Spatial Games with Adaptive Tit-for-Tats

Elpida S. Tzafestas

Institute of Communication and Computer Systems, Electrical and Computer Engineering
Department, National Technical University of Athens, Zographou Campus, Athens
15773, GREECE
brensham@softlab.ece.ntua.gr
http://www.softlab.ece.ntua.gr/~brensham

Abstract. This paper presents an adaptive tit-for-tat strategy and a study of its behavior in spatial IPD games. The adaptive tit-for-tat strategy is shown elsewhere to demonstrate high performance in IPD tournaments or individual IPD games with other types of strategies, and obtains higher scores than the pure tit-for-tat strategy. In spatial IPD games, the strategy exhibits stability and resistance to perturbations, and those properties are more pronounced in variations of the spatial game model that induce some degree of "noise" : probabilistic winning, spatial irregularity and continuous time. The adaptive tit-for-tat strategy is also compared to pure tit-for-tat and found to be more stable and predominant in perturbed environments.

1 Introduction

A major research theme in theoretical biology and evolutionary computation is the emergence and evolution of cooperative behavior between selfish agents. The cooperation problem states that each agent has a strong personal incentive to defect, while the joint best behavior would be to cooperate. This problem is traditionally modeled as a special two-party game, the Iterated Prisoner's Dilemma (IPD).

At each cycle of a long interaction process, the agents play the Prisoner's Dilemma. Each of the two may either cooperate (C) or defect (D) and is assigned a payoff defined by the following table.

Agent	Opponent	Payoff
C	C	3 (= Reward)
C	D	0 (= Sucker)
D	C	5 (= Temptation)
D	D	1 (= Punishment)

Usual experiments with IPD strategies are either tournaments or ecological experiments. In tournaments, each strategy plays against all others and scores are summed in the end. In ecological experiments, populations of IPD strategies play in tournaments and successive generations retain the best strategies in proportions analogous to their score sums.

The first notable behavior for the IPD designed and studied by Axelrod [1][2] is the Tit For Tat behavior (TFT, in short) :

> *Start by cooperating,*
> *From there on return the opponent's previous move*

This behavior has achieved the highest scores in early tournaments and has been found to be fairly stable in ecological settings. TFT demonstrates three important properties, shared by most high scoring behaviors in IPD experiments.

- *It is good (it starts by cooperating)*
- *It is retaliating (it returns the opponent's defection)*
- *It is generous (it forgets the past if the defecting opponent cooperates again).*

Further strategies include stochastic ones ([11]), the Pavlov strategy ([13]) that cooperates when it has played the same move as its opponent etc. In the literature we may also find studies in an evolutionary perspective ([6]), theoretical or applied biological studies ([3][5][10]) and studies of modified IPD versions ([14]).

We have designed an adaptive tit-for-tat strategy that is analyzed in detailed elsewhere ([15]) and is shown to score better than pure tit-for-tat and most of the other known strategies. Our initial motivation for this work was to find a strategy that would be able to cooperate with cooperative behaviors and defect against defective behaviors, while maintaining reversible, not permanent memory. This strategy would demonstrate behavioral gradualness and possess the potential for stability in front of changing worlds. In what follows, we are studying the issue of stability of the adaptive tit-for-tat strategy in spatial games of the sort introduced by Nowak and May ([12]).

2   The Adaptive Tit-for-Tat Strategy

An adaptive tit-for-tat strategy should be essentially tit-for-tat in the sense of being good, retaliating and forgiving. Moreover, it should demonstrate some behavioral gradualness that would show as fewer oscillations between C and D. To this end, it should use an estimate of the opponent's behavior, whether cooperative or defecting, and should react to it in a tit-for-tat manner. The estimate should be continuously updated throughout the interaction with the opponent. The above may be modeled with the aid of a continuous variable, the world's image, ranging from 0 (total defection) to 1 (total cooperation). Intermediate values will represent degrees of cooperation and defection. The adaptive tit-for-tat model can then be formulated as a simple linear model :

