ELSEVIER

Physica D 125 (1999) 155–166

PHYSICA D

# Evolutionary games with two timescales

Vivek S. Borkar[a,*], Sanjay Jain[b,1], Govindan Rangarajan[c,2]

[a] *Department of Computer Science and Automation, Indian Institute of Science, Bangalore 560 012, India*
[b] *Centre for Theoretical Studies, Indian Institute of Science, Bangalore 560 012, India*
[c] *Department of Mathematics and Centre for Theoretical Studies, Indian Institute of Science, Bangalore 560 012, India*

Received 23 April 1998; received in revised form 1 July 1998; accepted 14 August 1998
Communicated by A.M. Albano

**Abstract**

We consider a two timescale model of learning by economic agents wherein active or 'ontogenetic' learning by individuals takes place on a fast scale and passive or 'phylogenetic' learning by society as a whole on a slow scale, each affecting the evolution of the other. The former is modelled by the Monte Carlo dynamics of physics, while the latter is modelled by the replicator dynamics of evolutionary biology. Various qualitative aspects of the dynamics are studied in some simple cases, both analytically and numerically, and its role as a useful modelling device is emphasized. ©1999 Elsevier Science B.V. All rights reserved.

*Keywords:* Evolutionary games; Monte Carlo dynamics; Replicator dynamics; Singularly perturbed differential equations

## 1. Introduction

One of the classical problems in game theory and economics is that of equilibrium selection. Having concurred that Nash equilibrium is the most appropriate equilibrium concept in noncooperative situations, game theorists are faced with the problem of selecting the right one from many candidate equilibria in a typical situation. For several years, they worked on 'static' refinements of Nash equilibrium whereby one narrows down the choice by imposing additional intuitively reasonable conditions. To mention just a few, there were notions such as subgame perfect equilibrium, evolutionarily stable equilibrium, etc. (an excellent account of these appears in [1]), culminating in the classic work of Harsanyi and Selten [2]. In the past few years, however, the trend has been towards constructing dynamic models of disequilibrium and focus on those Nash equilibria that arise as stable equilibrium points of the proposed dynamics. The dynamics are intended to model economic agents in an interactive environment where they 'learn' from experience and adapt their strategies in real time. Going one step further, many such dynamics employ stochastic models that incorporate noisy inputs due to errors or exploratory moves by the agents. The large

---

[*] Corresponding author; e-mail: borkar@csa.iisc.ernet.in
[1] E-mail: jain@cts.iisc.ernet.in
[2] E-mail: rangaraj@math.iisc.ernet.in

PII: S0167-2789(98)00239-5

time asymptotics of the stochastic evolution is studied in the small noise limit to identify 'stochastically stable' equilibria. An excellent overview of these developments appears in [3] where further references may be found.

There are two major strands of research in this class of models. One of them considers the strategy profile of the population as a whole with its dynamics described by the replicator dynamics of evolutionary biology [4, 5]. In this, the population itself adapts by incrementally reinforcing or discouraging (implicitly, through birth–death mechanisms) the strategies that give, respectively, higher and lower payoffs than the current average, see [5, pp. 72–73]. To draw a parallel with physics, one may compare this with the ensemble picture of Gibbs et al. in classical mechanics where one studies the time evolution of the overall distribution of particles in the phase space without tagging them individually. In contrast, the second strand of research in this area is closer to the Newton–Hamilton 'particle picture' of classical mechanics, insofar as one retains the identity of individual agents in a finite collection and postulates adaptation rules for individual learning, which in turn lead to the aggregate behaviour being studied [6–9]. See, however, [10, 11] for a different interpretation. A few salient features of some or all of these models are given below.

*Bounded rationality.* Keeping in mind that the agents have limited memory and computational ability, the adaptation mechanisms use simple intuitive rules and do not perform any exact, explicit optimization using all the information available.

*Noisy evolution.* This accounts for noisy inputs from the environment which may render the payoffs random, errors made by the agents because of noisy observations or otherwise, random mutations (these are sometimes modelled separately), and last but not the least, noise used as an exploratory probe by the agents, a feature quite familiar to adaptive control and artificial intelligence communities.

*Inertia.* The adaptation is incremental in the sense that only small adjustments are made to the strategy profile, accounting for societal inertia or risk aversion by individuals.

Recalling the analogy with classical mechanics, one notes that the Liouville equation in the ensemble picture is derived from Hamilton's equations for particle motion. Analogously, one would like to derive replicator-like dynamics from evolution of individual behaviour – 'macrobehaviour from micromotives', to borrow a phrase from Schelling [12]. This programme has indeed been carried out with some success [13, 14]. Nevertheless, the original motivations for the two classes of models are distinct: they represent two different modes of learning. Individual learning captures the active or ontogenetic learning in biological parlance wherein an individual adapts to his/her environment. Population dynamics such as the replicator dynamics, on the other hand, model the passive or phylogenetic learning wherein the species profile gets modulated by the selection pressure.

In biology, this distinction is stark, because ontogenetic learning is not inherited. Any belief to the contrary amounts to Lamarckism, a strict taboo for biologists. There is no reason, however, why economic Lamarckism should be taboo, given the rags to riches (or vice versa) stories in the lives of individuals or firms, the after-effects of which are enjoyed (or rued) by their progeny. One distinction between the two modes of learning, however, must be carried over to economics, viz., their different timescales. Ontogenetic learning takes place over a single lifetime while phylogenetic learning takes place, in principle, over several lifetimes.

