization to an N–person game is straightforward. A player choosing C receives a payoff of $m \cdot R + (n - m - l) \cdot S$ (assuming that he has played simultaneously with all others and that of these m chose C). A player choosing D receives $m \cdot T + (N - m - l) \cdot P$. In our experiments automata played a Six-person Prisoner's Dilemma based on the following matrix (table 2). The payoffs $C(m)$ and $D(m)$ calculated from this matrix presented in table 3.

Automata did not "know" expected values of their payoffs. Let us stipulate that reward occurring means receiving one (1) prize and punishment occurring means receiving zero (0) prize. Because an automaton can be receptive only to binary inputs, the payoffs from environment were transformed to binary values.

First, let us pay attention to the purely auxiliary question: how does this transformation work ? Due to linearity of the *C(m)* and *D(m)* functions it is sufficient to use a linear transformation. Preference structure for players don't change under linear transformation of payoff functions, so the resulting game is the same. Let $W_j(m)$ be the expected value of the payoff to the jth automaton in the stage when m players choose cooperation. We can use any normalization that will provide $0 < W_j(m) < 1$. In our model $W_j(m)$ after normalization equals:

$$0.8(C(m) - C_{min})/(D_{max} - C_{min}) + 0.1, \text{ if } j\text{'s action was C},$$
$$0.8(D(m) - C_{min})/(D_{max} - C_{min}) + 0.1, \text{ if } j\text{'s action was D}.$$

Thus, $W_j(m)$ can be regarded as the probability of receiving a reward after choosing C or D when m players choose C. In other words, if all players fix their actions, then the arithmetical mean of the rewards received by the jth automaton after long period of time will be near to

m	0	1	2	3	4	5	6
C(m)		125	170	215	260	305	350
D(m)	200	250	300	350	400	450	

Table 3: Payoffs in a Six-person Prisoner's Dilemma with T=90, R=70, P=40, S=25.

$W_j(m)$ (according to the law of large numbers). It seems that interactions of the automata society within the stationary random environment under consideration can be described by a Markov chain [15, 13]. To achieve ergodicity of this Markov chain we added small steps from 0 and 1 during normalization, so that for every j: $0 < W_j(m) < 1$. The transformed expected payoffs are shown in table 4.

The second technical question is the quantization of payoffs during computer simulation. Every round and for every automaton a uniformly distributed random variable on interval (0,1] is generated by means of the standard $rand()$ library function. It is obvious that if the jth automaton receives 1 only when the generated value is less than or equal to $W_j(m)$, then its expected payoff will be $W_j(m)$.

The choice of action on each play was determined by a deterministic Markov strategy, a function that maps the set (CC, CD, DC, DD) onto C or D. Symbolically,

$$f_i : (A(t-1), M(t-1)) \mapsto A(t)$$

where f_i is *ith* Markov strategy, $A(t-1)$ - the player own choice on the previous play, $M(t-1)$ - the choice (either C, or D) of the majority of the other players, $A(t)$ - the new action. If each automaton knew all actions chosen, it could calculate the most preferable action of the other automata in the society. In the computer program the number of players was even and, hence, this "average action" played by the majority of the other automata can be calculated unequivocally (C or D). Because the participants of SD are inclined to hide their actual choices

m	0	1	2	3	4	5	6
C(m)		0.1	0.21	0.32	0.43	**0.54**	**0.65**
D(m)	0.285	0.41	0.53	0.65	0.78	0.9	

Table 4: The probabilities of reward for an automaton when m players chose cooperation

we relax this assumption and suppose only that each player knows preferable choice of its partners somewhere from.

Thus, the repertoire of strategies at the disposal of each player consisted of the 16 strategies shown in table 5.

As was mentioned above, memory of some finite span is joined each strategy. If we conceive the memory as a cell stack, then the memory span is the number of these cells. Let us note that the span of memory joined to each strategy is a constant throughout all automata.

