EUROPHYSICS LETTERS 1 August 2003

Europhys. Lett., **63** (3), pp. 319–325 (2003)

Optimal strategies in collective Parrondo games

L. Dinis and J. M. R. Parrondo

Grupo Intedisciplinar de Sistemas Complejos (GISC) and
Departamento de Física Atómica, Molecular y Nuclear
Universidad Complutense de Madrid - 28040 Madrid, Spain

(received 17 March 2003; accepted in final form 26 May 2003)

PACS. 02.50.-r – Probability theory, stochastic processes, and statistics.
PACS. 02.50.Ey – Stochastic processes.
PACS. 05.40.-a – Fluctuation phenomena, random processes, noise, and Brownian motion.

Abstract. – We present a modification of the so-called Parrondo's paradox where one is allowed to choose in each turn the game that a large number of individuals play. It turns out that, by choosing the game which gives the highest average earnings at each step, one ends up with systematic losses, whereas a periodic or random sequence of choices yields a steadily increase of the capital. An explanation of this behavior is given by noting that the short-range maximization of the returns is "killing the goose that laid the golden eggs". A continuous model displaying similar features is analyzed using dynamic programming techniques from control theory.

The physics of Brownian motors has recently inspired the discovery of a counterintuitive phenomenon in gambling games, which is attracting considerable attention. Ajdari and Prost showed that a one-dimensional Brownian particle in a flashing asymmetric potential experiments a net motion in a given direction [1]. The motion persists even against a small force. Consequently, one can have the following startling situation: the particle moves in the direction of the force if the potential is on or if it is off, whereas it moves in the opposite direction if the potential is flashing. The translation of the dynamics of this Brownian particle to gambling games constitutes the so-called *Parrondo's paradox* [2,3]: two losing games yield, when alternated, a winning game. The effect is obtained with two games, A and B, that mimic the behavior of the Brownian particle in a flat and a ratchet potential, respectively.

In game A, the capital $X(t)$ of the player increases by one unit with probability $p_1 = 1/2-\epsilon$, where ϵ is a small positive real number, and decreases by one unit with probability $1-p_1$. In the following, we will interpret the game as a bet on the toss of a slightly biased coin.

Game B is played with two coins depending on $X(t)$: if $X(t)$ is not a multiple of three, we use coin 2, with a probability to win $p_2 = 3/4-\epsilon$ and a probability to lose $1-p_2$; if $X(t)$ is a multiple of three, we use coin 3, with a probability to win $p_3 = 1/10-\epsilon$ and a probability to lose $1-p_3$. It can be proved that the combination of the "good" coin 2 and the "bad" coin 3 yields a fair game when $\epsilon = 0$, and a losing game when $\epsilon > 0$ [3]. By fair, losing and winning here we mean that the average capital $\langle X(t)\rangle$ is a constant, decreasing or increasing function of t, respectively.

We then have two games, A and B, which are fair (losing) if $\epsilon = 0$ ($\epsilon > 0$). The aforementioned counterintuitive effect is that the alternation of A and B, in some given random or periodic sequences, is a winning game.

The phenomenon indicates that the alternation of stochastic dynamics can result in a behavior which differs qualitatively from that exhibited by each of the dynamics, and therefore, could in principle be relevant in a variety of situations, ranging from economics to physics, where the constraints or the dynamics of a system switch between two arrangements [3].

However, the paradox loses all its interest if one is allowed to choose the game to play in each turn. In this case, the trivial strategy is to choose A when $X(t)$ is a multiple of three and B otherwise. The resulting game is clearly winning and this strategy performs better than any periodic or random alternation of games A and B. We could call these latter strategies "blind", since they do not use any information about the state of the system.

In a similar way, if one has some information about the position of the particle in a flashing ratchet, it is possible to switch the potential on and off in such a way that energy is extracted from a single thermal bath. This is nothing but a Maxwell demon [4].

