On the Evolution of Sex Determination and the Sex Ratio in Haplodiploid Populations*

ILAN ESHEL[†] AND MARCUS W. FELDMAN

*Department of Biological Sciences, Stanford University,
Stanford, California 94305*

DEDICATED TO PROFESSOR LUCA CAVALLI-SFORZA
ON HIS SIXTIETH BIRTHDAY

INTRODUCTION

Sex determination by ploidy, either of the entire genome (in haplodiploid populations) or of special chromosomes (XY or XO systems) is a common occurrence. It is commonly held that observed sex ratios are an evolutionary consequence of the sex-determination system; for example, that the XY system helps produce and/or maintain sex ratios to 1:1 (Edwards, 1961; Williams, 1975). It should be noted, however, that factors such as segregation distortion in the sex chromosomes (see e.g. Curtsinger and Feldman, 1980, and references therein), sex-dependent fitness components (see e.g. Thomson and Feldman 1975, Charlesworth, 1977) as well as hermaphroditism, parthogenesis, and sex conversion (see e.g. Williams, 1975, Chapter 10) are all expected to affect the sex ratio in those XY systems for which they are relevant. On the other hand, sex ratios in species with other sex-determining systems may also be close to 1:1 (Scudo, 1964; Spieth, 1974; Charlesworth, 1977). Thus the cytogenetic pattern can be regarded as neither a necessary nor a sufficient factor for the explanation of observed sex ratios. Nevertheless, the connection between sex determination and the sex ratio remains compelling to evolutionary theorists.

The evolution of a sex ratio close to 1:1 (or rather the evolution of sex determination which should determine such a ratio) was first explained by Fisher (1930) as a consequence of maximization of the number of grand-offspring through optimal allocation of parental investment in male and female offspring (see also Bodmer and Edwards, 1960). A crucial evolutionary assumption in Fisher's argument is that each sex must supply

half the genetical material of all subsequent generations. This is valid for autosomal genes in diploid populations.

Nur (1974) and Uyenoyama and Bengtsson (1979) have studied models in which the sex ratio among offspring is determined by their *parents'* genotypes at an autosomal locus with two alleles. Eshel (1975) considered the case where the probability that an individual is male is governed by its *own* genotype at an autosomal locus with two alleles. Eshel and Feldman (1982) generalized these treatments to an arbitrary number of alleles. Using the criterion of invasibility by a new sex-ratio-controlling allele they suggested that Fisher's proposition is valid in the long-term evolutionary sense even though for a given fixed set of alleles, the stable sex ratio in all of these exact genetical treatments may not be 1:1.

The situation is different when we consider sex-linked genes (or a haplodiploid species). Theoretical difficulties in the application of Fisher's argument to sex-linked genes have been pointed out by Shaw (1958), Hamilton (1967), and Bengtsson (1977). Although Fisher's argument can be extended to cover the *XY* situation (see e.g. Shaw and Mohler, 1953; Hartl and Brown, 1970, and below) the predictions depend on whether the considered genes affect the individual's sex, its offspring's sex, or it sibs' sex. For example, Hamilton (1967) suggests that any mutation which increases the probability of males in the offspring will, if it occurs on the *Y* chromosome, become established in the population. Selection of *X*-linked genes which determine the sex of offspring produces a stable sex ratio of 1:1 (Hartl and Brown, 1970; Charlesworth, 1977; Uyenoyama and Bengtsson, 1979, 1981). For *X*-linked genes in systems where the sister determines the sex ratio, Trivers and Hare (1976), Oster *et al.* (1977), Charnov (1978), MacNair (1978), and Benford (1978) predict a 1:3 sex ratio, although Uyenoyama and Bengtsson (1981) point out that this prediction is subject to some analytical complications. For all of these predictions it is assumed that haploid (or *XY*) individuals are necessarily male and diploids (or *XX*) are necessarily female. (The opposite would be assumed in considering birds of *Lepidoptera*.) Examples of departure from this assumption of complete determination by ploidy, although rare (see e.g. Kalela and Oksala, 1966) raise questions concerning the evolution of genes which affect the sex of their haploid (*XY*) or diploid (*XX*) carriers in a haplodiploid (sex-linked) population. Of particular interest is the evolution of the sex ratio in of a system in which sex is determined by ploidy.

