

Adaptive Dynamics and Evolutionary Stability

J. HOFBAUER AND K. SIGMUND

Institut für Mathematik, Universität Wien and IIASA, Laxenburg

(Received June 1990)

Abstract. A dynamics for frequency dependent selection is proposed and applied to several biological examples. The relation with game dynamics and evolutionary stability is analyzed.

Game theory has been used successfully to model the evolution of biological traits whose advantage depends on their frequency in the population (see Maynard Smith [7]). The basic notion of evolutionary stability singles out solutions which, while safe from future change, need not be likely outcomes of past adaptations. The first to point this out was Eshel ([3], see also [4]). Recently, Taylor [11] and Nowak [8] stressed the independence of (a) being able to resist any invasion and (b) being able to invade everything. In this note, we propose a dynamics to model the effect of adaptation, relate it with the stability of equilibria, and discuss some examples concerning, in particular, iterated interactions and gamete sizes.

The phenotypic traits we consider are determined by *continuous* variables (like size, sex ratio or the probability of some behaviour). We assume that the population is essentially *monomorphic* (all in some state \mathbf{x} except for an occasional mutant in a near-by state $\mathbf{y} = \mathbf{x} + \mathbf{h}$.) The fitness of such an individual is denoted by $A(\mathbf{y}, \mathbf{x})$, its relative advantage $A(\mathbf{y}, \mathbf{x}) - A(\mathbf{x}, \mathbf{x})$ by $W(\mathbf{h}, \mathbf{x})$. The state $\hat{\mathbf{x}}$ is said to be evolutionarily stable, if $A(\mathbf{y}, \hat{\mathbf{x}}) \leq A(\hat{\mathbf{x}}, \hat{\mathbf{x}})$ for all \mathbf{y} , with $A(\hat{\mathbf{x}}, \mathbf{y}) > A(\mathbf{y}, \mathbf{y})$ if $A(\mathbf{y}, \hat{\mathbf{x}}) = A(\hat{\mathbf{x}}, \hat{\mathbf{x}})$ for some $\mathbf{y} \neq \hat{\mathbf{x}}$. This implies that a population $\hat{\mathbf{x}}$ cannot be invaded by a small population \mathbf{y} . But it offers no argument for the *establishment* of $\hat{\mathbf{x}}$. It can happen that an evolutionarily stable state $\hat{\mathbf{x}}$ is a 'garden of Eden' configuration, i.e., without a supplanted predecessor: near-by states evolve *away* from it. Since a slight perturbation of the environment can make $\hat{\mathbf{x}}$ lose its evolutionary stability, subsequent adaptations will lead it further astray.

The proposed adaptive dynamics (first version) is

$$\dot{x}_i = \left. \frac{\partial}{\partial y_i} A(\mathbf{y}, \mathbf{x}) \right|_{\mathbf{y}=\mathbf{x}} \quad i = 1, \dots, n. \quad (1)$$

The rationale is that a few mutants \mathbf{y} close to \mathbf{x} test out alternatives, and that the whole population evolves in the direction of the most promising one. This is supposed to mimic an evolution favouring individual fitness. It can also be used for learning models, under the assumption that trials are *myopic*, i.e., explore only the vicinity of the current state. If A is frequency independent, i.e., depends only on \mathbf{y} , we get the usual hill-climbing dynamics leading to local optimization.

In the one-dimensional case, (1) is

$$\dot{x} = \frac{\partial W}{\partial h}(0, x). \quad (2)$$

Following Taylor [11], a state \hat{x} will be said to be *h-stable* if $W(h, \hat{x}) < 0$ for all small $h \neq 0$, and *x-stable* if $W(h, x)$ has the sign of $h(\hat{x} - x)$ for small h . Thus *h-stability* means

that \hat{x} is a strict (local) Nash equilibrium, and hence locally evolutionarily stable, while x -stability means that if the population is in a state x close to \hat{x} , then any mutations in the right direction (towards \hat{x}) will succeed. In each case, \hat{x} is an equilibrium of (2). Generically, an equilibrium is x -stable iff $\frac{\partial^2 W}{\partial x \partial h} < 0$ and h -stable iff $\frac{\partial^2 W}{\partial h^2} < 0$.

