A MODEL OF CULTURAL EVOLUTION IN THE CONTEXT OF STRATEGIC CONFLICT

MISHA PEREPELITSA
Department of Mathematics
University of Houston
4800 Calhoun Rd, Houston, TX

(Communicated by Dehua Wang)

Abstract. We consider a model of cultural evolution for a strategy selection in a population of individuals who interact in a game theoretic framework. The evolution combines individual learning of the environment (population strategy profile), reproduction, proportional to the success of the acquired knowledge, and social transmission of the knowledge to the next generation. A mean-field type equation is derived that describes the dynamics of the distribution of cultural traits, in terms of the rate of learning, the reproduction rate and population size. We establish global well-posedness of the initial-boundary value problem for this equation and give several examples that illustrate the process of the cultural evolution.

1. Introduction. Evolutionary game theory, pioneered by Maynard Smith and Price [12] is a powerful tool that explains dominance of some behavioral traits as being uninvadable by other traits in the competition for Darwinian fitness points, when fitness is frequency dependent. A deterministic dynamic process that selects a stable behavioral trait can be described by the replicator equation, see Taylor and Jonker [17], Hofbauer et al. [8], Zeeman [18]. The replicator equation also governs the dynamics of reinforcement learning in repeated play of a game, see Borgers and Sarin [1], Fudenberg and Levine [4], Krishnedu et al. [3], Perepelitsa [13].

Behavioral traits can be under control of an individual as a part of the adaptation mechanism. Its implementation often relies on learning about the environment in which the individual operates. If the environment includes other individuals then the situation is best described as learning in a strategic conflict (game). Learning in games is an integral part of game theory that goes back to works of Robinson [15] and Shapley [16]. One of its mainstays is fictitious play or statistical learning. The learning by fictitious play in large populations can be described by an ODE, called the best-response equation, see Gilboa and Matsui [6], Gaunersdorfer and Hofbauer [5], Hofbauer [7] and Hofbauer and Sigmund [9]. The best-response equation describes changes in the mean statistical frequencies about the opponent actions, and its stationary points are Nash equilibria.

In this paper we consider an evolutionary process that combines the concept of “the survival of fittest” from biological evolution with individual learning through
fictitious play, when a state of learning is transmitted to the next generation of players. The examples of this type of processes are furnished by the cultural evolution theory. Consider the following hypothetical situation for strategic conflict in the context of the famous Hawk-Dove-Retaliator (HDR) game, see Table 1. Suppose that three large prehistoric tribes move in into a new territory and prior to that, each tribe was exposed to an environment (other tribes), where they had developed a specific mixed or pure strategy of Hawk, Dove, or Retaliator as their responses in the competition over available resources. Let’s call these strategies \( \bar{p} = (p_1, p_2, p_3) \), \( \bar{q} = (q_1, q_2, q_3) \) and \( \bar{r} = (r_1, r_2, r_3) \), respectively. On the new territory, if tribes stick to their strategies, under natural assumption that individuals interact with a sufficiently big random sample of individuals, reproduce at the rate proportional to the average accumulated fitness point, and the population are large, the dynamics can be described by the replicator equation that takes point

\[
A = w_1 \bar{p} + w_2 \bar{q} + w_3 \bar{r},
\]

where \( w_1, w_2 \) and \( w_3 \) are population proportions of each tribe, and moves it to an evolutionarily stable state (ESS) \( R \) or a mixed ESS \( \frac{1}{2}H + \frac{1}{2}D \), depending on the location of point \( A \) in the strategy space.

We will take into account that the cognitive abilities of tribal members high enough to allow for simple statistical assessment of actions of opponents and to adjust their strategies as the best reply to collected statistics. More importantly, we assume that accumulated statistics (and so the best response strategies) are passed to the subsequent generations, unaltered. In this context, a mixed strategy represents a “culture” which is transmitted vertically, according to the terminology of cultural evolution theories, see Hoppitt and Laland [10], Richerson and Boyd [14].

The outcome of this type of evolution will depend on the rate at which interactions take place, and thus how quickly individuals learn about the actions of others, and on the rate of reproduction, as well on some additional parameters might be needed to completely specify the problem.

The goal of the paper is to develop a mathematical model that takes as input the initial distribution of cultural traits in a population, the above mentioned rates of interaction and reproduction, and outputs the distribution of cultural traits at any moment \( t \) in future. As we will see from the examples of section 3, it is essential for an accurate description of the dynamics that the model specifies the whole distribution of traits and not just some statistical averages, such as the mean and the variance. The model is derived as a mean-field approximation of the distribution density of a Markov process describing the interaction of agents, and is described by a kinetic equation with non-linear transport velocities. Both, the replicator and the best-response equation are contained as special cases of this kinetic model.

Our main result, stated in section 2, establishes global well-posedness of the initial-boundary value problem for the kinetic equation. Due to the discontinuities of the best-response function, solutions of this equation are intrinsically weak. The space of functions of finite total variation appears to be natural for this problem. Our approach is based on a priori estimates in \( BV \) and \( L^p \) spaces and a fixed-point argument, in order to prove existence of weak solutions. We also address the problem of long-time asymptotic behavior of the weak solutions. In general, determining asymptotic state for an arbitrary game is problematic due to complicated
dynamics and absence of an entropy functional. We obtain some partial results in this direction for a model with zero reproduction rate, for which we show that if two statistical averages, the mean empirical frequencies and mean best-response converge to some values (not necessarily the same), then the empirical frequencies of every agent in the population converge to a Nash equilibrium of the game.

In section 3 we discuss several examples of cultural evolution in Hawk-Dove-Retaliator game, described above, that illustrate the dynamics of cultural evolution for different learning and reproduction parameters, when the population consists of several groups of agents. We show that, compared to purely biological evolution (no learning), the cultural evolution can completely alter the selection of behaviors and/or change the outcome of the inter-group competition. In another example of HDR game we show that an exploding heterogeneous population can lock the cultural evolution in a sub-optimal pure strategy for some of the participants.