In this study we analyze a one-locus *n*-allele system of individual sex determination in a haplodiploid population. We suggest that there is a sex ratio which is an evolutionary "optimum" in the sense that new sex-ratio alleles enter the population only if their long-term evolutionary consequence is to bring the population sex ratio closer to this optimum. If the sex ratio is at this optimum value then as in Eshel and Feldman (1982), we expect any

sex-ratio alleles that enter the population to rapidly reestablish the optimum sex ratio. This value is different, in general, from 1:1 and depends on the fraction of haploids in the population. With haploid males the system of sex determination by ploidy is shown to be stable to perturbation in the diploids only if additional features are imposed on the model.

Model

The diploid genotypes in the population are $A_i A_j$ and the haploids A_i $(i, j = 1, 2,..., n)$. The frequencies of the genotypes $A_i A_i$, $A_i A_j$, and A_i are, respectively, X_{ii}, $2X_{ij}$, and Y_i. The fraction $\alpha = \sum_{i=1}^{n} Y_i$ of haploids in the population is assumed to be fixed by a mechanism extrinsic to the A locus (e.g., by material or sister behavior). Finally, the probabilities that genotypes $A_i A_j$ and A_i are male are m_{ij} and l_i, respectively. Thus, in organisms such as mammals or *Drosophila* we would have $m_{ij} = 0$, $l_i = 1$ for all i, j, in which case the male to female sex ratio would be $\alpha : 1 - \alpha$. One objective of the present study is to determine the stability of this sex ratio within a more general framework of sex determination in a haplodiploid population. With the above definitions the frequency of males in the population is

$$M = \sum_i \sum_j{}' m_{ij} X_{ij} + \sum_i l_i Y_i. \tag{1}$$

It is assumed that diploid offspring are produced by matings between males and females of each ploidy, for example, between male diploids and female haploids and between male and female haploids. Haploids, on the other hand, can be produced only from unfertilized females of either ploidy. From these assumptions we have the recursions for diploids and haploids in the next generation in terms of those in the present:

$$X'_{ij} = \frac{(1-\alpha)}{2M(1-M)} \left\{ \left[\sum_{k=1}^{n} m_{ik} X_{ik} + l_i Y_i \right] \left[\sum_{r=1}^{n} (1-m_{jr}) X_{jr} + (1-l_j) y_j \right] \right.$$
$$\left. + \left[\sum_{r=1}^{n} m_{jr} X_{jr} + l_j Y_j \right] \left[\sum_{k=1}^{n} (1-m_{ik}) X_{ik} + (1-l_i) Y_i \right] \right\} \tag{2}$$

$$Y'_i = \frac{\alpha}{1-M} \left\{ \sum_{k=1}^{n} (1-m_{ik}) X_{ik} + (1-l_i) Y_i \right\}. \tag{3}$$

These recursions specify the evolution of the genotype frequencies for $i, j = 1, 2,..., n$.

Remark. In the above model it is assumed that a haploid individual always receives all of its genes from a parent of one sex which we call female. This would be the case for true haplodiploid populations as well as for X-linked genes in an XY system. In XO systems, however, a haploid

(XO) offspring receives its genes from its male parent. The above model becomes applicable to such systems by changing the labels of males and females.

Analysis of the Model

Let

$$p_i = \sum_k X_{ik} + Y_i \tag{4}$$

and

$$r_i = \left(\sum_k m_{ik} X_{ik} + l_i Y_i \right) \Big/ M \tag{5}$$

be the frequencies of A_i in the population and among males, respectively. Then using (2) and (3) in (4) and (5) we have for $i = 1, 2, ..., n$,

$$p_i' = \sum_{k=1}^{n} X_{ik}' + Y_i'$$

$$= \{\alpha(p_i - Mr_i) + (1-\alpha)(p_i + r_i - 2Mr_i)/2\}/(1-M) \tag{6}$$

$$r_i' = \frac{1}{M'} \left\{ \frac{\alpha l_i}{1-M} (p_i - Mr_i) \right.$$

$$\left. + \frac{1-\alpha}{2(1-M)} \sum_k m_{ik}(p_i r_k + p_k r_i - 2Mr_i r_k) \right\}, \tag{7}$$

where $M' = \sum_i \sum_j m_{ij} X_{ij}' + \sum_i l_i Y_i'$.

