PHYSICAL REVIEW E **69**, 036110 (2004)

Self-organization in a simple model of adaptive agents playing 2×2 games with arbitrary payoff matrices

H. Fort[1] and S. Viola[2]

[1]*Instituto de Física, Facultad de Ciencias, Universidad de la República, Iguá 4225, 11400 Montevideo, Uruguay*

[2]*Instituto de Física, Facultad de Ingeniería, Universidad de la República, Julio Herrera y Reissig 565, 11300 Montevideo, Uruguay*

(Received 3 October 2003; published 23 March 2004)

We analyze, both analytically and numerically, the self-organization of a system of ''selfish'' adaptive agents playing an arbitrary iterated pairwise game (defined by a 2×2 payoff matrix). Examples of possible games to play are the *prisoner's dilemma* (PD) game, the *chicken* game, the *hero* game, etc. The agents have no memory, use strategies not based on direct reciprocity nor ''tags'' and are chosen at random, i.e., geographical vicinity is neglected. They can play two possible strategies: cooperate (C) or defect (D). The players measure their success by comparing their utilities with an estimate for the expected benefits and update their strategy following a simple rule. Two versions of the model are studied: (1) the deterministic version (the agents are either in definite states C or D) and (2) the stochastic version (the agents have a probability c of playing C). Using a general master equation we compute the equilibrium states into which the system self-organizes, characterized by their average probability of cooperation c_{eq}. Depending on the payoff matrix, we show that c_{eq} can take five different values. We also consider the mixing of agents using two different payoff matrices and show that any value of c_{eq} can be reached by tuning the proportions of agents using each payoff matrix. In particular, this can be used as a way to simulate the effect of a fraction d of ''antisocial'' individuals—incapable of realizing any value to cooperation—on the cooperative regime hold by a population of neutral or ''normal'' agents.

DOI: 10.1103/PhysRevE.69.036110 PACS number(s): 89.75.Fb, 87.23.Ge, 89.65.Gh

I. INTRODUCTION

Complex systems pervade our daily life. They are difficult to study because they do not exhibit simple cause-and-effect relationships and their interconnections are not easy to disentangle.

Game theory has been demonstrated to be a very flexible tool to study complex systems. It coalesced in its *normal form* [1] during the second World War with the work of von Neumann and Morgenstern [2] who first applied it in economics.

Later, in the 1970s, it was the turn of biology mainly with the work of Maynard-Smith [3], who showed that the game theory can be applied to various problems of evolution, and proposed the concept of evolutionary stable strategy (ESS), as an important concept for understanding biological phenomena. Following rules dictated by game theory to attain an ESS requires neither consciousness nor a brain. Moreover, a recent experiment found that two variants of a RNA virus seem to engage in two-player games [4].

This opens a new perspective, perhaps the dynamic of very simple agents, of the kind we know in physics which can be modeled by game theory providing an alternative approach to physical problems. For instance, energies could be represented as payoffs and phenomena such as phase transitions understood as many-agent games. As a particular application of this line of thought we have seen recently a proliferation of papers addressing the issue of *quantum games* [5–8] which might shed light on the hot issue of quantum computing. Conversely, physics can be useful to understand the behavior of adaptive agents playing games used to model several complex systems in nature. For instance, in some interesting works Szabó *et al.* [9,10] applied the sophisticated techniques developed in nonequilibrium statistical physics to spatial evolutionary games.

The most popular exponent of game theory is the *prisoner's dilemma* (PD) game introduced in the early 1950s by Flood and Dresher [11] to model the social behavior of ''selfish'' individuals—individuals who pursue exclusively their own self-benefit.

The PD game is an example of a 2×2 game in normal form: (i) there are two players, each confronting two choices—to cooperate (C) or to defect (D) (ii) with a 2×2 matrix specifying the payoffs of each player for the four possible outcomes: [C,C],[C,D],[D,C], and [D,D][1], and (iii) each player makes his choice without knowing what the other will do. A player who plays C gets the ''reward'' R or the ''sucker's payoff'' S depending if the other player plays C or D respectively, while if he plays D he gets the ''temptation to defect'' T or the ''punishment'' P depending if the other player plays C or D, respectively. These four payoffs obey the relations

$$T>R>P>S \tag{1}$$

and

$$2R>S+T. \tag{2}$$

Thus independently of what the other player does, by Eq. (1), defection D yields a higher payoff than cooperation C ($T>R$ and $P>S$) and is the *dominant strategy*. The outcome

[1][X,Y] means that the first player plays X and the second player plays Y (X and Y=C or D).

[D,D] is thus called a Nash equilibrium [12]. The dilemma is that if both defect, then both do worse than if both had cooperated $(P<R)$. Condition (2) is required in order that the average utilities for each agent of a cooperative pair (R) are greater than the average utilities for a pair exploitative exploiter $[(T+S)/2]$.

Changing the rank order of the payoffs—the inequalities (1)—gives rise to different games. A general taxonomy of 2×2 games (one-shot games involving two players with two actions each) was constructed by Rapoport and Guyer [13]. A general 2×2 game is defined by a payoff matrix M^{RSTP} with payoffs not necessarily obeying the conditions[2] (1) or (2)

$$M^{RSTP}=\begin{pmatrix} (R,R) & (S,T) \\ (T,S) & (P,P) \end{pmatrix}. \tag{3}$$

The payoff matrix gives the payoffs for *row* actions when confronting with *column* actions.

Apart from the PD game there are other some well studied games. For instance, when the damage from mutual defection in the PD is increased so that it finally exceeds the damage suffered by being exploited,

$$T>R>S>P, \tag{4}$$

the new game is called the *chicken* game. Chicken is named after the car racing game. Two cars drive towards each other for an apparent head-on collision. Each player can swerve to avoid the crash (cooperate) or keep going (defect). This game applies thus to situations such that mutual defection is the worst possible outcome (hence an unstable equilibrium).

