



## Continuous Stability and Evolutionary Convergence

ILAN ESHEL,\* UZI MOTRO† AND EMILIA SANSONE‡

\* *Department of Statistics, School of Mathematical Sciences, Tel Aviv University, Ramat-Aviv, Tel-Aviv 69978, Israel;* † *Department of Evolution, Systematics and Ecology, and Department of Statistics, The Hebrew University of Jerusalem, Jerusalem 91904, Israel;* and the ‡ *Department of Mathematics and Its Applications, Università degli Studi di Napoli, Naples, Italy*

(Received on 21 February 1996, Accepted in revised form on 6 November 1996)

A stochastic process of long-term evolution due to mutation and selection is defined over an asexually reproducing population, with selection according to a population game with a one-dimensional continuity of pure strategies. Limiting the analysis to mutations of small effect, it is shown that long-term dynamic stability in such a process is equivalent to continuous stability in the relevant population game. In the case of a one-dimensional strategy set (but not necessarily if the strategy set is multi-dimensional), this result is virtually independent of the distribution of mutations.

© 1997 Academic Press Limited

### 1. Introduction

The concept of Continuous Stability (CS or CSS for a Continuously Stable Strategy) has been suggested (Eshel & Motro, 1981) for symmetric population games in which there is a continuum of pure strategies, available to the players. As has been maintained, in such a case there are two evolutionarily relevant requirements for stability which, contrary to first intuition, do not necessarily follow from one another. One requirement, suggested by Maynard Smith & Price (1973), is the well-known ESS condition which requires that a strategy, once almost fixed in the population, will be advantageous over any single mutant strategy when in a low enough frequency. Another natural requirement is that small deviations of the entire population from the ESS will end up with selective advantage to mutations that render the population back to the ESS but not to mutations towards the other side. A strategy which satisfies these two requirements has been given the name CSS. The second condition alone was later given the name *m*-stability by Taylor (1989) and convergence stability by Christiansen (1991).

More specifically, assume a continuous-state population game in which the pure strategies are identified with values on the real line. Let  $V(x, y)$  be the pay-off of a player in this game who chooses the pure strategy  $x$  against an opponent who chooses the pure strategy  $y$ . Allowing only for mutations of small enough effect, it is easy to see that fixation on the pure strategy  $x^*$  is an ESS if

$$V_x(x^*, x^*) = 0 \tag{1}$$

and

$$V_{xx}(x^*, x^*) < 0. \tag{2}$$

[A necessary condition for ESS is, indeed, that (1) will hold with (2) as a weak inequality.] An additional condition, guaranteeing continuous stability is

$$V_{xx}(x^*, x^*) + V_{xy}(x^*, x^*) < 0 \tag{3}$$

[Eshel 1983; a necessary condition for CSS is that (1) will hold with (2) and (3) as weak inequalities.] Note that since  $V_{xx}(x, x) + V_{xy}(x, x) = d/dx V_x(x, x)$ , (1) together with (3) guarantee that for  $x > x^*$  at the vicinity of  $x^*$ ,  $V_x(x, x) < 0$  and vice versa for  $x < x^*$ . In both cases, this means that for  $y$  close enough to

$x$ ,  $V(y, x) > V(x, x)$  if and only if  $y$  deviates from  $x$  in the direction of  $x^*$ .

This, indeed, seems a reasonable requirement for stability but one should recall that the concept of stability actually depends both on the specific dynamics we are dealing with and, not less importantly, on the topology assumed about the space of strategies or, in this case, on how we define evolutionary closeness.

Thus, the concept of dynamic stability traditionally employed in population dynamics, corresponds to changes of frequencies within a given, finite set of genotypes, potentially existing in the present population. We refer to this as short-term evolutionary dynamics. This, to be distinguished from long-term evolutionary dynamics (Eshel, 1991), in which all sorts of mutations are introduced into the population at random, each being repelled from or established in the population according to the short-term dynamics of natural selection.

The concepts of ESS and CSS, on the other hand, by their definition assume the dynamics of long-term evolution [this, by the way, may account for the various contradictions between results, obtained by analysis of ESS and those based on so-called exact genetic models; see Eshel & Feldman (1984) and Eshel (1996)], yet they correspond to different aspects of evolutionary closeness. The concept of ESS, like that of short-term dynamics stability, is based on the Euclidean topology of distances between probability vectors, according to which the only distributions of strategies close to the fixation on a given phenotypic value  $x$ , are the almost fixations on  $x$ . In the case of a discrete set of pure strategies this is, actually, the only evolutionarily relevant measure of distance. However, in the case of a continuous change of a quantitative parameter during the long course of evolution, a different sort of "closeness", and therefore of stability, is required. In this case, the neighborhood of a fixation on a phenotypic value  $x$  must include not only "almost fixations on  $x$ ", but also "fixations (or almost fixations) on almost  $x$ ", which means that populations fixed on morphological traits close to each other, are regarded as evolutionarily close even if they have no type in common.

With this notion of evolutionary closeness, the natural requirement for long-term stability of a specific phenotypic value  $x^*$  is that, starting from any distribution of values, each being close enough to  $x^*$ , the long-term evolutionary dynamics of mutation and selection will render the population ever closer to it.

As we have seen (Eshel & Motro, 1981; Eshel, 1983), the ESS condition (1) and (2) alone is insufficient for this type of long-term stability. But are

the CSS conditions (1), (2) and (3) really sufficient for it? In the most general case, in which we start from any distribution of strategies at the vicinity of the CSS, even this is not that clear, since the CSS requirement is directly applied only to deviant fixations of pure strategies. Even less clear, on the face of it, is whether the CSS condition, as given above, is really necessary for long-term convergence. Thus, it has been speculated by several authors that the conditions (1) and (3) alone, or maybe some variant of them (say  $m$ -stability or convergence stability) is sufficient for what we refer to as long-term stability.

We see that, not surprisingly, this is indeed not true for the short-term evolutionary dynamics. We, therefore, start this work by analysing the long-term evolutionary dynamics at the vicinity of any given pure strategy  $x^*$ . To this end, we first characterize all distributions with a support at a certain vicinity of  $x^*$ , which are stable in the short-term dynamics. For any such short-term stable distribution  $F$  we then characterize the set of small-effect mutations under which the selection forces determined by the distribution  $F$  of strategies are initially selected. We shall see that any mutation within this set, once established in the population, starts a new short-term process of natural selection which converges to a new short-term stable distribution at the vicinity of  $x^*$ . This procedure determines a stochastic structure of transition probabilities among all short-term-stable distributions at the vicinity of  $x^*$ . This stochastic process represents the long-term evolutionary dynamics. We see that it converges to  $x^*$  with probability one if and only if  $x^*$  is a CSS. Moreover, if  $x^*$  is not a CSS (even if it is  $m$ -stable), then the long-term evolutionary dynamics does not converge to  $x^*$  with probability one.