> *Adaptive tit-for-tat basic model*
> *If (opponent played C in the last cycle) then*
>   *world = world + r*(1-world), r is the adaptation rate*
> *else*
>   *world = world + r*(0-world)*
> *If (world >= 0.5) play C, else play D*

The usual tit-for-tat model corresponds to the case of r=1 (immediate convergence to the opponent's current move). Clearly, the use of fairly small r's will allow more gradual behavior and will tend to be more robust to perturbations. We have observed that the adoption of two different rates, one for cooperation (r_c) and one for defection (r_d), allows us to assume the right behavior against retaliating or irrational strategies. Retaliating strategies are those that basically seek cooperation, but start by exploring the opponent's reaction to a few initial D moves. For example, the suspicious tit-for-tat (or STFT) strategy starts by defecting, and then plays usual tit-for-tat. On the

contrary, irrational strategies are those that do not employ any feedback from the game and play blindly using some innate law. For example, periodic strategies repeat patterns of C's and D's, such as CDD, CCD, CDCD etc. We have shown ([15]) that a high cooperation rate and a low defection rate allow us to converge to cooperation against a retaliating agent, while the opposite setting allows us to converge to total defection against an irrational agent, in both cases achieving maximum score. Thus, we need a method for the adaptive tit-for-tat agent to discover whether the opponent uses a retaliating behavior or is just irrational, so as to adopt the proper rate setting. We have designed and examined several such variants for estimating the opponent's irrationality and we have finally found the following rule :

> *Adaptive tit-for-tat model*
> *If (opponent played C in the last cycle) then*
>   *world = world + r_c*(1-world), r_c is the adaptation rate for cooperation*
> *else*
>   *world = world + r_d*(0-world), r_d is the adaptation rate for defection*
> *If (world >= 0.5) play C, else play D*
> *Throughout an observation window, record how many times (n) the agent's move has coincided with the opponent's move. At regular intervals (every "window" steps) adapt the rates as follows :*
> *If (n>threshold) then*   *$r_c = r_{min}$, $r_d = r_{max}$*
>   *else*   *$r_c = r_{max}$, $r_d = r_{min}$*

The rule may be translated as :

> *If (the world is cooperative)* then*   *$r_c = r_{min}$, $r_d = r_{max}$*
> *else*   *$r_c = r_{max}$, $r_d = r_{min}$*
> *(*) recall that "my move = opponent's move" is the so-called pavlovian criterion of cooperation (Nowak and Sigmund [13])*

Note that the agent drops its cooperation rate when the world is assumed cooperative, and increases it otherwise, that is, it uses negative feedback at the rate regulation level. Another alternative interpretation for a cooperative world is a world that tries to manipulate the agent (so as to get it to respond with the same value found in the world). In this case it makes sense to drop the cooperation (potential manipulation) rate and become less adaptive to the world.

We have shown in simulations ([15]) that the adaptive tit-for-tat strategy with this meta-regulation mechanism converges to the proper behavior and achieves the highest score against both retaliating and irrational strategies, such as STFT and periodic strategies, respectively.

The adaptive tit-for-tat strategy manages to differentiate between a retaliating strategy and an irrational one that has initially the same behavior. For example it manages to converge to total defection against CDCD, that resembles STFT in the beginning.

The meta-regulated model is insensitive to the initial value of its world variable, provided that it is at least equal to 0.5 (remember that a tit-for-tat like behavior should start by cooperating). However, even for a defective initial value of the world variable, the adaptive agent may converge to cooperative behavior against the tit-for-tat agent. The model is also insensitive to the exact values of r_{max} and r_{min}. Different values for the two rates will only result in scaling or stretching of the resulting curve for the world variable, the qualitative performance remaining intact. The same thing applies

to the values of the observation window and the threshold, although they must be constrained so that the window will be sufficiently large and the threshold sufficiently small compared to the window.

3   Spatial Games I : Basic Model

We have run spatial games ([12][9]) in two-dimensional grids consisting of diverse agents, i.e., agents having different IPD strategies. In these games, all agents run a finite IPD game against their neighbors at each cycle. After all games are finished, the agent at each place assumes the winner strategy among itself and its neighbors, i.e., the strategy whose score sum is greatest. The spatial game continues in the next cycle with the new generation of strategies and so on. Such experiments show the evolution of different types of strategies in spatial grids. In our work, we assume the classical IPD game with the typical payoff matrix already presented and we study the behavior of our adaptive tit-for-tat strategy and its potential for stability.