With the foregoing in mind, we propose a two-tier model of learning. At the ontogenetic level, i.e., on the faster timescale, the individual updates his strategy to pick moves that yield higher payoffs with higher probability, retaining at the same time 'expensive' moves, albeit with a low probability. The latter may be attributed to random exploration or errors. This behaviour is captured by the Monte Carlo dynamics of statistical physics [15] with one difference. The energy landscape on which our particle ($\approx$ economic agent) is moving is itself time-varying, as it depends on the time-varying strategy profile of the entire ensemble. We then pass over to the corresponding ensemble evolution equation, the 'macroevolution at the microscale', that describes how the strategy profile of the entire population evolves under ontogenetic learning. Later on we superimpose on this, on a slower timescale, the replicator dynamics, describing the evolution of a conglomeration of economic agents under selection pressure from the environment (which may include other such conglomerations). The two dynamics interact through the payoff matrices that enter their velocity fields. The payoff matrix for ontogenetic learning, indicative of 'utility' as seen by individual agents, gets modulated by the population behaviour on the slower timescale. Conversely, the payoff matrix for phylogenetic learning, reflecting 'fitness' parameters for the species, gets affected by the dynamics on the faster scale. This leads to a two timescale evolution, the macroevolution at macroscale, which can then be cast in the framework of singularly perturbed ordinary differential equations (ODEs) [16].

The paper is organized as follows. Section 2 describes the Monte Carlo dynamics for ontogenetic learning. As this seems to be new, it is analysed in some detail, particularly for the 2 × 2 symmetric case, which, to evolutionary game theorists, is what *'drosophila'* is to molecular biologists. Section 3 introduces the replicator dynamics and formulates the two timescale evolutionary picture. Section 4 presents simulation results for the 2 × 2 case. These show that even in such a simple case, two dynamics interacting at two different timescales lead to a far richer variety of qualitative behaviour. One is that in which the two dynamics counteract each other so as to have the system poised in between two distinct kinds of equilibria, reminiscent of (but distinct from) self-organized criticality [17]. Another is that in which under similar situations, one obtains sustained oscillations between two equilibria, a phenomenon reminiscent of (but distinct from) noise-induced transitions [18, Chapter 6] on one hand, and heteroclinic cycles [19, pp. 45–46], on the other. Appendix A takes care of some mathematical details glossed over in Section 2. It should be noted that learning at different timescales has also been studied by Binmore et al. [20]. Their motivation, formulation and analysis is quite different from ours. Two timescale paradigms have also been proposed in artificial life literature – see, e.g., [21].

## 2. Ontogenetic learning

Let there be $N$ behaviourally isomorphic agents and let $d$ denote the number of strategies each one of them has at his disposal when paired off at random with another agent to play a two person game. Without loss of generality, we identify the strategy set with $\{1, 2, \ldots, d\}$. Let $A = [[a_{ij}]]$, $1 \leq i, j \leq d$, denote the payoff ('utility') matrix where $a_{ij}$ is the payoff received on playing strategy '$i$' when the opponent plays '$j$'. Let $X_i(t)$ be the strategy being played by agent $i$ at time $t$. Define

$$I\{X_i(t) = j\} = 1 \text{ if } X_i(t) = j \quad \text{and} \quad 0 \text{ otherwise}$$

and

$$N_j(t) = \sum_{i=1}^{N} I\{X_i(t) = j\},$$

the number of agents playing $j$ at time $t$. Then $p_j(t) = N_j(t)/N$ is the proportion thereof and $p(t) = [p_1(t), \ldots, p_d(t)]$ a probability vector that gives the 'histogram' or strategy profile at time $t$. We postulate that an agent playing $i$ at time $t$ responds in immediate future to the empirical average payoff $\sum_{j=1}^{d} a_{ij} p_j(t)$. The justification is that this conforms to one's economic intuition that in a random interaction, in the absence of any special prior knowledge about the current opponent, an agent will base his action on the gross behavioural patterns he sees. The specific adaptation mechanism we hypothesize is the Monte Carlo dynamics [15]. If the agent is playing strategy $i$ at time $t$, then at time $t + \Delta t(\Delta t$ 'small') he changes to $j$, $j \neq i$, with probability $q_{ij}(p(t))\Delta t + o(\Delta t)$, where

$$q_{ij}(p(t)) = \exp\left(-\beta\left(\sum_k a_{ik} p_k(t) - \sum_l a_{jl} p_l(t)\right)^{+}\right), \tag{1}$$

and remains at $i$ with the remaining probability. Here $x^+ = \max(x, 0)$ and $\beta > 0$ is a parameter we call the inverse temperature by analogy with statistical mechanics. The idea is that he shifts his strategy to one paying a higher payoff with higher probability, while at the same time allowing low probability shifts to strategies with lower payoffs. In standard parlance of contemporary evolutionary biology, he makes 'exploitatory' uphill moves on the utility landscape with a high probability and 'exploratory' downhill moves with low probability. The difference with the standard Monte Carlo dynamics, of course, is that the 'landscape' itself depends on $p(t)$ and hence can vary with time. Note also that the dynamics does not change if one changes any column of $A$ by an additive constant. That is, the differences in payoffs matter, not the absolute payoffs.

This is the proposed microdynamics at the microscale. The macrodynamics at the microscale then describes the evolution of $p(t)$ in the $N \to \infty$ limit and is given by the ODE

$$\dot{p}(t) = F(p(t)) \tag{2}$$

for

$$F(\cdot) = [F_1(\cdot), \dots , F_d(\cdot)],$$

$$F_i(p) = \sum_{j \neq i} q_{ji}(p)p_j - p_i \sum_{j \neq i} q_{ij}(p), \quad p = [p_1, \dots , p_d], \quad i \geq 1.$$

This is simply the 'master equation' (albeit a nonlinear one) associated with the stochastic dynamics (1). The derivation of (2) from (1) goes along standard lines and is sketched in Appendix A.