The automata followed linear tactics. Upon receiving a punishment in state with minimal depth φ_1^i, a player switched from the strategy f_i to other strategy f_ℓ $(i, \ell = 1, ..., 16;\ i \neq \ell)$. On one condition, RND (random switching), the new strategy was chosen randomly from among the 16. In another condition, CHMP (Champion), the player switched to the strategy currently used by the "Champion", that is, the automaton which had accumulated the largest total payoff up to the play in question. Let us remind: when the automaton begins play a new strategy its memory state is the deepest one.

Our experiment differs from Axelrod's tournaments [2] in two ways. On the one hand, the strategies available to the players were much simpler than those submitted to Axelrod's tournaments. On the other hand, our players could switch from one strategy to another during the iterated plays. Note that our strategy 9 (cf., table 5) corresponds to TIT FOR TAT (**TFT**), the winner in both of Axelrod's tournaments (it was suggested by Anatol Rapoport). TFT is the most systematic strategy based on the principal of reciprocal altruism: a player begins by cooperating and then chooses on trial t the same action the other player has made on trial $t-1$.

A M	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
C C	C	C	C	C	D	C	C	D	C	D	D	C	D	D	D	D
C D	C	C	C	D	C	C	D	D	D	C	C	D	C	D	D	D
D C	C	C	D	C	C	D	D	C	C	C	D	D	D	C	D	D
D D	C	D	C	C	C	D	C	C	D	D	C	D	D	D	C	D

Table 5: Markov strategies

5 Results.

We set out to investigate which strategies would "survive" and which would be eliminated in the simulated learning process and what sort of behavior would emerge as a result of this selection. We were also interested in comparing the process under the RND and the CHMP switching conditions. Note, that the latter makes more "information" available to the players, namely, information about the "Champion's" current strategy.

Having a finite memory depth automata can come to a rather stable pattern of strategies that results in the persistence of the correspondent choices. Nevertheless, there is always a non-zero probability that a given play pattern could be destroyed in future. To stress this stochastic nature of stability we define *quasi-stable situation* as the persistence of a automata choices for a "long time," where "long time" in an iterated game is defined as a sequence of plays an order of magnitude greater than the memory span. For example, a "long time" means hundreds of iterations if the memory span is about 10. Let us note, the game situation where the majority of players chose D option can not be quasi-stable because in this situation the probability of punishment is too high and automata will go very quickly to new alternatives.

We define *stable cooperation* to be a quasi-stable situation, whose total amount of cooperative actions is essentially greater than 50%. It's clear that the "liveness" of a stable cooperation depends on the automaton's memory span. Although, the probability of reward is greater than probability of punishment when all automata cooperate (recall, that the probabilities of reward for a **6** automata society calculated in accordance with table 2 are presented in table 4), there is a non zero probability that an automaton receives a long series of punishments resulting in its abandoning its current strategy and, hence, destroying the cooperation.

It was established experimentally that stable cooperation is not likely to emerge if the memory span is shorter than 10. On the other hand, automata with long memory spans achieve cooperation very slowly because of the "inertia" of their behavior.

The simulations were carried out on an *IBM PC AT* computer [2]. To make results better comparable we started all games under the same initial conditions: all automata use linear (Tsetlin's) tactics, select the unconditional defective strategy (16) and defective actions (D). All experiments were performed for 6 automata, with payoff functions corresponding to table 4, and

[2] The program is distributed under certain conditions upon request.

memory span is varied in the ranges 20-35 (random switching) and 15-20 (champion switching). The surviving strategies of several runs are shown in table 6 below. [3]

The basic level of our simulation is a game with random switching between strategies. In this case players essentially do not communicate. It was shown in experiments that automata can demonstrate asymptotically cooperative behavior even in this primitive case. All they need for this is a memory which is not very short and a rather long relaxation time.

On the next level of the simulation one kind of society experience becomes available for the automata. We assume in this case the automata are informed about the champion's strategy. It was shown, that such a small element of human style behavior deeply affects automata behavior. They find a stable cooperative play much faster and under shorter memory depths.

One might suppose that because the rules of behavior of each automaton were the same and all were immersed in the same environment, their asymptotically adopted strategies would be identical, especially in the variant in which they all switched to the Champion's strategy. This, however, turned out not to be the case. In each simulation a "spectrum" of strategies emerged, and the "spectra" themselves differed. That is, "cooperation" emerged but not on the basis of adopting the same strategy by all automata. Let us see why.