2. Model. We consider a series of plays of a symmetric 2-player game between randomly selected agents in a large population. There are $d$ strategies available to agents and the payoffs are given by matrix $A = \{a_{ij}\}_{i,j=1}^d$, which we assume to have non-negative entries. The game defines multi-valued best response function $\text{BR}(p) : \Delta^{d-1} \to \mathcal{P}(\Delta^{d-1})$, where $\Delta^{d-1}$ is $d-1$ dimensional simplex. This function outputs a set of best strategies for a player when his/her opponent uses strategy $p$. We will also use its single valued representative $b(p) \in \text{BR}(p/\sum_i p_i)$. We refer the reader to Appendix section A for details. We start with the case when there is no reproduction and the population size is $N$.

We will record the change of the state of agents that occur at discrete epochs, labeled by $t$. The state of agent $i$ at epoch $t$ is a $d + 1$ dimensional vector

$$X^t_i = (P^t_i, S^t_i),$$

where

$$P^t_i = (p^t_{i,1}, \ldots, p^t_{i,d}),$$

is a vector of empirical frequencies (unscaled) and $S^t_i$ is averaged fitness.

An interaction is a round of the game between two random agents, say $i$ and $j$, who play according to their best response to empirical frequencies $P^t_i$ and $P^t_j$. Based on that, they earn fitness points and update the empirical frequencies. To describe the update rule we will use the following parameters: $h$ – the characteristic learning and fitness increment and $\delta$ – time increment. The rule takes the form

$$P^{t+\delta}_i = P^t_i + hb(P^t_j)$$

$$P^{t+\delta}_j = P^t_j + hb(P^t_i)$$

$$S^{t+\delta}_i = (1 - \mu h)S^t_i + \mu ha(b(P^t_i), b(P^t_j))$$

$$S^{t+\delta}_j = (1 - \mu h)S^t_j + \mu ha(b(P^t_j), b(P^t_i))$$

where $a(b(P^t_i), b(P^t_j)) = \sum_{k,l} a_{kl} b_k(P^t_i) b_l(P^t_j)$ is the fitness earned by agent $i$. In this formulation the fitness $S^t_i$ is averaged over the history of payoffs, in particular, it cannot grow without a bound. One can think of $\mu$ as being a recency parameter. Large values of $\mu$ put more weight on more recent payoffs.

We will derive an equation for $f(p, s, t)$ – the density of the distribution of agents over the space of empirical frequencies and fitness $(p, s) \in \mathbb{R}_+^d \times \mathbb{R}_+$. In the following derivation we use the convention that

$$x_i = (p_i, s_i), \quad \bar{x} = (x_1, \ldots, x_N) \in (\mathbb{R}_+^d \times \mathbb{R}_+)^N,$$
where $\bar{x}$ parametrizes the state of the whole population.

Let $w(\bar{x}, t)$ be the density of the distribution of empirical frequencies and fitness for the whole population. This function implicitly depends on parameters such as $h$, and $\mu$, but we suppress them from notation for convenience of presentation. The update rule can be expressed as a moment relation with a smooth test function $\phi$,

$$\int \phi(\bar{x}) w(\bar{x}, t + \delta) \, d\bar{x} = \sum_{i \neq j} (N(N - 1))^{-1} \int \phi(\bar{x}) \bigg|_{\bar{x}_i = \hat{x}_i, \bar{x}_j = \hat{x}_j} w(\bar{x}, t) \, d\bar{x},$$

where $\hat{x}_i = (p_i + hb(p_j), (1 - \mu h)s_i + \mu ha(b(p_i), b(p_j)))$, and symmetrically for $\hat{x}_j$.

The equation expresses the law of total probability. It can also be written as

$$\int \phi(\bar{x}) |w(\bar{x}, t + \delta) - w(\bar{x}, t)| \, d\bar{x} = \sum_{i \neq j} (N(N - 1))^{-1} \int [\phi(\bar{x})]_{\bar{x}_i = \hat{x}_i, \bar{x}_j = \hat{x}_j} - \phi(\bar{x})|w(\bar{x}, t) \, d\bar{x}. \tag{1}$$

Function $f(x, t)$, where $x = (p, s)$ is related to the multi-agent distribution $w(\bar{x}, t)$ through the rule:

$$f(x, t) = \sum_k N^{-1} \int w(\bar{x}, t) \bigg|_{\bar{x}_k = x} \, d\bar{x}_k, \quad x \in \mathbb{R}_+^d \times \mathbb{R}_+,$$

where $\bar{x}_k$ is a $(d + 1)(N - 1)$ dimensional vector of all coordinates, $x_1, ..., x_N$, excluding $x_k$. This is one-particle distribution function. In the formulas to follow we need to use two-particle distribution function

$$g(x, y, t) = \sum_{i \neq j} (N(N - 1))^{-1} \int w(\bar{x}, t) \bigg|_{x_i = x, x_j = y} \, d\bar{x}_{ij},$$

where $\bar{x}_{ij}$ is the $(d + 1)(N - 2)$ dimensional vector of all coordinated excluding $x_i$ and $x_j$.

Function $g$ is symmetric in $(x, y)$ and is related to $f$ by the formulas

$$f(x, t) = \int g(x, y, t) \, dx = \int g(x, y, t) \, dy.$$

The moments of function $f$ and $g$ are computed from the moments of $w$:

$$\int \psi(x) f(x, t) \, dx = \sum_k N^{-1} \int \psi(x_k) w(\bar{x}, t) \, d\bar{x},$$

and

$$\int \omega(x, y) g(x, y, t) \, dxdy = \sum_{i \neq j} (N(N - 1))^{-1} \int \omega(x_i, x_j) w(\bar{x}, t) \, d\bar{x}.$$
Now, we use (1) to obtain an integral equation of the change of function $f$. For that we select $\phi(\bar{x}) = \psi(x_k)$, sum over $k$ and take the average. We get

$$\int \psi(x)[f(x, t + \delta) - f(x, t)] \, dx = N^{-1} \sum_{i \neq j} \frac{N(N - 1)}{2} \int [\psi(\hat{x}_i) - \psi(x_i)] w(\bar{x}, t) \, d\bar{x}$$

$$= N^{-1} \sum_{i \neq j} (N(N - 1))^{-1} \left( \int [\psi(\hat{x}_i) - \psi(x_i)] w(\bar{x}, t) \, d\bar{x} + \int [\psi(\hat{x}_j) - \psi(x_j)] w(\bar{x}, t) \, d\bar{x} \right)$$

$$= \frac{2}{N} \int \int [\psi(\hat{x}) - \psi(x)] g(x, y, t) \, dx dy,$$  \hfill (2)

where $x = (p, s), y = (p', s')$ and

$$\hat{x} = (p + h\bar{b}(p'), (1 - \mu h)s + \mu h a(b(p), b(p')))$$.

