PHYSICAL REVIEW E **70**, 055101(R) (2004)

Error-driven global transition in a competitive population on a network

Sehyo Charley Choe,[1] Neil F. Johnson,[1] and Pak Ming Hui[2]
[1]*Clarendon Laboratory, Physics Department, Oxford University, Oxford OX1 3PU, United Kingdom*
[2]*Department of Physics, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong*
(Received 3 May 2004; published 10 November 2004)

We show, both analytically and numerically, that erroneous data transmission generates a global transition within a competitive population playing the "Minority Game" on a network. This transition, which resembles a phase transition, is driven by a "temporal symmetry breaking" in the global outcome series. The phase boundary, which is a function of the network connectivity p and the error probability q, is described quantitatively by the crowd-anticrowd theory.

DOI: 10.1103/PhysRevE.70.055101 PACS number(s): 02.50.Le, 05.40.−a, 87.23.Kg

The study of complex adaptive systems is enabling physics to expand its boundaries into a range of nontraditional areas within the biological, informational, and socioeconomic communities. Since these areas are often rich in empirical data, they also offer physics a new testing ground for theories of complex systems, and for nonequilibrium statistical mechanics in general [1]. It is widely recognized [2] that Arthur's multiagent "El Farol bar problem" (EFBP) [3] embodies the interacting, many-body nature of many real-world complex systems. In particular, a binary representation of the EFBP, namely, the "Minority Game" (MG) [4–8], plays the role of a new "Ising Model" for theoretical physics.

Despite widespread interest among physicists in biological, informational and socio-economic networks [9], researchers have only just started considering the effect of such networks in the MG and EFBP [7,8]. It has so far been assumed that any information shared between agents is always perfectly accurate. However, real-world complex systems do not operate at such levels of perfection. Furthermore, informational networks might be used by agents to spy and mislead rather than to benefit others. This raises the following question: What is the effect of networks within a competitive population such as the MG, where the information transmitted is corrupted?

Here we show analytically and numerically that erroneous data transmission generates an abrupt global transition within a competitive, networked population playing the Minority Game. This phaselike transition is driven by a "temporal symmetry breaking" in the global outcome series. The crowd-anticrowd theory, which accounts for the many-body (i.e., many-agent) correlations inherent in the system, provides a quantitative yet physically intuitive explanation of this phase transition.

Our model consists of N objects or "agents" who repeatedly compete to be in a minority: for example, commuters striving to choose the least crowded of two routes. The agents can be any form of adaptive object, e.g., biological or mechanical, and our general setup has potential application to a wide range of problems in the biological, informational and social sciences. The minority rule can also be generalized [5]. At each time step t, each agent decides between action +1, meaning to choose option "1," and action −1, meaning to choose option "0." The winning (i.e., minority) outcome at each timestep is 0 or 1. Each agent decides his actions in light of (i) *global information* which takes the form of the history of the m most recent global outcomes, and (ii) *local information* obtained via the cluster to which he is connected, if any. Such connections may be physically tangible (e.g., a telephone or Internet link, or biological structure) or physically intangible (e.g., a wireless communication channel, or biochemical pathway). Adaptation is introduced by randomly assigning S strategies to each agent. Each of the 2^{2^m} possible strategies is a bitstring of length 2^m defining an action (+1 or −1) for each of the 2^m possible global outcome histories $\mu(t)$ [4,5]. For $m=2$ for example, there are $2^{2^{m=2}}=16$ possible strategies and $2^2=4$ possible global outcome histories: $\mu(t)=00$, 01, 10 and 11. Strategies which predicted the winning (losing) action at a given time step are assigned (deducted) one point.

Agents use the connections they have, if any, to gather information from other agents. For simplicity we assume a random network between agents, fixed at the beginning of the game. The connection between any two agents exists with a probability p, hence each agent is on average connected to $p(N-1)$ others. At a given time step t, and with a given global history $\mu(t)$, each agent takes the action predicted by the highest-scoring strategy among his own *and* those of the agents to whom he is connected. The parameter q is the probability that an *error* arises in the information he gathers from his cluster. Alternatively, q can be viewed as the weight an agent places on the information gathered from his cluster. For example, if the actionof the best strategy in his cluster is +1, the agent records this as a −1 with probability q (and vice versa for a best action −1). The information transmission has been corrupted with probability q. Any agent with a higher-scoring strategy than those of his neighbors at a given time step, is unaffected by this error—the only source of stochasticity which might affect him is the standard coin-toss used to break any ties between his own strategies [5]. In contrast to the agents' "on-site" stochastic strategy selection arising in the "Thermal Minority Game" (TMG) [6,10], the stochasticity associated with q in our game depends on the agents' connectivity.