When the reward of mutual cooperation in the chicken game is decreased so that it finally drops below the losses from being exploited,

$$T>S>R>P, \tag{5}$$

it transforms into the *leader* game. The name of the game stems from the following every day life situation: Two car drivers want to enter a crowded one-way road from opposite sides, if a small gap occurs in the line of the passing cars, it is preferable that one of them takes the lead and enters into the gap instead of that both wait until a large gap occurs and allows both to enter simultaneously.

In fact, every payoff matrix, which at a first glance could seem unreasonable from the point of view of selfish individuals, can be applicable to describe real life situations in different realms or contexts. Furthermore, ‘‘unreasonable’’ payoff matrices can be used by minorities of individuals which depart from the normal ones (assumed to be neutral) for instance, absolutely D individuals incapable of realizing any value to cooperation or absolutely C ‘‘altruistic’’ individuals (more on this later).

In one-shot or nonrepeated games, where each player has a dominant strategy, as in the PD, then generally these strategies will be chosen. The situation becomes more interesting when the games are played repeatedly. In these *iterated games* players can modify their behavior with time in order to maximize their utilities as they play, i.e., they can adopt different strategies. In order to escape from the noncooperative Nash equilibrium state of social dilemmas it is generally assumed either memory of previous interactions [14] or features (‘‘tags’’) permitting cooperators and defectors to distinguish one another [15]; or spatial structure is required [16].

Recently, a simple model [17] of selfish agents was proposed without memory of past encounters, without tags and with no spatial structure playing an arbitrary 2×2 game, defined by a general payoff matrix such as Eq. (3). At a given time t, each of the N_{ag} agents, numbered by an index i, has a probability $c_i(t)$ of playing C $[1-c_i(t)$ of playing D]. Then a pair of agents are selected at random to play. All the players use the same measure of success to evaluate if they did well or badly in the game which is based on a comparison of their utilities U with an estimate of the expected income ϵ and the arithmetic mean of payoffs $\mu\equiv(R+S+T+P)/4$. Next, they update their $c_i(t)$ in consonance, i.e., a player keeps his $c_i(t)$ if he did well or modifies it if he did badly.

Our long term goal is to study the quantum and statistical versions of this model, that is, on one hand to compare the efficiency and properties of quantum strategies vs the classical ones for this model in a spirit similar to that of Ref. [5]. On the other hand, we are also interested in the effect of noise, for instance by introducing a Metropolis Monte Carlo temperature, and the existence of power laws in the space of payoffs that parametrize the game, of the type found in Refs. [9,10], for a spatial structured version of this model. Before embarking on the quantum or statistical mechanics of this model, the objective in this paper is to complete the study of the simplest nonspatial mean field version. In particular, to present an analytic derivation of the equilibrium states for any payoff matrix, i.e., for an arbitrary 2×2 game using elemental calculus, both for the deterministic and stochastic versions. In the first case the calculation is elementary and serves as a guide to the more subtle computation of the stochastic model. These equilibrium states into which the system self-organizes, which depend on the payoff matrix, are of three types: ‘‘universal cooperation’’ or ‘‘all C,’’ of intermediate level of cooperation and ‘‘universal defection’’ or ‘‘all D’’ with, respectively, $c_{eq}=1.0$, $0<c_{eq}<1.0$ and 0.0. We also consider the effect of mixing players using two different payoff matrices. Specifically, a payoff matrix producing $c_{eq}=0.0$ and the canonical payoff matrix are used to simulate, respectively, absolutely D or ‘‘antisocial’’ agents and ‘‘normal’’ agents.

II. THE MODEL

We consider two versions of the model introduced in Ref. [17]. First, a deterministic version, in which the agents are always in definite states either C or D, i.e., ‘‘black and white’’ agents without ‘‘gray tones.’’ Nevertheless, it is often remarked that this is clearly an oversimplification of the behavior of individuals [18]. Indeed, their levels of cooperation exhibit a continuous gamma of values. Furthermore, com-

[2]We will maintain the letters R,S,T, or P to denote the payoffs in order to keep the PD standard notation.

pletely deterministic algorithms fail to incorporate the stochastic component of human behavior. Thus, we consider also a stochastic version, in which the agents only have probabilities for playing C. In other words, the variable c_i, denoting the state or "behavior" of the agents, for the deterministic case takes only two values $c_i = 1$ (C) or 0 (D) while for the stochastic case c_i is a real variable $\in [0,1]$.

The pairs of players are chosen randomly instead of being restricted to some neighborhood. The implicit assumptions behind this are that the population is sufficiently large and the system connectivity is high. In other words, the agents display high mobility or they can experiment interactions at a distance (for example, electronic transactions, etc.). This implies that N_{ag}, the number of agents, needs to be reasonably large. For instance, in the simulations presented in this work the population of agents will be fixed to $N_{ag} = 1000$.

The update rule for the c_k of the agents is based on comparison of their utilities with an estimate. The simplest estimate ϵ_k that agent number k can make for his expected utilities in the game is provided by the utilities he would make by playing with himself[3] that is:

$$\epsilon_k^{RSTP}(t) = (R - S - T + P) c_k(t)^2 + (S + T - 2P) c_k(t) + P, \tag{6}$$

where c_k is the probability that in the game the agent k plays C. From Eq. (6) we see that the estimates for C agents ($c_k = 1$) ϵ_C and D agents ($c_k = 0$) ϵ_D are given by

$$\epsilon_C = R, \qquad \epsilon_D = P. \tag{7}$$

The measure of success we consider here is slightly different from the one considered in Ref. [17]: To measure his success each player compares his profit $U_k(t)$ with the maximum between his *estimate* $\epsilon_k(t)$, given by Eq. (6), and the arithmetic mean of the four payoffs given by $\mu \equiv (R + S + T + P)/4$.[4] If $U_k^{RSTP}(t) \geqslant (<) \max\{\epsilon_k^{RSTP}, \mu\}$ the player assumes he is doing well (badly) and he keeps (changes) his $c_k(t)$ as follows: if player k did well he assumes his $c_k(t)$ is adequate and he keeps it. On the other hand, if he did badly he assumes his c_k is inadequate and he changes it (from C to D or from D to C in the deterministic version).