For a Brownian particle, it is known that the acquisition of this information, or its subsequent erasure from a memory device, has some unavoidable entropy cost [5], which prevents any violation of the Second Law of Thermodynamics. On the other hand, in other contexts, like economics, there are no such limitations and it is unlikely that blind strategies could be of any interest.

However, the model that we present in this letter shows that this is not the case. It is a modification of the original Parrondo's paradox in which blind strategies are winning whereas a strategy which chooses the game with the highest average return is losing. Moreover, we will identify the mechanism underlying this counterintuitive behavior and show with a second model that it can also appear in simple deterministic systems.

The two models presented are also of interest in control theory. The choice of a strategy maximizing some quantity is a problem widely treated by optimal control theory, which has been proven to be a powerful tool in a number of disciplines including engineering, physics, chemistry, economics, social sciences, medicine and biology [6, 7]. The counterintuitive phenomenon discussed in this letter, up to our knowledge, has not been reported before and can be relevant in optimization problems involving dynamical systems.

The first model consists of a large number N of players. At each turn of the game, a fraction γ of these players is randomly selected. We are told how much money every player has and we are then allowed to choose a game, A or B, which will be played by *all* γN players in the subset. Our goal is to choose each turn between A or B in order to maximize the average earnings of the players. We consider three different strategies:

– *Periodic strategy*: the game is selected by following a given periodic sequence, for example ABBABB...

– *Random strategy*: the game is chosen randomly with equal probability for both A and B.

– *Short-range (SR) optimal strategy*: the game that will yield the highest average return is chosen.

As we will see below, the third strategy makes use of the available information whereas the periodic and random strategies are blind, in the sense defined above. Surprisingly, these blind strategies produce a systematic winning whereas the SR optimal strategy is losing, as is shown in fig. 1.

A detailed analysis of our model will reveal the underlying mechanism causing this unexpected phenomenon. The key magnitude for this analysis is $\pi_0(t)$, the fraction of players

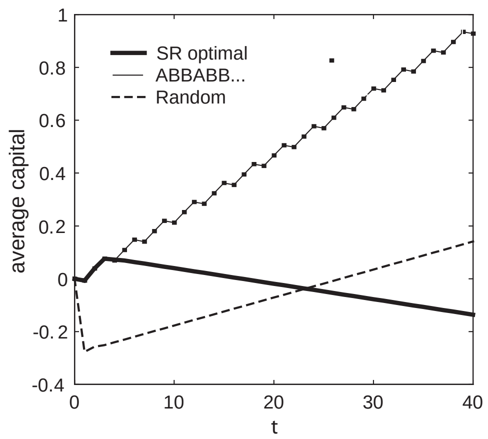

Fig. 1 – Evolution of the average money of an infinity number of players for $\gamma = 0.675$ and $\epsilon = 0.005$ and the three strategies discussed in the text.

whose money is a multiple of three in turn t. From $\pi_0(t)$, it is not difficult to calculate the average fraction of players that would win in each game:

$$
\begin{aligned}
p_{\mathrm{winA}} &= p_1, \\
p_{\mathrm{winB}} &= \pi_0 p_3 + \big(1 - \pi_0\big) p_2.
\end{aligned}
\tag{1}
$$

The SR optimal strategy chooses the game which gives the highest return in one turn. Comparing p_{winA} and p_{winB} we get the following prescription:

$$
\begin{aligned}
&\text{play A} \quad \text{if } \pi_0(t) \geq \pi_{0c}, \\
&\text{play B} \quad \text{if } \pi_0(t) < \pi_{0c},
\end{aligned}
\tag{2}
$$

with $\pi_{0c} \equiv (p_1 - p_2)/(p_3 - p_2) = 5/13$.

Let us focus now on the behavior of $\pi_0(t)$ for $\epsilon = 0$. On the one hand, $\pi_0(t)$ tends to $1/3$ if A is played a large number of turns, because, under the rules of A, the capital $X(t)$ is a symmetric and homogenous random walk. On the other hand, if B is played repeatedly, $\pi_0(t)$ tends to $5/13$, *i.e.*, to π_{0c}. This can be proved by analyzing game B as a Markov chain [3]. Notice also that this coincidence was expected since B is fair game for $\epsilon = 0$ and π_{0c} has been obtained by solving $p_{\mathrm{winB}} = p_{\mathrm{winA}} = 1/2$.