At equilibrium, $p_i' = p_i$, $r_i' = r_i$ and (6) is easily reduced to

$$(1 - \alpha - 2M) p_i = (1 - \alpha - 2M) r_i. \tag{8}$$

Hence, at equilibrium, either

$$M = (1 - \alpha)/2 = M^*, \tag{9}$$

say, or

$$\hat{p}_i = \hat{r}_i. \tag{10}$$

In the latter case, (7) reduces at equilibrium to

$$\left[M - \alpha l_i - (1-\alpha) \sum_j m_{ij} p_j \right] p_i = 0 \tag{11}$$

$(i = 1, 2, ..., n)$. Solutions of (11) include fixation equilibria with $X_i = Y_i = 1$ for some i and $X_j = Y_j = 0$ for $j \neq i$. The solutions of (11) will be termed

symmetric equilibria (see also Eshel, 1975; Uyenoyama and Bengtsson, 1979, 1981).

Stability of Symmetric Equilibria

Assume that a population with alleles $A_1, A_2, ..., A_n$ has reached a stable symmetric equilibrium, (11). A new allele A_{n+1} appears in low frequency and we ask when will A_{n+1} invade the population. This will occur if the original equilibrium is unstable in the higher-dimensional space of $A_1, A_2, ..., A_{n+1}$. We seek the conditions determining *local* stability of this original equilibrium. To this end assume that p_{n+1} and r_{n+1} are sufficiently small that quadratic terms of (6) and (7) are negligible. Then the local linear transformation in p_i and r_i can be written

$$\begin{pmatrix} p'_{n+1} \\ r'_{n+1} \end{pmatrix} = \mathbf{S} \begin{pmatrix} p_{n+1} \\ r_{n+1} \end{pmatrix}, \tag{12}$$

where the local stability matrix $\mathbf{S}$, determined from (6) and (7), is

$$S = \begin{bmatrix} (1+\alpha)/2(1-M) \\ [2\alpha l_{n+1} + (1-\alpha)\bar{M}_{n+1}]/2M(1-M) \end{bmatrix}$$

$$\begin{matrix} (1-\alpha-2M)/2(1-M) \\ [(1-\alpha)\bar{M}_{n+1}(1-2M) - 2\alpha l_{n+1}M]/2M(1-M) \end{matrix} \bigg] \tag{13}$$

and $\bar{M}_{n-1} = \sum_{j=1}^{n} m_{n+1j} \hat{p}_j = \sum_j m_{n+1j} \hat{r}_j$. The stability of (12) is determined by the eigenvalue of $\mathbf{S}$. It is a matter of algebra to show that a necessary and sufficient condition for the largest eigenvalue of $\mathbf{S}$ to be greater than unity is

$$(1-\alpha-2M)[M - \alpha l_{n-1} - (1-\alpha)\bar{M}_{n+1}]/M(1-M) < 0. \tag{14}$$

Thus (14) is sufficient in the usual sense for the instability of this equilibrium. It is informative to examine (14) closer; (14) is satisfied if either

$$M < (1-\alpha)/2 = M^* \tag{15a}$$

and

$$\alpha l_{n+1} + (1-\alpha)\bar{M}_{n+1} > M \tag{15b}$$

or

$$M > (1-\alpha)/2 = M^* \tag{16a}$$

and

$$\alpha l_{n+1} + (1-\alpha)\bar{M}_{n+1} < M. \tag{16b}$$

Now $\alpha l_{n+1} + (1-\alpha)\bar{M}_{n+1}$ is the frequency of males in the new mutant carrying genotypes. Thus we may restate the above as:

RESULT 1. (i) *Any equilibrium at which the proportion of males M is less than $(1 - \alpha)/2$ is unstable to new mutations whose marginal frequency of males in the population is larger than M and is stable to mutations whose marginal frequency of males is less than M, where the marginal frequency involves the resident alleles.*

(ii) *The opposite results hold if M is greater than $(1 - \alpha)/2$.*

These results follow from (15a), (15b), and (16a), (16b) and the fact that if $M \neq M^*$ then the equilibrium for $A_1, A_2,..., A_n$ must be symmetric in the sense that $\hat{p}_i = \hat{r}_i$ $(i = 1, 2,..., n)$. It appears that in the long run the success of A_{n+1} satisfying (15) or (16) results in an $(n + 1)$-allele equilibrium with the sex ratio closer to $(1 + \alpha) : (1 + \alpha)$ than to $M : 1 - M$.

Properties of an Asymmetric Equilibrium, $M = (1 - \alpha)/2$

Again suppose that the alleles $A_1, A_2,..., A_n$ are present at an equilibrium of the form $M = (1 - \alpha)/2$. We investigate the local stability of such an equilibrium to the initial increase of a new allele A_{n+1}. As before, assume p_{n+1} and r_{n+1} are sufficiently small that quadratic terms of (6) and (7) can be neglected. Then, since $M = (1 - \alpha)/2$ we have

$$p'_{n+1} = p_{n+1} + o(p_{n+1}, r_{n+1}) \tag{17}$$

$$r'_{n+1} = p_{n+1} \left\{ 4\alpha l_{n+1}/(1 - \alpha^2) + 2 \sum_{j=1}^{n} m_{n+1j}\hat{r}_j/(1 + \alpha) \right\}$$

$$+ r_{n+1} \left\{ 2 \sum_{j=1}^{n} m_{n+1j}[\hat{p}_j - (1 - \alpha)\hat{r}_j] - 2\alpha l_{n+1} \right\} \bigg/ (1 + \alpha). \tag{18}$$

The eigenvalues are therefore $\lambda_1 = 1$ and

$$\lambda_2 = 2 \left\{ \sum_{j=1}^{n} m_{n+1j}[\hat{p}_j - (1 - \alpha)\hat{r}_j] - \alpha l_{n+1} \right\} \bigg/ (1 + \alpha). \tag{19}$$

Since $m_{n+1j} \leqslant 1$ and $l_{n+1} \leqslant 1$ with $\hat{p}_j$ and $\hat{r}_j$ intrinsically non-negative, it is obvious that $1 > \lambda_2$. In addition,

$$1 + \alpha \geqslant 2\alpha l_{n+1} + 2 \sum_{j=1}^{n} m_{n+1j} \left\{ \sum_{k=1}^{n} m_{jk}X_{jk} + l_j Y_j \right\}.$$

Hence $\lambda_2 > -1$. Thus the larger eigenvalue of the local linear transformation (17), (18) is unity and we cannot infer the local stability from this linear analysis. In the same way, however, as in the companion paper (Eshel and Feldman, 1982), there appears to be a curve of sex ratio $1 - \alpha : 1 + \alpha$ which appears to rapidly attract new alleles in this case.

The Special Case When Haploids Are Male

If all haploids are necessarily male, then for any i (including $i = n + 1$), $l_i = 1$. Suppose that all diploids of genotype A_iA_j with $i, j = 1, 2,..., n$ are female, i.e., $m_{ij} = 0$ for these i, j. We ask is this state stable to the initial increase of an allele A_{n+1} which is such that for at least some i, $m_{in+1} > 0$. In other words, is the state at which all diploids are female stable to invasion by an allele which converts some diploids to male? This is a special case of the analysis of the previous sections. The assumptions entail that

$$M = \sum_{r=1}^{n} \sum_{j=1}^{n} m_{ij}X_{ij} + \sum_{i=1}^{n} l_iY_i = \sum_{i=1}^{n} Y_i = \alpha. \tag{20}$$