We are interested in measuring the average probability of cooperation c vs time, and in particular in its value of equilibrium c_{eq}, after a transient which is equivalent to the final fraction of C agents f_C.

[3] One might consider more sophisticated agents which have "good" information (statistics, surveys, etc.) from which they can extract the average probability of cooperation at "real time" $c(t)$ to get a better estimate of their expected utilities. However, the main results do not differ from the ones obtained with this simpler agents.

[4] The reason to include the mean μ is to cover a wider range of situations than the ones permitted by the so-called *Pavlov's* rule. Pavlov strategy consists in sticking to the former move if it earned one of the two highest payoffs but to switch in the contrary case. The measure considered here reduces to it when $R > \mu > P$.

III. COMPUTATION OF THE EQUILIBRIUM STATES

A. Deterministic version

For the deterministic case the values of c_{eq} are obtained by elementary calculus as follows. Once equilibrium has been reached, the transitions from D to C, on average, must equal those from C to D. Thus, the average probability of cooperation c_{eq} is obtained by equalizing the flux from C to D, J_{CD}, to the flux from D to C, J_{DC}. The players who play C either they get R (in [C,C] encounters) or S (in [C,D] encounters), and their estimate is $\epsilon_C = R$; thus, according to the update rule, they change to D if $R < \mu$ or $S < \max\{R, \mu\}$, respectively. For a given average probability of cooperation c, [C,C] encounters occur with probability c^2 and [C,D] encounters with probability $c(1-c)$. Consequently, J_{CD} can be written as

$$J_{CD} \propto a_{CC} c^2 + a_{CD} c(1-c), \tag{8}$$

with

$$a_{CC} = \theta(\mu - R) \quad \text{and} \quad a_{CD} = \theta(\max\{R, \mu\} - S), \tag{9}$$

where $\theta(x)$ is the step function given by

$$\theta(x) = \begin{matrix} 1 & \text{if} & x \geqslant 0 \\ 0 & \text{if} & x < 0. \end{matrix} \tag{10}$$

On the other hand, the players who play D either they get T (in [D,C] encounters) or P (in [D,D] encounters) and their estimate is $\epsilon_D = P$; thus, according to the update rule, they change to C if $T < \max\{\mu, P\}$ or $P < \mu$, respectively. As [D,C] encounters occur with probability $(1-c)c$ and [D,D] encounters with probability $(1-c)^2$, J_{CD} can be written as

$$J_{DC} \propto a_{DC} (1-c) c + a_{DD} (1-c)^2, \tag{11}$$

with

$$a_{DD} = \theta(\mu - P) \quad \text{and} \quad a_{DC} = \theta(\max\{P, \mu\} - T). \tag{12}$$

In equilibrium

$$J_{CD}(c_{eq}) = J_{DC}(c_{eq}), \tag{13}$$

and thus we get a set of second-order algebraic equations for c_{eq}:

$$(a_{CC} - a_{CD} + a_{DC} - a_{DD}) c_{eq}^2 + (a_{CD} - a_{DC} + 2a_{DD}) c_{eq} - a_{DD} = 0. \tag{14}$$

As there are two possibilities for each coefficient a_{XY}, we have a total of $2^4 = 16$ different equations governing all the possible equilibrium states (actually there are 15 since this includes the trivial equation $0 \equiv 0$). The roots[5] of these equations are

[5] The real roots $\in [0,1]$.

$$0$$
$$\frac{3-\sqrt{5}}{2}$$
$$1/2$$
$$\frac{\sqrt{5}-1}{2}$$
$$1. \tag{15}$$

In addition, we have to take into account the case when

$$a_{CC}=a_{DD}=0$$
$$a_{CD}=a_{DC}=1. \tag{16}$$

In this case we can see from Eq. (8) and (11) that $J_{CD} \equiv J_{DC}$ *identically*, so we have that $p_{eq}\equiv c_o$ (c_o being the initial mean probability), whatever the initial conditions are.

For instance, for the canonical payoff matrix we have $a_{CC}=0=a_{DC}$ and $a_{CD}=1=a_{DD}$, therefore we get

$$c_{eq}(1-c_{eq})=(1-c_{eq})^2, \tag{17}$$

with the root $c_{eq}=1/2$ corresponding to the stable dynamic equilibrium in which the agents change their state in such a way that, on average, half of the transitions are from C to D and the other half from D to C.

B. Stochastic version

In the case of a continuous probability of cooperation c_k, the calculation is a little bit more subtle: now the estimate ϵ_k for the agent k is not only R or P, as it happened in the discrete case, but it can take a continuum of values as the probability c_k varies in the interval [0,1]. From now on we will use the estimate as given in Eq. (6), but instead of a ϵ_k as a function of time we will use a generic ϵ that is a function of the cooperation probability (and implicitly of time, of course), that is,

$$\epsilon^{RSTP}(c)=(R-S-T+P)c(t)^2+(S+T-2P)c(t)+P. \tag{18}$$

So we have

$$\epsilon_k^{RSTP}(t)=\epsilon^{RSTP}(c_k(t)). \tag{19}$$

To calculate c_{eq} we begin by writing a balance equation for the probability $c_i(t)$. The agents will follow the same rule as before: they will keep their state if they are doing well (in the sense explained earlier) and otherwise they will change it. If two agents i and j play at time t, with probabilities $c_i(t)$ and $c_j(t)$, respectively, then the change in the probability c_i, provided he knows $c_j(t)$, would be given by

$$\begin{aligned} c_i(t+1)-c_i(t)=&-c_i(t)c_j(t)[1-\theta(R-\epsilon^{RSTP}(c_i(t)))\,\theta(R-\mu)]-c_i(t)[1-c_j(t)]\\ &\times[1-\theta(S-\epsilon^{RSTP}(c_i(t)))\,\theta(S-\mu)]\\ &+[1-c_i(t)]c_j(t)\,[1-\theta(T-\epsilon^{RSTP}(c_i(t)))\,\theta(T-\mu)]\\ &+[1-c_i(t)][1-c_j(t)][1-\theta(P-\epsilon^{RSTP}(c_i(t)))\,\theta(P-\mu)], \end{aligned} \tag{20}$$