To proceed we make an assumption of statistical independence of the states of two randomly selected agents:

$$g(x, y, t) = f(x, t)f(y, t).$$

The plausibility of this condition is partially justified if the population is large, so that same agents are rarely matched together, and the information about the interaction is not shared between other agents.

Then, expanding $\psi(\hat{x})$ in Taylor series and integrating by parts, we obtain

$$\int \psi(x)[f(x, t + \delta) - f(x, t)] \, dx = -\frac{2}{N} \int \int \psi(p, s) \text{div}_{p,s} \left( (h\bar{b}(p'), \mu hs - \mu h a(b(p), b(p'))) f(x, t) \right) f(y, t) \, dy dx$$

$$+ O(h^2)$$

$$= -\frac{2}{N} \int \psi(p, s) \text{div}_{p,s} \left( (h\bar{b}(t), \mu hs - \mu h a(b(p), t)) f(x, t) \right) \, dx$$

$$+ O(h^2),$$  \hfill (3)

with the mean best response

$$\bar{b}(t) = \int \int b(p) f(p, s, t) \, dp ds,$$  \hfill (4)

and the mean fitness for using strategy $b(p)$ :

$$\bar{a}(b(p), t) = \int \int a(b(p), b(p')) f(p', s', t) \, dp' ds' = \sum_{i,j} a_{ij} b_i(p) \bar{b}_j(t).$$  \hfill (5)

Dividing equation (3) by $\delta$ and ignoring higher order terms we arrive at Fokker-Planck equation for density $f(p, s, t)$ :

$$\partial_t f + \frac{2h}{N\delta} \text{div}_p \left( \bar{b}(t) f \right) + \frac{2h \mu}{N\delta} \partial_s ((\bar{a}(b(p), t) - s) f) = 0.$$  \hfill (6)
Here we replaced the weak formulation with classical one, assuming that solution is regular enough. Moreover, in passing from a discrete to continuous time model we are assuming \( \delta, h \) are small, \( N \) is large, so that ratios
\[
\alpha_p = \frac{2h}{N\delta}, \quad \alpha_s = \frac{2\mu h}{N\delta}
\]  
(7)
are of finite order. Note that \( (N\delta)^{-1} \) can be interpreted as a number of interactions per agent, per unit of time, which we assume to be of order \( h^{-1} \gg 1 \).

Now we extend the model to variable size populations, by allowing agents to reproduce at the rate proportional their level of fitness. At this point we proceeding heuristically, leaving out the details of the derivation. With reproduction, the Fokker-Planck equation must be appended by a source term proportional to \((s - \bar{r}(t))f(p, s, t)\) on the right-hand side of (6), where \( \bar{r}(t) \) is mean population fitness
\[
\bar{r}(t) = \sum_{i,j} a_{ij} \bar{b}_i(t) \bar{b}_j(t).
\]  
(8)
The final model reads:
\[
\partial_t f + \alpha_p \text{div}_p (\bar{b}(t)f) + \mu \alpha_p \partial_s ((\bar{a}(b(p), t) - s)f) = \alpha(s - \bar{r}(t))f, \quad (9)
\]
with \( \bar{b}(t), \bar{a}(b(p), t) \) and \( \bar{r}(t) \) given by (4), (5), and (8), respectively, and \( \alpha > 0 \).

2.1. **Singular limit of recency parameter \( \mu \).** In the reproduction scenario described by (9), children acquire not only statistical empirical frequencies \( p \) of parents but also their averaged, accumulated fitness \( s \). Hypothetically, this might be a valid assumption in some situations, however, it seems more relevant to consider the case that it is only knowledge \( p \) that eventually determines the fitness of offspring.

This can easily be achieved in the framework of models (7)-(9) be taking the \( \mu \) to be large (of the order \( h^{-1} \)) which overweights the stimulus obtained from most recent encounters. For the derivation of the new model we proceed informally. Dividing equation (9) by \( \mu \) and passing to the limit \( \mu \rightarrow \infty \), we obtain
\[
\partial_s ((\bar{a}(b(p), t) - s)f) = 0.
\]
Since \( f \) is non-negative, this equation can be true only if for all \( p \in \mathbb{R}^d_+ \), and \( t > 0 \), \( f \) is a delta-function concentrated on value \( \bar{a}(b(p), t) \):
\[
f(p, s, t) = \delta(s - \bar{a}(b(p), t)).
\]
That is, fitness equals to the expected payoff for an agent using strategy \( b(p) \) against the population strategy profile \( \bar{b}(t) \):
\[
s = \bar{a}(b(p), t) = \sum_{ij} a_{ij} \bar{b}_i(p) \bar{b}_j(t).
\]
Now, the dimension of the problem can be reduced, as we can integrate (9) in \( s \), and find an equation for moment \( \int_{-\infty}^{\infty} f(p, s, t) ds \), which, with slight abuse of notation, we still call \( f \). The equation reads:
\[
\partial_t f + \alpha_p \text{div}_p (\bar{b}(t)f) = \alpha(\bar{a}(b(p), t) - \bar{r}(t))f
\]
\[
= \alpha \left( \sum_{ij} a_{ij} \bar{b}_i(p) \bar{b}_j(t) - \sum_{ij} a_{ij} \bar{b}_i(t) \bar{b}_j(t) \right) f. \quad (10)
\]
This is the equation of our main interest, for which we will establish global well-posedness. To complete the mathematical setup for equations (10) and (13) it remains to add the initial conditions for the density

\[ f(p, 0) = f_0(p), \quad p \in \mathbb{R}^d_+ . \]  

(11)

and boundary conditions (zero influx of probability):

\[ f(p, t) = 0, \quad p \in \partial \mathbb{R}^d_+, \quad t \geq 0 . \]  

(12)

Note that velocity vector \( \bar{b}(t) \) is always directed into \( \mathbb{R}^d_+ \), and the problem is not over-determined.