## 2. The Model—Long-term Evolution and Long-term Convergence

We consider a symmetric, pairwise-encounter population-game with a continuous, one-dimensional set of pure strategies. We assign to these strategies the real values  $x, y, \dots$  etc. A general, mixed strategy in such a game is given by any probability distribution  $F$  on the real line. Thus, the pay-off obtained by an  $F$ -player when playing against a population-strategy  $G$  is given by

$$V(F, G) = \iint V(x, y) dF(x) dG(y), \quad (4)$$

where  $V(x, y)$  is the pay-off to an  $x$ -player when encountering a  $y$ -player. We assume that the function  $V(x, y)$  has continuous second derivatives. For any specific value  $x^*$  we write, for

convenience,  $V_x^* = V_x(x^*, x^*)$ ,  $V_{xx}^* = V_{xx}(x^*, x^*)$  and  $V_{xy}^* = V_{xy}(x^*, x^*)$ .

DEFINITIONS

(i) A distribution  $F$  is said to be a short-term equilibrium if for all  $x$  in the support of it

$$V(x, F) = V(F, F), \tag{5}$$

namely, if  $F$  is an invariant of either the continuous time dynamics or the discrete time asexual population-game dynamics with no mutation.

(ii) A short-term equilibrium  $F$  is said to be short-term stable if it is stable (i.e. Lyapunov stable) either in the continuous time dynamics or the discrete time asexual population-game dynamics, when perturbations from  $F$  are limited to the set of strategies within the support of  $F$ , namely those strategies which are already present in positive density within an  $F$ -distributed population. (Note that any fixation of a single pure strategy is, in a trivial way, a short-term stable equilibrium.)

Assume now the dynamic process  $\{F_t\}$ , determined for all  $t \geq 0$  by either the continuous time dynamics or the discrete time asexual population-game dynamics.

Definition

If  $\lim F_t = F$  we say that  $F_0$  is short-term converging to  $F$ .

Remark

The definitions of short-term convergence and short-term stability are, indeed, slightly different for the continuous time dynamics or the discrete time asexual population-game dynamics. We see, however, that all results obtained in this work will be applicable to both.

Assume now any short-term stable equilibrium  $F$ . We know that  $F$  is stable against small deviations, caused by the introduction of a small frequency of any pure strategy, already present on its support. One can ask, though, which mutant strategies outside the support of  $F$  can be successfully established in an  $F$ -distributed population. Dealing with a pure-strategy mutant  $x$ , it is easy to see that either in the continuous time dynamics or the discrete time asexual population-game dynamics, such a mutant will have an initial success in an  $F$  population if and only if it satisfies one of the following two conditions:

$$V(x, F) > V(F, F) \tag{6}$$

or

$$V(x, F) = V(F, F) \quad \text{and} \quad V(x, x) > V(F, x). \tag{7}$$

We now restrict our analysis to mutations of small effect. More specifically, we assume a small positive value  $h > 0$  such that the phenotypic change, caused by any single mutation, cannot surpass the value  $h$ . This means that, starting from any distribution  $F$  of strategies in the population, the only mutant strategies to occur must be in the  $h$ -vicinity of the support of  $F$ .

Definition

A strategy  $x$  in the  $h$ -vicinity of the support of a short-term stable equilibrium  $F$ , satisfying either (6) or (7), is called an  $F$ -intruder.

Once an  $F$ -intruder is initially established with a small frequency  $\epsilon > 0$  in the population, a new distribution,  $(1 - \epsilon)F + \epsilon I_{[x, \infty)}$  (where  $I_{[x, \infty)}(\xi) = 1$  for  $\xi \geq x$  and  $I_{[x, \infty)}(\xi) = 0$  for  $\xi < x$ ), becomes a subject to the forces of the relevant short-term population dynamics which may lead to the establishment of a new short-term stable equilibrium.

Definition

If  $F$  is a short-term stable equilibrium,  $x^*$  is an  $F$ -intruder, and if, for small enough value  $\epsilon > 0$ ,  $(1 - \epsilon)F + \epsilon I_{[x, \infty)}$  is short-term converging to the distribution  $G$ , then we say that *the passage from  $F$  to  $G$  is possible*. We denote it by  $F \rightarrow G$ .

A priori it is not guaranteed that the short-term process of natural selection, when starting from an arbitrary distribution of the form  $(1 - \epsilon)F + \epsilon I_{[x, \infty)}$ , will converge. In the next section, however, we shall see that this is always the case if only both the support of the short-term stable equilibrium  $F$  and the mutation  $x$  lie within some well-defined vicinity of the point  $x^*$  in question. We shall see, moreover, that if the support of  $F$  lies within this vicinity when  $x^*$  is a CSS, then there are no  $F$ -invaders outside this vicinity, and the process of long-term evolution must pass then from one short-term stable equilibrium to another within this vicinity. In such a case, the probability-distribution of such passages depends on the distribution of the mutations, because for any short-term stable equilibrium  $F$  there may be many  $F$ -invaders, each leading to a passage to a different new equilibrium; hence the probability of a passage from  $F$  to any measurable set  $B$  of equilibria is exactly the probability that the first  $F$ -invader mutation to appear will initiate a passage to a state in  $B$ .

Assume, for a moment, that we know the distribution of the mutations to which any pure strategy is exposed. (We do not, in reality, and

therefore we are looking for results which are independent of this distribution.)

*Definition*

Let  $L$  be any set of pure strategies. Consider a process, either stochastic or sub-stochastic, its states all being short-term stable equilibria with a support on  $L$ . The transition probability  $P(F, B)$  from a state  $F$  to a measurable set of states  $B$  is defined to be the probability that the first  $F$ -intruder mutation to occur, will lead to short-term convergence to a state in  $B$ . We call such a process *long-term evolution on  $L$* .