θ being the step function. The equation of evolution for $c_j(t)$ is obtained by simply exchanging $i \leftrightarrow j$ in Eq. (20). Certainly, the assumption that each agent knows the probability of cooperation of his opponent is not realistic. Later, when we perform the simulations, we will introduce a procedure to estimate the opponent's probability (more on this in Sec. V B)

In Eq. (20) if at time t the payoff obtained by agent i, X $(=R,S,T,$ or $P)$, is less than $\max\{\epsilon^{RSTP}(c_i(t)),\mu\}$, the first two terms on the right-hand side decrease the cooperation probability of agent i, while the two last terms increase it. The terms give no contribution if the payoff X is greater than or equal to $\max\{\epsilon^{RSTP}(c_i(t)),\mu\}$.

We will use the canonical payoff matrix M^{3051} to illustrate how the above equation of evolution for $c_i(t)$ works. In this case, the estimate function is, by Eq. (18)

$$\epsilon^{3051}(c)=-c^2+3c+1, \tag{21}$$

thus it is easy to see that

$$\begin{aligned} &\theta(3-\epsilon^{3051}(c))=1 \quad \forall\, c\in[0,1],\\ &\theta(0-\epsilon^{3051}(c))=0 \quad \forall\, c\in[0,1],\\ &\theta(5-\epsilon^{3051}(c))=1 \quad \forall\, c\in[0,1],\\ &\theta(1-\epsilon^{3051}(c))=0 \quad \forall\, c\in(0,1]. \end{aligned} \tag{22}$$

In addition we have for this case $\mu=2{,}25$, thus

$$\begin{aligned} &\theta(3-\mu)=1,\\ &\theta(0-\mu)=0,\\ &\theta(5-\mu)=1,\\ &\theta(1-\mu)=0. \end{aligned} \tag{23}$$

We can then write, to a very good approximation [we are assuming that the last line of Eq. (22) is valid for $c=0$ also],

$$c_i(t+1)-c_i(t)=-c_i(t)[1-c_j(t)]+[1-c_i(t)][1-c_j(t)]$$
$$=[1-c_j(t)][1-2c_i(t)] \quad \forall i\neq j. \tag{24}$$

Defining the mean probability of cooperation as

$$c=\frac{1}{N_{ag}}\sum_{i=1}^{N_{ag}} c_i, \tag{25}$$

summing Eq. (24) over i and j leads to

$$c(t+1)-c(t)=[1-c(t)][1-2c(t)]=1-3c(t)+c(t)^2, \tag{26}$$

within an error of $O(1/N_{ag})$ since Eq. (24) is valid $\forall\, i\neq j$ but we are summing over all the N_{ag} agents.

Thereof we can calculate the equilibrium mean probability of cooperation c_{eq}:

$$0=1-3c(t)+c(t)^2, \tag{27}$$

obtaining the two roots

$$c_{eq}=\begin{cases}1\\ 1/2,\end{cases} \tag{28}$$

$c_{eq}=1/2$ being the stable solution. Hence we obtain the same result as that in the deterministic case.

Using analog reasoning for the general case, we can conclude that if

$$X\notin[\mu,\epsilon_{max}^{RSTP}] \tag{29}$$

or

$$\mu>\epsilon_{max}^{RSTP} \tag{30}$$

the results for the mean cooperation probability for the deterministic version and the stochastic version are the same.

There is an easy way to evaluate ϵ_{max}^{RSTP} in practice. It can be seen—see the Appendix—that if

$$S+T>2\max\{R,P\} \quad \Rightarrow \epsilon_{max}^{RSTP}=P-\frac{1}{4}\frac{(S+T-2P)^2}{(R-S-T+P)} \tag{31}$$

while, if

$$S+T\leqslant 2\max\{R,P\} \quad \Rightarrow \epsilon_{max}^{RSTP}=\max\{R,P\}. \tag{32}$$

When there is a payoff X such that

$$X\in[\mu,\epsilon_{max}^{RSTP}] \tag{33}$$

things can change because agents who get X update in general their probability of cooperation $c_i(t)$ differently depending whether $X<\epsilon(c_i)$ or $X\geqslant\epsilon(c_i)$. So as the probability takes different values in the interval $[0,1]$, we have different equations of evolution, which somehow ‘‘compete’’ against each other in order to reach the equilibrium. The different equations that can appear are of course restricted to the ones generated by the coefficients a_{XY} as they appear in Eq. (14). It is reasonable to expect then that the final equilibrium value for the mean probability will be somewhere in between the original equilibrium values for the equations competing. We will analyze some particular cases of this type in Sec. V B to illustrate this point.

Although at first sight one may think that the universe of possibilities fulfilling condition (33) is very vast, it happens that no more than three different balance equations can coexist. This can be seen as follows: from Eqs. (31) and (32), $\epsilon_{max}^{RSTP}\geqslant\max\{R,P\}$, and besides we know that the estimate never could be greater than all the payoffs, so there is at least one X such that $\epsilon_{max}^{RSTP}<X$. So this leaves us with only two payoffs that effectively can be between μ and ϵ_{max}^{RSTP}, and this results in at most three balance equations playing in a given game.