In the next two sections we show that the well-known best-response equation of Gilboa and Matsui [6], and the replicator equation of Taylor and Jonker [17], are contained as special cases in our model.

2.2. Relation to the best-response equation. Recall that in the model of fictitious play in heterogeneous populations, see [6], the agents meet for pairwise interactions and update their empirical frequencies \( P_n \) according to the rule we used above. Moreover, in that model, the size of the population \( N \) is large but fixed. This corresponds exactly with conditions under which equation (10) was derived, assuming in addition zero reproduction rate, i.e. that the population consists of the same agents. From equation (10) with \( \alpha = 0 \), we obtain:

\[ \partial_t f + \alpha_p \text{div}(\bar{b}(t) f) = 0, \]  

(13)

where the mean best response

\[ \bar{b}(t) = \int b(p) f(p, t) \, dp. \]

Using equation (13) we can compute the equation for the mean empirical frequencies vector \( P(t) = \int (p/ \sum_j p_j) f(p, t) \, dp : \)

\[ \frac{dP_i}{dt} = \alpha_p \int \frac{1}{\sum_j p_j} \left( \bar{b}_i(t) - \frac{p_i}{\sum_j p_j} \right) f(p, t) \, dp, \quad i = 1..d, \]

since \( \sum_j \bar{b}_j(t) = 1 \). If one postulates that all agents have the same, or approximately the same, empirical frequencies

\[ p(t) = (P_1(t), ..., P_d(t)), \]

then the above equation reduces to a variant of the best response dynamics equation:

\[ \frac{dP_i}{dt} = \frac{\alpha_p}{\sum_j P_j(t)} \left( \bar{b}_i(P) - P_i \right), \quad i = 1..d. \]

Notice, also, the positive factor on the right-hand side of the equations. For a learning processes in which empirical frequencies become large, the learning rate slows down.

2.3. Relation to the replicator equation. With zero learning rate \( \alpha_p = 0 \) model (10) is simply the replicator equation written in terms of the distribution function \( f(p, t) \). Indeed, in this case each agent uses a fixed strategy \( b(p) \), so that the population is split into at most \( d \) groups, each using a particular strategy, and each reproducing at the rate proportional to the averaged fitness obtained from interacting with whole population. Formally, one obtains the replicator equation by integrating (10) over sets \( \{ p : b(p) = e_k \}, \quad k = 1..d. \)
2.4. Existence of weak solutions. In this section we establish our main result, theorem 1. Let \( \Omega = \mathbb{R}^d_+ \), and \( C^1_0(\Omega) \) be a space of continuously differentiable functions with compact support in \( \Omega \). We adopt standard notation for \( L^p(\Omega) \) spaces and the space of functions of locally bounded variation \( BV_{loc}(\Omega) \). The latter consists of all measurable and locally integrable functions \( f \) such that for any ball \( B_r \),

\[
\| f \|_{TV(B_r \cap \Omega)} = \sup \left\{ \int_{B_r \cap \Omega} f(\operatorname{div} \psi) \, dp : \psi \in C^1_0(B_r \cap \Omega), \sup_p |\psi| \leq 1 \right\} < +\infty.
\]

For such functions, the distributional derivative \( \partial_i f, i = 1, \ldots, d \), is a signed Radon measure. One can find the information on these spaces and the results from functional analysis that we use below, for example, in a book by Brezis [2].

**Theorem 1.** Let \( f_0 \in C^1_0(\Omega) \) be a non-negative function with unit mass. There is a unique weak solution \( f \) of (10), (11), (12) such that

\[
f \in C([0, T]; L^1(\Omega)) \cap L^\infty([0, T]; BV(B_r \cap \Omega)), \quad \forall r, T > 0.
\]

For any \( t > 0 \), \( f(p, t) \geq 0 \), a.e. in \( \Omega \) and \( \int f(p, t) \, dp = 1 \).

**Proof.** From the definition of function \( b(p) \) and properties of \( BR(p) \) it follows that for any ball \( B_r \), \( b(p) \) has finite total variation on \( B_r \cap \Omega \), and there is \( C = C(r) \), but not depending on the center of the ball, such that

\[
\| b \|_{TV(B_r \cap \Omega)} \leq C. \tag{14}
\]

Equation (10) can be written in non-conservative form as

\[
\partial_t f + \bar{b}(t) \nabla f = \left( \sum a_{ij}(b_i(p) \bar{b}_j(t) - \bar{b}_i(t) \bar{b}_j(t)) \right) f, \tag{15}
\]

where for simplicity we set \( \alpha = 1 \). Given a continuous function \( \bar{b}(t) \) we solve this equation by the method of characteristics. For a mapping \( X^t : \mathbb{R}^d \to \mathbb{R}^d \), defined as

\[
X^t(p) = p + \int_0^t \bar{b}(\tau) \, d\tau,
\]

\( f \) is expressed through the formula

\[
f(X^t(p), t) = f_0(p) \exp \left\{ \int_0^t \sum a_{ij}(b_i(X^\tau(p)) \bar{b}_j(\tau) - \bar{b}_i(\tau) \bar{b}_j(\tau)) \, d\tau \right\},
\]

or as

\[
f(p, t) = f_0(p) - \int_0^t \bar{b}(\tau) \, d\tau
\]

\[
\times \exp \left\{ \int_0^t \sum a_{ij}(b_i(p) - \int_\tau^t \bar{b}_i(s) \, ds) \bar{b}_j(\tau) - \bar{b}_i(\tau) \bar{b}_j(\tau)) \, d\tau \right\}. \tag{16}
\]

Let \( g \in C([0, T]; L^1(\Omega)) \) be a non-negative function such that \( g(p, 0) = f_0(p) \), and \( \int g(p, t) \, dp = 1 \), for all \( t \in [0, T] \). We denote this subset of functions as \( K \). It is a closed, convex subset of \( C([0, T]; L^1(\Omega)) \).