We now investigate the effects of this *microscopic* connection-driven data error on the system's *macroscopic* dynamics. We shall build an analytic theory based on the crowd-anticrowd theory [5] which incorporates the many-agent (many-body) correlations arising in the system's strategy space as a result of the dynamics in the history space. To a very good approximation, we can replace the full strategy space by a reduced strategy space (RSS) [4] which provides a minimal basis set of strategies for the system [5]. The appropriate choice for the RSS depends on the relative frequency of visits to the 2^m histories. With all histories visited equally often, the RSS comprises a total of $2P=2\times 2^m$ strategies [4]. We will examine how the interplay between p and q affects the fluctuations in the excess demand $D(t)$. As emphasized in previous MG works, small fluctuations in $D(t)$ are signatures of an efficient self-organization within the population [4–7]. The excess demand $D(t)$ at time step t is given by

$$D(t) = n_{+1}(t) - n_{-1}(t) = \sum_{K=1}^{K=2P} a_K n_K(t), \tag{1}$$

where $n_{+1}(t)$ $[n_{-1}(t)]$ is the number of agents taking action $+1$ (-1); $a_K=\pm 1$ is the action predicted by strategy K in response to history $\mu(t)$ and $n_K(t)$ is the number of agents using strategy K at time t. K labels the Kth highest scoring strategy while $\overline{K}=2P+1-K$ labels the anticorrelated strategy. In the non-networked MG at low m, $D(t)$ exhibits large, crowd-driven fluctuations while $\mu(t)$ follows a quasideterministic Eulerian trail in which all histories $\{\mu(t)\}$ are visited equally [11]. Hence, the time averages $\langle n_{-1}(t)\rangle=\langle n_{+1}(t)\rangle$, yielding $\langle D(t)\rangle=0$, which is the optimal value for $D(t)$. We continue this focus on small m here, since we are interested in the effect of q on these crowd-driven fluctuations. We will assume that the combined effect of averaging over t for a given Ψ (where Ψ is a given realization of the initial strategy allocation matrix [5]), *and* averaging over Ψ, will have the same effect as averaging over all histories. This is true for the non-networked MG, and produces a mean $D(t)$ of zero. Hence the fluctuation (i.e., standard deviation) of the excess demand, σ_D, is given by

$$\sigma_D^2 = \left\langle \sum_{K=1}^{P} [n_K(t) - n_{\overline{K}}(t)]^2 \right\rangle_{t,\Psi} \approx \sum_{K=1}^{P} (n_K^{\text{mean}} - n_{\overline{K}}^{\text{mean}})^2. \tag{2}$$

We have used the orthogonality properties of the vectors with elements a_K where $K=1,2,\ldots,2P$ [5]. Since $n_K(t)$ will generally fluctuate around some mean value n_K^{mean}, we have also written $n_K(t)=n_K^{\text{mean}}+\epsilon_K(t)$ and assumed that the fluctuation terms $\{\epsilon_K(t)\}$ are uncorrelated stochastic processes. In Eq. (2),

$$n_K^{\text{mean}} = n_K + n_{\to K}^q - n_{K\to}^q - n_K^q \tag{3}$$

and similarly for $n_{\overline{K}}^{\text{mean}}$, where

(1) n_K is the mean number of agents whose own best strategy is actually the Kth highest scoring strategy in the game [5]

$$n_K = N\left[\left(1 - \frac{K-1}{2^{m+1}}\right)^S - \left(1 - \frac{K}{2^{m+1}}\right)^S\right]. \tag{4}$$

(2) $n_{\to K}^q$ is the mean number of agents who only possess strategies worse (i.e., lower scoring) than K, but who will use strategy K due to connections they have to one or more agents who each possess strategy K but no better

$$n_{\to K}^q = (1-q)n_{\to K} + qn_{\to\overline{K}}, \tag{5}$$

where

$$n_{\to K} = \left[\sum_{J>K} n_J\right]\left[(1-p)^{\Sigma_{G<K}n_G}\right]\left[1-(1-p)^{n_K}\right] \tag{6}$$

with $n_{\to\overline{K}}$ being obtained from Eq. (6) by setting $K\to\overline{K}=2P+1-K$.