Note that among the sixteen strategies (cf. table 5), the following map CC on C: 1, 2, 3, 4, 6, 7, 9 and 12. That is to say, an automaton adopting any of these strategies and only these would continue to cooperate after a cooperative outcome. Hence a stable cooperative situation has to be based on this set of "cooperation preserving" strategies (they were called in [2] *nice* strategies).

Cooperation can be established due to at least two different reasons. Either it happened to arise by lucky chance or it could be rationally produced as a result of strategies interaction. First way is very simple. Let us suppose that one automaton comes to strategy 1 (unconditional cooperation). Certainly, it will not be very successful with this strategy if other automata defect. It will have a lot of punishments. But once an automaton selects this strategy it has to follow it at least during the time equal to its memory span. If during that time other automata happened to come to the same strategy or to the other nice strategies (1, 2, 3, 4) by lucky chance, the total

[3] We perform 15 computer experiments for Random switching with 250,000 iterations each. A cooperation was not established during 2 experiments. Calculating the mean value we took relaxation time for them as (at least) 250,000. The resulting average value is (at least) 132500.

Strategies distribution. Random switching																
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	relax time
1	1	1		1				2								50000
1			1	1				3								75000
			2			1		1			2					50000
	2		1	1	1			1								12000
1		2	1					2								207000
		1	2					2	1							87000
1	1	2						1	1							150000
1			1		1			2					1			166000
2	2		1										1			148000
	2		2				1	1								40000
1	1		2					1	1							150000
1			1			2		1	1							2000
	1	2						1			2					237000
stable cooperation did not emerge																250000
stable cooperation did not emerge																250000
average																132500
Strategies distribution. Champion switching																
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	relax time
		3						2				1				9000
																15000
		3						3								12000
			3					2	1							1200
	1	2				1		2								1800
	1		4						1							6300
			3					2		1						10400
3	3															3800
1	4						1									10900
	3	1	1						1							23400
			1			3		2								750
			2					4								4800
			1			3		2								10000
2	1		1					2								2500
			3		1			2								23000
	3					3										16000
			1			4							1			34000
		1	3					2								52300
2								4								19000
average																13500

Table 6: "Surviving" strategies and corresponding relaxation time before the stable cooperation.

cooperation would emerge in a collective. It's clear that probability of such type of cooperation arising depends on the relation between the number of automata and their memory depth. A small collective having a long memory has a good chance to come to a stable cooperation. But the probability will diminish if the collective become greater.

To explain the second mechanism note that several combinations of cooperation preserving strategies produce a sequence of outcomes leading to CC and hence to the emergence of stable cooperation. They do this by providing a "peace initiative" trigger. Consider a pair of automata, each using strategy 9 (TIT FOR TAT). If they happen to start with CC, they will continue to cooperate indefinitely. If, however, they happen to start with DD and there is no other strategy to switch to, they will both continue to defect indefinitely. Finally, if they start with either CD or DC, and there is no other strategy to switch to, they will be trapped in a vendetta cycle: $(C, D) \longleftrightarrow (D, C)$. They can get out of these traps only if a player switches from D to C and the other follows suit or continues to cooperate, in other words responds to a "peace initiative".

H.R. Alker and R. Hurwitz [1] and H. Moulin [7] pointed out that in conditions of information exchange, TFT can provide a "deterrence" effect. In our case such information exchange was excluded. It follows that to produce stable cooperation in every case, TFT must be accompanied by those strategies from the cooperation preserving set, which will provide **peace initiative** trigger. Consider, for example, strategies 4 and 7. Any combination of these two produces CC following DD and preserves it thereafter. Further, a combination of strategies 2 and 4 produces the sequence $(D, D) \longmapsto (D, C) \longmapsto (C, D) \longmapsto (C, C)$. The same is true for 2 and 9. Some combinations from the cooperation preserving set do not lead to stable cooperation, e.g., strategies 4 and 9 or 7 and 9: $(D, D) \longmapsto (C, D) \longmapsto (D, C) \longmapsto (C, D)$... Note, however, that strategy 3, although it rewards the co-player's defection and punishes its cooperation, complements TFT, producing the sequence $(D, C) \longmapsto (D, D) \longmapsto (C, D) \longmapsto (C, C)$. Note further that the demanding strategy 12 ("I will cooperate only after cooperation has been established") and the stubborn strategy 6 ("I will continue doing what I have been doing, regardless") cannot serve as "peace initiative" triggers.