We shall first consider $d = 2$. Since $p_1(\cdot) + p_2(\cdot) = 1$, this reduces (2) to a one-dimensional problem. For simplicity, let $a_{11} = a, a_{12} = b, a_{21} = c, a_{22} = e$. Let $\alpha = a + e - c - b, \delta = b - e$. The ODE for $p_1(\cdot)$ becomes

$$\dot{p}_1(t) = \exp(-\beta(\alpha p_1(t) + \delta)^-)(1 - p_1(t)(1 + \exp(-\beta(\alpha p_1(t) + \delta))), \tag{3}$$

where we use the notation $x^- \equiv -\min(0, x)$. Since a one-dimensional ODE with bounded trajectories cannot have asymptotic behaviour other than convergence to an equilibrium point, we need to look only for these. These are found by setting the right-hand side of (3) equal to zero, leading to

$$p^* = [1 + \exp(-\beta(\alpha p^* + \delta))]^{-1}. \tag{4}$$

The solution can be obtained by looking at the intersection of the graphs of the maps $p \to p$ and $p \to g(p) = (1 + \exp(-\beta(\alpha p + \delta)))^{-1}$ mapping interval [0,1] to itself. The following proposition is easily established.

**Proposition 1.** The following hold:

*Case 1.* $a > c, b > e$ (i.e., strategy 1 dominates strategy 2).

In this case,

$$\alpha p + \delta = (a - c)p + (b - e)(1 - p) > 0, \quad p \in [0, 1].$$

Thus $g(p) > 1/2$ always, implying $p^* > 1/2$. As $\beta \to \infty$ (the zero temperature limit, corresponding to no exploration or errors), $p^*$ approaches 1 as expected. Thus the dominating strategy is stochastically stable in the sense of Foster and Young [22].

The case $a < c, b < e$ (strategy 2 dominates) can be handled similarly.

*Case 2.* $a > c, b < e$ (the coordination problem).

In this case, $\alpha > 0, \delta < 0$. The function $g(\cdot)$ increases from a value below 1/2 at $p = 0$ to a value above 1/2 at $p = 1$, with $g(-\delta/\alpha) = 1/2$ for all $\beta$. Thus $p^* > 1/2$ or $< 1/2$ depending on whether $-\delta/\alpha < 1/2$ or $> 1/2$, respectively. As $\beta \to \infty$, $p^* \to 1$, respectively, 0. (We ignore the case $-\delta/\alpha = 1/2$, $p^* = 1/2$ for all $\beta$, as being nongeneric.)

Thus in the limit one obtains the risk dominant equilibrium in the sense of Harsanyi and Selten [2].

*Case 3.* $a < c, b > e$ (no pure strategy equilibrium).

For this case, $\alpha < 0, \delta > 0$. Thus the function $g(\cdot)$ decreases from a value above 1/2 at $p = 0$ to a value below 1/2 at $p = 1$, with $g(-\delta/\alpha) = 1/2$ for all $\beta$. Ignoring the nongeneric solution $-\delta/\alpha = 1/2$, we have: as $\beta \to \infty$, $p^*$ tends to the mixed equilibrium $[p_1, p_2]$ with $p_1 = -\delta/\alpha$.

These observations coincide with those arising from the alternative dynamics of Kandori et al. [7] and Fudenberg and Harris [23], etc., which is hardly surprising.

The general case $d > 2$ is much more complicated. Something can be said, though, when $A$ is symmetric. By analogy with statistical mechanics, define the 'free energy'

$$H(p) = -\frac{1}{2}\sum_{i,j} a_{ij} p_i p_j + \beta^{-1}\sum_{1} p_j \ln p_i$$

for $p = [p_1, \dots, p_d]$. Differentiating $H(\cdot)$ w.r.t. $p_i$'s and setting it equal to zero, we obtain the following necessary condition for a local minimum $p^*$:

$$p_l^* = Z^{-1}\exp\left(\beta\sum_k a_{lk} p_k^*\right), \quad 1 \le l \le d, \tag{5}$$

where $Z = \sum_i \exp(\beta\sum_k a_{ik} p_k^*)$. The Brouwer fixed point theorem ensures at least one solution for (5). Also, it is easily verified by direct substitution that $p^*$ satisfies (5) if and only if it is an equilibrium point for (2). This holds true regardless of whether $A$ is symmetric.

**Proposition 2.** As $t \to \infty$, $p(t)$ converges to the solution set of (5).

**Proof.** Observe that

$$\frac{\mathrm{d}}{\mathrm{d}t} H(p(t)) = \frac{1}{\beta}\sum_{i,j} p_i(t) q_{ij}(p(t)) \ln\left(\frac{p_j(t) q_{ji}(p(t))}{p_i(t) q_{ij}(p(t))}\right).$$

By Jensen's inequality the right-hand side is nonpositive and is strictly negative unless $p(t)$ satisfies (5). Thus $H(\cdot)$ serves as a Lyapunov function for (5). Standard Lyapunov theory (see, e.g., [24]) then leads to: $p(t) \to \{p^* \mid p^*$ satisfies (5)$\}$.

In other words, the entire population appears to collude in order to minimize the overall payoff $H(p)$. For $A$ not necessarily symmetric, (5) will again have at least one solution by Brouwer fixed point theorem, which is then an equilibrium point for (2). But it is no longer clear that $p(t)$ converges to the set of equilibrium points. That the foregoing argument fails can be seen by considering skew-symmetric $A$. Then $pAp^{\mathrm{T}} = 0$ for all $p$ and $H(p) = \beta^{-1}\sum_i p_i \ln p_i$, which is minimized by the uniform distribution. Hence, if $p(0)$ is uniform, $H(p(t))$ cannot be decreasing.