Experiments starting with randomly placed agents of three kinds (ALLC, ALLD and adaptive TFT) have been found to quickly settle on a stable configuration consisting of high percentages of adaptive TFT and very low percentages of ALLC with occasional small stable territories occupied by ALLD. Figure 1 shows such a stable configuration with a stable ALLD territory and the corresponding experiment's evolution of strategies proportions in the population. As is obvious from figure 1, ALLD's safe territories lie always between ALLCs (that ALLD can exploit) and adaptive TFTs (that ALLD cannot cheat).
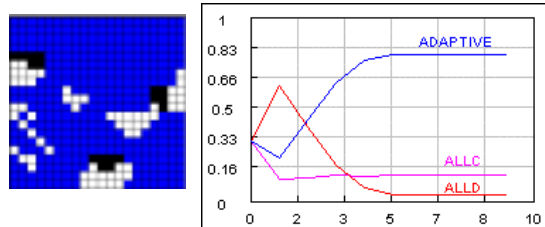
Fig. 1. (a) A stable configuration of adaptive TFT, ALLC and ALLD, that has occurred after a few cycles in a spatial IPD game (20x20 grid). Blue (Gray) = Adaptive TFT, White = ALLC, Black = ALLD. ALLD has built a safe territory. Adaptive TFT has dominated to the detriment of ALLD and ALLC has been invaded by ALLD. (b) Corresponding evolution of strategies.


The experiment has been repeated with populations of various strategies and it has been found that most strategies die out quickly, i.e., in a few generations, while ALLC, ALLD, pure TFT and adaptive TFT persist.

We have also repeated the experiments for initially regular configurations of strategies and we have obtained similar results. Figures 2 to 5 show characteristic results for some regular initial configurations. In figure 2, ALLD builds a stable territory, while in figure 3 (with similar initial population composition) ALLD dies out. Figure 4 demonstrates the process of safe ALLD territory emergence through "chasing" by adaptive TFT agents. Finally, figure 5 gives a type of almost regular configuration that

may lead to two completely different final stable configurations : either to a 100% ALLD population, or to a 100% adaptive TFT population. In all cases, we have observed the emergence of stable configurations of varying composition and form, which may include ALLD blocks next to cooperative strategies.

In order to examine the potential of adaptive TFT as far as stability and resistance to perturbations is concerned, we have performed two kinds of perturbation studies. First, we have performed the "injection test", where, after the population has stabilized, we inject a number of ALLD agents in random places on the grid. In this case, we have found that the system restabilizes to a new configuration, where the adaptive TFT strategy has risen to higher values than its former stable value, and this to the detriment of ALLC. Figure 6a gives the results of a typical run of this experiment.

The second kind of test that we have performed is the "shuffle test", where, after the population has stabilized, we shuffle the agents of the population, i.e. we move each one of them in a random place on the grid. In this case as above, we have found that the system restabilizes to a new configuration, where the adaptive TFT strategy has risen to higher values than its former stable value, and this to the detriment of ALLC. Figure 6b gives the results of a typical run of this experiment. As is obvious from the figure, the shuffling of the agents has the potential of leading ALLD agents to total extinction. The conclusion from both these tests is that a perturbation of a stable population -either in the form of injection of irrational agents or in the form of shuffling of the agents- results in a rise of the adaptive tit-for-tat population. This happens to the detriment of other strategies and most notably to the detriment of the ALLC strategy.
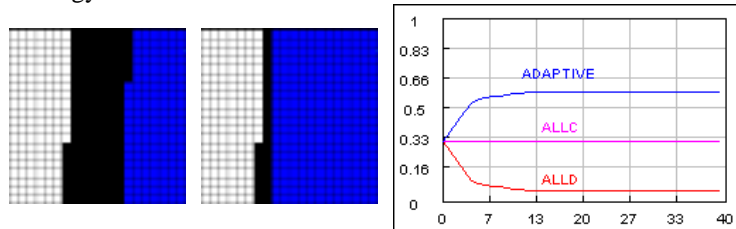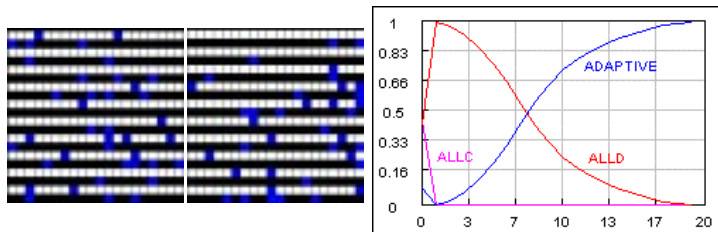

