AN ITERATED MULTIPLAYER GAME PLAYED BY LEARNING AUTOMATA

Ath. Kehagias

I. Introduction

In a previous paper (Kehagias, 1994) we have studied an iterated version of a two-player game, the celebrated *Prisoner's Dilemma* (henceforth PD), played by *Learning Automata* (henceforth LA). We have presented some computer simulation results which suggest that, under fairly general conditions, cooperation between the learning automata will be established. At the end of that paper we considered the possibility that similar results might hold for multiplayer versions of PD. Here we present some new computer experiments towards this direction, involving several automata playing a *multiplayer* game. This game, the so called Threshold Game (henceforth TG), is similar to the PD in that it involves the dilemma of cooperation and competition. However, TG has its own characteristics and should not be considered as just a multiplayer version of PD. While PD has been discussed extensively as a model of the conflict between competition and cooperation, or between individual and collective rationality, TG appears to be a new game, not previously studied in the game theoretic or learning literature. In this paper we use computer simulation to study an iterated version of TG, played by LA's. The main conclusion is that cooperation is a more viable and persistent alternative than competition in TG. This is an intuitively satisfying result; however it must be viewed with caution since, strictly speaking, it only pertains to the particular version of TG, played by LA's, presented here. A more general analysis would require mathematical tools which fall outside the scope of this journal. At any rate, even this limited analysis may offer useful insights to more general versions of TG and to the general competition - cooperation problem.

Our analysis combines elements from Game Theory and the theory of Learning Automata. From a game theoretic point of view, TG is an *N-person, nonzero-sum game*. For an exposition of the relevant concepts, see (Rapoport, 1966), (Rapoport, 1968) and (Raiffa & Luce, 1985). In particular, for a nontechnical but lucid look into the problems of competition and cooperation, see (Hofstadter, 1984). As for LA's, they are an important paradigm of Artificial Intelligence, offering a simple way to describe the learning process of several interacting agents. In this sense they are a very suitable model for the evolution of game playing strategies. A very detailed discussion of the theory of LA's and their applications can be found in (Narendra & Thathatchar, 1989).

II. The Threshold Game

The Threshold Game will first be presented in its simplest possible form. It involves two players and three possible moves for each player. Generalizations to N players and M moves are obvious and will be discussed at the end of this section.

Consider two players playing the following game: each one chooses an number among the integers 1, 2 and 3; this is the player's bid. The referee sums the bids and, if the sum is less than or equal to 4, pays each player the bid in dollars. If the sum is greater than 4 (in other words, if it is 5 or 6) both players receive nothing. The game can be described concisely by the following table:

	1	2	3
1	1,1	1,2	1,3
2	2,1	2,2	0,0
3	3,1	0,0	0,0

Table 1

Here the first column denotes the bids of player 1, the first row denotes the bids of player 2 and every other entry at the table is a pair of numbers which shows the payoffs collected by player 1 and player 2.

We assume that each player plays selfishly and tries to maximize his payoff. A possible strategy to this end is to bid the highest number possible, i.e. 3. However, if this strategy is followed by both players, it will defeat its purpose, since a (3,3) bid will yield zero payoff to both players. A (suboptimal) alternative is conservative play, for instance a bid of (1,1) which yields a payoff of one dollar for each player; this is clearly not the best they can do. It is also possible that that the players establish some form of (implicit or excplicit) cooperation: for instance, a (2,2) bid seems fair and yields to each player a two dollars payoff. While this is not the maximum each player can get, it is fairly high and, in addition appears to be the maximum each player can get under a fair division of the collective four dollars payoff. Finally, one of the players might choose an aggresive strategy of consistently bidding 3, hoping that the other player will retreat to a bid of 1. At that point the bids (and returns) will become (3,1). This is an equilibrium, although it appears unfair. In short the problem of competition and cooperation (familiar from the PD situation) appears here as well.