To draw a further analogy with physics, note that $A$ can be written as the sum of a symmetric ('collusion') part and a skew-symmetric ('zero sum') part. This is analogous to the decomposition (locally) of vector fields into gradient (dissipative) fields and divergence free (conservative) fields in classical mechanics (see, e.g., [25]). Unfortunately the corresponding flows do not decompose as such and it is precisely the interplay between the two that leads to complex dynamics. One expects similar phenomena here.

Finally, consider the 'small noise limit' of (5) as $\beta \to \infty$. We no longer require that $A$ be symmetric.

**Proposition 3.** As $\beta \to \infty$, solutions of (5) converge to the set of symmetric Nash equilibria.

**Proof.** If $p^*(\beta)$ denotes a solution to (5) for a given $\beta$, let $\bar{p}$ denote a limit point of $\{p^*(\beta)\}$ as $\beta \to \infty$. (Recall that the simplex of probability vectors is compact.) Then it is easy to see that if for some $i \neq j$:

$$\sum_k a_{ik}\bar{p}_k > \sum_k a_{jk}\bar{p}_k,$$

then the r.h.s. of the $j$th equation in (5) decays to zero exponentially as $\beta \to \infty$, leading to $\bar{p}_j = 0$. It follows that if $q$ is another probability vector in $\mathbb{R}^d$:

$$\bar{p}A\bar{p}^{\mathrm{T}} \ge qA\bar{p}^{\mathrm{T}}.$$

That is, $\bar{p}$ is a symmetric Nash equilibrium.

If, in addition, one had

$$q \neq \bar{p}, \quad \bar{p}A\bar{p}^{\mathrm{T}} = qA\bar{p}^{\mathrm{T}} \Rightarrow qAq^{\mathrm{T}} < \bar{p}Aq^{\mathrm{T}},$$

then it would be an evolutionarily stable equilibrium in the sense of Maynard Smith and Price [26]. But this in general need not hold: An evolutionarily stable equilibrium may not exist at all [1].

We conclude this section with some important observations about our model:

(i) In our model, all transitions to a strategy with payoff higher than the current one are equally likely. The dynamics does not favour one over the other even when the former fetches a higher payoff. The mechanism underlying this is as follows: At time $t$, the agent using strategy $i$ 'samples' another strategy $j$ with equal 'selection' probability $\Delta t + o(\Delta t)$. She is not, however, obliged to switch to it, though she will do so if it fetches a higher payoff. If not, she will switch only with a further 'selection' probability of $\exp[-\beta(\sum_k a_{ik} p_k(t) - \sum_l a_{jl} p_l(t))]$. In particular, she 'rejects' the sample with some nonzero probability. This is exactly the paradigm underlying Monte Carlo algorithms and their variants such as 'simulated annealing' used for optimization [27]. There are, however, alternative dynamics that do weigh all strategies explicitly according to their payoffs, whether they are below the current average or not. One such dynamics is given by (see [28, pp. 94])

$$q_{ij}(p(t)) = 1 \Big/ \left\{ 1 + \exp\left[ \beta \left( \sum_k a_{ik} p_k(t) - \sum_l a_{jl} p_l(t) \right) \right] \right\}.$$

It is easily verified that this dynamics leads to the same equilibria as the dynamics above and therefore all the results above remain unchanged. It is likely, however, that nonpoint attractors of the two may differ. We do not pursue this dynamics here. It is worth noting that as models of bounded rationality, our model makes less demands of 'rational thinking' (read 'computation') on the agent than the alternate dynamics, as she is required to rank-order two strategies only for the exploratory downhill moves, not otherwise.

(ii) Many models of learning tend to be 'payoff-monotonic', i.e., favour or disfavour strategy alternatives explicitly in accordance to the rank-ordering of their payoffs. Because we allow exploratory 'downhill' moves, our model is not payoff-monotonic. Thus, in particular, dominated strategies can persist with small probability. Nevertheless, it leads to similar conclusions in the $\beta \to \infty$ limit, which partially restores payoff monotonicity to the extent that only payoff improving strategies are now in contention. In fact, taking $\beta \to \infty$ limit in the dynamics (2) itself permits only moves to payoff-improving strategies. This is not necessarily advantageous for the same reason that one has $\beta < \infty$ in Monte Carlo algorithms for optimization: The exploratory moves permit escape out of local peaks (maxima).

It has been observed [29] that absence of payoff monotonicity itself can lead to an interesting dynamic behaviour, like oscillations. This is not an issue in the $2 \times 2$ examples we consider, wherein the dynamics is essentially one-dimensional and thus the only possible $\Omega$-limit sets are point attractors. Nevertheless, in higher dimensions, this has to be kept in mind as an additional source of variation in qualitative dynamics, over and above what our two-tier model may lead to due to the interaction of the two timescales. Note that the $\beta \to 0$ limit leads to all strategy changes being equally likely, i.e., a pure random walk.

(iii) We derived (2) in the $N \to \infty$ limit. This fact is crucial here. For finite populations, each agent should, ideally speaking, subtract out her own contribution from the average payoffs to strategies under consideration. Once again, this 'self-exclusion' feature in itself can lead to an interesting phenomena. In the context of a somewhat different dynamics, this has been studied in [10, 11].