IV. AN EXAMPLE OF COEXISTENCE OF AGENTS USING DIFFERENT PAYOFF MATRICES: COOPERATION IN PRESENCE OF ‘‘ALWAYS D’’ AGENTS

Let us analyze now the situation where there is a mixing of agents using two different payoff matrices, each leading separately to a different value of c_{eq}. For simplicity we consider the deterministic version but the results for the stochastic version are similar. We call antisocial individuals those for whom cooperation never pays and thus, although they can initially be in the C state, after playing they turn to state D and remain forever in this state. They can be represented by players using a payoff matrix that always updates c_i to 0; for instance M^{1053}. Notice that these individuals are basically different from those which use a payoff matrix fulfilling conditions (1) and (2) who, even though they realize the value of cooperation, i.e., $R>P$ and $2R>T+S$, often may be tempted to ‘‘free ride’’ in order to get a higher payoff. However, with the proposed mechanism—which implies a sort of indirect reciprocity—when D grows above 50% it punishes, on average, this behavior more than C favoring thus a net flux from D to C. Conversely, if C grows above 50% it punishes, on average, this behavior more than D favoring thus the opposite flux from C to D. In other words, small oscillations around $f_C=0.5$ occur. On the other hand, agents using M^{1053} are ‘‘immune’’ to the former regulating mechanism. Let us analyze the effect they have on cooperation when they ‘‘contaminate’’ a population of neutral agents (using the canonical payoff matrix). In short, the two types of individuals play different games (specified by different payoff matrices) without knowing this fact, a situation which does not seem too far from real life.

The asymptotic average probabilities of cooperation can be obtained by simple algebra combining the update rules for M^{3051} and M^{1053}. The computation is completely analogous to the one which leads to Eq. (17). We have to calculate J_{DC} and J_{CD} as a function of the variable c and the parameter d and by equalizing them at equilibrium we get the equation for c_{eq}. To J_{DC} only contribute the fraction $(1-d)$ of normal players using the canonical payoff matrix who play D against a player who also plays D (normal or antisocial). That is, J_{DC} is given by

$$J_{DC}\propto(1-d)(1-c)^2. \tag{34}$$

TABLE I. c_{eq} for agents using M^{3051} contaminated by a fraction d of antisocial agents using M^{1053}.

$c_{eq}(d=0.0)=0.5000$
$c_{eq}(d=0.1)=0.4538$
$c_{eq}(d=0.2)=0.4123$
$c_{eq}(d=0.3)=0.3727$
$c_{eq}(d=0.4)=0.3333$
$c_{eq}(d=0.5)=0.2929$
$c_{eq}(d=0.6)=0.2500$
$c_{eq}(d=0.7)=0.2029$
$c_{eq}(d=0.8)=0.1492$
$c_{eq}(d=0.9)=0.0845$
$c_{eq}(d=1.0)=0.0$

On the other hand, contributions to J_{CD} come from one of these three types of encounters: (i) [C,D] no matter if agents are neutral or antisocial, (ii) [C,C] of two antisocial agents, and (iii) [C,C] of a neutral and antisocial agent (the neutral agent remains C and the antisocial, who started at $t=0$ playing C and has not played yet, changes from C to D). The respective weights of these three contributions are $c(1-c)$, d^2c^2, and $\frac{1}{2}2d(1-d)c^2$. Therefore, J_{CD} is given by

$$J_{CD}\propto c(1-c)+d^2c^2+d(1-d)c^2=c(1-c)+dc^2. \tag{35}$$

In equilibrium $J_{DC}=J_{CD}$ and the following equation for c_{eq} arises:

$$(1-d)(2c_{eq}^2-2c_{eq}+1)+c_{eq}=0, \tag{36}$$

and solving it we get

$$c_{eq}=\frac{3-2d\pm\sqrt{-4d^2+4d+1}}{4(1-d)}. \tag{37}$$

We must take the roots with the "$-$" sign because those with "$+$" are greater than 1 for non-null values of d. We thus get Table I for c_{eq} for different values of the parameter d.

V. SIMULATIONS

A. Deterministic version

In this section we present some results produced by simulation for the deterministic version. Different payoff matrices were simulated and it was found that the system self-organizes, after a transient, in equilibrium states in total agreement with those calculated in Eq. (15).

The update from $c_i(t)$ to $c_i(t+1)$ was dictated by a balance equation of the kind of Eq. (20). The measures are performed over 1000 simulations each and $\overline{c}_{eq}$ denotes the average of c_{eq} over these milliard of experiments. In order to show the independence from the initial distribution of probabilities of cooperation, Fig. 1 shows the evolution with time of the average probability of cooperation for different initial proportions of C agents f_{C0} for the case of the canonical payoff matrix M^{3051} (i.e., $R=3,S=0,T=5$, and $P=1$).

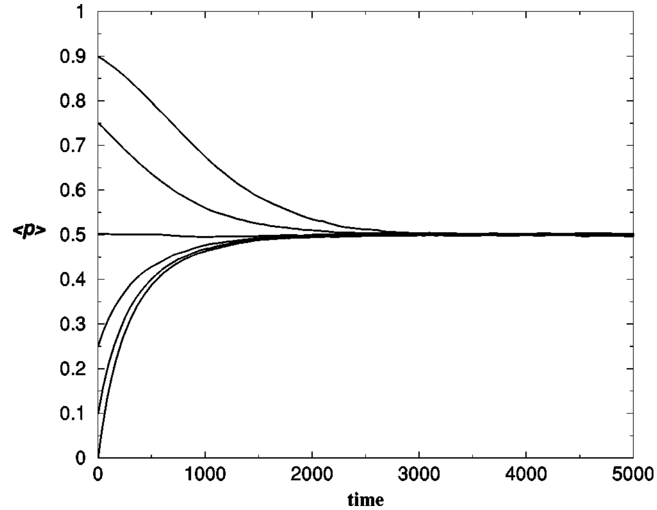

FIG. 1. $\overline{c}$ vs time, for different initial values of f_{C0}, for the canonical payoff matrix.