Figure 2 represents schematically the evolution of $\pi_0(t)$ under the action of each game, as well as the prescription of the SR optimal strategy given by eq. (2). Now we are ready to explain why the SR optimal strategy yields worse results than the periodic and random sequences.

Consider an initial distribution of the capital such that $\pi_0(0) < \pi_{0c} = 5/13$. The SR optimal strategy chooses B and, consequently, π_0 increases. If $\pi_0(1)$ is still under $5/13$ (and this is the case for γ small enough), the SR optimal strategy chooses B again. We see that, as far as $\pi_0(t)$ does not exceed $5/13$, the SR optimal strategy chooses B in every turn. However, this choice, although it is the one which gives the highest returns in each turn, drives $\pi_0(t)$ towards $5/13$, *i.e.*, towards values of $\pi_0(t)$ where the gain is small. For instance, if $\gamma = 1/2$, the SR optimal strategy chooses B forty times in a row before switching to game A. This will make π_0 approximately equal to $\pi_{0c} = 5/13$ at almost every turn, as can be seen in fig. 3

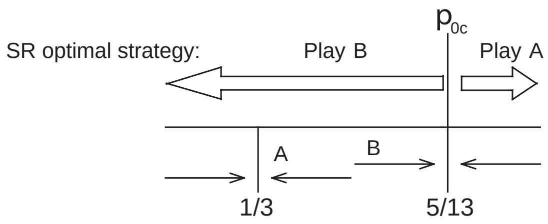

Fig. 2 – Schematic representation of the evolution of $\pi_0(t)$ under the action of game A and game B. The prescription of the SR optimal strategy is also represented.

(left). The same figure (right) shows that, as long as $\pi_0(t)$ is close to π_{0c}, the average capital remains approximately constant.

On the other hand, the random and the periodic strategies choose game A even when $\pi_0 < \pi_{0c}$. This will not produce earnings in this specific turn, but will take π_0 away from π_{0c} and make the corresponding average money grow faster than that for the SR optimal strategy, as can be seen in the figure.

In other words, the SR optimal strategy, by choosing B too many times, is "killing the goose that laid the golden eggs", and to perform better than this strategy one must sacrifice short-term profits for higher returns in the future, as the blind strategies considered here do.

The introduction of ϵ has two consequences: it turns A and B into losing games when played alone, and it decreases the stationary value for game B, which now is *below* $\pi_{0c} = 5/13$ (the critical value π_{0c} does not change). Due to the latter, the system may get trapped playing game B forever when following the SR optimal strategy and, since B is now a losing game, the average money will decrease. This is, for instance, the situation already presented in fig. 1 for $\epsilon = 0.005$ and $\gamma = 0.675$.

Now we present a continuous and deterministic model which displays some of the features