Two players can play the game with more than three possible moves, for instance each player can choose any number among the integers 1, 2, ... , M. The threshold may be kept at 4, or changed, and in general we take it to be L. For instance, the two player game with L=7, M=4 is characterized by the following table.

	1	2	3	4	5
1	1,1	1,2	1,3	1,4	1,5
2	2,1	2,2	2,3	2,4	2,5
3	3,1	3,2	3,3	3,4	0,0
4	4,1	4,2	4,3	0,0	0,0
5	5,1	5,2	0,0	0,0	0,0

Table 2

The game can be further generalized by having N, rather than two players. This game we call the discrete (L, M, N) Threshold Game. We will not give the table form of this game, since it is difficult to present the moves of more than two players in a two dimensional table. However, the rules remain exactly the same. Each player chooses a number between 1, 2, ... , M, the referee receives N bids and, if their sum is less than or equal to L, each player is paid his or her bid. In case the sum of bids exceeds the threshold L, every player receives zero payoff.

A final generalization of the TG allows involves a threshold L and N players, but in this case each player can bid any number (not necessarily integer) between 0 and M. This will be called the *Continuous* (L, M, N) Threshold Game; in the rest of the paper we will deal exclusively with this game.

III. The Threshold Game as a Model of Human Behavior

The TG game can be considered as a simplified model of several real world situations. For example, consider the following enviromental problem. The water resources of a city are running low. There is a steady but low inflow of water, which is not sufficient for the needs of all the population. To simplify matters, assume that the city has a population of only two citizens. Each one of them has the choice of conserving water (which will be inconvenient but not unbearable) or consuming at a high rate. If both citizens consume water at a high rate, pretty soon the city reservoirs will be empty and the citizens thirsty; this corresponds to high bids (above threshold) in the TG. If both conserve, the water resources will suffice for their (restricted) needs; this corresponds to low bids. Finally, if only one citizen conserves, the other can use all the water he wants without triggering a crisis; this corresponds to the TG situation of one high and one low bid. The "game" can also be played by several "players" (citizens). Each player has an incentive to consume, no matter what the others do, but if everybody follows this greedy strategy the outcome will be catastrophic. This analysis is not limited to water; clean air, oil and a number of other resources (even parking space at city center) could be used instead.

Similar examples can be found in many other areas where several entities share resources: economics (the players are firms), biology (the players are organisms) and so on. In all cases, the crux of the problem is the choice between a cooperative strategy, where each player bids his fair share (approximately L / N, the threshold divided by the number of players) and a competitive strategy where each player bids his maximum amount, hoping to scare the rest of the players into lower bids.

While the TG is similar to PD, in that both highlight the competition - cooperation theme, it is not equivalent to it. For example, in the case of two players TG does not reduce to a PD, because it does not satisfy certain inequalities between payoffs . For a detailed explanation of the payoff sructure of PD, see (Kehagias, 1994). At any rate, we find TG to be of interest in itself, as a model of human behavior.

IV. Learning a Bidding Strategy

If we consider the TG as a model of human behavior, it will be immediately obvious that TG, just like PD, is rarely played only once. The usual situation involves an iterated TG, played many times in succession, each player remembering his own and the other players' past moves. Playing TG in this mode may promote cooperation, both in a negative and a positive way. For example, in the first round of TG all players bid the maximum amount allowed; as a result they all receive zero payoffs. This is negative reinforcement of lower bids; the players may remember their loss and bid smaller amounts in the next round. Conversely, suppose all players cooperate (even by accident) once, by bidding small amounts; consequently the sum of bids is below the threshold and every player gets a nonzero payoff. This is positive reinforcement of low bids. Of course, if the players realize they were well below the threshold they may consider increasing their bids in the next round. In this way competition may be promoted. One expects that some equilibrium will be found between bidding small and large amounts, so that the total amount bid will be equal or slightly less than the threshold. A further desideratum would be *justice* : each players bids (and receives) the same amount.