Note that the process we refer to as long-term evolution on  $L$  is a well-defined stochastic process if and only if for any short-term stable equilibrium  $F$  with a support on  $L$ , the short-term process of selection, following the establishment of any  $F$ -intruder, converges to another short-term stable equilibrium with a support on  $L$ . Indeed, any realization of such a process is a sequence  $\{F_n\}$ , finite or infinite, of short-term stable equilibria.

REMARKS

(i) Note that in real situations, we never know the distribution of the effect of a single mutation. Hence, we never know the real transition probabilities of the process which we refer to as long-term evolution. As we shall see, we can, nevertheless, characterize all the possible realizations of such a process, starting with a support at a given vicinity of a pure strategy  $x^*$ , namely all sequences  $\{F_n\}$  of short-term stable equilibria such that for all  $n = 0, 1, 2, \dots, F_n \rightarrow F_{n+1}$ , with the support of  $F_0$  at a given vicinity of  $x^*$ . All results, obtained in this work, are based exclusively on this information, drawn from the structure of the game  $V(x, y)$  alone.

(ii) In the most general case, the process of long-term evolution (Eshel, 1991, 1996) can, in principle, be defined for any population dynamics, either asexual or, more interestingly, sexual, and over all possible states at which the population can be found as a result of natural selection without mutation. In theory, these include also states of permanent cycling or of chaos. A generalization of the definition of long-term evolution, given above, can be obtained, in theory, by first characterizing all mutations that can successfully enter a population in such a state, and then identifying the new state, determined by the introduction of any such mutation into the population. Analytically, however, this is a very difficult project and, as we see, it is not required for the definition and analysis of long-term convergence in an asexual population.

*Definition*

A pure strategy  $x^*$  is said to be long-term stable if there is a positive constant  $C > 0$  such that, starting from any distribution  $F_0$  of pure strategies within the interval  $(x^* - C, x^* + C)$ , and allowing for mutations of small enough effect, the long-term process of evolution over  $(x^* - C, x^* + C)$  is a well-defined stochastic process, and any realization  $\{F_n\}$  of it converges to a fixation of  $x^*$ .

**3. Long-term Dynamic Stability and CSS**

In this section we prove that the CSS conditions (1–3) are sufficient, and with (2) and (3) as weak inequalities, they are also necessary conditions for long-term convergence. In order to show it, we first characterize the short-term stable equilibria with a support at the vicinity of a given pure strategy  $x^*$ .

LEMMA 1

(i) If  $V_x^* \neq 0$ , then there exists a constant  $C > 0$  so that the only short-term stable equilibria with a support in  $(x^* - C, x^* + C)$  are fixations of pure strategies within this region.

(ii) If  $V_x^* = 0$  and  $V_{xx}^* \neq 0$ , then there exists a constant  $C > 0$  so that the only short-term stable equilibria with a support in  $(x^* - C, x^* + C)$  are either fixations of pure strategies or, maybe, pairs of strategies  $(x, y)$ .

*Proof*

Let  $F$  be a short-term stable equilibrium, then we know from (5) that for any  $x$  on the support of  $F$ ,  $V(x, F) = V(F, F)$ .

(i) Suppose  $V_x^* \neq 0$ . Without loss of generality assume  $V_x^* > 0$ . In this case there exists a constant  $C > 0$  so that for all  $x, y$  with  $x^* - C < x, y < x^* + C, V_x(x, y) > 0$ . Hence, for any distribution  $F$  with a support in  $(x^* - C, x^* + C)$  and for any value  $x$  in this region,  $V_x(x, F) = \int V_x(x, y)dF(y) > 0$ , in contradiction to (5), unless the support of  $F$  consists of one point.

(ii) Suppose now  $V_x^* = 0$  and  $V_{xx}^* \neq 0$ , say  $V_{xx}^* > 0$ , then it follows from the continuity of  $V_{xx}(x, y)$  that there exists a constant  $C > 0$  so that for all  $x$  and  $y$  in  $(x^* - C, x^* + C), V_{xx}(x, y) > 0$ . This means that for any distribution  $F$  with a support in  $(x^* - C, x^* + C)$  and for any value  $x$  in this region,  $V_{xx}(x, F) = \int V_{xx}(x, y)dF(y) > 0$ , in contradiction to (5), unless the support of  $F$  consists of two points, at the most, over which  $V(x, F)$  is the same.

*Notation*

We denote by  $(p \cdot x, q \cdot y)$  the distribution of strategies  $x$  and  $y$  where  $x$  is chosen with probability  $p$  and  $y$  is chosen with probability  $q = 1 - p$ .

If the distribution  $(p \cdot x, q \cdot y)$  is short-term stable for a given value  $0 < p < 1$ , we say, for brevity, that the pair  $(x, y)$  is stable.

We now define the function

$$H(p, x, y) = V(x, (p \cdot x, q \cdot y)) - V(y, (p \cdot x, q \cdot y)) \\ = p[V(x, x) - V(y, x)] + q[V(x, y) - V(y, y)]. \quad (8)$$

It is easy to see that

$$H(p, x, y) = -H(1 - p, y, x). \quad (9)$$

By expanding  $H(p, x, y)$  to a two-dimensional Taylor series of the second order around the values  $x = x^*$ ,  $y = x^*$ , and ignoring terms of the third order, we get

$$H(p, x, y) = p[(s - t)V_x^* + \frac{1}{2}(s^2 - t^2)V_{xx}^* \\ + s(s - t)V_{xy}^*] + q[(s - t)V_x^* + \frac{1}{2}(s^2 - t^2)V_{xx}^* \\ + t(s - t)V_{xy}^*] = (s - t)V_x^* + \frac{1}{2}(s^2 - t^2)V_{xx}^* \\ + (s - t)(ps + qt)V_{xy}^*, \quad (10)$$

where  $s = x - x^*$ ,  $t = y - x^*$ .

LEMMA 2

Fixation on the pure strategy  $y$  is stable against a mutation  $x$ ,  $x$  and  $y$  both being at the vicinity of  $x^*$ , if

$$H(0, x, y) < 0, \quad (11)$$

and only if (11) holds at least as a weak inequality.

*Proof*

From (8) it follows that  $H(0, x, y) = V(x, y) - V(x, x)$ , and the proof of the lemma follows immediately from (6) and (7).