## 3. A two-tier model for learning

Let $B = [[b_{ij}]]$, $1 \leq i, j \leq s$, denote the payoff (fitness) matrix at the phylogenetic level. We do not insist on $B = A$, because what is good for the individual in immediate future may be bad for the species, e.g., excessive consumption of resources. (This is the well-known dichotomy between individual vs. social optimality.) The standard model for phylogenetic adaptation, as already noted, is the replicator dynamics

$$\dot{p}(t) = G(p(t)). \tag{6}$$

Here

$$G(\cdot) = [G_1(\cdot), \dots, G_s(t)], \qquad G_i(p) = p_i[e_i Bp^{\mathrm{T}} - pBp^{\mathrm{T}}], \quad 1 \leq i \leq s,$$

where $p$ is an $s$-dimensional probability vector and $e_i$ the unit vector in the $i$th coordinate direction, both written as row vectors. The intuition behind (6) is as follows: $p_i(t)$ denotes the population share of strategy $i$ at time $t$. Its rate of change is proportional to the current population share $(p_i(t))$ times the excess of its current average payoff from the population $(e_i Bp(t)^{\mathrm{T}})$ over the population's average payoff at time $t(p(t)Bp(t)^{\mathrm{T}})$. This equation and its variants have been studied extensively and several book-length treatments are available [4, 5]. We refer the reader to these references for further details.

The way we interface the ontogenetic and phylogenetic learning is by hypothesizing a combined 'strategy vector' $p(t) = [p_1(t) : p_2(t)] \in \mathbb{R}^{d+s}, t \geq 0$, where $p_1(t) \in \mathbb{R}^d$ is the strategy profile for ontogenetic learning and $p_2(t) \in \mathbb{R}^s$ is that for phylogenetic learning. The combined dynamics is described by the coupled differential equations (our 'macrodynamics at the macroscale'):

$$\epsilon \dot{p}_1(t) = F_{p_2(t)}(p_1(t)), \tag{7}$$

$$\dot{p}_2(t) = G_{p_1(t)}(p_2(t)). \tag{8}$$

Here, $F_p(\cdot) = [F_{p_1}(\cdot), \dots, F_{p_d}(\cdot)] : \mathbb{R}^d \to \mathbb{R}^d$, $p \in \mathbb{R}^s$, and $G_p(\cdot) = [G_{p_1}(\cdot), \dots, G_{p_s}(\cdot)] : \mathbb{R}^s \to \mathbb{R}^s$, $p \in \mathbb{R}^d$, are given by

$$F_{p_i}(x) = \sum_{j \neq i} q^p_{ji}(x) x_j - x_i \sum_{j \neq i} q^p_{ij}(x),$$

with

$$q^p_{ji}(x) = \exp\left(-\beta\left(\sum_k a_{ik}(p)x_k - \sum_l a_{jl}(p)x_l\right)^+\right)$$

for $x = [x_1, \dots, x_d] \in \mathbb{R}^d$, $p \in \mathbb{R}^s$, and $A(\cdot) = [[a_{ij}(\cdot)]] : \mathbb{R}^s \to \mathbb{R}^{d \times d}$ a prescribed Lipschitz continuous maps. Similarly,

$$G_{p_i}(x) = x_i(e_i B(p)x^{\mathrm{T}} - xB(p)x^{\mathrm{T}})$$

for $x \in \mathbb{R}^s$, $p \in \mathbb{R}^d$ and $B(\cdot) = [[b_{ij}(\cdot)]] : \mathbb{R}^d \to \mathbb{R}^{s \times s}$ a prescribed Lipschitz continuous map. (We need view only those $p, x$ that are probability vectors.) Finally, $\epsilon > 0$ is a small parameter.

Eq. (7) is the same as (2) except that the utility matrix $A$ is now a function of the phylogenetic profile vector $p_2$. The presence of the small 'scale parameter' $\epsilon > 0$ on the left-hand side ensures that (7) operates on a much faster timescale than (8). In turn, (8) is the same as (6) with the proviso that the fitness matrix $B$ is now a function of the ontogenetic profile vector $p_1$.

A few comments regarding this model are in order here. Though we projected ontogenetic learning as learning in a single lifetime, there is no explicit modelling of birth and death in (2). Effectively, each agent that dies is being instantly replaced by a newborn clone. Admittedly oversimplistic in this respect, we still opt for this model because even this simple model exhibits a rich variety of qualitative behaviour. Including 'demographic' and even more so, spatiotemporal aspects would lead to an even more complex repertoire of qualitative behaviour, but we do not pursue this here.

It is also important to keep in mind here that both $p_1(\cdot)$ and $p_2(\cdot)$ refer to the same population: They describe the strategy profiles of the same population in different contexts, viz., ontogenetic and phylogenetic. As a toy example, consider payoff matrices $A$ and $B$ with only two choices; either to exhaust resources rapidly or save and replenish for the future. The first would be 'dominant' in the short term and $A$ can be structured accordingly. $B$, on the other hand, may have the second strategy as dominant since this strategy may be better for the population as a whole. The long term behaviour may then correspond to what our numerical experiments reported later suggest.

Mathematically, (7) and (8) form a singularly perturbed differential equation and may be analysed as such [16]. This is an extensively studied topic and we shall limit ourselves to giving the broad intuition behind the analysis, which goes as follows. Eq. (7), being on a faster timescale, views $p_2(\cdot)$ as quasi-static, since $p_2(\cdot)$ will be changing very, very slowly. Then as a first approximation, one may replace (7) by

$$\dot{\tilde{p}}_1(1) = F_{\bar{p}_2}(\tilde{p}_1(t)), \tag{9}$$

where $\bar{p}_2$ is held fixed as a parameter. Since $\tilde{p}_1(\cdot)$ lies in the closed, bounded simplex of probability vectors, it has a closed, bounded $\Omega$-limit set $\Omega$. Furthermore, for any continuous function $h : \mathbb{R}^d \to \mathbb{R}$, the time average,

$$\frac{1}{t}\int_0^t h(\tilde{p}_1(s))\mathrm{d}s,$$

will converge to the average of $h$ w.r.t. an invariant probability measure $\mu$ supported on $\Omega$, for all initial conditions in a set of $\mu$-probability one. This follows from standard ergodic theory [24, Chapter X]. Note, however, that neither $\Omega$ nor $\mu$ are unique. Also, they depend on the 'parameter' $\bar{p}_2$ entering (9).