The rest of this paper deals with formulating a simple model of learning automata playing the threshold game and exploring the emergence and properties of equilibria. We use as our testbed the continuous (L,1,N) TG. This means there are N automata; at time t, automaton n bids a number xn(t), between 0 and 1. The threshold L is a number between 0 and 1, for instance 0.5. The referee sums the bids x1(t), x2(t), ... , xN(t). If x1(t)+...+xN(t) < L, then each automaton is awarded its bid, xn(t) and updates its bid for the next time t+1, to xn(t+1), given by the formula

(1) $xn(t+1) = (1-a)*xn(t)+a$,

where a, some number between 0 and 1, is a *learning rate* [1]. This results to a net increase of the bid of each automaton. On the other hand, if x1(t)+...+xN(t) > L, then each automaton is awarded nothing, and updates its bid for the next time t+1, to xn(t+1), given by the formula

(2) $xn(t+1) = (1-a)*xn(t)$,

which results in a net decrease of the bid. One hopes that as t increases, all automata will reach the fair level of bidding L/N. We test this conjecture by running several computer experiments; the results are presented in the next section.

V. Computer Experiments

Our experimentation plan is the following. Each experiment is characterized by several parameters, namely number of players N, initial bids x1(0), ... , xN(0), learning rate a and threshold L. We choose specific values of these parameters and run a computer simulation of equations (1) and (2) for a a large number of time steps *t* =1, 2, ... Enough time steps must be taken to ensure that the learning process is completed and each player's bid reaches equilibrium. A typical learning curve can be seen in Fig.1. This refers to an experiment with 3 players (automata), starting with inital bids 0.5, 0.8 and 0.1 respectively, learning rate a=0.01 and threshold L=1.5=3*0.5. We observe that in the initial 50 or so time steps bids change by large amounts, but from around time t=51 until final time t=350 bids settle down to approximately the same equal bid of 0.5. A similar result is illustrated in Fig.2, with all parameters the same, except that learning rate is now a=0.001 (resulting in slower learning and a longer time until equilibrium is reached). Finally, in Figs. 3 and 4 similar experiments are illustrated, but with a threshold L=3*0.2=0.6. In all cases we observe that sooner or later equilbirum is reached, with all automata bidding approximately L/N, in other words their fair share of the maximum total payoff possible. Thus it appears that the automata have learned fair bidding. However, there is a complication that mars this optimistic conclusion. Careful observation of Figs. 1 - 4 reveals that, while all bids converge to the same level, this level oscillates slightly above and below the value L/N. Thus the sum of bids oscillates slightly above and below threshold L, with the results that every few turns the threshold will be exceeded and the automata will receive zero payoff. Of course in the next move each automaton will lower its threshold and the sum bid will be below threshold, resulting in nonzero payoffs. Nevertheless, while the strategy of the automata is fair, since every

[1] The term "learning rate" refers to the fact that large values of a result in large changes of .xn(t+1) relative to xn(t).

automaton bids the same amount, it is not optimal, since in some turns the threshold is exceeded and potential "profit" is lost.

We have run several experiments along these lines, trying various combinations of N, x1(0), ... , xN(0), a and L values. The results are presented in Table 3 (varying initial bids x1(0), ... , x1(N)), Table 4 (varying threshold L) and Table 5 (varying learning rate a). In all cases we use N=3 and N=10 players. The reader will observe that in every table only the first player's final bid, x1(350) is listed. The reason for this is simple and pleasing: the bids of all automata converge to the same value, which is approximately equal to L/N, just like in Figs.1 - 4. The convergence of bids means that for all parameter combinations tried, all automata learn that their "best" and "fairest" bid is equal to L/N. We refer to this strategy as *just cooperation*. Hence every automaton receives its fair share of the payoff. However the strategy learned is not *optimal*, since the maximum possible average payoff is not obtained. This can be seen from the last two columns in Tables 3, 4 and 5, where we list the payoff averaged over the duration of the game, and over the final, equilibrated part of the game. In both cases the payoff is less than the maximum possible (which is L). The suboptimality of the strategy is dues to the oscillation phenomeno discussed in the previous paragraph. It should be note that this oscillation is minimzed for small learning rates (e.g. a=0.001).