LEMMA 3

A necessary and sufficient condition for the stability of the pair  $(x, y)$  is that

$$H(0, x, y) > 0 > H(1, x, y). \quad (12)$$

*Proof*

The pair  $(x, y)$  is stable, by definition, if and only if there exists a value  $0 < p < 1$  for which  $(p \cdot x, q \cdot y)$  is a short-term stable distribution. This means that

$$H(p, x, y) = 0 \quad (13)$$

and

$$H_p(p, x, y) < 0. \quad (14)$$

Condition (13) means that  $V(x, (p \cdot x, q \cdot y)) - V(y, (p \cdot x, q \cdot y)) = 0$  so that  $(p \cdot x, q \cdot y)$  is a short-term equilibrium. Condition (14) means that  $(p \cdot x, q \cdot y)$  is stable against deviations from the proportion  $p$  of the strategy  $x$ . But  $H$  is a linear function of  $p$ , hence the condition (13) and (14) is equivalent to (12).

LEMMA 4

If  $V_{xy}^* \geq 0$  then there exists a constant  $C > 0$  so that the only short-term stable distributions with a support in  $(x^* - C, x^* + C)$  are fixations of pure strategies.

*Proof*

From (10) it follows that  $H_p(p, x, y) = (x - y)^2 V_{xy}^*$ . Thus, if  $V_{xy}^* \geq 0$  then (14) cannot possibly be satisfied.

LEMMA 5

(i) If  $V_{xx}^* < 0$ , then there exists a constant  $C > 0$  so that for any stable pair of strategies  $(x, y)$  with  $x^* - C < x < y < x^* + C$ , a new mutation  $u$  at the region  $(x^* - C, x^* + C)$  will be successfully established in the population if and only if  $x < u < y$ .

(ii) If  $V_{xx}^* > 0$ , then there exists a constant  $C > 0$  so that for any stable pair of strategies  $(x, y)$  with  $x^* - C < x, y < x^* + C$ , a new mutation  $u$  at the region  $(x^* - C, x^* + C)$  will be successfully established in the population if and only if it is outside the interval  $(x, y)$ .

*Proof*

Suppose  $(x, y)$  is a stable pair of strategies, then for some  $0 < p < 1$ ,  $F = (p \cdot x, q \cdot y)$  is a short-term stable equilibrium and we know from (5) that

$$V(x, F) = V(y, F) = V(F, F). \quad (15)$$

(i) Suppose  $V_{xx}^* < 0$ . In this case we already know (see proof of Lemma 1) that there exists a constant  $C > 0$  so that for any distribution  $F$  with a support in  $(x^* - C, x^* + C)$  and for any value  $u$  in this interval,  $V(u, F)$  is a concave function. As a special case, if  $x^* - C < x, y < x^* + C$ , this is true for  $F = (p \cdot x, q \cdot y)$ . Thus, from (15) and the concavity of  $V(u, F)$  on the interval  $(x^* - C, x^* + C)$  it follows that for all  $u$  in this interval,

$$V(u, F) > V(F, F) \quad \text{if and only if} \quad x < u < y. \quad (16a)$$

(ii) If, on the other hand,  $V_{xx}^* < 0$ , then it follows in exactly the same way that

$$V(u, F) > V(F, F)$$

if and only if  $u$  is outside the interval  $(x, y)$ . (16b)

We now prove first that the CSS condition is sufficient for long-term convergence.

PROPOSITION 1

The CSS conditions (1–3) are sufficient for long-term convergence.

*Proof*

With the condition (1), (10) can be written as

$$H(p, x, y) = \frac{1}{2}(s^2 - t^2)V_{xx}^* + (t - s)(ps + qt)V_{xy}^* \quad (17)$$

and as a special case we get:

$$\begin{aligned} H(0, x, y) &= (s - t)\left[\frac{1}{2}sV_{xx}^* + t(V_{xy}^* + \frac{1}{2}V_{xx}^*)\right] \\ &= \frac{1}{2}(s - t)V_{xx}^*(s + At), \end{aligned} \quad (18)$$

where

$$A = 1 + 2V_{xy}^*/V_{xx}^*, \quad (19)$$

and, as we recall [see (10)],  $s = x - x^*$ ,  $t = y - x^*$ .

We now distinguish between two cases,  $V_{xy}^* \geq 0$  and  $V_{xy}^* < 0$ .

CASE I  $V_{xy}^* \geq 0$

In this case it follows from Lemma 4 that there exists a constant  $C > 0$  so that the only short-term stable equilibria with a support in  $(x^* - C, x^* + C)$  are fixations of pure strategies. Suppose  $y$  is such a fixation. Without loss of generality we can assume  $t = y - x^* > 0$ . From Lemma 2 and (18) it follows that  $y$  can be replaced by the mutant strategy  $x$  if and only if

$$H(0, x, y) = \frac{1}{2}(s - t)V_{xx}^*(s + At) > 0. \quad (20)$$

Having assumed, in this case,  $V_{xy}^* \geq 0$  and  $V_{xy}^* + V_{xx}^* < 0$ , it is easy to see that  $-1 < A < 1$ . Thus, with the condition (2), namely  $V_{xx}^* < 0$ , the condition (20) readily becomes

$$t > s > -At. \quad (21)$$

In this case, as  $-1 < A < 1$ , we see that the distance of any fixation from  $x^*$  will only decrease from one passage to another and we get a monotone long-term convergence to  $x^*$ .

CASE II  $V_{xy}^* < 0$

In this case we know from Lemma 1 that short-term stable equilibria at the vicinity of  $x^*$  can be either fixations or stable pairs of strategies. Thus, there might be four possible passages from one

short-term stable distribution at the  $C$ -vicinity of  $x^*$  to another:

- (i) a passage from fixation to fixation;
- (ii) a passage from fixation to a stable pair of strategies;
- (iii) a passage from a stable pair of strategies to a stable pair of strategies;
- (iv) a passage from a stable pair of strategies to fixation.

For any short-term stable equilibrium  $F$  with support at the  $C$ -vicinity of  $x^*$  we now assign a real value  $w(F)$  in the following way:

For fixation on a pure strategy  $x$  we define

$$w(x) = A(x - x^*)^2, \quad (22)$$

where  $A$  is given by (19). Note, though, that in this case where both  $V_{xx}^* < 0$  and  $V_{xy}^* < 0$  we get  $A > 1$ .

