Learning to Coordinate in a Complex and Nonstationary World

M. Marsili,[1] R. Mulet,[2,*] F. Ricci-Tersenghi,[2] and R. Zecchina[2]

[1]*Istituto Nazionale per la Fisica della Materia (INFM), Unità di Trieste-SISSA, I-34014 Trieste, Italy*
[2]*The Abdus Salam International Center for Theoretical Physics, Condensed Matter Group,*
Strada Costiera 11, P.O. Box 586, I-34100 Trieste, Italy

(Received 22 May 2001; published 24 October 2001)

We study analytically and by computer simulations a complex system of adaptive agents with finite memory. Borrowing the framework of the minority game and using the replica formalism we show the existence of an *equilibrium* phase transition as a function of the ratio between the memory λ and the learning rates Γ of the agents. We show that, starting from a random configuration, a *dynamic* phase transition also exists, which prevents agents from reaching optimal coordination. Furthermore, in a nonstationary environment, we show by numerical simulations that the phase transition becomes discontinuous.

DOI: 10.1103/PhysRevLett.87.208701 PACS numbers: 02.50.Le, 05.20.Dd, 64.60.Ht, 87.23.Ge

Social interactions pose many coordination problems to individuals. Generally social agents face problems of sharing and distributing limited resources in an optimal way. Examples range from the use of public roads and the Internet to exchanging what we produce with what we consume. A solution to a problem of this kind invokes the intervention of a public authority who finds the social optimum and imposes or suggests the optimal behavior to agents. While such a solution may be easy to find, its implementation may be difficult to enforce in practical situations.

Self-enforcing solutions—where agents achieve optimal allocation of resources while pursuing their self-interests, without explicit communication or agreement with others—are of great practical importance. Competitive markets are the prototypical example of such a solution: With everybody maximizing his own profit and no one really caring for global optimality, competitive markets perform the remarkable task of leading to system-wide optimality.

Microeconomics and game theory [1] have gone quite far in explaining what equilibria one can expect in social interactions. However, most of these studies have focused on cases with either few players or with many, but identical, agents. Also, the analysis is restricted to the equilibria which deductively rational players would agree upon. Such an approach seems unrealistic in cases involving many individuals with different goals and characteristics. The computational complexity required by deductive rationality may easily go far beyond the capabilities of agents. It has been argued [2] that bounded rationality and inductive thinking provide more suitable descriptions of how real people behave. A growing effort has been made recently in understanding under what conditions bounded inductively rational agents may reach optimal outcomes. Several learning rules have been found to lead to optimal outcomes when a single agent "plays" against nature [3] or in simple games with few players [4].

In this Letter we address the problem of how many heterogeneous adaptive agents learn to coordinate in a complex, eventually nonstationary, world. We draw inspiration from recent work on the minority game (MG) [5], in order to model a typical situation where a large number of agents pursue different individual goals, using a certain number of distributed resources. Optimal use of resources then becomes a complex coordination problem.

We focus on agents with finite memory and finite learning rates. We find that, when agents need to "learn" collectively a fixed structure of interactions, they can attain a close to optimal coordination, provided that their memory extends far enough into the past. As the memory decreases, the system undergoes a phase transition to a state where agents are unable to learn and play in a random way.

More interestingly we find situations where the agents are unable to coordinate and the game ends in a stationary regime with no cooperation. This is a completely dynamical effect which prevents the system from a proper convergence to equilibrium. In such cases the game theoretic approach, based on the analysis of Nash equilibria—which are those states in which each player's strategy is optimal, given the current strategy of all other players [1]—is useless: Even though Nash equilibria are stable states the dynamics will not converge to them.

This is further clear evidence of the relevance of tools and ideas of statistical mechanics in the study of complex socioeconomic systems; indeed dynamical transitions are well-known phenomena in statistical mechanics [6].

The model we study is closely related to the minority game. The reason for this choice is that this allows us to benefit from the detailed understanding which has been recently uncovered by the statistical mechanics approach [7,8]. On one hand, we can make reference to exact results; on the other, we can extend our understanding of this keystone model of complex adaptive systems.