Eq. (8) in turn views (7) as 'fast', i.e., near-equilibrated. That is, as $p_2(t)$ changes, $p_1(t)$ will closely track the equilibrium behaviour of (9) corresponding to $\bar{p}_2 \approx p_2(t)$. Therefore as a first approximation, (8) may be replaced by

$$\dot{\tilde{p}}_2(t) = G^{\tilde{p}_2(t)}(\tilde{p}_2(t)), \tag{10}$$

where $G^{\bar{p}_2}(\cdot)$, for a probability vector $\bar{p}_2$ in $\mathbb{R}^s$, denotes the average

$$\int_\Omega G_p(\cdot)\mu(\mathrm{d}p)$$

for the appropriate $\Omega$, $\mu$ which depend on $\bar{p}_2$ (among other things) because of the dependence of (9) on $\bar{p}_2$. Note that (10) is, in principle, a self-contained equation for $\tilde{p}_2(\cdot)$ which can be solved analytically or numerically. (*Remark:* The 'other thing' $\mu$ will depend on is $p_1(0)$. We do not keep track of this explicitly because it is fixed once and for all).

One particularly simple qualitative behaviour that may emerge is an asymptotically stable equilibrium. Suppose $p_1^*$, $p_2^*$ are probability vectors in $\mathbb{R}^d$, $\mathbb{R}^s$, resp., such that $p_1^*$ is an asymptotically stable equilibrium for

$$\dot{p}_1(t) = F_{p_2^*}(p_1(t))$$

and $p_2^*$ is an asymptotically stable equilibrium for

$$\dot{p}_2(t) = G_{p_1^*}(p_1(t)).$$

Then the pair $(p_1^*, p_2^*)$ is an equilibrium for (7) and (8). This is the situation where in both ontogenetic and phylogenetic behaviours reinforce each other. The equilibrium may then be deemed to be 'phylogenetically stable'.

Though desirable, this is a very special situation. (An example is when $A$, $B$ each have, say, $i = 1$ as the single dominant strategy when the other does.) In general, one has a rich variety of qualitative behaviour. We illustrate this through the numerical simulations presented in Section 4, for the simple case of $s = d = 2$. For $A(\cdot)$, $B(\cdot)$, we assume the simple structure

$$A = p_2(1)A_0 + p_2(2)A_1, \qquad B = p_1(2)B_0 + p_1(1)B_1,$$

where $p_1 = [p_1(1), p_1(2)]$, $p_2 = [p_2(1), p_2(2)]$, and $A_0$, $B_0$, $A_1$, $B_1$ are prescribed $2 \times 2$ matrices. The qualitative behaviour also depends on the parameters $\beta$ (cf. (1)) and $\epsilon$.

## 4. Numerical experiments

For the above simple situation, the following were the observed behaviour from numerical simulations:

First we chose $A_0$, $B_0$ such that strategy 2 dominates and $A_1$, $B_1$ such that strategy 1 dominates. If $\beta$ is small ($\approx 10$), a single globally attracting fixed point was observed regardless of the value of $\epsilon$. For large $\beta$ ($\approx 1000$), $p_1(1)$ and $p_2(1)$ tend to oscillate around this point, the amplitude of oscillations decreasing with increasing $\epsilon$.

Next we considered $A_0$, $B_0$ that exhibited coordination ('*Case* 2' above). $A_1$, $B_1$ were chosen to correspond to the mixed strategy case ('*Case* 3' above). For both small and large $\beta$, more than one attractor was found, one of which was always characterized by oscillations around a point, the amplitude thereof decreasing with increasing $\epsilon$ as before.

The third case was with $A_0$, $B_0$ in the 'coordination' situation with $-\delta/\alpha > 1/2$, whereas $A_1$, $B_1$ the same with $-\delta/\alpha < 1/2$. In this case, the attractors were simple point attractors with no oscillations.

We also studied situations where the roles of $A_0$, $B_0$ and $A_1$, $B_1$ are interchanged as compared to the above cases. Several other combinations of $A$ and $B$ matrices were investigated – (i) $A_0$, $B_0$ with one strategy dominant, $A_1$, $B_1$ with coordination behaviour, (ii) $A_0$, $B_0$ with one strategy dominant, $A_1$, $B_1$ with mixed strategy behaviour, and (iii) all matrices have one strategy dominant. All these studies lead to one or more of the behaviours that we have already mentioned earlier.

This is just a sample of the rich variety of qualitative behaviours one obtains in the two timescale dynamics. At present, we do not have a systematic classification of the same, but wish to underscore the point that many interesting dynamic behaviours one observes in social or biological systems may be modelled more elegantly by fairly simple two timescale dynamics, whereas capturing them by a dynamics on a single timescale may call for a much more complicated model. Also, as argued in Section 1, such models are also intuitively justified on factual grounds.