For a stable pair of strategies  $(x, y)$  we define

$$w(x, y) = (x - x^*)(x^* - y). \quad (23)$$

From Lemma 3 and (18) it follows that  $(x, y)$  is a stable pair of strategies if and only if

$$\frac{1}{2}(s - t)V_{xx}^*(s + At) > 0 > \frac{1}{2}(s - t)V_{xx}^*(t + As).$$

This means that  $s + At$  and  $t + As$  must be of different signs. But, as in the case we are dealing with,  $A > 0$  (we know, in fact, that  $A > 1$ ), this is possible only if  $t = y - x^*$  and  $s = x - x^*$  are of different signs. As a result, we know that the r.h.s. of (23) is always positive. This is indeed true for the r.h.s. of (22), provided  $x \neq x^*$ . Hence we see that  $w(F)$  is non-negative for any short-term stable distribution  $F$  of strategies in the  $C$ -vicinity of  $x^*$  and it is zero only for a fixation on  $x^*$ . We shall see now that it is monotone decreasing for any passage from one short-term stable equilibrium to another.

(i) A passage from fixation on a pure strategy  $y$  to fixation on a pure strategy  $x$  requires not only that the fixation on  $y$  will be unstable against the mutation  $x$ , but also that the fixation on  $x$  will be stable, in turn, against a mutation  $y$ . Otherwise it follows from Lemma 3 that  $x$  and  $y$  will be established, instead, as a stable pair. Employing again (18) and Lemma 2, this second requirement means that

$$H(1, x, y) = -H(0, x, y) = \frac{1}{2}(s - t)V_{xx}^*(t + As) > 0,$$

and with the assumption  $V_{xx}^* < 0$ , this implies

$$(s - t)(t + As) < 0. \quad (24)$$

Without loss of generality, assume again  $t = y - x^* > 0$ . One can readily verify that (24) is satisfied if and only if

$$-t/A < s < t,$$

namely,  $x$  must lie either between  $y$  and  $x^*$ , or to the other side of  $x^*$ , at a distance not further than  $1/A$  of the distance between  $y$  and  $x^*$ . As in the case  $A > 1$ , we get:

$$w(x) = A(x - x^*)^2 < A(y - x^*)^2 = w(y). \quad (25)$$

(ii) In order that a passage from fixation on a pure strategy  $y$  to a stable pair of strategies  $(x, y)$  will occur, it is first necessary that the fixation on  $y$  will be unstable against the mutation  $x$ . From Lemma 3 and (18) it follows that this is the case if and only if

$$\frac{1}{2}(s - t)V_{xx}^*(s + At) = H(0, x, y) > 0.$$

With the assumption  $V_{xx}^*/2 < 0$ , this implies

$$(s - t)(s + At) < 0,$$

which is satisfied if either  $t > 0$  and  $-At < s < t$ , or  $t < 0$  and  $t < s < -At$ . This means that

$$-st < At^2, \quad (26)$$

hence

$$\begin{aligned} w(x, y) &= (x - x^*)(x^* - y) < A(x^* - y)^2 \\ &= w(y). \end{aligned} \quad (27)$$

(iii) From Lemma 5 it follows that, with the assumption  $V_{xx}^* < 0$ , a passage from a stable pair of strategies  $(x, y)$  to another stable pair of strategies  $(u, y)$  is possible only when  $u$  lies within the open interval  $(x, y)$ . Otherwise the mutation  $u$  will not be established in the population to start with. In this case we know that in order to be maintained in a stable pair with  $y$ ,  $u$  and  $y$  must be of different sides of  $x^*$ . This means that  $u$  must lie between  $x$  and  $x^*$  and we immediately get:

$$\begin{aligned} w(u, y) &= (u - x^*)(x^* - y) < (x - x^*)(x^* - y) \\ &= w(x, y). \end{aligned} \quad (28)$$

(iv) Again it follows from Lemma 5 that a passage from a stable pair of strategies  $(x, y)$  to fixation on a mutant strategy  $u$  is possible only if  $u$  lies within the open interval  $(x, y)$ . In this case, however, it follows from Lemma 3 that the fixation of  $u$  must be stable against both an  $x$  and a  $y$ -mutation. Assume, without loss of generality, that  $u$  lies between  $x^*$  and  $x$ , that is, on the other side of  $y$ . We already know that in this case  $u$  will be stable against a  $y$ -mutation if and only if

$$|u - x^*| < 1/A|y - x^*|.$$

This means

$$\begin{aligned} w(u) &= A(u - x^*)^2 < (u - x^*)(x^* - y) \\ &< (x - x^*)(x^* - y) = w(x, y). \end{aligned} \quad (29)$$

From (25), (27), (28) and (29) it thus follows that for any possible passage from a short-term stable distribution  $F$  with a support in the vicinity of  $x^*$  to another short-term stable distribution  $G$ , we get  $w(G) < w(F)$  and the process converges to the fixation of  $x^*$ , which is the only short-term stable equilibrium over which  $w$  is 0.

We have, therefore, proved that the CSS conditions (1–3) are sufficient for long-term convergence.

We now prove that each of the CSS conditions (1–3), the latter two holding as weak inequalities, is also necessary for long-term convergence.

PROPOSITION 2

The condition (1), namely  $V_x^* = 0$ , is necessary for the long-term stability of the pure strategy  $x^*$ .

*Proof*

Assume, without loss of generality,  $V_x^* > 0$ . In this case we know that there exists a constant  $C > 0$  so that for any distribution  $F$  with a support in  $(x^* - C, x^* + C)$ ,  $V(x, F)$  is an increasing function of  $x$  over the entire interval  $(x^* - C, x^* + C)$ . As a special case, if  $x^* + C > y > x > x^* - C$ , then for all  $p$ ,

$$H(p, x, y) = V(x, (p \cdot x, q \cdot y)) - V(y, (p \cdot x, q \cdot y)) = 0.$$

Hence, the possible passages will always be from fixation on one value to fixation on a larger value until the long-term process will, inevitably, pass the value  $x^* + C$  and never return back to  $x^*$ .

PROPOSITION 3

The weaker version of condition (3), namely  $V_{xx}^* + V_{yy}^* \leq 0$ , is necessary for the long-term stability of the pure strategy  $x^*$ .