The model is precisely defined as follows [5,7]: Agents live in a world which can be in one of P states, labeled by an integer $\mu = 1, \dots, P$. Each agent $i = 1, \dots, N$ can choose between two personal strategies, labeled by a spin variable s_i, which prescribe an action $a^{\mu}_{s_i,i}$ for each state μ. These actions are drawn from a bimodal distribution for all i, s, and μ, such that there are two possible actions, do something ($a^{\mu}_{s_i,i} = 1$) or do the opposite ($a^{\mu}_{s_i,i} = -1$).

The payoff received by an agent who plays strategy s_i, while her opponents take strategies $s_{-i} = \{s_j, \forall j \neq i\}$, is, in the state μ,

$$u_i^{\mu}(s_i, s_{-i}) = -a^{\mu}_{s_i,i} A^{\mu}, \tag{1}$$

where $A^{\mu} = \sum_j a^{\mu}_{s_j,j}$. The total payoff to agents is always negative: The majority of agents receives a negative payoff, whereas only the minority of them gains.

The game is repeated many times; as in [9] the state μ is drawn from a uniform distribution $\rho^{\mu} = 1/P$ at each time and agents try to estimate, on the basis of past observations, which of their strategies is the best one. More precisely, if $s_i(t)$ is the strategy played by agent i at time t, we assume as in [10] that

$$\mathrm{Prob}[s_i(t) = s] \propto \exp[\Gamma U_{s,i}(t)], \tag{2}$$

where $U_{s,i}(t)$ is the *score* of strategy s at time t, and Γ is a positive constant [11]. Each agent monitors the scores $U_{s,i}(t)$ of each of her strategies s by

$$U_{s,i}(t+1) = (1 - \lambda/N) U_{s,i}(t) + u_i^{\mu}[s, s_{-i}(t)]/N, \tag{3}$$

where the last term is the payoff agent i would have received if she had played strategy s at time t [see Eq. (1)] against the strategies $s_{-i}(t) = \{s_j(t), \forall j \neq i\}$ played by her opponents at that time.

In other words, Eqs. (2) and (3) model agents who play more likely strategies which have performed better in the past. Equations (2) and (3) belong to a class of learning models which has received much attention recently [12].

The relevant parameter [13] is the ratio $\alpha = P/N$ between the "information complexity" P and the number of agents, and the key quantity we shall look at is σ^2 defined as the time average of $(A^{\mu})^2$ in the stationary state. σ^2 is a measure of the inefficiency of agents' coordination because, per Eq. (1), the total payoff to agents is $-(A^{\mu})^2$. Hence optimal states correspond to minima of σ^2.

This model differs from the MG [5] by two important aspects: First, agents compute correctly the payoff for strategies $s \neq s_i(t)$ which they did not play. In the MG, agents account for only the explicit dependence of u_i^{μ} on s which arises from $a^{\mu}_{s,i}$ [see Eq. (1)], whereas they neglect the fact that if they had taken a different decision A^{μ} would have also changed. This seems reasonable at first sight because A^{μ} is an aggregate quantity and its dependence on each individual agent is weak. A more careful analysis [7,8], however, shows that if agents properly account for their impact on A^{μ} as in Eq. (3) a radically different scenario arises: Rather than converging to a unique stationary state as in the MG, the dynamics (with $\lambda = 0$) converges to one of exponentially many (in N) Nash equilibria, which are characterized by an optimal coordination. This change emerges in the statistical mechanics approach with the breakdown of replica symmetry (RS): While the minority game is described by a replica symmetric theory, Nash equilibria are described by a full replica symmetry broken (RSB) phase [8]. Our aim is precisely that of studying the coordination of adaptive agents in a complex world with exponentially many optimal states (Nash equilibria).

The second key feature is that previous work has explored only the dynamics of learning with an infinite memory [14], i.e., with $\lambda = 0$ in Eq. (3), and for a fixed structure of interactions, i.e., with fixed (quenched) disorder $a^{\mu}_{s,i}$. Our goal is to clarify the role of different time scales involved in the learning dynamics. We shall first study the case where the structure of interactions is fixed—which corresponds to $a^{\mu}_{s,i}$ being the usual quenched disorder—and then move to the more realistic case where the structure of interactions changes over long time scales.