For example, if x is the sex-ratio, then $A(y, x) = \frac{y}{x} + \frac{1-y}{1-x}$ and hence (2) yields

$$\dot{x} = \frac{1-2x}{x(1-x)}.$$

The equilibrium $\hat{x} = \frac{1}{2}$ is both x -stable and evolutionarily stable, but only 'weakly h -stable' since $W(h, \hat{x}) = 0$ for all h . Eshel [3] gave some ingenious examples of equilibria which are h - but not x -stable. Taylor [11] states that he has yet to see biologically plausible equilibria which are x - but not h -stable (or not evolutionarily stable). We believe that they can be found for iterated interactions.

The Iterated Prisoner's Dilemma is a paradigm for the evolution of cooperation [2]. In each round of the interaction, the choice is between E_1 (cooperation) and E_2 (defection) and the payoff is given by the 2×2 matrix a_{ij} satisfying

$$a_{21} > a_{11} > a_{22} > a_{12} \quad \text{and} \quad 2a_{11} > a_{22} + a_{12}. \quad (3)$$

Tit For Tat (start with E_1 , then do whatever the other did last) proved amazingly successful in computer tournaments. But in biological applications, there is always the probability of an error which turns a mistaken defection into the starting point of an expensive feud. In this context, it is interesting to study *stochastic reactive strategies* defined by the conditional probability x_i to play E_1 , given that the other's last move was E_i . For simplicity, we consider the *infinitely* repeated case. The initial move, then, does not matter (but see [9] for the general situation). A simple computation yields

$$A(y, x) = \sum_{i,j=1}^2 a_{ij} c_i d_j$$

with

$$c_1 = \frac{y_2 + (y_1 - y_2)x_2}{1 - (y_1 - y_2)(x_1 - x_2)} \quad d_1 = \frac{x_2 + (x_1 - x_2)y_2}{1 - (y_1 - y_2)(x_1 - x_2)}$$

and $c_2 = 1 - c_1$, $d_2 = 1 - d_1$. Hence (1) yields

$$\dot{x}_1 = \frac{x_2}{(1-r)^3(1+r)} F(x_1, x_2) \quad \dot{x}_2 = \frac{1-x_1}{(1-r)^3(1+r)} F(x_1, x_2)$$

with

$$F(x_1, x_2) = ax_2(1+r) + ((a_{12} - a_{22}) + (a_{21} - a_{22})r)(1-r)$$

where $a = a_{11} - a_{21} + a_{22} - a_{12}$ and $r = x_1 - x_2$. Now suppose there is no variation in x_2 : it is fixed at some small value ε (the probability of a mistake). If $a < 0$, there is a unique \hat{x}_1 with $F(\hat{x}_1, \varepsilon) = 0$. This \hat{x}_1 is h - but not x -stable (see Fig. 1). On the other hand, if $x_1 = 1 - \varepsilon$ is fixed at some value close to 1, there exists for $a > 0$ an \hat{x}_2 such that $F(1 - \varepsilon, \hat{x}_2) = 0$, which is x - but not h -stable (see Fig. 2). Is this biologically plausible? The fixation of $x_1 = 1 - \varepsilon$ can be viewed as a consensus to cooperate as long as the other does. The evolutionary interesting variable is then the probability $1 - x_2$ to retaliate after a defection. It should offer no incentive for exploitation, but avoid needless recrimination after a mistake.

A similar example holds for the repeated game of *Chicken* (or *Hawk-Dove*): E_2 means now escalation of the fight, E_1 avoidance of it. Instead of (3) one has $a_{21} > a_{11} > a_{12} > a_{22}$. If $x_1 = \varepsilon$ is fixed at some small value (if the adversary chickened out last time, one gets bold), the question is to find the right probability to escalate following an escalation of the adversary. It is given by the value \hat{x}_2 of $F(\varepsilon, \hat{x}_2) = 0$. Again, \hat{x}_2 is x -stable (the adaptive dynamics leads towards it), but not h -stable (see Fig. 3).