*Proof*

Let us assume  $V_{xx}^* + V_{yy}^* > 0$ . Without loss of generality assume further that  $V_x^* = 0$ , otherwise we already know from Proposition 2 that  $x^*$  is not long-term stable. Recall that fixation on the pure strategy  $y$  is stable against a mutation  $x$  if and only if

$$H(0, x, y) = \frac{1}{2}(s - t)V_{xx}^*(s + At) > 0,$$

where  $A = 1 + 2V_{xy}^*/V_{xx}^*$ . We distinguish between three cases:

Case I:  $V_{xx}^* < 0$  (thus,  $x^*$  is an ESS) and  $V_{xy}^* > -V_{xx}^* > 0$  (in which case  $A < -1$ ).

Case II:  $V_{xx}^* > 0$  and  $V_{xy}^* > 0$  (in which case  $A > 1$ ).

Case III:  $V_{xx}^* > 0$  and  $-V_{xx}^* < V_{xy}^* < 0$  (in which case  $-1 < A < 1$ ).

In Cases I and II we know from Lemma 4 that the only short-term stable equilibria at the  $C$ -vicinity of  $x^*$  are the fixations of pure strategies. Without loss of generality we assume, in all cases, that  $y > x^*$  or, equivalently,  $t > 0$ .

#### CASE I

In this case we already know from Lemma 2 and (18) that, as  $V_{xx}^* < 0$ , a mutant  $x$  is successfully established in a  $y$ -population (in which case it, inevitably, replaces it as a fixation) if and only if

$$(s - t)(s + At) < 0.$$

But as in this case  $A < -1$ , this condition is satisfied if and only if  $t < s < -At$ , which means that  $x = x^* + s$  is further apart from  $x^*$  than  $y = x^* + t$ , and  $x^*$  must be unstable in the long run.

#### CASE II

In this case, as  $V_{xx}^* > 0$ , it follows from Lemma 2 and (18) that a mutant  $x$  is successfully established in a  $y$ -population and inevitably replaces it as a fixation, if and only if

$$(s - t)(s + At) > 0.$$

This is satisfied for either  $s > t$  or  $s < -At$ . But as, in this case,  $A > 1$ , we always get  $|x - x^*| = |s| > |t| = |y - x^*|$ , and again  $x^*$  must be unstable in the long run.

#### CASE III

In this case it follows from Lemma 3 and (20) (given  $V_{xx}^* > 0$ ) that  $(x, y)$  is a stable pair of strategies if and only if

$$(s - t)(s + At) > 0 > (s - t)(s - At).$$

For  $t > 0$  this is possible if and only if

$$s < 0 \quad \text{and} \quad -(1/A)t < s < -At, \quad (30)$$

which is impossible for  $A < 0$ . But in the case we are dealing with, we already know that  $-1 < A < 1$ , hence (30) is possible only when  $0 < A < 1$ . It is impossible if  $-1 < A < 0$ . We, therefore, distinguish between these two sub-cases:

#### Case IIIa

$0 < V_{xx}^* < -2V_{xy}^*$  and, thus,  $0 < A < 1$ . In this case, short-term stable equilibria at the vicinity of  $x^*$  can be either fixations or stable pairs of strategies. Thus, as we have seen in the proof of Proposition 2, there might be four possible passages from one short-term stable equilibrium at the vicinity of  $x^*$  to another:

- (i) a passage from fixation to fixation;
- (ii) a passage from fixation to a stable pair of strategies;
- (iii) a passage from a stable pair of strategies to a stable pair of strategies;
- (iv) a passage from a stable pair of strategies to fixation.

For any short-term stable distribution  $F$  of strategies in the  $C$ -vicinity of  $x^*$  we again use the real function  $w(F)$  as defined above in (22) and (23), namely

$w(x) = A(x - x^*)^2$  for fixation on a pure strategy  $x$ ,

$$w(x, y) = (x - x^*)(x^* - y)$$

for a stable pair of strategies  $(x, y)$ .

Note, though, that in this case, unlike in the case dealt with in Proposition 1,  $0 < A < 1$ . We now show, moreover, that for any possible passage from one short-term stable equilibrium  $F$  to another, say  $G$ , the value of  $w$  increases:

- (i) a passage from fixation on  $y$  to fixation on  $x$  is possible only if  $s < -(1/A)t$ , otherwise it follows from Lemma 2 that  $x$  cannot be fixed in the presence of  $y$ . Hence  $|s| > |t|$  and, therefore,  $w(x) > w(y)$ ;
- (ii) a passage from fixation on  $y$  to a stable pair of strategies  $(x, y)$  is possible if and only if  $-(1/A)t < s < -At$  and again, it follows from (22), (23) and the condition  $0 < A < 1$  that  $w(x, y) > w(y)$ ;
- (iii) from Lemma 5 it follows that a passage from a stable pair of strategies  $(x, y)$  to another stable pair of strategies with a new mutant strategy  $u$  is possible only if  $u$  is outside the interval  $(x, y)$ , in which case it follows from the definition of the function  $w$  that its value must thereby increase;
- (iv) a passage from a stable pair of strategies  $(x, y)$  to fixation on a new strategy  $u$  is possible if and only if  $u$  is outside the interval  $(x, y)$ , say, without loss of generality,  $u < x$  and, in addition,

$$(u - x^*) < -(1/A)(y - x^*).$$

In this case, again,

$$w(u) = A(u - x^*)^2 > (x - x^*)(y - x^*) = w(x, y).$$

We therefore conclude that also in Case IIIa,  $x^*$  must be unstable in the long run.

Finally we have:

*Case IIIb*

$V_{xx}^* > -2V_{xy}^* > 0$  and, thus,  $-1 < A < 0$ . In this case we know that, as in Cases I and II, the only possible passages are from a fixation on  $y$  to another fixation, say on  $x$ . We know, further, that, as  $V_{xx}^* > 0$ , the condition for this is, again,  $(s - t)(s + At) > 0$ . It is easy to see that in this case, when  $-1 < A < 0$ , this is possible if and only if

$$\text{either } s > t \text{ or } s < -At. \quad (31)$$

In this case, unlike all the previous ones, we, thus see that also mutations  $x$  which are closer to  $x^*$  than  $y$  (in addition to all mutations which are further apart from  $x^*$ ) can be established and then become fixed in the population. In fact, this always includes even the value  $x^*$  itself (which is, however, unstable to any new mutation). A necessary condition for this is, however, that the effect of the new mutation will be larger than  $(1 + A)t$ . (Note that in this case,  $1 > 1 + A > 0$ .) Hence, when  $t > Kh$ , where  $K = 1/(1 + A)$ , the only possibly successful mutations are those which render their carrier further apart from  $x^*$ . Furthermore, for any fixation  $y$  at the  $Kh$ -vicinity of  $x^*$  there is a positive probability to be replaced by a mutation further apart from  $x^*$  and this probability is bounded from below by a positive value, because it is a continuous function of  $y$ , and it is 1 for  $y = x^* + Kh$  and for  $y = x^* - Kh$  (in fact, it is also 1 for  $y = x^*$ ). The system can, therefore, stay only for a finite time at the interval  $(x^* - Kh, x^* + Kh)$ , and afterwards it must escape  $x^*$  in a monotone way. We therefore conclude that also in this case  $x^*$  must be strictly unstable in the long run, and this completes the proof of Proposition 3.