(3) $n_{K\to}^q$ is the mean number of agents who possess strategy K, but who will nevertheless use a strategy better than K due to connections

$$n_{K\to}^q = (1-q)n_{K\to} + qn_{\overline{K}\to}, \tag{7}$$

where

$$n_{K\to} = n_K\left[1-(1-p)^{\Sigma_{G<K}n_G}\right] \tag{8}$$

and similarly for $n_{\overline{K}\to}$.

(4) n_K^q accounts for the situation in which an agent is connected to other agents with the same highest scoring strategy K as him. q therefore gives the probability that this agent will take the opposite action to strategy K

$$n_K^q = qn_K\left[1-(1-p)^{n_K}\right] - qn_{\overline{K}}\left[1-(1-p)^{n_{\overline{K}}}\right]. \tag{9}$$

Figure 1 compares the numerical and analytical results for σ_D, which is the standard deviation in excess demand. The agreement is remarkable given the complexity of σ_D as a function of p and q. As the "noise" level q increases, the system undergoes a change in regime at a critical connectivity p defined by the critical boundary $C_{\text{crit}}(q,p)$. Moving across $C_{\text{crit}}(q,p)$, the symmetry in the global outcome string is spontaneously broken in a manner reminiscent of a phase transition. Specifically, the global outcome series changes from the low-q phase where it resembles the period-4 Eulerian trail…00110011…, to a high-q phase where it comprises *two* distinct branches [see Fig. 1(a)]. Figure 2 shows individual runs near the critical noise threshold. The higher branch corresponds to a period-2 global outcome series…1010… which is *antiferromagnetic* if we denote 0 (1) as a *spatial* spin up (down) as opposed to a *temporal* outcome. The lower branch corresponds to the period-1 series of "frozen" outcomes…0000…or…1111…, i.e., *ferromagnetic*. In this high-q phase, the system will choose one of these two global outcome branches spontaneously, as a result of the type and number of links each agent has. This symmetry breaking of the global outcome series along the channel of minimum fluctuation in Fig. 1, $C_{\text{crit}}(q,p)$, originates in the internal coupling between the history dynamics, the strategy space and the individual agent networks. Because of the initial strategy allocation and connections, many agents will have a built-in bias towards one of the two possible actions and hence act in a deterministic or "decided" way at a given time step. However, there exist a few "undecided" agents who need to toss an unbiased coin to decide between the equally balanced signals they gather from their local net-

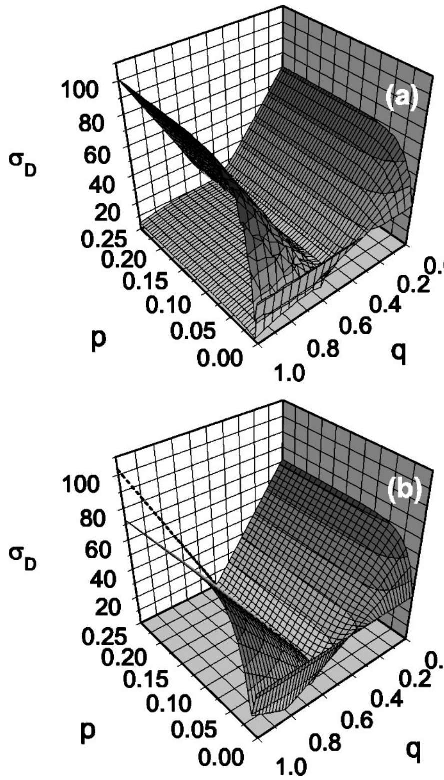

FIG. 1. Fluctuation in excess demand, σ_D, as a function of the error probability q and the network connectivity p. (a) Numerical results averaged over 300 runs, each with 10^5 iterations. (b) Analytic crowd-anticrowd theory. At high q, the two branches in (a) correspond to different dynamical attractor states, while the single branch in (b) represents an effective average (see text). The dotted line in (b) at $p=0.25$ illustrates the modified analytical results for the upper branch if one assumes some knowledge of this branch's global output series. Parameters: $m=1$, $S=2$, and $N=101$.

work. It is the fluctuations of these few undecided agents who then push the system onto a particular branch.