Following the self-consistent approach of Ref. [15], we find that, for $N \gg 1$ and $\Gamma/N \ll 1$, the long time dynamics of $y_i(\tau) = \Gamma[U_{+,i}(t) - U_{-,i}(t)]/2$ in the rescaled continuum time $\tau = \Gamma t/N$ is well approximated by

$$\begin{aligned} \frac{dy_i}{d\tau} &= -\frac{\lambda}{\Gamma} y_i - h_i - \sum_{j\neq i} J_{i,j} \tanh(y_j) + \eta_i(\tau), \\ h_i &= \frac{1}{P} \sum_{\mu=1}^{P} \sum_{j=1}^{N} \frac{a^{\mu}_{+,i} - a^{\mu}_{-,i}}{2} \frac{a^{\mu}_{+,j} + a_{-,j}\mu}{2}, \\ J_{i,j} &= \frac{1}{P} \sum_{\mu=1}^{P} \frac{a^{\mu}_{+,i} - a^{\mu}_{-,i}}{2} \frac{a^{\mu}_{+,j} - a^{\mu}_{-,j}}{2}, \end{aligned} \tag{4}$$

with $\eta_i(\tau)$ being a white noise with zero mean and correlations

$$\langle \eta_i(\tau) \eta_j(\tau') \rangle \simeq \frac{\Gamma \sigma^2}{N} J_{i,j} \delta(\tau - \tau'). \tag{5}$$

The explicit derivation follows the same steps as Ref. [15].

Let us first neglect stochastic fluctuations induced by η_i and consider the deterministic dynamics—which by Eq. (5) is legitimate only for $\Gamma \ll 1$. As in Ref. [7], one finds that the dynamics minimizes the function

$$H = \sigma^2 + \frac{\lambda}{\Gamma} \sum_i [\log(1 - m_i^2) + 2 m_i \tanh^{-1}(m_i)], \tag{6}$$

where $m_i = \langle s_i \rangle = \tanh y_i$. For $\Gamma \ll 1$, one finds [15]

$$\sigma^2 = H_0 + 2 \sum_i h_i m_i + \sum_{j \neq i} J_{i,j} m_i m_j,$$

with H_0 a being constant.

For $\lambda = 0$ the stationary state is described by the minima of σ^2. As shown in Refs. [7,8], σ^2 takes its minima for $m_i = \pm 1$—which correspond to $y_i \to \pm\infty$. These states are Nash equilibria. For $0 < \lambda \ll \Gamma$, coordination between agents persists: Indeed minima occur for $|y_i| \simeq \Gamma/\lambda \gg 1$ which means that agents converge to states close to Nash equilibria. However, when $\lambda/\Gamma \gg 1$ the minima of H are dominated by the second term, i.e., $m_i \approx y_i \approx 0$. In other words, when $\lambda \gg \Gamma$ the agents are unable to coordinate because their memory is too short for learning correctly the interaction structure.

The phase transition which takes place between these two regimes is captured, for $\Gamma \ll 1$, by the statistical mechanics approach of Ref. [7]: In order to study the minima of H we introduce an inverse temperature β, we compute the partition function and the free energy per agent, and then we take averages over the disordered variables $a_{s,i}^{\mu}$ with the replica method [16]. The free energy, within the RS ansatz, reads

$$f(q,r,Q,R) = \frac{\alpha}{\beta}\ln\left[1 + \frac{\beta(Q-q)}{\alpha}\right] + \frac{\alpha}{2}\,\frac{1+q}{\alpha+\beta(Q-q)} + \frac{1-Q}{2} - \frac{1}{\beta}\left\langle \ln\int_{-1}^{1} dm\, e^{-\beta V_z(m)}\right\rangle + \frac{\alpha\beta}{2}(RQ - rq), \qquad (7)$$

where $Q = \frac{1}{N}\sum_i (m_i)^2$ and $q = \langle m_i^a m_i^b\rangle$, with $a \neq b$ labeling different replicas of the systems; R and r arise as Lagrange multipliers and $V_z(m) = -\sqrt{\alpha r}\, mz + \frac{\alpha\beta}{2}(r - R)m^2 + \frac{\lambda}{\Gamma}[\log(1-m^2) + 2m\tanh^{-1}(m)]$. The ground state properties of H are obtained by solving the saddle point equations [7,16] in the limit $\beta \to \infty$.