Fig. 2. (a) Initial configuration of adaptive TFT, ALLC and ALLD in a spatial IPD game (20x20 grid). (b) Final configuration. (c) Evolution of strategies in the population.
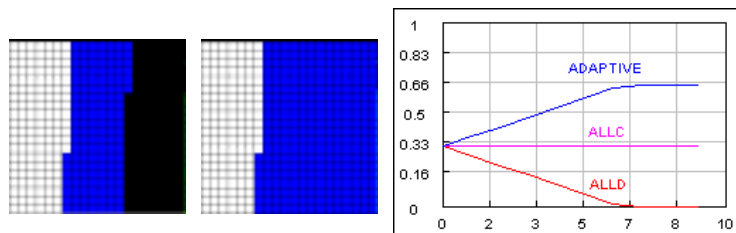

Fig. 3. (a) Initial configuration of adaptive TFT, ALLC and ALLD in a spatial IPD game (20x20 grid). Note that the relative positions of ALLD and adaptive TFT blocks are inversed, in comparison to the situation of figure 2. (b) Final configuration. (c) Evolution of strategies in the population.
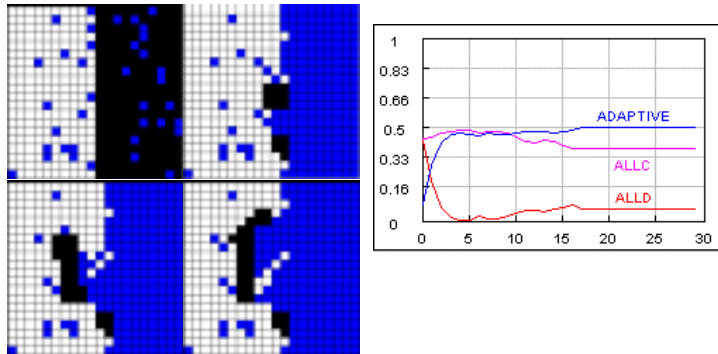
Fig. 4. Configuration of adaptive TFT, ALLC and ALLD in a spatial IPD game (20x20 grid). (a-d) Configuration at t=0, t=6. t=12, t=18. (e) Evolution of strategies in the population.

Fig. 5. A regular configuration of adaptive TFT, ALLC and ALLD in a spatial IPD game (20x20 grid). (a) Configuration leading to a 100% ALLD population. (b) Configuration leading to a 100% adaptive TFT population. (c) Evolution of strategies in the population of the configuration (b).
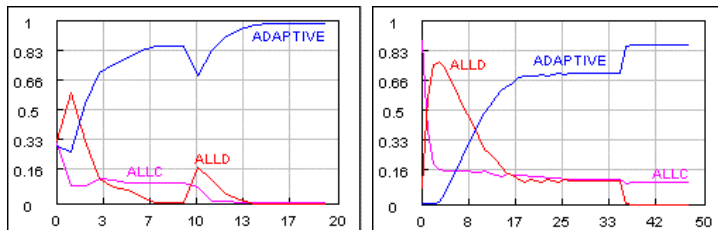
Fig. 6. Perturbation tests in a population of adaptive TFT, ALLC and ALLD in a spatial IPD game (20x20 grid). (a) Injection of ALLD agents at t=10, after the system has stabilized (t=7). (b) Agent shuffling at t=35, after the system has stabilized (t=28).