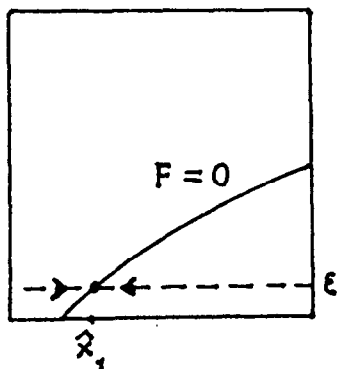


Figure 1

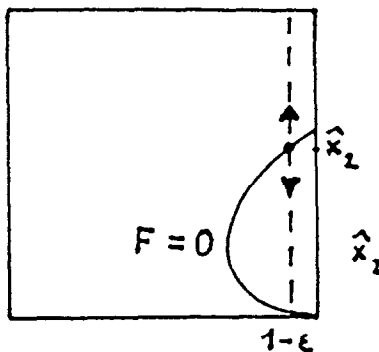


Figure 2

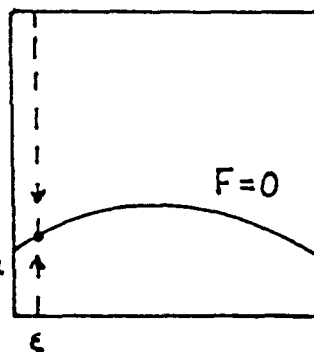


Figure 3

Let us consider now a sexually reproducing species with two mating types 1 and 2. A phenotype \mathbf{x} corresponds to the choice of gamete weights x_1 and x_2 when of type 1 or 2. The number of gametes is inversely proportional to their weight. The survival probability $V(m)$ of a zygote of mass m is increasing in m . Then, assuming sex ratio $\frac{1}{2}$ and mass action, we get the payoff

$$A(\mathbf{y}, \mathbf{x}) = \frac{1}{2} \left(\frac{V(y_1 + x_2)}{y_1 x_2} + \frac{V(y_2 + x_1)}{y_2 x_1} \right),$$

so that (1) becomes

$$\dot{x}_i = \frac{1}{x_1 x_2} \left(V'(x_1 + x_2) - \frac{V(x_1 + x_2)}{x_i} \right) \quad i = 1, 2.$$

Since

$$\dot{x}_1 - \dot{x}_2 = \frac{4V(x_1 + x_2)}{(x_1 x_2)^2} (x_1 - x_2),$$

we see that the flow points away from the (invariant) diagonal. The equilibria are on the diagonal (satisfying $V(2\hat{x}_1) = \hat{x}_1 V'(2\hat{x}_1)$). They are evolutionarily stable (uninvadable if one of the gamete sizes is fixed), if $V''(2\hat{x}_1) < 0$: but it is highly unlikely that evolution will be trapped at such a point. The adaptive dynamics leads to a dimorphism of gamete sizes (one mating type follows the *male* strategy of producing many small gametes, the other mating type the *female* one).

So far, we assumed that \mathbf{x} varies in an open subset of \mathbb{R}^n equipped with the Euclidean metric: fluctuations in every directions are equally likely. It may happen that another metric is more appropriate, e.g., if the genetic or developmental constraints render variations in one direction more likely than in another. It may also happen that \mathbf{x} is restricted to some subset of \mathbb{R}^n , for example to the simplex S_n , if the x_i are probabilities of some strategies, summing up to 1.

Again, in an adaptive dynamics, the prevalent state \mathbf{x} of the population will tend in the direction of the maximal local increase of the fitness $A(\mathbf{y}, \mathbf{x})$, i.e., the *y-gradient* of $A(\mathbf{y}, \mathbf{x})$ at $\mathbf{y} = \mathbf{x}$: $\dot{\mathbf{x}}$ will be proportional to the unit vector ξ maximizing $A(\mathbf{x} + \varepsilon\xi, \mathbf{x}) - A(\mathbf{x}, \mathbf{x})$, in the limit $\varepsilon \rightarrow 0$. Obviously, this notion of gradient, and therefore the adaptive dynamics associated with the fitness function $A(\mathbf{y}, \mathbf{x})$, depends on a *Riemannian metric* to be chosen on the state space.