Now since $M = \alpha$, (16b) is impossible and (15b) must hold, the condition for the *stability* of symmetric equilibria is

$$M = \alpha > (1 - \alpha)/2 = M^*,$$

which is equivalent to $\alpha > \frac{1}{3}$. A special case of this model occurs when the offspring sex ratio (determined by the mother, say) is 1:1. As shown above a new allele converting diploids into males will be initially advantageous if $\alpha < \frac{1}{3}$. In particular, this holds if $M = \alpha = \frac{1}{4}$, which has been predicted by Trivers and Hare (1976) to be the "optimum" sex ratio selected via sister control of the sex ratio in haplodiploids (see also Oster *et al.*, 1977; Uyenoyama and Bengtsson, 1981).

Interpretation of Above Results as an Extension of Fisher's Argument

The properties of the symmetric and asymmetric equilibria described above produce the conjecture that the value $(1 - \alpha)/2$ is an "optimal" fraction of males in the sense that there should be evolution toward this value. A rigorous proof of this conjecture appears to be quite difficult due to the lack of monotonicity in the sex-ratio trajectories over time.

The male fraction $(1 - \alpha)/2$ is the probability that an allele taken at random from a sexual cell of an individual (either haploid or diploid) of the present generation was transmitted by a male parent in the previous generation. Indeed, α is the probability that the allele is taken from a haploid parent, in which case, according to the model, it could not have been transmitted by a male. And $1 - \alpha$ is the probability that the allele is taken from a diploid individual in which case the conditional probability that it came from a male (given the individual's diploid state) is $\frac{1}{2}$. In a diploid population we have $\alpha = 0$ and $M^* = \frac{1}{2}$ and the result agrees with Fisher's argument for the evolution of a sex ratio of 1:1 in diploid populations.

The following is an extension of Fisher's argument to include sex determination by the carrier in a haploid population. If α is the fraction (deter-

mined by parents) of haploid offspring and M is the fraction of (haploid or diploid) males, then a fraction M of all individuals in the population (either haploid or diploid) contribute $M^*(\alpha) = (1 - \alpha)/2$ (i.e. $\frac{1}{2}$ in Fisher's argument) of all the alleles to the next generation. A proportion $1 - M$ of the population (females) contributes the rest, i.e., $1 - M^*(\alpha)$ of the alleles. The ratio between the average male contribution and the average of females is therefore

$$\frac{M^*(\alpha)}{M} : \frac{1 - M^*(\alpha)}{1 - M}.$$

This ratio is larger or smaller than unity if $M < M^*(\alpha)$ or $M > M^*(\alpha)$, respectively. Following Fisher's reasoning, male advantage is expected if $M < M^*(\alpha)$ and male disadvantage is expected if $M > M^*(\alpha)$. Heuristically, it is expected that natural selection should act to shift the sex ratio closer to $M^*(\alpha): 1 - M^*(\alpha)$. Our model lends some analytic credence to this argument.

DISCUSSION

Possible examples of the type of model studied here include systems of sex determination in the platyfish *Xiphophorus maculatus* (Kallman, 1965) and a number of other fish species, while specific environmental stimuli are known to affect sex differentiation in a number of amphibia. The evolutionary consequences of the possibility that a given genotype may be of either sex must be distinguished from those resulting from situations in which the haploid fraction is determined maternally or by worker sisters. We have pursued the former question in a haplodiploid context in which, in every generation the fraction of haploid offspring is the constant α. We suggest that alleles which render the male fraction to $(1 - \alpha)/2$ will eventually be favored and those causing further departure from $(1 - \alpha)/2$ will eventually be lost. In this sense, $M^* = (1 - \alpha)/2$ is an optimal sex ratio and extends Fisher's argument for the optimality of the 1:1 sex ratio in diploids to the haplodiploid situation. According to the definition of Eshel and Feldman (1982), $1 - \alpha : 1 + \alpha$ has the property of evolutionary genetic stability, EGS.