From figure 1, we may conjecture that in the case of random switching, strategies 1, 2, 3, 4 and 9 are conducive to emergence of cooperation. In the case of switching to the Champion's strategy (a more "sophisticated" form of behavior), combinations especially conducive to cooperation appear be 2, 4, 7, and 9. The role of strategy 7 (which is an important peaceful trigger)

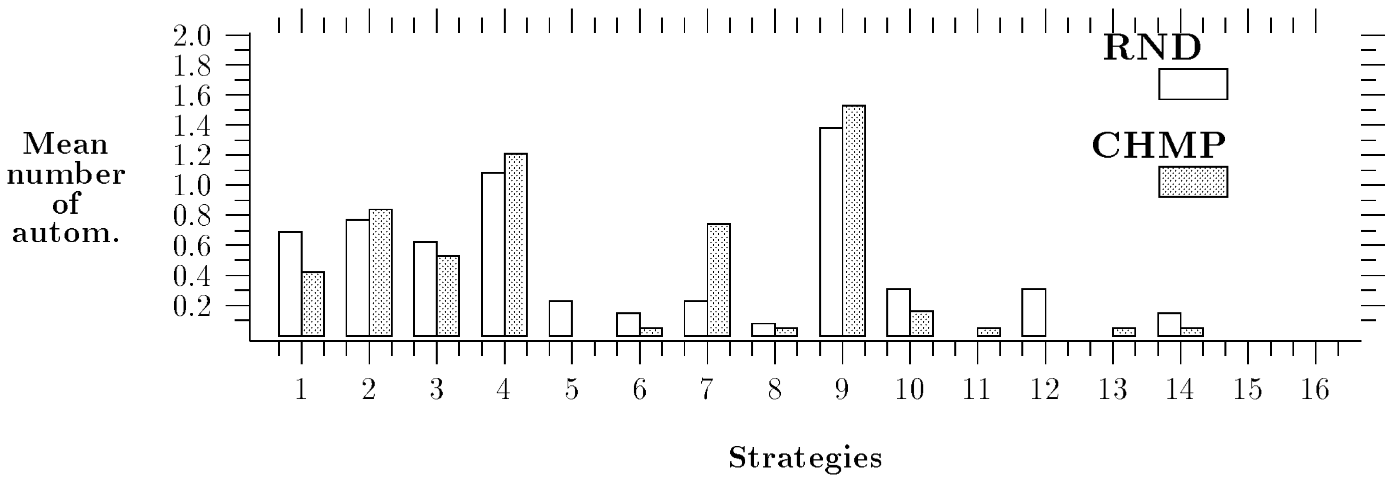

Figure 1: Mean number of automata using the survived strategies in stable cooperation.

considerably grows, while 6 and 12 (which can not play a role of triggers) are practically excluded. The strategy 1 has very little chance to acquire the status of champion at the relaxation stage of the game.

Analyzing table 6 and figure 1 we can conclude: the mechanism of strategies interaction plays considerably more important role in the case of Champion switching. One of the possible reason is the following: when the Champion switches to a nice strategy, it forces other switching automata choose the same nice strategy. This conclusion is confirmed by the several times faster relaxation and the changes in strategies distribution, in particular strategies 6, 7 and 12.

This observation has a good confirmation in the problem of what is the minimum memory span which is sufficient for emergence of stable cooperation when the number of players varied. Since the establishment of cooperation may take several steps, it appears that a sufficiently large memory span is required to realize it. Moreover, there is a correlation between the number of players and the probability of getting a cooperative play by chance. The set of experiments conducted when the payoff functions have been calculated from table 7 lead to table 8 as a result.