# 4   Spatial Games II : Model Extensions

We have also performed experiments with the modified game models of May et al. [9] because they represent some kind of "perturbed game models". First, we have experimented with games of probabilistic winning, where the winner strategy in an agent's neighborhood is not determined in an absolute way, as the one achieving the highest score. Instead, it is determined probabilistically according to its score (the

higher the score, the higher the probability of determining the strategy as winner). In this case as before, the system settles to a configuration, where the adaptive TFT strategy rises to high values to the detriment of ALLC. There are, however, three major differences.

- The system does not settle definitely to a configuration, but keeps changing all the time, while the relative proportions of the various strategies are stable around some value.
- The stabilization of the system is much slower than that of absolute winning and a little "noisy", which shows in irregularity of the resulting curves.
- The ALLD strategy gets always extinct, because of symmetry breaking : All ALLD agents are eventually replaced by suboptimal ALLC or adaptive TFT neighbors.

Figure 7 gives the results of probabilistic winning in a random and a regular initial configuration.
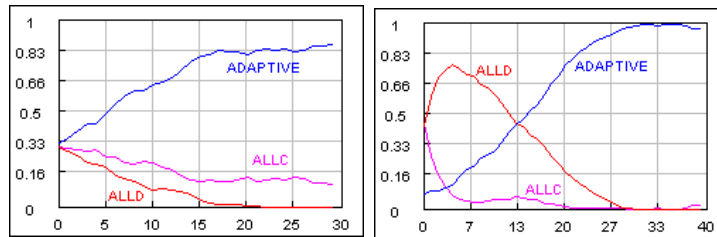
Fig. 7. Evolution of strategies in a population of adaptive TFT, ALLC and ALLD in a spatial IPD game (20x20 grid) with probabilistic winning. (a) A random initial configuration such as the one of figure 1. (b) A regular initial configuration such as the one of figure 5.

Next, we have experimented with games of spatial irregularity, where the opponents of an agent are not necessarily its immediate neighbors or not all of them. In our implementation, an agent has 50% probability of playing against each one of its immediate neighbors, so that practically at each round it plays against half of them, chosen randomly. In this case the general conclusions of previous game models apply as well. A detailed analysis in the same spirit as that of probabilistic winning reveals the following :

- The stabilization of the system is slower than that of absolute winning but quicker than that of probabilistic winning. The stabilization is steady, as that of absolute winning, and not "noisy", as that of probabilistic winning.
- The ALLD strategy gets always extinct, because every ALLD agent is eventually replaced by an ALLC or adaptive TFT neighbor that has achieved a higher score in its own different spatial tournament.

Figure 8 gives the results of spatial irregularity in a random and a regular initial configuration.

Finally, we have experimented with continuous-time games ([7][9]), where agents do not play against their opponents simultaneously and are not updated synchronously, but they play and are updated one after the other in random order. In this case the general conclusions of previous game models apply as well. A detailed analysis in the same spirit as that of probabilistic winning reveals the following :

- Stabilization is comparable in speed and "noisy", as in probabilistic winning.
- However, the ALLD strategy is not always extinct, but manages to build safe territories, although a little less often than in the default game model.