A corollary to our result refers to the case in which the only ambiguity in sex determination occurs in diploids (this is probably the most usual case of such ambiguity in nature). Here an equilibrium in which ploidy completely determines sex (i.e., all diploid or *XX* individuals are females) is stable as long as the fraction of haploids α is larger than $\frac{1}{3}$. This is likely to be the case in *XY* and *XO* vertebrates and insects as well as non-social *Hymenoptera* (Uyenoyama and Bengtsson, 1981). The situation, however,

may be different if, as in eusocial *Hymenoptera*, the sex ratio is determined by sister workers (see e.g., Trivers and Hare, 1976; Oster *et al.*, 1977; Uyenoyama and Bengtsson, 1981). In this case the male fraction may be close to $\frac{1}{4}$, which according to the previous remark would permit alleles causing diploid males to enter the population. Diploid males have been detected in Hymenoptera although they are, at least in the case of *Apis mellifera*, usually eaten as larvae by workers (Adams *et al.*, 1977).

We suggest that the evolution of sex determination and the sex-ratio in XY systems may be influenced by any or all of the following seven factors.

(i) Autosomal mutations affecting the sex ratio among offspring of their carriers.

(ii) Autosomal mutations affecting the sex determination of their carriers.

(iii) Sex-linked mutations affecting the sex ratio among offspring born to the carrier mother.

(iv) Mutations on the X chromosome which affect the sex ratio of offspring born to a carrier (XY) father.

(v) Y-linked mutations which affect the sex ratio among offspring of carriers.

(vi) Sex-linked mutations which affect the carrier's sex determination.

(vii) Mutations which affect the sex ratio may be pleiotiopic with effects on mortality, viability, mate choice, etc. in such a way that these influence the sex ratio.

Fisher's argument is expressed in terms of the passage of genes to the next generation through males and females. Factors (i) and (ii) are relevant here, and Fishers argument has evaluated in exact genetical terms (for one locus, two allele models in terms of genefrequency changes) by Uyenoyama and Bengtsson (1979) and Eshel (1975) for (i) and (ii), respectively. The 1:1 sex ratio in these cases occurs irrespective of whether the genes affect the sex determination of an individual of the sex ratio among its offspring. Oster *et al.* (1977) and Uyenoyama and Bengtsson (1981) have shown that factors of class (iii) also produce a 1:1 sex ratio. Factor (v), however, may cause substantial deviations from the 1:1 sex ratio (Hamilton, 1967; Thomson and Feldman, 1975) and the same can be said of factor (iv) which might be exemplified by the XY females in the wood lemming (Bengtsson, 1977; Uyenoyoma and Bengtsson, 1981).

REFERENCES

ADAMS, J., ROTHMAN, E. D., KERR, W. E., AND PAULINHO, Z. L. 1977. Estimation of the number of sex alleles and queen matings from diploid male frequencies in a population of *Apis mellifera*. *Genetics* **86**, 583–596.

BENFORD, F. A. 1978. Fisher's theory of the sex ratio applied to the social Hymenoptera. *J. Theor. Biol.* **72**, 701–727.

BENGTSSON, B. O. 1977. Evolution of the sex ratio in the wood lemming, *Myopus schisticolor*, *in* "Measuring Selection in Natural Populations" (F. B. Christiansen and T. M. Fenchel, Ed.), pp. 333–343, Springer-Verlag, New York.

BODMER, W. F., AND EDWARDS, A. W. F. 1960. Natural selection and the sex ratio. *Ann. Human Genet.* **24**, 239–244.

CHARLESWORTH, B. 1977. Population genetics, demography and the sex ratio, pp. 345–363, *in* "Measuring Selection in Natural Populations" (F. B. Christiansen and T. M. Fenchel, Eds.), pp. 345–363, Springer-Verlag, New York.

CHARNOV, E. L. 1978. Sex ratio selection in eusocial Hymenoptera. *Amer. Natur.* **112**, 317–326.

CURTSINGER, J. W. AND FELDMAN, M. W. 1980. Experimental and theoretical analysis of the "sex-ratio" polymorphism in *Drosophila pseudoobseura*, *Genetics* **94**, 445–466.

EDWARDS, A. W. F. 1961. The population genetics of "sex ratio" in *Drosophila pseudoobscura*. *Heredity* **16**, 291–304.