## 5. Conclusions

In this paper, we have proposed a two timescale model of evolutionary games. The 'fast' timescale represents the ontogenetic or active learning wherein individual agents modify their strategies in 'real time' based on the observed payoffs as per a prescribed payoff ('utility') matrix. On the slower scale, we have phylogenetic or passive learning wherein the population profile gets modulated by selection pressure, as reflected in the payoff ('fitness') matrix that enters the dynamics at the slower timescale. The two interact insofar as the payoff matrix on one timescale is modulated by the strategy profile on the other.

An important observation of the numerical experiments is the subtle interplay of the two free parameters, $\beta$ and $\epsilon$. The latter indicates separation of the timescales: The smaller the $\epsilon$, the more separated the timescales and the more accurate the intuition behind 'averaging' arguments. Smaller $\beta$ in the Monte Carlo dynamics, on the other hand, corresponds to a less peaked (i.e. more spread out) equilibrium distribution, but a faster convergence to the same. Thus viewing slower dynamics as 'quasi-static' for analysing the faster one is better justified when $\beta$ is high. The simulation results show that the two free parameters need to be considered together and not in isolation.

One of the phenomena that we observed numerically in many cases was that of an equilibrium point poised between the two equilibria dictated by two extreme cases under consideration. This suggests two timescale mechanism as an alternative mechanism leading to the population profiles getting poised at a 'critical' level between the two extremes.

The other observed phenomenon of oscillations is also noteworthy. In particular, we wish to contrast it with two other phenomena already mentioned in Section 1. One is that of noise-induced transitions leading to 'punctuated equilibria'. These are characterized by their inherently random (but stationary in statistical sense) nature and very slow timescale – they have to be 'rare' if the model is accurate. The other is heteroclinic cycles which have some symmetry lurking in the background and are characterized by the fact that ever increasing time is spent near the equilibria with a relatively rapid transition between two occurring at epochs of ever increasing lengths. The oscillations we observed are a third category. While being nonrandom like the latter, they also retain the qualitative aspect of the latter of spending relatively large time near equilibria with comparitively rapid transitions between them. At the same time, this behaviour is 'stationary' or 'time-homogeneous' unlike the heteroclinic cycles.

In conclusion, we note that our aim has been to explore the effect of *different* payoffs accruing on *different* timescales to the *same* strategy set for all agents. It is possible to conceive of models wherein the strategy set available to an agent at the ontogenetic level depends on the strategy being pursued by her at the phylogenetic level. We do not study this scenario here. See, however, [30] for an interesting dynamics with interdependent strategy preferences. Coming back to our present model, we would like to underscore a methodological implication of it: In many controversies regarding competing and equally plausible evolutionary mechanisms to explain observed phenomena, the resolution may not be that one is valid and the other is not, but that some or all of them are operating at different timescales and interacting with each other by modulating the payoff structures.

## Acknowledgements

The work of VSB was supported by grant no. III 5(12)/96-ET from the Department of Science and Technology, Government of India. SJ is also associated with the Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India and is an Associate Member of the Abdus Salam International Centre for Theoretical Physics, Trieste, Italy. The authors thank two anonymous referees for their astute observations which led to a significant improvement of our exposition.

## Appendix A

Here we briefly sketch the passage from (1) to (2). Recall that an agent playing strategy $i$ at time $t$ changes to strategy $j \neq i$ at time $t + \Delta t$ ($\Delta t$ 'small') with probability $q_{ij}(p(t))\Delta t + o(\Delta t))$. Mathematically, what we are saying here is that the process

$$\begin{aligned} M_{ij}(t) &= I\{X_i(t) = j\} - I\{X_i(0) = j\} \\ &\quad - \int_0^t \left( \sum_{k\neq j} q_{kj}(p(s)) I\{X_i(s) = k\} - I\{X_i(s) = j\} \sum_{k\neq j} q_{jk}(p(s)) \right) \mathrm{d}s, \qquad t \geq 0 \end{aligned}$$

is a zero mean martingale for all $i$, $j$. That is, it satisfies

$$E[M_{ij}(t)/X_l(y), y \leq s, 1 \leq l \leq N] = M_{ij}(s), \quad t \geq s.$$

In particular, it has uncorrelated increments and thus corresponds to one's intuition of 'noise'. The above equation then may be rewritten as the noise-driven system

$$I\{X_i(t) = j\} = I\{X_i(0) = j\} - \int_0^t \left( \sum_{k\neq j} q_{kj}(p(s)) I\{X_i(s) = k\} - I\{X_i(s) = j\} \sum_{k\neq j} q_{jk}(p(s)) \right) \mathrm{d}s + M_{ij}(t).$$

This is the proposed microdynamics at the microscale. Summing over $i$ and dividing by $N$, we get

$$p_j(t) = p_j(0) + \int_0^t \left( \sum_{k\neq j} q_{kj}(p(s)) p_k(s) - p_j(s) \sum_{k\neq j} q_{jk}(p(s)) \right) \mathrm{d}s + \bar{M}_j^N(t), \tag{A.1}$$

where $\bar{M}_j^N(t) = (1/N)\sum_{i=1}^N M_{ij}(t)$. This is the macrodynamics at the microscale, describing the ensemble evolution.

We shall be interested in the 'thermodynamic limit' of (A.1) as $N \to \infty$. For this purpose, consider $M_{ij}(t), t \geq 1, 1 \leq j \leq d$, as prescribed processes on a given probability space and construct succesively for $N = 1, 2, \ldots$, the corresponding $p(\cdot)$ by (A.1). We shall assume that for each $j$, $M_{ij}(\cdot), i = 1, 2, \ldots$, are independent. Standard techniques from the theory of stochastic processes can be invoked to show that for each $j$, as $N \to \infty$, the processes $\bar{M}_j^N(\cdot)$ converge to the 'process' that is identically zero. (This convergence is in the sense of a 'weak convergence of probability measures' on an appropriate space of trajectories, in this case the Skorohod space of right continuous functions with left limits – see, e.g. [31, Chapter 3]. For analogous results in different contexts, see [32, Chapters 4–6]). Thus we obtain (2) as a thermodynamic limit of (A.1).