Let \( \bar{b}_g(t) = \int b(p) g(p, t) \, dp \),

and define map \( f = \mathcal{L}(g) \) by evaluating (16) with \( \bar{b} = \bar{b}_g \). Notice that due to assumptions on \( g \), \( \sup_p |\bar{b}_g(t)| \leq 1 \). It follows that

\[
\sup_{p, t} f(p, t) \leq e^{CT} \sup_p f_0(p),
\]
for some $C > 0$ independent of $g$, and $\int f(p,t) \, dp = 1$. Moreover, the following lemma holds

**Lemma 1.** For any $r > 0$, $f \in L^\infty((0,T); BV(B_r \cap \Omega))$, and there is $C = C(r,T)$, independent of $g$ such that

$$\text{ess sup}_t \| f(\cdot,t) \|_{TV(B_r \cap \Omega)} \leq C(r,T).$$

**Proof.** Recall that $b(p)$ is a function of finite total variation that verifies estimate (14). Differentiating (16) in $p_k$, and using the chain rule we find that for any ball $B_r$,

$$\int_{B_r} |\partial_{p_k} f| \, dp \leq C(T) \int_\Omega |\partial_{p_k} f_0| \, dp$$

$$+ C(T) \sup f_0 \int_0^t \int_{B_r} |f_s b_i(p)| \, dp \, ds$$

$$\leq C(T) \int_\Omega |\partial_{p_k} f_0| \, dp + C(r,T) \sup f_0 \leq C(r,T),$$

(17)

where $|\partial_{p_k} b_i(p)|$ is a Borel measure. \hfill \Box

Using the argument of the last lemma one easily verifies that $f$ is Lipschitz continuous in time with values in $L^1(\Omega)$ :

**Lemma 2.** Let $t$ and $t + \delta \in [0,T]$. Then, there is $C = C(T)$, independent of $g$, such that

$$\int |f(p,t + \delta) - f(p,t)| \, dp \leq \|f_0\|_{C^1(\Omega)} C\delta.$$  

(18)

From the properties of $f = \mathcal{L}(g)$ that we have just established we see that $\mathcal{L}$ maps $K$ into itself. In addition, we now show that

**Lemma 3.** $\mathcal{L}[K]$ is pre-compact in $C([0,T]; L^1(\Omega))$.

**Proof.** Indeed, since $f$ has bounded total variation in $p$, we know that

$$\sup_{t \in [0,T]} \int |f(p+h,t) - f(p,t)| \, dp \leq C(T)h.$$  

The support of functions $f(\cdot,t)$ for all different $t$'s and $g$'s is contained in some fixed ball $B_r$ because $X^t$ is an uniform translation with a continuous vector $\int_0^t b_g(\tau) \, d\tau$. By Kolmogorov-Riesz-Frechet theorem, for all $t \in [0,T]$, set

$$\{ \mathcal{L}(g) \}_{g \in K}$$

is pre-compact in $L^1(\Omega)$. Using Lipschitz continuity in time, this also implies that

$$\{ \mathcal{L} \}_{g \in K}$$

is pre-compact in $C([0,T]; L^1(\Omega))$. \hfill \Box

Thus, $\mathcal{L}$ is a compact mapping from $K$ into itself. By Schauder fixed point theorem, there is a fixed point $f = \mathcal{L}(f)$ in $K \subset C([0,T]; L^1(\Omega))$. Clearly, it verifies all estimates that we have derived. Moreover, it can be shown that $f$ is a weak solution of pde (10).
Uniqueness of solutions follows from a stronger property, stability estimate. Let \( f_1, f_2 \) be two solutions of (10)–(12) with initial conditions \( f_0,1, f_0,2 \). Such solutions verify the formula (16), from which we find that

\[
\int |f_1(p,t) - f_2(p,t)| \, dp \leq C(T) \int |f_{0,1}(p) - f_{0,2}(p)| \, dp \\
+ C \int_0^t \int |f_1(p,\tau) - f_2(p,\tau)| \, dp \, d\tau.
\]

Thus, according to Gronwall’s inequality

\[
\int |f_1(p,t) - f_2(p,t)| \, dp \leq C(T) \int |f_{0,1}(p) - f_{0,2}(p)| \, dp.
\]

Now we collect information on the support of solutions of (10) that will be used in the proof of theorem 2.

**Lemma 4.** Suppose that \( \text{supp} \, f_0 \subset \text{Interior}(\Omega) \). Then, for any \( t > 0 \),

a. \( \text{supp} \, f(\cdot, t) \subset \text{Interior}(\Omega) \);

b. for any \( p \in \Omega \),

\[ |p + \int_0^t \bar{b}(\tau) \, d\tau| \geq t/d^2; \]

c. if \( \text{supp} \, f_0 \subset B_r(p_0) \), for some \( r \) and \( p_0 \in \Omega \), then

\[ \text{supp} \, f(\cdot, t) \subset B_r \left( p_0 + \int_0^t \bar{b}(\tau) \, d\tau \right). \]

**Proof.** Since for any \( t, \bar{b}(t) \in \Delta^{d-1} \subset \Omega \) and \( \Omega \) is a cone, it follows that \( \int_0^t \bar{b}(\tau) \, d\tau \in \Omega \) and for any \( p \in \text{Interior}(\Omega) \), \( p + \int_0^t \bar{b}(\tau) \, d\tau \in \text{Interior}(\Omega) \). Moreover, the distance from \( p + \int_0^t \bar{b}(\tau) \, d\tau \) to \( \partial \Omega \) is no less than the distance from \( p \) to \( \partial \Omega \). This proves part. a. Part b. follows from the fact that for any \( t > 0 \), \( \sum_{i=1}^d \bar{b}_i(t) = 1 \), and so, there is \( i_0 \), and there is \( \bar{b}_0 \subset [0, t] \), such that \( \bar{b}_{i_0}(t) \geq 1/d \), for all \( t \in \Delta_0 \), and \( |\Delta_0| \geq t/d \). Part c. follows immediately from (16).