In the inset of Fig. 1 we compare the analytical predictions for σ^2 and Q with simulation results. We focus on small α (i.e., $\alpha = 0.1$), where the effects we wish to discuss are more evident. Little discrepancies between numerical data and analytical curves might be due to RSB effects. Note that a phase transition occurs at $\lambda_c \simeq 0.46\Gamma$, where both σ^2 and Q change their analytical behavior. We studied this equilibrium phase transition in the $(\lambda, 1/\Gamma)$ plane, confirming the critical line $\lambda_c = 0.46\Gamma$: The open symbols in Fig. 1 refer to a *static* experiment, where we let the system equilibrate to a Nash equilibrium for $\lambda = 0$ and then we move it slowly along lines $\lambda\Gamma = \text{const}$. We locate the phase transition in the point where Q changes its analytic behavior along these lines.

Figure 1 shows that the analytic predictions, derived for $\Gamma \ll 1$, hold in a much wider range of values of Γ. For $\lambda/\Gamma \ll 1$, the stochastic force η_i is unable to contrast the deterministic drift towards Nash equilibria. The only effect of η_i is to induce small stochastic fluctuations of y_i around its average. The phase $\lambda/\Gamma \gg 1$, however, is dominated by the stochastic force η_i. The very lack of coordination—which results in large values of σ^2/N—enhances the noise strength by Eq. (5). This makes fluctuations in the uncoordinated state even stronger.

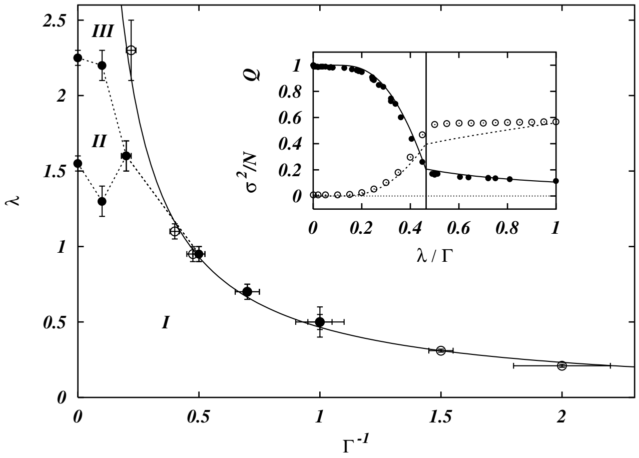

FIG. 1. Phase diagram: static (○) and dynamic (●) critical lines obtained from the simulation. The solid line represents the RS critical line. The dashed lines are guides to the eyes. Inset: Q (●) and σ^2/N (○) as a function of λ/Γ from simulations with $\lambda\Gamma = 0.1, 1, 10$, $\alpha = 0.1$, and $N = 10^3$. The lines represent the RS solution.

Agents may, however, fail to coordinate for $\Gamma \gg 1$ when they start from scratch [i.e., $U_{s,i}(0) = 0\ \forall\{s,i\}$] in each run. In such a situation the dynamics reaches a stationary regime different from the static one, which is characterized by larger fluctuations (i.e., larger σ^2). These dynamical effects make the phase diagram more complex in the $\lambda < \lambda_c$ region (see Fig. 1): In I the system always relaxes to the static equilibrium, in II it sometimes converges to equilibrium and sometimes get trapped in a metastable regime with large fluctuations, while in III it never reaches equilibrium. This dynamical transition is further evidence that an analysis in terms of Nash equilibria may not be enough to predict the collective behavior of a system. Agents may fail to coordinate on Nash equilibria because of purely dynamical effects.