Figure 9 gives the results of continuous-time games in a random and a regular initial configuration. The conclusion from the above three studies is that a perturbed game model results in higher adaptive tit-for-tat population proportions. This happens to the detriment of other strategies and most notably to the detriment of the ALLC strategy.
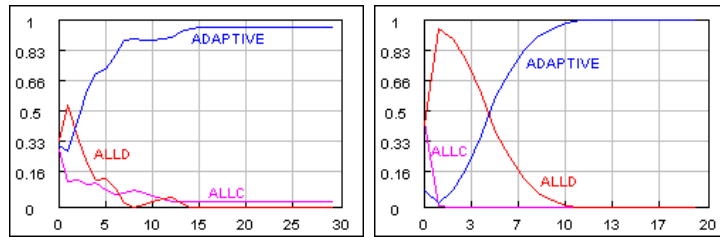
Fig. 8. Evolution of strategies in a population of adaptive TFT, ALLC and ALLD in a spatial IPD game (20x20 grid) with spatial irregularity. (a) A random initial configuration such as the one of figure 1. (b) A regular initial configuration such as the one of figure 5.
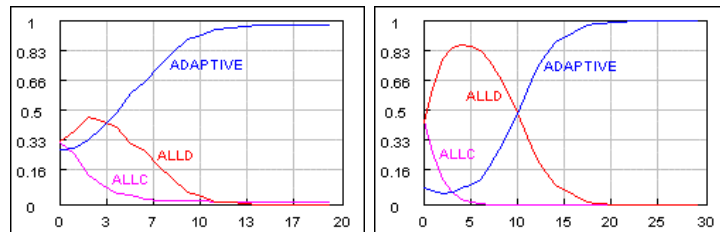
Fig. 9. Evolution of strategies in a population of adaptive TFT, ALLC and ALLD in a spatial IPD game (20x20 grid) with continuous time. (a) A random initial configuration such as the one of figure 1. (b) A regular initial configuration such as the one of figure 5.


5   Spatial Games III : Adaptive TFT versus Pure TFT

Because the adaptive TFT strategy has been shown better than pure TFT in individual games against other strategies as well as in tournaments ([15]), we have examined their relative strength in front of perturbations. There are two notable results in this direction. First, we have performed experiments with arbitrary initial populations including a fair amount of pure and adaptive TFT agents, in the basic game model. As has been already explained, these systems settle eventually to a stable configuration. After such systems stabilize, we inject at regular intervals a number of agents of arbitrary strategy. At each such "injection", the adaptive TFT strategy rises rapidly to the detriment of ALLC and pure TFT. In figure 10a, the initial configuration is a random one that contains 20% STFT, 20% ALLC, 20% CDCD, 20% ALLD, 10% pure TFT and 10% adaptive TFT agents. The system settles at t=5. At t=10, t=15 and t=20, we inject 10% CDCD, 10% lunatic (random) and 10% STFT agents, respectively. As is obvious from the figure, at each perturbation the proportion of the

adaptive TFT rises to the detriment of ALLC and pure TFT. Hence, perturbations strengthen the adaptive TFT and weaken the pure TFT strategy.

A second observation concerns the relative evolution of adaptive TFT and pure TFT in initial configurations consisting of ALLC, ALLD, pure TFT, adaptive TFT and irrational agents (such as CDCD and STFT). In such experiments, it has been observed that the pure and the adaptive TFT strategies stabilize to different values even if starting from the same initial value. More specifically, when the two strategies start from about the same initial value, they settle to a configuration such that the adaptive TFT strategy is of a significantly higher value than the pure TFT strategy, as figure 10b shows. This is an early indication of some sort of Baldwin effect ([4][8]) that deserves further study.
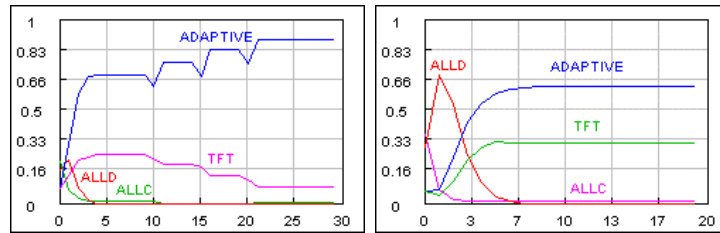
Fig. 10. (a) Evolution of strategies in an initial population that contains 20% STFT, 20% ALLC, 20% CDCD, 20% ALLD, 10% pure TFT and 10% adaptive TFT agents in a spatial IPD game (20x20 grid). At t=10, t=15 and t=20, we inject 10% CDCD, 10% lunatic (random) and 10% STFT agents, respectively. (b) Evolution of strategies in an initial population that contains 7% STFT, 38% ALLC, 7% CDCD, 32% ALLD, 8% pure TFT and 8% adaptive TFT agents in a spatial IPD game (20x20 grid). See text for comments on Baldwin effect.