2.5. **Asymptotic behavior in fictitious play.** Consider a model of statistical learning in a large population described by equation (13). An initial boundary-value problem (11), (12) with arbitrary \( f_0 \in C_0^1(\Omega) \), has a global unique solution, as was established in theorem 1. Denote population mean empirical frequencies by

\[ P(t) = \int \frac{p}{\sum_i p_i} f(p,t) \, dp, \]

and by \( \hat{f} \) the projection of \( f(p,t) \) onto the simplex \( \Delta^{d-1} \). That is,

\[ \hat{f}(\hat{p},t) = f(p,t), \quad \hat{p} = \frac{p}{\sum_i p_i} \in \Delta^{d-1}. \]

The next theorem shows that if the population averages \( P(t) \) and \( \bar{b}(t) \) converge to certain values, then these values must be the same and equal to a Nash equilibrium for the matrix game, and the empirical frequencies of every agent in the population converge to that Nash equilibrium.
Table 1. Hawk-Dove-Retaliator game.

|        | Hawk | Dove | Retaliator |
|--------|------|------|------------|
| Hawk   | -1   | 2    | -1         |
| Dove   | 0    | 1    | 0.9        |
| Retaliator | -1 | 1.1  | 1          |

**Theorem 2.** Suppose that \( \lim_{t \to \infty} P(t) = P_0 \) and \( \lim_{t \to \infty} \bar{b}(t) = b_0 \). Then,

\[
b_0 = P_0 \in \text{BR}(P_0),
\]

and \( \forall \epsilon > 0, \exists T(\epsilon) \) such that if \( t > T(\epsilon) \), then

\[
supp \hat{f}(\cdot, t) \subset B_\epsilon(P_0) \cap \Delta^{d-1}. \tag{19}
\]

**Proof.** Consider function \( \hat{f}(p, t) \) which is defined for \( p \in \Delta^{d-1} \). From the definition of \( P(t) \) it follows that \( P(t) \) belongs to the closed convex hull spanned by \( supp \hat{f}(\cdot, t) \). At time \( t = 0 \), the support of \( f_0 \) is separated from the origin, and thus, by properties b. and c. of lemma 4 (it applies to solutions of (13) as well), support of \( f(\cdot, t) \) will be contained in a ball of fixed radius and the center diverging to infinity. This means that the diameter of the support of projection \( \hat{f} \) decreases to zero. At the same time, since it contains point \( P(t) \) accumulating at \( P_0 \), statement (19) follows.

To prove the first statement, notice that for sufficiently small \( \epsilon \) and large \( t \), all mass of \( \hat{f} \) is near \( P_0 \) so that \( \bar{b}(t) \) is a convex combination of values of of \( \text{BR}(p) \) in polytops adjacent to point \( P_0 \), and so (see (25) from Appendix), is an element of \( \text{BR}(P_0) \). On the other hand \( P_0 \) must be equal to \( b_0 \), because of the transport structure of the kinetic equation (13).

3. Examples.

3.1. **Cultural evolution in Hawk-Dove-Retaliator game.** In this section we illustrate the dynamics of learning and its dependence of parameters of model (10) when the interactions between agents are described by the Hawk-Dove-Retaliator game from Table 1 that we take from Maynard Smith [11] and Zeeman [18].

The game has two ESS: a pure strategy Retaliator \( R \), and an equal mix of Hawk and Dove \( (0.5H + 0.5D) \). In the absence of learning \( (\alpha_p = 0, \alpha > 0) \) the replicator dynamics will proceed to one of ESS’s as shown in Figure 1, bottom plot. The same strategies are also asymptotically stable points for the best-response dynamics, which describes the statistical learning (fictitious play) in this game, in the absence of reproduction \( (\alpha_p > 0, \alpha = 0) \). Figure 1 shows the basins of attraction for each of the ESS’s, for the best-response and the replicator dynamics, respectively.

Consider a situation when the population consisting of three groups, whose initial empirical frequencies (learning priors) are located in three best-response polygons. With finite number of sub-populations the model reduces to a system of ODEs. In this particular example the density function

\[
f(p, t) = \sum_{i=1}^{3} w_i(t) \delta(p - p_i(t)), \quad \sum_{i=1}^{3} w_i(t) = 1.
\]
where functions $p_i$ and $w_i$ are solutions of
\[
\begin{align*}
\partial_t p_i &= \alpha_i \bar{b}(t), \quad i = 1, 3, \\
\partial_t w_i &= \alpha w_i \left( \sum_{kl} a_{kl} b_k(p_i(t)) \bar{b}_l(t) - a_{kl} \bar{b}_k(t) \bar{b}_l(t) \right), \quad i = 1, 3,
\end{align*}
\]
and
\[
\bar{b}(t) = w_1(t)b(p_1(t)) + w_2(t)b(p_2(t)) + w_3(t)b(p_3(t)).
\]  
(20)

Notice that all empirical frequencies $p_i(t)$ change in the direction of the mean best response $\bar{b}(t)$ (when projected to $\Delta^2$, this means that $p_i(t)$ moves toward $\bar{b}(t)$), and the weights $w_i$ change according to the performance of empirical frequencies $p_i$. The above system combines the best-response and the replicator equations, coupled through the relation (20). The qualitative behavior of solutions of this system depends on learning-to-reproduction ratio $\alpha_p/\alpha$.

We select the following set of initial data for the population proportions and empirical frequencies: $w_1(0) = 0.3$, $w_2(0) = 0.2$, $w_3(0) = 0.5$, $p_1(0) = (0.1, 0.8, 0.1)$, $p_2(0) = (0.7, 0.2, 0.1)$, $p_3(0) = (0.05, 0.25, 0.7)$. With such initial data, the mean best response $\bar{b}(0) = (0.3, 0.2, 0.5)$. In the discussion below we use the solutions of the above ODEs, with such initial data, and various values of $(\alpha, \alpha_p)$. Solutions were obtained numerically, using explicit Euler method, see Appendix for details.