Depending on the payoff matrix the equilibrium asymptotic states can be of three types: of "all C" ($\overline{c}_{eq}=1.0$, "all D" ($\overline{c}_{eq}=0.0$ or something in between ($0<\overline{c}_{eq}<1$).

We have seen that the canonical payoff matrix M^{3051} provides an example of matrix which gives $\overline{c}_{eq}=0.5$.

Let us see examples of payoff matrices which produce other values of $\overline{c}_{eq}$. A payoff matrix which produces $\overline{c}_{eq}=1.0$ is obtained simply by permuting the canonical values of S (0) and T (5), i.e., M^{3501}. For this matrix we have, by inspection of Eqs. (9) and (12),

$$a_{CC}=a_{CD}=0,\ \ a_{DC}=a_{DD}=1. \tag{38}$$

Hence, after playing the PD game the pair of agents always ends [C,C] since $J_{CD}\equiv 0$ by Eq. (8).

On the other hand, a payoff matrix which leads $\overline{c}_{eq}=0.0$ is obtained simply by permuting the canonical values of R (3) and P (1), i.e., M^{1053}, for which

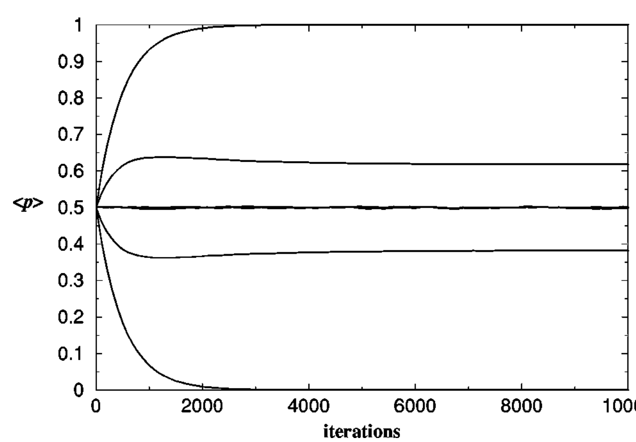

FIG. 2. Curves of $\overline{c}$ vs time for different payoff matrices producing the five possible values of c_{eq} (from below to above): payoff matrices M^{3501} with $c_{eq}=1$, M^{2091} with $c_{eq}\simeq 0.62$, M^{3051} with $c_{eq}=0.5$, M^{2901} with $c_{eq}\simeq 0.38$, and M^{1035} with $c_{eq}=0$.

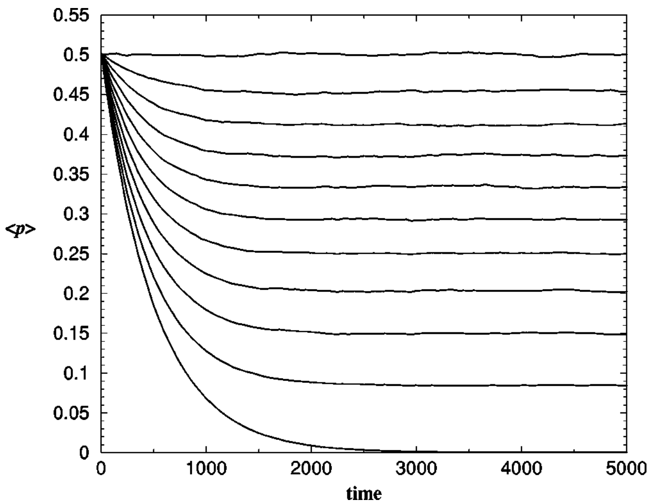

FIG. 3. The evolution of $\overline{c}$ with time, for different values of the fraction d "antisocial" agents (using M^{1053}) embedded in a population of neutral agents (using the canonical payoff matrix).

$$a_{CC}=a_{CD}=1, \quad a_{DC}=a_{DD}=0. \tag{39}$$

That is, all the changes are from C to D since in this case $J_{DC}\equiv 0$.

The rate of convergence to the possible values of $\overline{c_{eq}}$ depends on the values of J_{CD} and J_{DC}.

Figure 2 shows the approach of the average probability of cooperation for different payoff matrices to their final five equilibrium values.

Finally, we simulated the mixing of agents using payoff matrices M^{3051} and M^{1053}. The evolution to equilibrium states for different fixed fractions d of agents using M^{1053} is presented in Fig. 3. The results are in complete agreement with the asymptotic probabilities of cooperation which appear in Table I.

B. Stochastic version

In this case simulations were made updating the probability of cooperation according to Eq. (20). However, as we anticipated, we have to change slightly this equation to reflect reality: two agents i and j interact and they obtain the payoffs X_i and X_j, respectively. For each of them there is no way, from this only event, to know the probability of cooperation c_k of his opponent. What they can do then is to (roughly) estimate this c_k as follows. The player i average utility in an encounter at time t with agent j is given by

$$U_{ij}(t)=R\, c_i(t)c_j(t)+S\, c_i(t)[1-c_j]+T\,[1-c_i(t)]c_j(t) + P\,[1-c_i(t)][1-c_j(t)]. \tag{40}$$

When he plays he gets the payoff X_i, so his best estimate $\tilde{c}_j^i$ for the probability of agent j is obtained by replacing $U_{ij}(t)$ for X_i in Eq. (40). Then he will have

$$\tilde{c}_j^i(t)=\frac{X_i-P+c_i(t)(P-S)}{c_i(t)(R-S-T-P)+T-P}. \tag{41}$$

Exchanging i for j in this equation gives the estimate of the probability $c_i(t)$ that makes agent j. Equation (41) can retrieve any value of $\tilde{c}_j^i(t)$ and not just in the interval $[0,1]$, so it is necessary to make the following replacements:

$$\text{if} \quad \tilde{c}_j^i(t)>1 \;\Rightarrow\; \tilde{c}_j^i(t)=1$$

$$\text{and if} \quad \tilde{c}_j^i(t)<0 \;\Rightarrow\; \tilde{c}_j^i(t)=0. \tag{42}$$

When this happens, the agent is making the roughest approximation, which is to assume that the other player acts like in the deterministic case.