It appears that in the case or random switching the minimum memory span sufficient for emergence of stable cooperation increases with the number of players, while in the case of Champion switching the minimum memory span appears to be independent of group size for groups larger than 2.

6 Short Conclusion.

It was shown earlier that automata of very simple construction (set of actions and memory associated with them) are capable of demonstrating cooperative behavior in a stochastic environment. Most of them can be described by Markov chains theory. In our study automata switch between "beliefs" (rules of behavior) instead of simple actions. This innovation demonstrates that "expedient" cooperative behavior can be based not only on pure chances, but also on more "intelligent" strategies interaction. The diversity of strategies available for each participant facilitates the emergence of cooperation. It was demonstrated that a little information control can affect strongly the way by which automata look for a proper strategy, resulting in fast relaxation instead of slow random search.

7 Future Work.

We plan to investigate the behavior of automata governed by stochastic Markov strategies. For instance, if an automaton uses strategy 9 (TIT FOR TAT) and a proportion p of the other players choose C, while proportion 1-p choose D, then the automaton would choose C with probability p and D with probability l-p. We plan also to study the behavior of large numbers of automata (say of the order of 100) playing CD games. It would seem that emergence of cooperation would be considerably less likely in such "large societies".

1/2	C	D
C	70 70	30 90
D	90 30	51 51

Table 7: Second matrix

	number of players				
switching type	2	4	6	8	10
champion	15	30	30	30	30
random	15	30	45	60	75

Table 8: Memory depth enough for stable cooperation

References

[1] Alker H.R., Hurwitz R.*"Resolving Prisoner's Dilemmas"*, MIT, (Student's Manual), 1981, 132p.

[2] Axelrod R.*"The evolution of cooperation"*. Basic Books, 1984, N.Y.

[3] Dawes R.M. *"Formal models of dilemmas in social decision-making"*. In: M.F.Kaplan, S Schwartz (eds.) "Human judgment and decision processes", N.Y., 1975, p.88-107.

[4] Gurvich E.T. *"Method for asymptotic investigation of games between automata"*. In: Automation and Remote Control, 1975, v.36, N2, p.257-270.

[5] Hamburger H. *"N-person prisoner's dilemma"*. In: J. of Mathematical Sociology, 1973, v.3, N1, p.27-48.

[6] Hardin G. *"The tragedy of the commons"*. In: Science, 1968, v.162, N3859, p.1243-1248.

[7] Moulin H. *"Theorie des Jeux Pour L'Economie et la Politique"*. Hermann, Paris, 1981.

[8] Pilisuk M. *"Experimenting with the arms race"*. In: J. of Conflict Resolution, 1984, v.28, N2, p.296-315.

[9] Rapoport A., Chammah A.M. *"Prisoner's dilemma. A study in conflict and cooperation"*. Ann Arbor, 1965, 258p.

[10] Robbins H.A. *"A sequential decision problem with finite memory"*. In: Proceedings of the National Academy of Science of USA, 1956, v. 42, N 3.

[11] Schelling Th. *"Hockey helmets, concealed weapons, and daylight-saving: a study of binary choices with externalities"*. In: J. of Conflict Resolution, 1973, v.17, N3, p.381-428.

[12] Snyder G.H., Diesing P. *"Conflict among nations"*. Princeton, 1977, 528p.

[13] Sragovich V.G. *"Adaptive control"*. Moscow, "Nauka", 1981 (in Russian).

[14] Soutchanski M.E. (Suchanskiy M.E.) *"Adaptive algorithm for determination of weakly efficient variant under randomness"*. In: Soviet J. of Computer and Systems Sciences, 1987, v.25, N3, p.148-157

[15] Tsetlin M.L. *"Automaton theory and modeling of biological systems"*. (ser. Mathematics in Science and Engineering, v. 102), Academic Press, N.Y., 1973, 288p.

The paper "Automata Simulation of N–Person Social Dilemma Games" published in the **Journal of Conflict Resolution**, 1994, vol. 38, N1 (March) is different from this version because the editorial board did not receive the final improved copy of the manuscript. At the time when this paper has been written, both authors were employed at the Laboratory of Structural Analysis and Modeling of Decision Making in the Institute of the USA and Canada Studies.