PROPOSITION 4

The weaker version of local ESS condition (2), namely

$$V_{xx}^* \leq 0, \quad (2')$$

is necessary for the long-term stability of the pure strategy  $x^*$ .

*Proof*

Let us assume  $V_{xx}^* > 0$ . Without loss of generality assume further that  $V_x^* = 0$  and  $V_{xx}^* + V_{xy}^* \leq 0$ , otherwise we already know from Propositions 2 and 3 that  $x^*$  is unstable. In this case it is easy to verify that  $A = 1 + 2V_{xy}^*/V_{xx}^* \leq -1$ .

From (9), (12) and (20) we infer that  $(x, y)$  is a stable pair of strategies if and only if

$$\begin{aligned} \frac{1}{2}(s - t)V_{xx}^*(s + At) &= H(0, x, y) > 0 > H(1, x, y) \\ &= -H(0, y, x) = \frac{1}{2}(s - t)V_{xx}^*(t + As). \end{aligned}$$

Assuming, without loss of generality,  $t > s$  (that is,  $y > x$ ) and employing  $V_{xx}^* > 0$ , this becomes

$$s + At < 0 < t + As. \quad (32)$$

Since  $A$ , in this case, is negative, (32) indeed holds for all  $s < 0 < t$ , which means that all pairs  $(x, y)$  with values on both sides of  $x^*$  are stable pairs.

But, as  $V_{xx}^* > 0$ , we know from Lemma 5 that there exists a constant  $C > 0$  so that for any stable pair of strategies  $(x, y)$  with  $x^* - C < x < 0 < y < x^* + C$ , a new mutation  $u$  at the region  $(x^* - C, x^* + C)$  will be successfully established in the population if and only if  $u$  is outside the interval  $(x, y)$ . Thus,  $u$  must be further apart from  $x^*$  than the strategy of the pair which is on its side. In this case we further know that  $u$  must replace this strategy, and establish a stable pair with the other. Hence, the process can pass only to a new stable pair of strategies, say  $(x', y')$ , with  $(y' - x^*)(x^* - x') > (y - x^*)(x^* - x) > 0$ , and the long-term process must, therefore, diverge out of the interval  $(x^* - C, x^* + C)$ .

REMARK

As we learn from Proposition 4, contrary to previous conjectures, the conditions (1) and (3) alone (say,  $m$ -stability or, in another terminology, convergence stability) do not guarantee long-term convergence to  $x^*$ . If the weak local ESS condition (2') is not satisfied, then  $x^*$  cannot possibly be long-term stable and, moreover, the population cannot even remain in the vicinity of  $x^*$ . Indeed, based on local analysis of the pay-off function  $V(x, y)$  at the vicinity of  $x^*$ , one cannot possibly predict the behavior of the long-term process, once outside this vicinity. In fact, one can easily produce examples in which the variance of the process then increases to infinity. In other examples the process may converge to fixation on another CSS.

It is true, though, that even without the local ESS condition (2) or even (2'), the condition (3) with (1) alone guarantees that, starting from fixation at the vicinity of  $x^*$ , any small mutation in the direction of  $x^*$  will be successfully established in the population. Employing (10) and (11), it can further be shown that starting from a fixation on a value  $y = x^* + t > x^*$ , a mutant  $x = x^* + s$  will successfully enter the population if and only if either  $s < t$  (i.e. if  $x$  is in the direction of  $x^*$ ) or  $s > -At$ , where, as we recall [given (3) as a strict inequality]  $-A > 1$ . This means

that, in addition to mutations in the direction of  $x^*$ , mutations in the other direction can be also successful, if of sufficiently large effect, say more than  $-(1+A)y$ . Thus, with a small value  $h > 0$  as a maximal effect of a single mutation, the second condition  $s > -At$  can never be satisfied unless  $y$  is in the  $Kh$ -vicinity of  $x^*$ , where  $K = -1/(A+1) > 0$ .

Thus, starting from a fixation on a value  $y$  which is not too close to  $x^*$ , say  $Kh < y < x^* + C$ , the only possible passages will be to a fixation on a value  $x$ , closer to  $x^*$ . The process of long-term evolution must, then, converge to the  $Kh$ -vicinity of  $x^*$ . This stands, at least partly, in agreement with a conjecture of Christiansen (1991), that  $m$ -stability (or, in his terminology, convergence stability) alone should guarantee convergence to some small vicinity of  $x^*$ , with small variance being maintained by the repelling, non-ESS property of  $x^*$ . However, from any fixation at the  $Kh$ -vicinity of  $x^*$ , the process can reach, with a finite number of possible passages, a stage in which, by overshooting, a stable pair of strategies is established. Moreover, as the process cannot possibly get out of the  $(K+1)h$ -vicinity of  $x^*$  without passing through a stage of a stable pair of strategies, the event of establishing such a pair must, sooner or later, occur in probability 1 and then, as we have seen, the process must diverge, ending up outside the interval  $(x^* - C, x^* + C)$ .

#### 4. Summary and Discussion

Necessary and sufficient conditions for long-term evolutionary stability in a pairwise, random encounter population-game with a continuity of pure strategies have been analysed. The process of long-term evolution at the vicinity of fixation on any pure strategy  $x^*$  (except for the singular case in which  $V_x^* = V_{xx}^* = 0$ ) has been defined as follows: First we have characterized all distributions with support at a certain vicinity of  $x^*$ , which, with the absence of mutations, are stable under natural selection [assuming that natural selection is, indeed, determined by the population game  $V(x, y)$ ]. We refer to such distributions as short-term stable. For any short-term stable distribution  $F$  we have, then, characterized the range of small-effect mutations which, under the selection forces determined by the distribution  $F$  of strategies, are selected for, and thus can be initially established in the population. We called such mutations  $F$ -intruders. Any such a mutation, once established in the population, starts a new process of natural selection. This has been shown to converge to a new short-term stable distribution, say  $G$ . In this case we say that the passage from  $F$  to  $G$  is *possible*.