As we mentioned above, with $\alpha_p = 0$, the dynamics coincides with that of replicator equation and empirical frequencies do not change: $p_i(t) = p_i(0)$. As the initial mean best response $\bar{b}(0)$ is located above line HOD, see Figure 1, $b(t)$ will converge toward state $R$, meaning that the group that using strategy $R$ out evolves the other two groups.

Consider now the pair of values $(\alpha, \alpha_p)$ with ratio $\alpha_p/\alpha = 1/5$. The dynamics of the empirical frequencies (projected on $\Delta^2$) and the mean best response is shown in Figure 2. Curves starting at points A, B and C represent what each group learns about the actions of all agents by time $t$. The curves move in the direction of the mean best response $\bar{b}(t)$, which is simply the mixed strategy in the whole population. As long as subgroups remain in their decision polygons, the agents in each subgroups play the same strategy $H$, $D$ or $R$. The evolution of $\bar{b}(t)$ proceeds according to the replicator equation and shows how the proportions of agents playing $H,D$ and $R$ (proportions of sizes of each subgroup, if all located in different decision polygons) change in the population. During this process, see Figure 2, a subgroup may move to an adjacent decision polygon, changing abruptly the proportions of agents adopting given strategies. This leads to a discontinuous change of $\bar{b}(t)$. Here, due to learning, all three groups end up using strategy $R$, and none outcompetes the others.

Now we increase learning-to-reproduction ratio to $\alpha_p/\alpha = 1/3$, keeping all other data unchanged. Figure 2 shows the corresponding dynamics. After a certain transient time, the group located initially at C moves into the adjacent decision polygon, causing the mean best response to change discontinuously to point $Q1$. After that, the dynamics of empirical frequencies and that of $\bar{b}(t)$ converge to a mixed strategy $0.5H + 0.5D$, so that half of the population is using playing $H$, and the other $D$.

Let us mention without presenting the simulations that the initial data $p_i(0)$ and $w_i(0)$ can be found such that if the evolution without learning proceeds to $0.5H + 0.5D$, then the addition of learning alters it to $R$. 

The above examples illustrate that learning results in significant changes to the course of the biological evolution, either with the respect the group competition, or the evolution final outcome.

Figure 1. Phase portraites for the best-response (top) and the replicator (bottom) equations for Hawk-Dove-Retaliator game in table 1. On the top plot, three polygonal regions, formed by lines $OM$, $ON$ and $O(0.5H + 0.5D)$ are the regions where the best response function a single value: $H$, $D$, or $R$, respectively. Line $KON$ separates the basins of attraction of states $R$ and $0.5H + 0.5D$. On the bottom plot line $HOD$ separates the basins of attraction of the same states $R$ and $0.5H + 0.5D$. The plots show two trajectories for the best-response and the replication equations that converge to $R$ and $0.5H + 0.5D$, respectively.

3.2. A model with constant interaction rate. In this section we consider a variant of model (10) which is obtained from a different scaling in (7). We will
Figure 2. Cultural evolution: small learning-to-reproduction ratio $\alpha_p/\alpha$. Three sub-populations starting at $A$, $B$, and $C$, respectively, move toward Retaliator strategy in empirical frequencies $p_i(t)$. Blue line is the trajectory of the mean best response (population strategy profile) $\bar{b}(t)$. It changes discontinuously when one of the sub-populations moves into an adjacent decision polygon. When all sub-groups move into the upper polygon, $\bar{b}(t) = R$.

Figure 3. Cultural evolution: large learning-to-reproduction ratio $\alpha_p/\alpha$. Same initial conditions as in Figure 2. All groups sharply change their dynamics and converge to $0.5H + 0.5D$ after the group that started at $C$ moves to the adjacent polygon. Blue line is a discontinuous trajectory of the mean best response $\bar{b}(t)$ (strategy profile). It starts at point $Q$, evolves continuously and then jumps to point $Q1$ and then proceed continuously to $0.5H + 0.5D$. 
assume that the total number of interactions among agents, per unit of time, measured by $\delta^{-1}$ is large and inversely proportional to stimulus increment $h \ll 1$, but is independent of the population size $N(t)$. In this case, the effective learning rate

$$\alpha_p = \frac{2h}{N(t)\delta} = \frac{\alpha_1}{N(t)},$$

where $\alpha_1$ is a constant. The system of equations governing the evolution is:

$$\partial_t f + \frac{\alpha_1}{N(t)} \text{div}_p(\bar{b}(t)f) = \alpha f \left( \sum_{ij} a_{ij} b_i(p) \bar{b}_j(t) - a_{ij} \bar{b}_i(t) \bar{b}_j(t) \right),$$

(21)

$$\partial_t N(t) = \alpha N \sum_{ij} a_{ij} \bar{b}_i(t) \bar{b}_j(t),$$

(22)

with $\bar{b}(t)$ given by (4). The main difference here compared to model (10) is an additional equation for the population size $N(t)$ which is coupled with the equation for $f$.

In this model the learning slows down with the increase of the population size. For sufficiently fast growth, for example exponential, this may prevent any significant learning and the evolution will proceed according to the fitness of strategies adopted by the agents, that is, according to the replicator dynamics. We illustrate this behavior on the Hawk-Dove-Retaliator game from section 3.1.

We will use the same initial data for $p_i(0)$ and $w_i(0)$, $i = 1..3$, and the initial population $N(0) = 10$, so that there are 3, 2 and 5 agents in groups 1, 2 and 3, respectively. The the model parameters $\alpha_1 = 1$ and $\alpha = 0.5$. The system (21), (22) reduces to a system of 5 ODEs for $w_i(t)$, $p_i(t)$, $N(t)$, $i = 1..3$.

The dynamics of empirical frequencies $p_1(t)$, $p_2(t)$, $p_3(t)$, and the mean best response $\bar{b}(t)$, shown in Figure 4, obtained from solving the system of ODEs numerically. In this example, which was run for sufficiently long time, empirical frequencies $p_i(t)$ asymptotically approach some locations in the same decision polygons where they have started, producing no significant learning. This means that the each sub-population uses the same best-response strategy as it had initially, and the sub-population proportions $w_i(t)$ change according to the replicator equation (blue line in Figure 4), asymptotically approaching state $R$.