Note that the independence of $M_{ij}(\cdot), i \geq 1$, is, strictly speaking, an additional assumption here. Our original assumption concerning the agents was merely that they be behaviourally isomorphic, which translates into the weaker requirement that $M_{ij}(\cdot), i \geq 1$, be 'exchangeable'. In other words, any joint distributions of finite subcollections of random variables $\{M_{ij}(t), i \geq 1, t \geq 0\}$ be invariant with respect to permutations on the index $i$. Under this weaker condition, one allows for a certain amount of 'coherence' in the behaviour of agents. An extreme case is $M_{ij}(\cdot) = M_{kj}(\cdot) k \neq i, j$, when they move in step if we assume the same initial condition. In general, under mere exchangeability, the limit of $M_j^N(\cdot)$ as $N \to \infty$ need not be the zero process. This suggests that in societies in which the behaviour of individuals is strongly coherent, one may expect larger levels of 'systemic' noise even in large populations.

## References

[1] E. Van Damme, Stability and Perfection of Nash Equilibria, 2nd ed., Springer, Berlin, 1991.

[2] J. Harsanyi, R. Selten, A General Theory of Equilibrium in Games, MIT Press, Cambridge, MA, 1988.

[3] F. Vega Redondo, Games, Evolution and Ecocnomic Behaviour, Oxford University Press, Oxford, 1996.

[4] J. Hofbauer, K. Sigmund, The Theory of Evolution and Dynamical Systems, Cambridge University Press, Cambridge, 1988.

[5] J.W. Weibull, Evolutionary Game Theory, MIT Press, Cambridge, MA, 1995.

[6] L. Blume, The statistical mechanics of strategic interaction, Games Eco. Behaviour 5 (1993) 387–424.

[7] M. Kandori, G. Mailath, R. Rob, Learning, mutation and long run equilibrium in games, Econometrica 61 (1993) 29–56.

[8] M. Kandori, R. Rob, Evolution of equilibria in the long run: a general theory and applications, J. Eco. Theory 65 (1995) 383–414.

[9] H.P. Young, The evolution of conventions, Econometrica 61 (1993) 57–84.

[10] V.S. Borkar, S. Jain, G. Rangarajan, Dynamics of individual specialization and global diversification in communities, Complexity 3 (1998) 50–56.

[11] V.S. Borkar, S. Jain, G. Rangarajan, Collective behaviour and diversity in economic communities: Some insights from an evolutionary game, in: J. Kertesz, I. Kondor (Eds.), Econophysics, Kluwer Academic Publisher, Dordrecht, 1998, in press.

[12] T. Schelling, Micromotives and Macrobehaviour, W.W. Norton, New York, 1978.

[13] J. Bjornerstedt, J. Weibull, Nash equilibrium and evolution by imitation, in: K. Arrow, E. Colombatto (Eds.), Rationality in Economics, McMillan, New York, 1993.

[14] T. Borgers, R. Sarin, Learning through reinforcements and replicator dynamics, J. Eco. Theory 77 (1997) 1–14.

[15] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. 21 (1953) 1087–1092.

[16] F. Verhulst, Nonlinear Differential Equations and Dynamical Systems, Springer, Berlin, 1990.

[17] P. Bak, How Nature Works, Springer, New York, 1996.

[18] M. Freidlin, A.D. Wentzell, Random Perturbations of Dynamical Systems, Springer, New York, 1989.

[19] J. Guckenheimer, P. Holmes, Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields, Springer, New York, 1983.

[20] K.G. Binmore, L. Samuelson, R. Vaughan, Musical chairs: modeling noisy evolution, Games Eco. Behaviour 11 (1995) 1–35.

[21] D.H. Ackley, M.L. Littman, Learning from natural selection in an artificial environment, in: in: C.G. Langton (Ed.), Artificial Life II Video Proceedings, Addison-Wesley, Redwood City, CA, 1992.

[22] D. Foster, H.P. Young, Stochastic evolutionary dynamics, Theor. Population Biol. 38 (1990) 219–232.

[23] D. Fudenberg, C. Harris, Evolutionary dynamics with aggretate shocks, J. Eco. Theory 57 (1992) 420–491.

[24] G. Sell, Topological Dynamics and Ordinary Differential Equations, Van Nostrand, New York, 1971.

[25] R. Vilela Mendes, J.T. Duarte, Vector fields and neural networks, Complex Syst. 6 (1992) 21–30.

[26] J.M. Smith, G. Price, The logic of animal conflict, Nature 246 (1973) 15–18.

[27] E. Aarts, J. Korst, Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing, Wiley, New York, 1989.

[28] R. Kindermann, J. Laurie Snell, Markov Random Fields and their Applications, Contemporary Mathematics, vol. 1, American Mathematical Society, Providence, RI, 1980.

[29] R. Sethi, Strategy-specific barriers to learning and monotonic selection dynamics, Games Eco. Behavior 23 (1998) 284–304.

[30] L. Kockesen, E. Ok, R. Sethi, Evolution of interdependent preferences in aggregative games, Working paper, Department of Economics, New York University and Department of Economics, Columbia University, May 1998.

[31] S. Ethier, T.G. Kurtz, Markov Processes: Characterization and Convergence, Wiley, New York, 1986.

[32] M. Aoki, New Approaches to Macroeconomic Modeling, Cambridge University Press, Cambridge, 1996.