Finally, we call long-term evolution, the process in which a population, subject to mutation and selection pressures, passes from one short-term stable distribution to another. Thus, any realization of such a process is an infinite sequence  $\{F_n\}$  of short-term stable equilibria such that for all  $n = 0, 1, 2, \dots$ , the passage from  $F_n$  to  $F_{n+1}$  is possible.

Unlike the well studied short-term process of changes in genotype frequencies due to selection over a fixed set of genotypes (i.e., short-term evolution), we see that long-term evolution is a stochastic process, its states being short-term stable distributions. Unfortunately, the transition probabilities from one state to another depend on the probability distribution of the mutation in nature, a distribution about which we do not have much information. However, except for rudimentary information about the range of the mutation, the analysis suggested in this work assumes only knowledge about which transitions are possible and which ones are not, knowledge that can be drawn directly from the structure of the game.

Under the only assumption of a small upper bound to the maximal effect of a single mutation, it has been shown that the CSS conditions (1–3) are sufficient and almost necessary for long-term convergence, the necessary condition being the same with (2) and (3) as weak inequalities.

It has been argued (e.g., Christiansen 1991) that the second CSS condition ( $m$ -stability or convergence stability) is sufficient to guarantee convergence, at least to some small vicinity of the  $m$ -stable state. The argument goes as follows: (i) If  $x^*$  is  $m$ -stable, then starting from a population state close to it, natural selection will favor mutations in the direction of  $x^*$ , hence the population will come ever closer to  $x^*$ . (ii) If  $x^*$  is not an ESS, the population cannot become fixed on  $x^*$ , since natural selection will initially favor intruding mutations. (iii) The population will, however, remain in the vicinity of the  $m$ -stable state because, once shifting from  $x^*$ , natural selection again favors mutations in its direction.

As we have seen, however, this argument is incorrect, since  $m$ -stability alone *does not* guarantee the advantage of mutations in the direction of  $x^*$ . This is true only for one-sided deviations from  $x^*$  (provided they are small enough). The situation is different in cases of two-sided deviations. In such cases, we have shown that if  $x^*$  is not an ESS, then not only natural selection will favor mutations which shift the phenotype further away from the  $m$ -stable phenotype, it will operate to eliminate all phenotypes in an ever increasing vicinity of  $x^*$ .

Quite surprisingly, although Continuous Stability, as well as Evolutionary Stability, is a genetic-free

deterministic concept, it corresponds to long-term evolution, a process that is stochastic, and, moreover, whose transition probabilities depend on the genetic structure (Eshel & Feldman, 1984; Eshel, 1991) or at least on the distribution of the possible effects of a single mutation. As has been shown elsewhere (Eshel & Feldman, 1984; Liberman, 1988), long-term convergence to ESS in a population game with a finite number of pure strategies, even in a diploid, multilocus random mating population, is, within a wide range of assumptions, independent of the genetic structure. This is in contrast to short-term convergence under these conditions. In this work we have seen that at least in an asexual population-game dynamics with a one-dimensional continuity of pure strategies, long-term convergence to a CSS is deterministic, as well as independent on the distribution of the effect of the mutation, provided it is bounded by a small enough value. This result, however, cannot be generalized to the case of a two-dimensional continuum of pure strategies. In that case it has been shown by Matessi & Di Pasquale (1996) that long-term stability does depend on the distribution of the mutations, or at least on the ratio between the rates of mutation in the two dimensions. Thus, a generalization of the concept of CSS to a two-dimensional continuum of strategies, as suggested by Matessi & Di Pasquale, is not determined by the game structure alone.

After this paper had been accepted for publication, we became aware of new work by Friedman & Yellin (unpublished work), that had a similar but more positive result, analysing a general short-term process of evolution on the distribution of pure strategies in a continuous, state population game.

It seems to us that as far as the important process of short-term selection is concerned, all inaccuracies and ambiguities mentioned in the introduction to our paper, have been fully resolved by Friedman & Yellin.

An anonymous referee drew our attention to the work by Metz *et al.* (1996), which deals with a somewhat related problem, using a different approach..

#### REFERENCES

- CHRISTIANSEN, F. B. (1991). On conditions for evolutionary stability for a continuously varying character. *Amer. Natur.* **138**, 37–50.
- ESHEL, I. (1983). Evolutionary and continuous stability. *J. theor. Biol.* **103**, 99–111.
- ESHEL, I. (1991). Game theory and population dynamics in complex genetical systems: the role of sex in short term and in long term evolution. In: *Game Equilibrium Models. Vol. I: Evolution and Game Dynamics* (Selten, R. ed.), pp. 6–28. Berlin: Springer Verlag.
- ESHEL, I. (1996). On the changing concept of population stability as a reflection of a changing problematics in the quantitative theory of evolution. *J. Math. Biol.* **34**, 485–510.
- ESHEL, I. & FELDMAN, M. W. (1984). Initial increase of new mutants and some continuity properties of ESS in two locus systems. *Amer. Natur.* **124**, 631–640.
- ESHEL, I. & MOTRO, U. (1981). Kin selection and strong evolutionary stability of mutual help. *Theor. Popul. Biol.* **19**, 420–433.
- LIBERMAN, U. (1988). External stability and ESS: criteria for initial increase of a mutant allele. *J. Math. Biol.* **26**, 477–485.
- MATESSI, C. & DI PASQUALE, C. (1996). Long-term evolution of multilocus traits. *J. Math. Biol.* **34**, 613–654.
- MAYNARD SMITH J. & PRICE, G. R. (1973). The logic of animal conflict. *Nature* **246**, 15–18.
- METZ, J. A. J., GERITZ, S. A. H., MESZÉNA, G., JACOBS, F. J. A. & VAN HEERWAARDEN, J. S. (1996) Adaptive dynamics: a geometrical study of the consequences of nearly faithful reproduction. In: *Dynamical Systems and their Applications* (van Strien, S. J. & Verduyn Lunel, S. M., eds) pp. 183–231. Amsterdam: North Holland.
- TAYLOR, P. (1989). Evolutionary stability in one-parameter models under weak selection. *Theor. Popul. Biol.* **36**, 125–143.