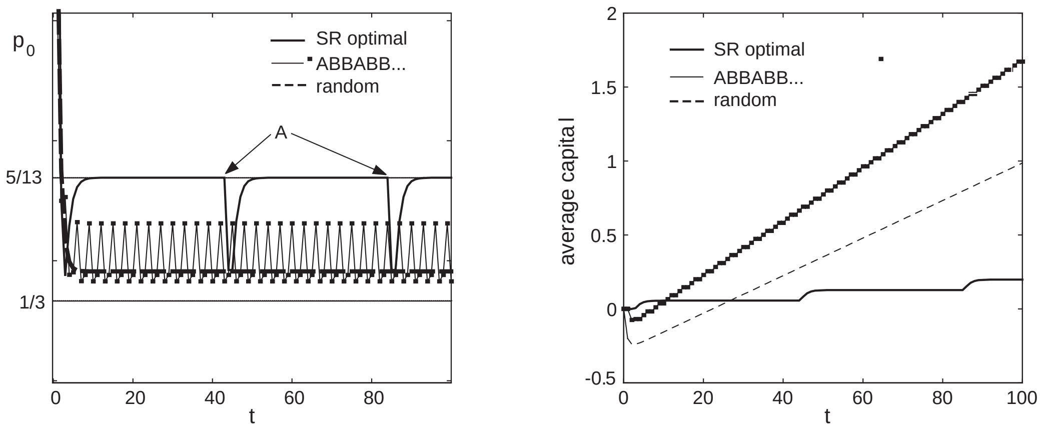

Fig. 3 – Evolution of π_0 (left) and the average capital (right) for $N = \infty$, $\gamma = 0.5$, $\epsilon = 0$ and the three different strategies. The arrows in the left figure show the turns where the short-range optimal strategy chooses game A and they coincide with the steps of the curve representing the capital.

of the previous one. Consider the following dynamical system:

$$\begin{aligned}\dot{y}(t) &= \alpha(t)x(t),\\ \dot{x}(t) &= -\frac{1}{\tau}\big[x(t) - x_{\rm fc}(1-\alpha(t))\big]\end{aligned} \tag{3}$$

with $\alpha(t) = 0$ or 1. Our task is to find $\alpha(t)$ that maximizes $y(T)$. These equations are a rather generic model of a system which produces some output like, for instance, a production plant. $y(t)$ is the total cumulative output of the plant up to time t. We can decide to switch the plant on and off at every time t by setting $\alpha(t) = 1$ or $\alpha(t) = 0$, respectively. Finally, $x(t)$ is the productivity of the plant, which decreases exponentially when the plant is working and goes back to its full capacity value $x_{\rm fc}$ when the plant is off, τ being the characteristic time of these relaxations. If we are allowed to use the plant up to a time T, the problem is to find the protocol or *policy* $\alpha(t)$ maximizing the total output $y(T)$.

A naive approach to the problem consists of maximizing $\dot{y}(t)$ at every time t:

$$\alpha(t) = \begin{cases} 0, & \text{if } x(t) < 0,\\ 1, & \text{if } x(t) \geq 0. \end{cases} \tag{4}$$

However, it is not hard to see that this set-up will keep $y(t) = y(0)$ for all t if initially the productivity $x(0)$ is negative, or make $y(t)$ tend to $y(0) + \tau x(0)$ exponentially if $x(0) > 0$. In either case, $y(t)$ will attain a saturation value. The criterion (4) prescribes making the plant work whenever the productivity is positive, and this is equivalent to the short-range optimal strategy in our previous model, which dictated to play game B if $\pi_0(t) < \pi_{0c}$. If the productivity is positive, we of course get more when the plant is working, but then we also get a decrease of the productivity which will end up exhausting the system. With the prescription given by (4) we are also killing the goose that laid the golden eggs, and it is again possible to do better by letting the plant rest even when the productivity $x(t)$ is positive.

Surprisingly enough and despite the linearity and simplicity of our system, the precise optimal policy $\alpha(t)$ is not easy to find. Some of the techniques provided by control theory fail in this case, such as the Euler-Lagrange equations and the Pontryagin principle [6]. The only way to completely solve the problem is to optimize a discrete version such as

$$\begin{aligned}x_{k+1} &= x_k - h\big(x_k - 1 + \alpha_k\big),\\ y_{k+1} &= y_k + h\alpha_k x_k,\end{aligned} \tag{5}$$

where h is a small time step, and we have taken $x_{\rm fc} = \tau = 1$ for simplicity. The optimization of this discrete system can be done by applying the so-called *Bellman's Optimality Criterion* [6], which states that in the optimal policy, the final decisions α_k $(k = n, \ldots, N)$ are optimal *given* the state resulting from the first decisions, α_k $(k = 0, \ldots, n-1)$. That is, we can find the α_{N-n} maximizing $J_n \equiv y_{N+1} - y_{N-n}$ as a function of x_{N-n}, y_{N-n}, recursively from $n = 0$ to $n = N$ [8].