When the external world is nonstationary, i.e., changes with time, the adaptation task becomes still harder. We mimic the external world dynamics as follows: Every τ time steps a state of the world μ is randomly chosen and replaced by a new one. This means that the corresponding values of the strategies $a_{s,i}^{\mu}$ are redrawn at random for all i and s.

Here we focus on the results of the simulations done with $\tau = 10^3$, $\Gamma = \infty$, $NP = 10^4$, and many λ values. The results are not dependent on the initial conditions.

Figure 2 shows how the system relaxes to equilibrium: We define a time dependent σ^2 as the average of $(A^{\mu})^2$ on logarithmic time bins. For $\lambda = 2.5$ (upper panel of Fig. 2), σ^2/N, which is initially $\simeq 1$, converges smoothly to its equilibrium value. With our choice $\tau = 10^3$, the system reaches cooperative behavior before the world starts changing. Hence for $\lambda = 2.5$ the system is robust with respect to slow changes of the world: Apart from occasional excursions to states with large σ^2, agents are able to adapt themselves to the evolving interaction structure.

In the lower panel we present the evolution of σ^2/N for $\lambda = 3.5$ (i.e., with shorter memory) in 50 different samples. The behavior is now completely different: After having reached a low value of σ^2/N (coordination), the system undergoes a sharp transition and σ^2/N jumps to a high value. The players are no longer able to adapt to

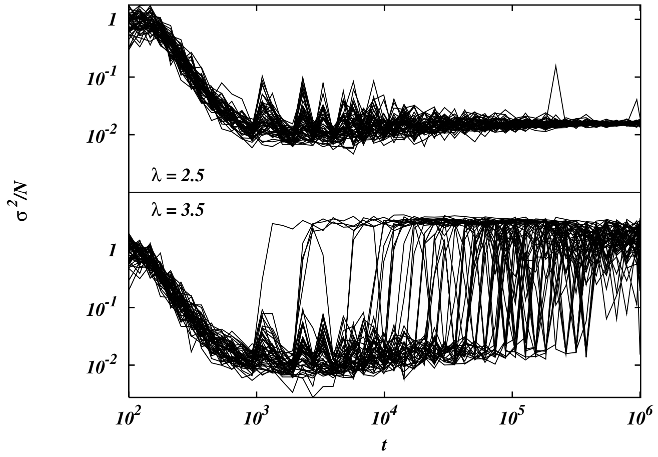

FIG. 2. A nonstationary world ($\tau = 10^3$) with the evolution of σ^2/N with the simulation time for 50 different samples and two values of λ ($NP = 10^4$, $\alpha = 0.1$, and $\Gamma = \infty$).

the changing world and they start playing the wrong way. Occasionally agents may achieve good coordination with small σ^2, but they eventually always go back to uncoordinated states with large σ^2.

For large times, the instantaneous values of σ^2/N have a roughly bimodal distribution: They are either low ($\sim 10^{-2}$) or high (~ 1). In Fig. 3 we plot the average of the low (○) and high (□) values (these averages can be defined in an unambiguous way thanks to the gap between low and high σ^2 values). In the inset we report the fraction of samples that spends the last decade in the high σ^2 regime. In a whole intermediate range around $\lambda_c \approx 3.3$ we find that coordinated states with small σ^2 coexist with wildly fluctuating states ($\sigma^2 > 1$).

It is worth noticing some facts in Fig. 3. The minimum of σ^2, corresponding to the best cooperation, is no longer located in $\lambda = 0$ (i.e., infinite memory). In other words, in a nonstationary environment the agents play better with a *finite* memory, which allows them to make a decision based more on the recent past rather than on the distant past. The minimum they can attain is very near to the σ^2/N value in an unchanging world (shown with a horizontal dotted line in Fig. 3). The second remarkable fact is that the transition from a coordinated state to a high σ^2 regime when λ increases—which was continuous in a fixed world—shows features of first order transitions such as discontinuities and phase coexistence.