Let G be a general Riemannian metric which associates (in a smooth way) to each \mathbf{x} a symmetric positive definite matrix $G(\mathbf{x}) = (g_{ij}(\mathbf{x}))$, such that the inner product at \mathbf{x} is given by

$$\langle \xi, \eta \rangle_{\mathbf{x}} = \xi^T G(\mathbf{x}) \eta = \sum_{ij} g_{ij}(\mathbf{x}) \xi_i \eta_j. \quad (4)$$

If the state space is (an open subset of) \mathbf{R}^n , then it is well-known that the G -gradient of a function $V(\mathbf{x})$ is given by $G(\mathbf{x})^{-1} \text{grad } V(\mathbf{x})$, where $\text{grad } V(\mathbf{x})$ is the Euclidean gradient. From (4) and the definition of adaptive dynamics, we have

$$\langle \eta, \dot{\mathbf{x}} \rangle_{\mathbf{x}} = \eta^T G(\mathbf{x}) \dot{\mathbf{x}} = \sum g_{ij}(\mathbf{x}) \eta_i \dot{x}_j = D_{\mathbf{y}} A(\mathbf{y}, \mathbf{x})|_{\mathbf{y}=\mathbf{x}}(\eta) \quad (5)$$

for all η .

Let us now consider the most important special case where $\mathbf{x} \in S_n$ and A is linear in \mathbf{y} , i.e.,

$$A(\mathbf{y}, \mathbf{x}) = \mathbf{y}^T \mathbf{f}(\mathbf{x}) = \sum_i y_i f_i(\mathbf{x}). \quad (6)$$

Then (5) must hold for all η in the tangent space \mathbf{R}_0^n , and we obtain

$$\sum_j g_{ij} \dot{x}_j = f_i(\mathbf{x}) + \Phi(\mathbf{x}) \quad (7)$$

for some function Φ which is determined by the constraint $\sum_i \dot{x}_i = 0$. If the matrix G is invertible, which can always be achieved by extending it in a positive definite way from \mathbf{R}_0^n to \mathbf{R}^n , we can transform (7), i.e., $G\dot{\mathbf{x}} = \mathbf{f} + \Phi \mathbf{1}$, into

$$\dot{\mathbf{x}} = G^{-1} \mathbf{f}(\mathbf{x}) + \mathbf{g} \Phi(\mathbf{x}), \quad (8)$$

with $\mathbf{g} = G^{-1} \mathbf{1}$, $\mathbf{1}$ denoting the n -vector of 1's. Then $\mathbf{1}^T \dot{\mathbf{x}} = \sum \dot{x}_i = 0$ implies $\mathbf{1}^T G^{-1} \mathbf{f} + \Phi \mathbf{1}^T \mathbf{g} = 0$ or $\Phi = \frac{\mathbf{g}^T \mathbf{f}}{\mathbf{1}^T \mathbf{g}}$, where $\mathbf{1}^T \mathbf{g} = \mathbf{1}^T G^{-1} \mathbf{1} > 0$. Hence, (7) is equivalent to the explicit form

$$\dot{x}_i = \sum_j c_{ij}(\mathbf{x}) f_j(\mathbf{x}) \quad (9)$$

with

$$C = G^{-1} - \frac{\mathbf{g} \mathbf{g}^T}{\mathbf{g}^T \mathbf{1}}. \quad (10)$$

It is not hard to see that C restricted to \mathbf{R}_0^n is positive definite. Conversely, if C is a symmetric $n \times n$ matrix, with $C \mathbf{1} = 0$ and $\xi^T C \xi > 0$ for $\xi \in \mathbf{R}_0^n \setminus \{0\}$, then we can define G as a generalized inverse of C and thus obtain a Riemannian metric on S_n , such that (9) is the corresponding adaptive dynamics.

We illustrate this with two examples: a) *Euclidean metric* on S_n . Here $g_{ij}(\mathbf{x}) = \delta_{ij}$, hence $\mathbf{g} = G^{-1} \mathbf{1} = \mathbf{1}$ and $C = I - \frac{1}{n} \mathbf{1} \mathbf{1}^T$, i.e., $c_{ij} = \delta_{ij} - \frac{1}{n}$, so that the adaptive dynamics reads

$$\dot{x}_i = f_i(\mathbf{x}) - \frac{1}{n} \sum_{k=1}^n f_k(\mathbf{x}).$$

b) The *Shahshahani metric* [1] on S_n is defined by $g_{ij} = \delta_{ij} \frac{1}{x_i}$. Here $\mathbf{g} = G^{-1} \mathbf{1} = \mathbf{x}$, so by (10), $c_{ij} = x_i \delta_{ij} - x_i x_j$ and the adaptive dynamics reads

$$\dot{x}_i = x_i (f_i(\mathbf{x}) - \bar{f}(\mathbf{x})), \quad \bar{f}(\mathbf{x}) = \sum_k x_k f_k(\mathbf{x}).$$