The optimal choice of $\alpha(t)$ in terms of the state $x(t)$ happens to be

$$\alpha(t) = \begin{cases} 0, & \text{if } x(t) < x^{\rm c}(t),\\ 1, & \text{if } x(t) \geq x^{\rm c}(t), \end{cases} \tag{6}$$

where $x^{\rm c}(t)$ is a critical value for the productivity which can be calculated by solving the resulting recurrence equations. In fig. 4 we show $x^{\rm c}(t)$ for $\tau = x_{\rm fc} = 1$, $h = 0.1$, and $T = 2$,

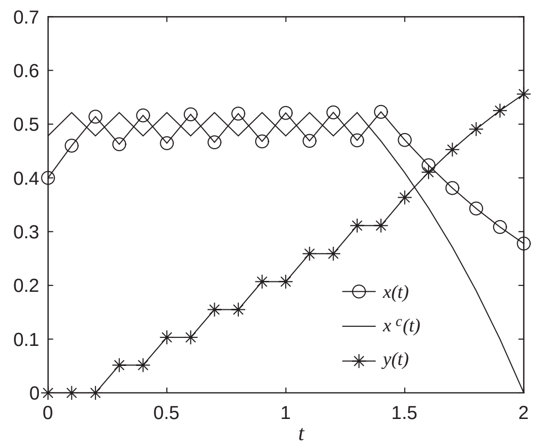

Fig. 4 – The cumulative output $y(t)$, the productivity $x(t)$, and the critical value $x^{\mathrm{c}}(t)$ for the discrete version of the system (3), $T = 2$, $h = 0.1$, and $\tau = x_{\mathrm{fc}} = 1$.

as well as the behaviour of $y(t)$ and $x(t)$ (we have chosen a relatively big time step to make the rapid changes in $\alpha(t)$ and $x(t)$) clear. We see that this optimal policy achieves a steadily increase of the output $y(t)$. Figure 5 shows a numerical computation of $x^{\mathrm{c}}(t)$ now for a shorter time step $h = 0.001$, and different values of the total time, $T = 2$, 3, and 4. Here we can see again that $x^{\mathrm{c}}(t) = 0.5$ until the final part of the interval $[0, T]$.

Consequently, the behaviour of the optimal policy $\alpha(t)$ is as follows. There is a first stage in which the productivity $x(t)$ goes to 0.5 by setting $\alpha = 0$ if $x(0) < 0.5$ and $\alpha = 1$ if $x(0) \geq 0.5$. Once $x(t)$ reaches the value 0.5, the optimal policy prescribes very rapid changes between $\alpha = 0$ and $\alpha = 1$ which keep $x(t) \simeq 0.5$. Finally, in the last part of the interval $[0, T]$, when $x^{\mathrm{c}}(t)$ starts to decrease, the optimal policy chooses $\alpha(t) = 1$.

The first two stages can be easily explained, even for arbitrary values of τ and x_{fc}. For this purpose, let us assume for a moment that $\alpha(t)$ is constant but can take any value within the interval $[0, 1]$. Then $x(t)$ reaches a stationary value $x_{\mathrm{st}} = x_{\mathrm{fc}}(1 - \alpha)$. This gives the following

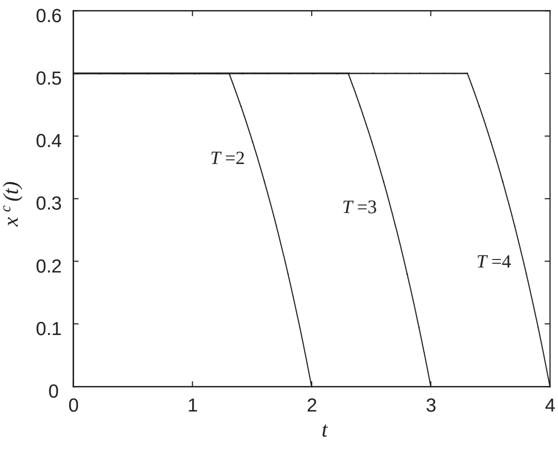

Fig. 5 – Numerical computation of $x^{\mathrm{c}}(t)$, for $h = 0.001$, $T = 2$, 3 and 4, and $\tau = x_{\mathrm{fc}} = 1$.