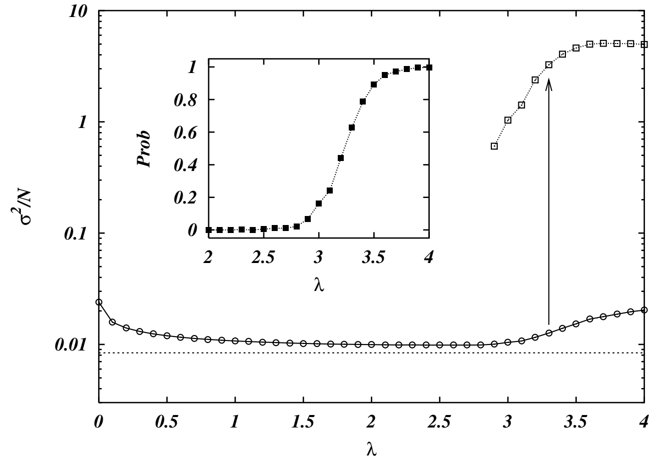

FIG. 3. Average low (○) and high (□) σ^2/N as a function of λ ($NP = 10^4$, $\alpha = 0.1$, $\Gamma = \infty$, and $\tau = 10^3$). The arrow indicates a transition from the cooperative to the noncooperative regime. The horizontal dotted line is the σ^2/N value with fixed world ($\tau = \infty$). Inset: probability of being in a noncooperative regime as a function of λ.

In conclusion, we have extended the replica solution of the minority game to the case where agents have finite memory and finite learning rates. We have proven that a phase transition between phases with low and high σ^2 exists as a function of λ/Γ. We have also shown, by means of computer simulations, that a dynamical phase transition exists for high values of λ (short memories), and that dynamical effects may be responsible for coordination failures. When the structure of the interactions changes slowly, agents with infinite memory behave worse than agents with a finite memory. In addition the transition to noncoordinated states for large λ becomes discontinuous.

*Permanent address: Superconductivity Laboratory, Physics Faculty-IMRE, University of Havana, CP 10400, La Habana, Cuba.

[1] D. Fudenberg and J. Tirole, *Game Theory* (MIT Press, Cambridge, MA, 1991).
[2] H. A. Simon, *Models of Bounded Rationality* (MIT Press, Cambridge, MA, 1982); see also W. B. Arthur, Am. Econ. Assoc. Papers Proc. **84**, 406 (1994).
[3] A. Rustichini, Games Econ. Behav. **29**, 244 (1999).
[4] D. Fudenberg and D. K. Levine, *The Theory of Learning in Games* (MIT Press, Cambridge, MA, 1998).
[5] D. Challet and Y.-C. Zhang, Physica (Amsterdam) **246A**, 407 (1997).
[6] J. P. Bouchaud *et al.,* in *Spin Glasses and Random Fields,* edited by A. P. Young (World Scientific, Singapore, 1998).
[7] D. Challet, M. Marsili, and R. Zecchina, Phys. Rev. Lett. **84**, 1824 (2000); M. Marsili, D. Challet, and R. Zecchina, Physica (Amsterdam) **280A**, 522 (2000).
[8] A. De Martino and M. Marsili, J. Phys. A **34**, 2525 (2001).
[9] A. Cavagna, Phys. Rev. E **59**, R3783 (1999).
[10] A. Cavagna, J. P. Garrahan, I. Giardina, and D. Sherrington, Phys. Rev. Lett. **83**, 4429 (1999).
[11] The probabilistic nature of Eq. (2) does not necessarily model irrational behavior. D. L. Mc Fadden [Ann. Econ. Soc. Measurement **5**, 363 (1976)] has shown that Eq. (2) also describes rational agents who maximize a random utility.
[12] C. Camerer and T.-H. Ho, Econometrica **67**, 827 (1999).
[13] R. Savit, R. Manuca, and R. Riolo, Phys. Rev. Lett. **82**, 2203 (1999).
[14] See, however, M. L. Hart *et al.,* cond-mat/0102384.
[15] M. Marsili and D. Challet, Phys. Rev. E (to be published).
[16] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).