We now discuss two technical details. First, there are many ties in strategy scores at very small m, and hence many tie-breaking coin tosses. This means that the fluctuation terms $\{\epsilon_K(t)\}$ can no longer be ignored. The $m=1$ surface in Fig. 1(b) was therefore produced by averaging over the $2\times 2^m=4$ time steps in the Eulerian trail [11]. In other words,

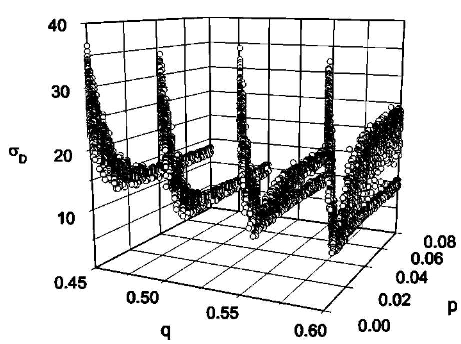

FIG. 2. Numerical results for individual runs, showing the fluctuation in excess demand, σ_D, around $q=0.5$. Parameters as in Fig. 1.

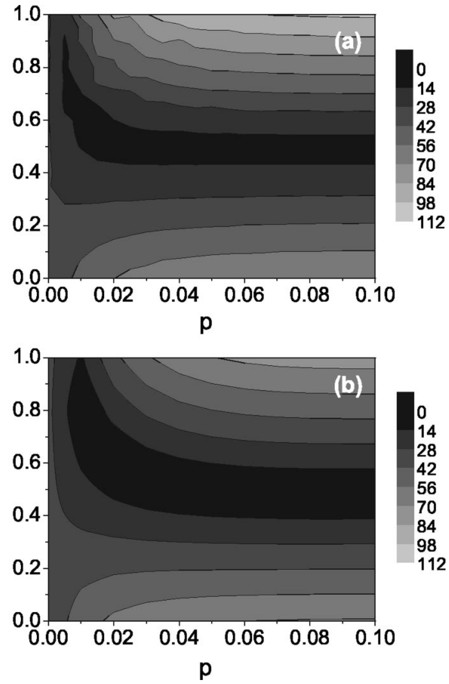

FIG. 3. Contour map version of Fig. 1. Contours correspond to a constant value of σ_D as a function of the error probability q and the network connectivity p. (a) Numerical results. (b) Analytic crowd-anticrowd theory. For clarity, only the upper branch of the numerical results is shown for the high q phase.

the double average in Eq. (2) was evaluated over the $2\times 2^m=4$ time steps in the Eulerian trail. When a tie break between the strategies $K=1$ and $K=2$ arises at one of the four timesteps, one replaces $n_{K=1}^{\rm mean}$ and $n_{K=2}^{\rm mean}$ by $\frac{1}{2}(n_{K=1}^{\rm mean}+n_{K=2}^{\rm mean})$ at that timestep, as one does, for tie breaks between any other K and K'. In this way, the average over the Eulerian trail is easily evaluated analytically. As m increases, there are more time steps over which one must average (i.e., $2P=2\times 2^m$ time steps). However, since ties also become less frequent as m increases, one can simply ignore them without significant loss of accuracy (see Ref. [5] in which good agreement is obtained for the non-networked MG for a wide range of m values without considering ties). Second, the theory has assumed the non-networked MG result that the dynamics follow the Eulerian trail. Only one branch therefore emerges in Fig. 1(b) at high q, appearing like some effective average over the global output series for all branches in Fig. 1(a). If instead one uses knowledge of the actual global output series for these separate branches (i.e., antiferromagnetic or ferromagnetic), then results even closer to Fig. 1(a) can be obtained. This is illustratedat one particular p by the dotted line in Fig. 1(b).

Figure 3 provides a contour plot of σ_D around the minimum. The black contour, centered around the critical curve $C_{\rm crit}(q,p)$, effectively separates the two different regimes of behavior. The low σ_D values around $C_{\rm crit}(q,p)$ can be easily understood using the physical picture provided by the crowd-anticrowd theory: the stochasticity induced by q (i.e., noise) breaks up the size of the crowds using a given strategy K, while simultaneously increasing the size of the anticrowds using the opposite strategy $\overline{K}$. It is remarkable that a linear increase in the noise q gives rise to such a non-linear variation in σ_D. Using the analytic expressions in this paper, an equation for $C_{\rm crit}(q,p)$ can be obtained. We do not include it, however, because it is cumbersome. As noted above, the theory neglects a full treatment of the dynamical fluctuations around $n_K^{\rm mean}$. Hence, the theory overestimates the crowd-anticrowd cancellation arising in Eq. (2) and thus slightly underestimates σ_D in the neighborhood of $C_{\rm crit}(q,p)$ [compare Figs. 3(a) and 3(b)]. As p increases, $C_{\rm crit}(q,p)$ becomes less dependent on the connectivity p since more and more agents join the same network cluster. For $p \gtrsim 0.05$ the system passes the percolation threshold and hence is dominated by a giant, common cluster.