6   Conclusion

We have presented an adaptive tit-for-tat strategy for the IPD game and have studied its stability and resistance to perturbations in spatial IPD games on a 2D grid. The adaptive TFT strategy uses a continuous estimate of the opponent's behavior that it regulates between two bounds. Moreover, at a meta level, it regulates the rates of the previous level regulation, so that it manages to achieve maximum scores in individual games against other strategies and in tournaments. The behavior of the adaptive TFT strategy in spatial IPD games is studied systematically and several perturbation studies are performed, in an attempt to identify the strategy's features and advantages. In typical spatial IPD games between adaptive TFT, ALLC and ALLD on a 2D grid, a population settles to a certain configuration, where adaptive TFT rises to high percentages in the population, while ALLC and ALLD fall to very low values. Surprisingly at first hand, ALLD manages occasionally to survive in such environments and build stable blocks between large ALLC and adaptive TFT areas. As far as the form of resulting final configurations is concerned, even slightly dissimilar initial configurations may lead to completely different final configurations (cf. fig. 5), thus exhibiting some form of chaotic behavior.

Similar results are also obtained in the case of perturbed or noisy game models (probabilistic winning, spatial irregularities and continuous-time games), with differences in convergence speed and steadiness. The relative strength and dominance of adaptive TFT compared to pure TFT is finally demonstrated in a number of experiments that involve subtle perturbation schemes. In all cases, persistent perturbations of various forms strengthen the adaptive TFT and weaken pure TFT and ALLC, while they lead to extinction of irrational strategies in the population. Adaptive TFT demonstrates thus a selective advantage in perturbed environments, because it is fairly stable and resistant to perturbations.

Future work includes a study of parameterized games, in the same spirit as May et al. [9] and an extension of the IPD spatial game model to a two-species game, each one having as a goal to dominate the whole grid.

References

1. Axelrod, R., and Hamilton, W.D.: The evolution of cooperation, Science 211 (1981) 1390-96
2. Axelrod, R.: The evolution of cooperation. Basic Books (1984)
3. Axelrod, R., and Dion, D.: The further evolution of cooperation, Science 242 (1988) 1385-90
4. Baldwin, J.M.: A new factor in evolution, American Naturalist, 30 (1896) 441-451
5. Feldman, M.W., and Thomas, E.A.C.: Behavior-dependent contexts for repeated plays of the prisoner's dilemma II: Dynamical aspects of the evolution of cooperation, Journal of Theoretical Biology 128 (1987) 297-315
6. Fogel, D.: Evolving behaviors in the iterated prisoner's dilemma, Evolutionary Computation 1 (1993) 77-97
7. Huberman, B.A., and Glance, N.S.: Evolutionary games and computer simulations, Proceedings of the National Academy of Sciences, 90 (1993) 7712-7715
8. Hinton, G.E. and Nowlan, S.J.: How learning can guide evolution, Complex Systems, 1 (1987) 495-502
9. May, R.M., Bohoeffer, S. and Nowak, M.A.: Spatial games and evolution of cooperation, Advances in Artificial Life, Proceedings of the 3rd European Conference on Artificial Life, by F. Morán, A. Moreno, J.J.Merelo and P. Chacón (eds.), Springer Verlag (1995)
10. Milinski, M.: Tit for tat in sticklebacks and the evolution of cooperation, Nature 325 (1987) 433-435
11. Nowak, M.A., and Sigmund, K.: Tit for tat in heterogeneous populations, Nature 355 (1992) 250-53
12. Nowak, M.A., and May, R.M.: Evolutionary games and spatial chaos, Nature 359 (1992) 826-29
13. Nowak, M.A., and Sigmund, K.: A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game, Nature 364 (1993) 56-58
14. Stanley, E.A., Ashlock, D., and Tesfatsion, L.: Iterated Prisoner's Dilemma with Choice and Refusal of Partners, Artificial Life III, Addison-Wesley (1994)
15. Tzafestas, E.: Toward adaptive cooperative behavior, Proceedings of the Simulation of Adaptive Behavior Conference, Paris, September (2000)