For the canonical payoff matrix, the result was the expected one as this is a matrix obeying condition (29): as predicted by the analytical calculation of Sec. III B, the value for the equilibrium mean probability is $c_{eq}=1/2$ as in the deterministic case, despite the change introduced in Eq. (41). Simulations for other payoff matrices satisfying conditions (29) or (30) were also made and in all the cases the deterministic results were recovered.

We will illustrate the case in which some

$$X\in[\mu,\epsilon_{max}^{RSTP}] \tag{43}$$

with two particular examples. One of them is the case of the normalized matrix M^{1S10}, with S varying from 1 to 2, both limiting cases in which condition (43) ceases to be valid. So for $S\leqslant 1$ the update equation is given simply by

$$c_i(t+1)-c_i(t)=[1-c_i(t)][1-c_j(t)] \tag{44}$$

with $c_{eq}=1$ in this case, while for $S>2$,

$$c_i(t+1)-c_i(t)=1-c_i(t)-c_i(t)c_j(t) \tag{45}$$

for which $c_{eq}=(\sqrt{5}-1)/2$ is the corresponding equilibrium value. When $S\in(1,2]$, both balance equations play a role; the general equation for the update follows from Eq. (20) applied to this particular case:

$$c_i(t+1)-c_i(t)=[1-c_i(t)][1-c_j(t)]+[c_j(t) -2c_i(t)c_j(t)][1-\theta(1-\epsilon)]. \tag{46}$$

So we can see that for $R=T=1\geqslant\epsilon$, Eq. (46) reduces to Eq. (44) while if $R=T=1<\epsilon$ we obtain Eq. (45). When the simulation takes place, c_j has to be replaced by $\tilde{c}_j^i$.

The same analysis can be done for the matrices M^{11T0}, with T varying from 1 to 2 also. In this case the other root competing with $c_{eq}=1$ is $c_{eq}=(3-\sqrt{5})/2$.

The results of the simulations for both cases are presented in Table II, and data for $S>2$ and $T>2$—for which Eq. (29) is valid—are also included.[6]

As it can be seen from the data, for $1<X\leqslant 2$, that is, when condition (43) is valid, the results for the stochastic case are the same as they would be if we were working with

[6] $\overline{c_{eq}}$ corresponds to the average of c_{eq} over 100 experiments.

TABLE II. Equilibrium values for different normalized payoff matrices.

X	$\overline{c_{eq}}$ for $X=S$	$\overline{c_{eq}}$ for $X=T$
1	1	1
1.5	1	1
1.9	1	1
2	1	1
2.1	0.617	0.383
4	0.581	0.370
8	0.556	0.403
16	0.530	0.467
1000	0.548	0.455

the deterministic model. This is a consequence of the estimate (41) together with conditions (42).

For values of T and S greater than 2, for which condition (43) does not hold anymore, we can observe what at first may seem a curiosity: for T or S near 2, the equilibrium values for the deterministic case are recovered as expected, but as we increase the values of T or S, the value of $c_{eq}=1/2$ is approached. After a little thought, it is clear that this is also a consequence of the estimation of Eq. (41), since it depends on the payoffs. It can be easily seen that in the case of M^{1S10},

$$\text{if } T \gg 1 \text{ then } \tilde{c}_j^i \simeq 0 \ \ \forall\ i,j \ \ (\text{for } X_i \neq T). \tag{47}$$

If we take then $c_j=0$ in Eq. (20), and remembering that $T\to\infty$ implies that $\mu\to\infty$, we will obtain that $c_{eq}=1/2$. In an analogous way for M^{1S10},

$$\text{if } S \gg 1 \text{ then } \tilde{c}_j^i \simeq 1 \ \ \forall\ i,j \ \ (\text{for } X_i \neq S) \tag{48}$$

which together with Eq. (20) again leads to $c_{eq}=1/2$. The encounters for which $X_i=S$ or T are responsible for that the exact value $c_{eq}=1/2$ is not attained. A similar analysis can be done when R or $P\to\infty$.

VI. SUMMARY AND OUTLOOK

The proposed strategy, the combination of measure of success and update rule, produces cooperation for a wide variety of payoff matrices.

In particular, we notice that the following facts.

(1) A cooperative regime arises for payoff matrices representing "*social dilemmas*" such as the canonical one. On the other hand spatial game algorithms such as the one of Ref. [16] produce cooperative states ($c_{eq}>0$) in general for the case of a "weak dilemma" in which $P = S = 0$ or at most when P is significantly below R.[7]

(2) Payoff matrices with $R=S=0$ which, at least in principle, one would bet that favor D, actually generate equilibrium states with $c_{eq}\neq 0$, provided that $P<\mu$—see Eqs. (8)-(13).

(3) Any value of equilibrium average cooperation can be reached in principle, even in the case of the deterministic model, by the appropriate mixing of agents using two different payoff matrices. This is an interesting result that goes beyond the different existent social settings. For instance we have in mind situations in which one wants to design a device or mechanism with a given value of c_{eq} that optimizes its performance.

(4) In this work we adopted a *mean field* approach in which all the spatial correlations between agents were neglected. One virtue of this simplification is that it shows the model does not require that agents interact only with those within some geographical proximity in order to sustain cooperation. Playing with fixed neighbors is sometimes considered as an important ingredient to successfully maintain the cooperative regime [16,19]. (Additionally, the equilibrium points can be obtained by simple algebra.)