This scenario could be different were the population starting at $C$ moved to the adjacent polygon before locking in. In that case (not shown), the mean best-response strategy would change discontinuously and proceed to state $(0.5H + 0.5D)$ as in Figure 3.

**Acknowledgments.** The author would to like to thank anonymous referees for critical assessment of the manuscript and numerous valuable suggestions for improving the presentation.

**Appendix A. Best response function BR(p).** Let $\Delta^{d-1}$ be the $d-1$ dimensional simplex $\{ p \in \mathbb{R}_+^d : \sum_j p_j = 1 \}$. Let $A = \{ a_{ij} \}$ represents payoff matrix in a symmetric game. We will assume that for no two indexes $i \neq j$,

$$\sum_k a_{ik} p_k = \sum_k a_{jk} p_k, \quad \forall p \in \Delta^{d-1}. $$

(23)
Figure 4. Cultural evolution with constant interaction frequency. Same initial conditions as in Figure 2. Sub-populations started at A, B and C do not leave their decision polygons. Trajectory starting at Q is the mean best-response (strategy profile) determined by the replication equations. Asymptotically it moves to R, meaning that the sub-population that started at point C out-evolves other sub-populations.

Denote by \( r_i(p) = \sum_k a_{ik}p_k \), the payoff to strategy \( i \) played against mixed strategy \( p \), and a set

\[
\mathcal{I}(p) = \left\{ i_0(p) \in 1..d : r_{i_0}(p) = \max_i r_i(p) \right\}.
\]

Denote the coordinate vectors \( e_i = (0,..0,1,0..0) \), with 1 in \( i^{th} \) position, and a multi-valued function

\[
\text{BR}(p) = \left\{ \text{convex hull of all } e_{i_0(p)}, \text{ such that } i_0(p) \in \mathcal{I}(p) \right\}.
\] (24)

Under hypothesis (23), \( \Delta^{d-1} \) is a union of finite number of polytops such that \( \text{BR}(p) \) is single-valued in the interior of each polytop \( P_k \), and at any point \( p \) on the boundary of \( P_k \), the best response \( \text{BR}(p) \) contains the value \( \text{BR}(p_1) \) from the interior of \( P_k \):

\[
\text{BR}(p_1) \in \text{BR}(p), \quad \forall p_1 \in \text{Interior}(P_k), \quad p \in \partial P_k.
\]

This condition can be re-phrased in an equivalent way, as a continuity condition: for any \( p \in \Delta^{d-1} \), there is \( \epsilon > 0 \), such that for any \( \epsilon_1 < \epsilon \), and any point \( p_1 \in B_{\epsilon_1}(p) \cap \Delta^{d-1} \),

\[
\text{BR}(p_1) \subseteq \text{BR}(p).
\] (25)

Finally, we select a single-valued representative of \( b(p) \) from the values of \( \text{BR}(p) \). If \( p \in \mathbb{R}_+^d \), then \( b(p) \) is one of the values of \( \text{BR}(p/\sum_i p_i) \). The selection can be, for example, the barycenter of the set of values of \( \text{BR}(p) \), which corresponds to the situation when agents are choosing one strategy at random (from an uniform distribution).
Appendix B. Numerical simulation. Numerical solution of system of ODEs were obtained using explicit Euler method with the step size $h < 0.01$, implemented in MATLAB R2017b. The code is available in the Supplementary Materials.

REFERENCES

[1] T. Börgers and R. Sarin, Learning through reinforcement and replicator dynamics, *J. Econom. Theory*, 77 (1997), 1–14.
[2] H. Brezis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, Universitext, Springer, New York, 2011.
[3] K. Chatterjee, D. Zufferey and M. A. Nowak, Evolutionary game dynamics in populations with different learners, *J. Theoret. Biol.*, 301 (2012), 161–173.
[4] D. Fudenberg and D. K. Levine, *The Theory of Learning in Games*, MIT Press Series on Economic Learning and Social Evolution, 2, MIT Press, Cambridge, MA, 1998.
[5] A. Gaunersdorfer and J. Hofbauer, Fictitious play, Shapley polygons and the replicator equation. Evolutionary game theory in biology and economics, *Games Econom. Behav.*, 11 (1995), 279–303.
[6] I. Gilboa and A. Matsui, Social stability and equilibrium, *Econometrica*, 59 (1991), 859–867.
[7] J. Hofbauer, From Nash and Brown to Maynard Smith: Equilibria, dynamics and ESS, *Selection*, 1 (2000), 81–88.
[8] J. Hofbauer, P. Schuster and K. Sigmund, A note on evolutionary stable strategies and game dynamics, *J. Theoret. Biol.*, 81 (1979), 609–612.
[9] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics*, Cambridge University Press, Cambridge, 1998.
[10] W. Hoppitt and K. Laland, *Social Learning: An Introduction to Mechanisms, Methods and Models*, Princenton University Press, 2013.
[11] J. Maynard Smith, *Evolution and the Theory of Games*, Cambridge University Press, New York, 1982.
[12] J. Maynard Smith and G. R. Price, The logic of animal conflict, *Nature*, 246 (1973), 15–18.
[13] M. Perepelitsa, Adaptive learning in large populations, *J. Math. Biol.*, 79 (2019), 2237–2253.
[14] P. Richerson and R. Boyd, *Not by Genes Alone: How Culture Transformed Human Evolution*, University of Chicago Press, 2004.
[15] J. Robinson, An iterative method of solving a game, *Ann. of Math. (2)*, 54 (1951), 296–301.
[16] L. S. Shapley, Some topics in two person games, in *Advances in Game Theory*, Princeton Univ. Press, Princeton, NJ, 1964, 1–28.
[17] P. D. Taylor and L. B. Jonker, Evolutionary stable strategies and game dynamics, *Math. Biosci.*, 40 (1978), 145–156.
[18] E. C. Zeeman, Dynamics of the evolution of animal conflicts, *J. Theoret. Biol.*, 89 (1981), 249–270.

Received June 2020; 1st revision November 2020; 2nd revision January 2021.

E-mail address: mperepel@central.uh.edu