N	$x_1(0)$	$x_2(0)$	$x_3(0)$	$x_1(350)$	$x_2(350)$	$x_3(350)$	a	L/N	aver. gain	aver. gain
3	0.100	0.100	0.100	0.199	0.199	0.199	0.01	0.200	0.219	0.199
3	0.100	0.400	0.950	0.189	0.198	0.214	0.01	0.200	0.107	0.167
3	0.100	0.500	0.900	0.195	0.207	0.219	0.01	0.200	0.107	0.171
3	0.500	0.500	0.500	0.207	0.207	0.207	0.01	0.200	0.151	0.205
3	0.800	0.500	0.300	0.208	0.199	0.193	0.01	0.200	0.172	0.221
3	0.900	0.900	0.900	0.206	0.206	0.206	0.01	0.200	0.116	0.205
3	0.219	0.047	0.679	0.201	0.196	0.215	0.01	0.200	0.159	0.196
10	0.100	0.100	0.100	0.199	0.199	0.199	0.01	0.200	0.219	0.199
10	0.100	0.400	0.900	0.198	0.207	0.222	0.01	0.200	0.169	0.195
10	0.100	0.500	0.900	0.189	0.201	0.213	0.01	0.200	0.101	0.167
10	0.500	0.500	0.500	0.199	0.199	0.199	0.01	0.200	0.116	0.188
10	0.800	0.500	0.300	0.219	0.210	0.204	0.01	0.200	0.215	0.238
10	0.900	0.900	0.900	0.206	0.206	0.206	0.01	0.200	0.116	0.205
10	0.679	0.935	0.384	0.209	0.216	0.200	0.01	0.200	0.179	0.223

Table 3
Experiments with initial bids variation

In all cases considered the final bids are close to the optimal bid, which is L/N=0.2; in other words, irrespective of whether the original bid of an automaton is above or below the *optimal fair* bid, eventually all automata learn the optimal fair bid 0.2, which maximizes total payoff and equidistributes it among all automata. The penultimate column of the table lists the average payoff received by automaton nr.1 over the 350 turns of the iterated TG. This number is lower than the optimal average payoff, which should be 350*0.2; this is due to the fact that in the initial stage of the TG, before learning has occurred, the automata many times exceed the threshold. In all cases considered the final bids are close to the optimal bid, which is 0.2. The last column of the table lists the payoff received by automaton nr.1 averaged over the last stage of the game, when learning has occurred. This is closer to the optimal payoff, but still not exactly equal to it, due to the oscillation effect discussed earlier.

N	$x_1(0)$	$x_2(0)$	$x_3(0)$	$x_1(350)$	$x_2(350)$	$x_3(350)$	a	L/N	aver. gain	aver. gain
3	0.100	0.500	0.900	0.195	0.207	0.219	0.01	0.200	0.107	0.171
3	0.100	0.500	0.900	0.485	0.497	0.509	0.01	0.500	0.384	0.467
3	0.100	0.500	0.900	0.884	0.896	0.908	0.01	0.900	0.741	0.865
3	0.800	0.500	0.300	0.208	0.199	0.193	0.01	0.200	0.172	0.221
3	0.800	0.500	0.300	0.505	0.496	0.490	0.01	0.500	0.555	0.523
3	0.800	0.500	0.300	0.901	0.892	0.886	0.01	0.900	0.947	0.915
10	0.100	0.500	0.900	0.189	0.201	0.213	0.01	0.200	0.101	0.167
10	0.100	0.500	0.900	0.489	0.501	0.513	0.01	0.500	0.374	0.464
10	0.100	0.500	0.900	0.882	0.894	0.906	0.01	0.900	0.736	0.862
10	0.800	0.500	0.300	0.219	0.210	0.204	0.01	0.200	0.215	0.238
10	0.800	0.500	0.300	0.514	0.505	0.499	0.01	0.500	0.641	0.533
10	0.800	0.500	0.300	0.866	0.866	0.866	0.01	0.900	0.874	0.865

Table 4
Experiments with threshold variation

In all cases considered the final bids are close to the optimal bid, which is L/N; in other words, irrespective of whether the original bid of an automaton is above or below the *optimal fair* bid, eventually all automata learn the optimal fair bid L/N, which maximizes total payoff and equidistributes it among all automata. This is true for every value of threshold L considered. The suboptimal average payoff discussed for Table 3, holds here, too, as can be seen in the last two columns of this table.