To conclude we mention different extensions and applications of this model as possible future work. We mentioned, at the beginning, "statistical mechanic" studies. For instance, by dividing the four payoffs between say the reward R reduces the parameters to three: $a=S/R$, $b=T/R$, and $d=P/R$, and we are interested in analyzing the dependence of c_{eq} on each one of these three parameters in the vicinity of a transition between two different values. It is also interesting to introduce noise in the system, by means of an inverse temperature parameter β, in order to allow irrational choices. The player i changes his strategy with a probability W_i given by

$$W_i=\frac{1}{1+\exp[\beta(U_i-\tilde{\epsilon}_i)]},$$

where $\tilde{\epsilon}_i\equiv\max\{\epsilon_i,\mu\}$.

We are planning also a quantum extension of the model in order to deal with players which use superposition of strategies $\alpha_C|C\rangle+\alpha_D|D\rangle$ instead of definite strategies.

The study of the spatial structured version and how the different agents lump together is also an interesting problem to consider. Results on that topic will be presented elsewhere.

Finally, a test for the model against experimental data seems interesting. In the case of humans the experiments suggest, for a given class of games (i.e., a definite rank in the order of the payoffs), a dependency of f_c with the relative weights of R,S,T, and P, which is not observed in the present model. Therefore, we should change the update rule in such a way to capture this feature. Work is also in progress in that direction.

APPENDIX: CALCULUS FOR THE MAXIMUM OF THE GAIN ESTIMATE FUNCTION IN THE STOCHASTIC CASE

We will now show in detail the calculus for the maximum of the gain estimate function $\epsilon^{RSTP}(c)$, restricted to the interval [0,1]. First we have to know if the function has a

[7]In particular, in a spatial game in which each player interacts with his four nearest neighbors, we have checked that the canonical payoff matrix leads to the an "all D" state with $c_{eq}=0$.

maximum in the open interval (0,1). This can be done by noticing that, by Eq. (18), for having negative concavity, we have the condition

$$R-S-T+P<0. \tag{A1}$$

By doing

$$\frac{d}{dc}\epsilon^{RSTP}=0 \tag{A2}$$

we find that the extremum of $\epsilon^{RSTP}(c)$ is attained at

$$c_o=-\frac{1}{2}\frac{(S+T-2P)}{(R-S-T+P)}. \tag{A3}$$

Imposing $c_o>0$, $c_o<1$, and using Eq. (A1) for consistency, we obtain

$$S+T>2P, \quad S+T>2R. \tag{A4}$$

Notice that the sum of these two conditions is equivalent to condition (A1). In turn, Eq. (A4) can be expressed as

$$S+T>2\max\{R,P\} \tag{A5}$$

so this inequality resumes Eqs. (A1) and (A4). It can be seen that if Eq. (A5) is fulfilled, $\epsilon^{RSTP}_{max}\geqslant\mu$ always.

So if condition (A5) holds, the maximum of the function $\epsilon^{RSTP}(c)$ takes place in the interval (0,1) and its value as a function of the parameters R,S,T, and P is

$$\epsilon^{RSTP}_{max}=P-\frac{1}{4}\frac{(S+T-2P)^2}{(R-S-T+P)}. \tag{A6}$$

On the other hand, if

$$S+T\leqslant 2\max\{R,P\} \tag{A7}$$

then

$$\epsilon^{RSTP}_{max}=\max\{R,P\} \tag{A8}$$

since $\epsilon^{RSTP}(0)=P$, $\epsilon^{RSTP}(1)=R$.

Similar conditions can be obtained for the minimum of the function $\epsilon^{RSTP}(c)$ within the interval [0,1]. It can be shown that if $S+T<2\min(R,P)$, the expression in (A6) is a minimum instead of a maximum, and that $\epsilon^{RSTP}_{\min}\leqslant\mu$. If $S+T\geqslant 2\min(R,P)$, then $\epsilon^{RSTP}_{min}=\min(R,P)$.

[1] J. Hofbauer and K. Sigmund, *The Theory of Evolution and Dynamical Systems* (Cambridge University Press, Cambridge, 1988).
[2] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior* (Princeton University Press, Princeton, 1944).
[3] J. Maynard-Smith, *Evolution and the Theory of Games* (Cambridge University Press, Cambridge, 1982).
[4] P.E. Turner and L. Chao, Nature (London) **398**, 441 (1999).
[5] D.A. Meyer, Phys. Rev. Lett. **82**, 1052 (1999).
[6] J. Eisert, M. Wilkens, and M. Lewenstein, Phys. Rev. Lett. **83**, 3077 (1999).
[7] C.F. Lee and N. Johnson, quant-ph/0207012.
[8] C.F. Lee and N. Johnson, Phys. Lett. A **301**, 343 (2002).
[9] G. Szabó and C. Tőke, Phys. Rev. E **58**, 69 (1998).
[10] G. Szabó, T. Antal, P. Szabó, and M. Droz, Phys. Rev. E **62**, 1095 (2000).
[11] M. Flood, Research Memorandum, RM-789-1, 1952, The RAND Corporation.
[12] J. Nash, Ann. Math. **54**, 286 (1951).
[13] A. Rapoport and M. Guyer, Gen. Sys. **11**, 205 (1966).
[14] R. Axelrod, *The Evolution of Cooperation* (Basic Books, New York, 1984).
[15] J. Epstein (unpublished).
[16] M.A. Nowak and R. May, Int. J. Bifurcation Chaos Appl. Sci. Eng. **3**, 35 (1993); M.A. Nowak and R. May, Nature (London) **359**, 826 (1992).
[17] H. Fort, Phys. Rev. E **68**, 026118 (2003).
[18] J. W. Weibull, *Evolutionary Game Theory* (MIT Press, Cambridge, MA, 1995).
[19] M.D. Cohen, R.L. Riolo, and R. Axelrod, Rationality and Society **13**, 5 (2001).