This is the *replicator equation* [6]. It is interesting to note that while the metric is defined only in $\text{int } S_n$, the dynamics extends continuously to the boundary, and leaves the boundary faces invariant. Hence this kind of adaptation does not lead to new strategies.

Let now $\hat{x} \in \text{int } S_n$ be an ESS for the fitness function $A(y, x) = y^T f(x)$, then by Pohley and Thomas [10]

$$\hat{x}^T f(x) > x^T f(x) \quad \text{for } x \text{ near } \hat{x}, x \neq \hat{x}. \quad (11)$$

We claim that \hat{x} is locally asymptotically stable for each adaptive dynamics (9). Let $V(x) = \langle x - \hat{x}, x - \hat{x} \rangle_{\hat{x}} = (x - \hat{x})^T G(\hat{x})(x - \hat{x})$, an approximation for the geodesic distance of x to \hat{x} . Then by (7)

$$\begin{aligned} \dot{V}(x) &= 2(x - \hat{x})^T G(\hat{x})\dot{x} \approx 2(x - \hat{x})^T G(x)\dot{x} \\ &= 2(x - \hat{x})^T (f(x) + \Phi 1) = 2(x - \hat{x})^T f(x) < 0. \end{aligned}$$

Hence, V is a local Ljapunov function, and \hat{x} is asymptotically stable. Therefore, near an ESS, the geodesic distance between orbits decreases monotonically for each adaptive dynamics. In particular, for the replicator equation, the distance induced by the Shahshahani metric, i.e., the arccos distance [1, p. 39], tends monotonically to 0, near an ESS. Actually, at least for linear payoff functions $f(x) = Ax$, this gives a remarkable relation between the notion of ESS and the Shahshahani metric: \hat{x} is an ESS, if and only if the geodesic distance decreases near \hat{x} in the replicator equation. A related result of Hines [5] shows some other sort of converse: If \hat{x} is stable for each adaptive dynamics (7), i.e., for each choice of C or G , then \hat{x} must (essentially) be an ESS.

REFERENCES

1. E. Akin, *The Geometry of Population Genetics*, Lecture Notes in Biomathematics 31, Springer, Berlin - Heidelberg - New York, (1979).
2. R. Axelrod, *The Evolution of Cooperation*, Basic Books, New York, (1984).
3. I. Eshel, Evolutionary and continuous stability, *J. Theor. Biol.* 103, 99-111 (1983).
4. I. Eshel and U. Motro, Kin selection and strong evolutionary stability of mutual help, *Theor. Pop. Biol.* 19, 420-433 (1981).
5. W.G.S. Hines, Three characterizations of strategy stability, *J. Appl. Prob.* 17, 333-340 (1980); , 21, 213-214 (1984).
6. J. Hofbauer and K. Sigmund, *The Theory of Evolution and Dynamical Systems*, Cambridge University Press, (1988).
7. J. Maynard Smith, *Evolution and the Theory of Games*, Cambridge University Press, (1982).
8. M. Nowak, An evolutionarily stable strategy may be inaccessible, *J. Theor. Biol.* (to appear).
9. M. Nowak and K. Sigmund, The evolution of reactive strategies in iterated games, *Acta Applicandae Math.* (submitted).
10. H.-J. Pohley and B. Thomas, Nonlinear ESS-models and frequency dependent selection, *Biosystems* 16, 87-100 (1983).
11. P.D. Taylor, Evolutionary stability in one-parameter models under weak selection, *Theor. Pop. Biol.* 36, 125-143 (1989).

Institut für Mathematik, Universität Wien, Strudlhofgasse 4, A-1090 Vienna, Austria
 IIASA, A-2361 Laxenburg, Austria