stationary slope for the cumulative output $y(t)$:

$$\dot{y}(t) = \alpha x_{\rm st} = x_{\rm fc}\big[\alpha(1-\alpha)\big], \tag{7}$$

which is maximum for $\alpha = 0.5$ and $x_{\rm st} = x_{\rm fc}/2$. Therefore, the optimal policy should try to drive $x(t)$ to $x_{\rm fc}/2$ and keep it there, *i.e.*, it should try to make the plant work at half of its full capacity. In our case, $\alpha(t)$ can only take values 0 and 1. However, by rapid oscillations one can get an effective value of $\alpha(t)$ equal to any real number between 0 and 1. The optimal policy implies a rapid variation which gives an effective value $\alpha = 0.5$, yielding a slope for $y(t)$ which is $\dot{y}(t) = x_{\rm fc}/4$, definitely better than the short-range optimization of $\dot{y}(t)$ which gave us a horizontal slope.

The final stage of the optimal policy $\alpha(t)$ has a clear intuitive explanation. We have to abandon the plant at time T. Therefore, the optimal policy when t is approaching T should set $\alpha(t) = 1$, since we want to get as much as possible and do not care if we leave the plant exhausted after T. This is exactly what happens to a middle-distance runner: she keeps a constant velocity which allows her to maintain a stationary regime but she sprints in the last meters of the race to use up all her strength. With this picture in mind, we call this last stage the "sprint". One can calculate the duration of the sprint, $t_{\rm sprint}$, in our model, assuming that $x(0) = x_{\rm fc}/2$, $\alpha(t) = 1/2$ for $t < T - t_{\rm sprint}$ and $\alpha(t) = 1$ for $t > T - t_{\rm sprint}$. Equation (3) can then be fully solved yielding:

$$y(T) = \frac{x_{\rm fc}}{4}\big(T - t_{\rm sprint}\big) + \frac{\tau x_{\rm fc}}{2}\big(1 - e^{-t_{\rm sprint}/\tau}\big). \tag{8}$$

From this expression one easily finds that $y(T)$ reaches its maximum for $t_{\rm sprint} = \tau \ln 2 \simeq 0.693\tau$, in agreement with the curves in fig. 5.

In conclusion, we have presented a stochastic model in which a short-range optimization yields to systematic loses, whereas blind strategies steadily win. We have found an explanation of this phenomenon based on the fact that the short-range optimal strategy is "killing the goose that laid the golden egg", and proven that the same mechanism can also arise in a linear deterministic system. In fact, similar mechanisms have been widely reported in the realm of economics and ecology, although mainly described qualitatively. The risks of overtaxing commerce and the overuse of natural resources are representative cases. We believe that the models presented here could inspire new quantitative approaches to these problems.

* * *

We are grateful to C. Van den Broeck for fruitful discussions. This work has been financially supported by grant BFM2001-0291-C02-02 from the Ministerio de Ciencia y Tecnología (Spain) and by a grant from *Del Amo Program* (Universidad Complutense).

REFERENCES

[1] Ajdari A. and Prost J., *C.R. Acad. Sci. Paris II*, **315** (1993) 1635.
[2] Harmer G. P. and Abbott D., *Nature*, **402** (1999) 846.
[3] Harmer G. P. and Abbott D., *Fluct. Noise Lett.*, **2** (2002) R71.
[4] Parrondo J. M. R. and Cisneros B. J., *Appl. Phys. A*, **75** (2002) 179.
[5] Leff H. S. and Rex A. F., *Maxwell's Demon. Entropy, Information, Computing* (Adam Hilger) 1990.
[6] Layton J. M., *Multivariable Control Theory* (Peter Peregrinus) 1976.
[7] Lewis F. L. and Syrmos V. L., *Optimal Control* (Wiley) 1995.
[8] Dinis L. and Parrondo J. M. R., in preparation.