N	$x_1(0)$	$x_2(0)$	$x_3(0)$	$x_1(350)$	$x_2(350)$	$x_3(350)$	a	L/N	aver. gain	aver. gain
3	0.100	0.500	0.900	0.195	0.207	0.219	0.01	0.200	0.107	0.171
3	0.100	0.500	0.900	0.201	0.201	0.201	0.05	0.200	0.187	0.206
3	0.100	0.500	0.900	0.247	0.247	0.247	0.10	0.200	0.211	0.225
3	0.100	0.500	0.900	0.173	0.173	0.173	0.20	0.200	0.210	0.218
10	0.100	0.500	0.900	0.189	0.201	0.213	0.01	0.200	0.101	0.167
10	0.100	0.500	0.900	0.220	0.220	0.220	0.05	0.200	0.187	0.211
10	0.100	0.500	0.900	0.274	0.274	0.274	0.10	0.200	0.211	0.225
10	0.100	0.500	0.900	0.220	0.220	0.220	0.20	0.200	0.187	0.211

Table 5
Experiments with learning rate variation

In all cases considered the final bids are close to the optimal bid, which is 0.2; in other words, irrespective of whether the original bid of an automaton is above or below the *optimal fair* bid, eventually all automata learn the optimal fair bid which maximizes total payoff and equidistributes it among all automata. The suboptimal average payof effect is again present, for all values of learning rate.

VII. Conclusion

Caution is necessary in interpreting the results of our experiments: they constitute only a partial analysis of our model. A fuller analysis by computer experiments would require a finer variation and more extensive combination of parameter values. An alternative method of analysis would involve a mathematical study of eqs. (1) and (2). Such an analysis requires rather sophisticated mathematical methods and belongs to a more specialized journal. Even a complete mathematical analysis will only give information about our particular model of playing iterated TG; many other models are possible, for instance one using *probabilistic* LA (Kehagias, 1994). Keeping all these qualifications in mind, we still have evidence for two conclusions, one optimistic and the other pessimistic.

The optimistic conclusion is that just cooperation will emerge in an iterated TG under very general conditions. In fact in every experiment conducted we obtained just cooperation. This is no accident: mathematical analysis of the model (which will be published elsewhere) reveals that playing TG using eqs. (1) and (2) will always yield just cooperation, or, in mathematical terms, $xn(t) \rightarrow L/N$, for $n=1, 2, \ldots, N$.

The pessimistic conclusion is that just cooperation is not optimal, since the automata do not learn to bid exactly L/N, and hence win their maximum fair payoff.

Conclusions of a moral character are left to the reader. The author, speaking with a mathematician's point of view, lists two directions for further research, both concerning the introduction of probability. This can be done either in the game itself (e.g. introducing a probabilistically varying threshold) or in the learning process (using probabilistic LA), or in both. It may be that such extensions will introduce situations where competition, rather than cooperation, prevails. This will be examined in a furture paper.

References

Axelrod, A. *The Evolution of Cooperation*. New York. Basic Books, 1984.

Hofstadter, D.R. *Metamagical Themas*. New York. Basic Books, 1985.

Luce, R.D. & Raiffa, H., *Games and Decisions*. New York, Dover, 1985.

Rapoport, A. *Two Person Game Theory*. Ann Arbor. Un. of Michigan Press, 1966.

Rapoport, A. *N-Person Game Theory*. Ann Arbor. Un. of Michigan Press, 1968.