We have discussed the case of small m, where the Eulerian trail acts as a quasiattractor at $(q=0,p=0)$. For large m, the Eulerian trail no longer provides such an attractor. At $(q=0,p=0)$, σ_D is smaller for large m than for small m because the typical crowd (anticrowd) sizes decrease (increase) as m increases. The regime $(q=0,p\neq 0)$ is discussed in Ref. [8] for general m, while the regime $(q\neq 0,p\neq 0)$ will be similar to the present results for sufficiently large p. A full discussion for larger m will be presented elsewhere.

Finally, we note that our results raise the interesting possibility whereby imperfect information transmission could be induced at the local level in order to achieve a desired change in the macroscopic fluctuations within biological, informational or socio-economic systems.

We are grateful to Sean Gourley for assistance with this work. P.M.H. acknowledges support from the Research Grants Council of the Hong Kong SAR Government (Grant No. CUHK4241/01P).

[1] See, for example, A. Soulier and T. Halpin-Healy, Phys. Rev. Lett. **90**, 258103 (2003).
[2] J. L. Casti, *Would-be Worlds* (Wiley, New York, 1997).
[3] W. B. Arthur, Science **284**, 107 (1999); W. B. Arthur, Pap. Proc. Annu. Meet. Am. Econ. Assoc. **84**, 406 (1994); N. F. Johnson, S. Jarvis, R. Jonson, P. Cheung, Y. R. Kwong, and P. M. Hui, Physica A **258**, 230 (1998).
[4] D. Challet and Y. C. Zhang, Physica A **246**, 407 (1997); D. Challet, M. Marsili, and R. Zecchina, Phys. Rev. Lett. **84**, 1824 (2000); R. Savit, R. Manuca, and R. Riolo, *ibid.* **82**, 2203 (1999); N. F. Johnson, P. M. Hui, R. Jonson, and T. S. Lo, *ibid.* **82**, 3360 (1999); S. Hod and E. Nakar, *ibid.* **88**, 238702 (2002).
[5] N. F. Johnson, and P. M. Hui, e-print cond-mat/0306516; N. F. Johnson, S. C. Choe, S. Gourley, T. Jarrett, P. M. Hui, in *Advances in Solid State Physics*, edited by B. Kramer (Springer, Heidelberg, 2004), Vol. 44, pp. 427–438; M. Hart, P. Jefferies, N. F. Johnson, and P. M. Hui, Physica A **298**, 537 (2001).
[6] A. Cavagna, J. P. Garrahan, I. Giardina, and D. Sherrington, Phys. Rev. Lett. **83**, 4429 (1999); M. Hart, P. Jefferies, N. F. Johnson, and P. M. Hui, Phys. Rev. E **63**, 017102 (2000).
[7] M. Anghel, Z. Toroczkai, K. E. Bassler, and G. Korniss, Phys. Rev. Lett. **92**, 058701 (2004); I. Caridi and H. Ceva, e-print cond-mat/0401372; M. Sysi-Aho, A. Chakraborti, and K. Kaski, Physica A **322**, 701 (2003); D. Challet, M. Marsili and G. Ottino, e-print cond-mat/0306445.
[8] S. Gourley, S. C. Choe, N. F. Johnson, and P. M. Hui, Europhys. Lett. **67**, 867 (2004).
[9] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. Lett. **85**, 5468 (2000); R. Albert and A. L. Barabasi, Phys. Rev. Lett. **85**, 5234 (2000).
[10] Defining $q=(1/2)[1-\tanh(1/T)]$ Ref. [6], one could choose to convert the parameter q into a physical temperature T which creates disorder in the information carried by the bonds (as opposed to the sites in the TMG Ref. [6]). However, since T must be positive to be physical, this transformation would restrict q to the range $0\leqslant q\leqslant 1/2$, thereby discarding the interesting high q behavior.
[11] The Eulerian trail corresponds to a particular cycle around the global information states, of duration $2P=2\times 2^m$ time steps, in which every possible transition between global information states is visited once. For $m=1$ it is given by …00110011…. See P. Jefferies, M. L. Hart, and N. F. Johnson, Phys. Rev. E **65**, 016105 (2002).