ESHEL, I. 1975. Selection on sex ratio and the evolution of sex-determination, *Heredity* **34**, 351–361.

ESHEL, I., AND FELDMAN, M. W. 1982. On evolutionary genetic stability of the sex ratio, *Theor. Pop. Biol.* **21**,

FISHER, R. A. 1958. "The Genetical Theory of Natural Selection,"

HAMILTON, W. D. 1967. Extraordinary sex ratios, *Science* **156**, 477–488.

HARTL, D. L., AND BROWN, S. W. The origin of male haploid gnetic systems and their expected sex ratio, *Theor. Pop. Biol.* **1**, 165–190.

KALELA, O., AND OKSALA, T. 1966. Sex ratio in the wood-lemming *Myopus Schistocolor*, *Ann. Univ. Turkuensis Ser. A. II* **39**, 1–24.

KALLMAN, K. D. 1965. Genetics and geography of sex determination in the pocillid fish *Xiphophorus maculatus, Zoologica* **50**, 151–190.

MACNAIR, M. R. 1978. An ESS for the sex ratio in animals, with particular reference to the social Hymenoptera, *J. Theor. Biol.* **70**, 449–459.

NUR, U. 1974. The expected changes in the frequency of alleles affecting the sex ratio, *Theor. Pop. Biol.* **5**, 143–147.

OSTER, G., ESHEL, I., AND COHEN, D. 1977. Worker-queen conflict and the evolution of social insects. *Theor. Popul. Biol.* **12**, 49–85.

SCUDO, F. M. 1964. Sex population genetics, *Ri. Sci.* **34**, II-B, 93–146.

SHAW, R. F. 1958. The theoretical genetics of the sex ratio, *Genetics* **43**, 149–163.

SHAW, R. F., AND MOHLER, J. D. 1953. The selective significance of the sex ratio, *Amer. Natur.* **87**, 337–342.

SPIETH, P. T. 1974. Theoretical considerations of unequal sex ratios, *Amer. Natur.* **108**, 837–849.

THOMSON, G. J. AND FELDMAN, M. W. 1975. Population of modifiers of meiotic drive. IV. On the evolution of sex-ratio distortion, *Theor. Popul. Biol.* **8**, 202–211.

TRIVERS, R. L., AND HARE, H. 1976. Haplodiploidy and the evolution of the social insects, *Science* **191**, 249–263.

UYENOYAMA, M. K., AND BENGTSSON, B. O. 1979. Towards a genetic theory for the evolution of the sex ratio, *Genetics* **93**, 721–736.

UYENOYAMA, M. K., AND BENGTSSON, 1981. Towards a genetic theory for the evolution of the sex ratio. II. Haplodiploid and diploid models with sibling and parental control of the brood sex ratio and brood size, *Theor. Pop. Biol.* **20**, 57–79.

WILLIAMS, G. C. 1975. "Sex and Evolution, Princeton Univ. Press, Princeton, N. J.

On Evolutionary Genetic Stability of the Sex Ratio*

ILAN ESHEL[†] AND MARCUS W. FELDMAN

*Department of Biological Sciences,
Stanford University, Stanford, California 94305*

Received August 1, 1981

DEDICATED TO PROFESSOR LUCA CAVALLI-SFORZA
ON HIS SIXTIETH BIRTHDAY

INTRODUCTION

Under the assumption that the costs involved in producing male and female offspring were equal, Fisher (1930) argued that a sex ratio of one to one would maximize the number of grand-offspring and would, therefore, be expected to be achieved under evolution (see also Bodmer and Edwards, 1960). Simple one-locus two-allele "exact" population genetic models of sex determination by autosomal genes carried either by the individual (Eshel, 1975; Scudo, 1944) or its parents (Nur, 1974; Uyenoyama and Bengtsson, 1979) produce the result that the stable equilibria either determine a sex ratio which is one to one, or, under certain well-defined restrictions, a sex ratio which can be interpreted as being as close as possible to one to one, depending on the domains of attraction and parameters of the model. These findings may be interpreted as supporting Fisher's prediction. It is well known that the prediction may fail for sex-linked factors affecting sex determination (Hamilton, 1967; Thomson and Feldman, 1975; Charlesworth, 1977; Uyenoyama and Bengtsson, 1981; Eshel and Feldman, 1982).

The present study considers three situations of sex determination by a single gene locus with multiple alleles, in all of which diploid genotypes determine the probabilities that an individual is male or female. The results of our analysis strongly suggest that no matter how many sex determining alleles are present in these models at equilibrium, new mutations which affect the sex ratio will succeed if they bring it closer to one to one and fail if they cause increasing departure from equality. To support this conclusion only formal population genetic modeling is used; there is no recourse made to any

430

criterion other than the local stability of an equilibrium involving alleles A_1, $A_2,...,A_n$, which affect the sex ratio, with respect to the initial increase of a new allele A_{n+1}.

Model I: Diploid Individual Sex Determination

Consider a single locus with n alleles $A_1, A_2,..., A_n$. Generations are nonoverlapping. The only function of the gene is to affect the probability that an individual is male. Thus m_{ij} is the probability that genotype A_iA_j is male. Let the proportion of A_iA_j in males and females be $2X_{ij}$ and $2Y_{ij}$ ($i \neq j$) with X_{ii} and Y_{ii} the corresponding proportions for the homozygotes A_iA_i. The proportions of A_iA_j and A_iA_i in the whole population are $2Z_{ij}$ and Z_{ii}, respectively. Random mating between males and females is assumed. Set $X_i = \sum_{j=1}^n X_{ij}$, $Y_i = \sum_{j=1}^n Y_{ij}$ for the proportions of allele A_i in males and females, respectively.

Among the newborn offspring of the next generation we have

$$Z'_{ij} = (X_i Y_j + X_j Y_i)/2 \tag{1}$$

of genotype A_iA_j. Of these, m_{ij} are males. The male frequency among the offspring is therefore

$$M = \sum_i \sum_j Z'_{ij} m_{ij} = \sum_i \sum_j X_i Y_j m_{ij}. \tag{2}$$

The proportion of male A_iA_j in the offspring is therefore

$$X'_{ij} = m_{ij} Z'_{ij}/M. \tag{3}$$

Hence

$$X'_i = \sum_{j=1}^n X'_{ij} = \left\{ X_i \sum_j m_{ij} Y_j + Y_i \sum_j m_{ij} X_j \right\} \Big/ 2M, \tag{4}$$

and

$$Y'_i = \left\{ X_i \sum_j (1 - m_{ij}) Y_j + Y_i \sum_j (1 - m_{ij}) X_j \right\} \Big/ 2(1 - M) \tag{5}$$

$$= (X_i + Y_i - 2MX'_i)/2(1 - M). \tag{6}$$

At equilibrium, $X'_i = X_i$, $Y'_i = Y_i$ and therefore

$$(1 - 2M) X_i = (1 - 2M) Y_i \qquad (i = 1, 2,..., n), \tag{7}$$

so that

$$M = \tfrac{1}{2} \quad \text{or} \quad \hat{X}_i = \hat{Y}_i \quad \text{for all } i = 1, 2,..., n.$$

We therefore have

RESULT 1. *There are two possible classes of equilibria,* (i) *symmetric equilibria at which* $\hat{X}_i = \hat{Y}_i$ ($i = 1, 2,..., n$) *and* (ii) *equilibria at which the proportion of males is* $M = \frac{1}{2}$. *Note that fixation on any single allele is a symmetric equilibrium.*

Evolutionary Genetic Stability in Model I

The stability of these two classes of equilibria can, in principle, be analyzed in the usual way, that is, with respect to the n alleles already present in the population. An equally interesting evolutionary question, however, concerns the stability of these equilibria to invasion by an arbitrary new allele; this is the question of evolutionary genetic stability. First consider a symmetric equilibrium with alleles $A_1, A_2,..., A_n$ and frequencies $\hat{X}_1 = \hat{Y}_1$, $\hat{X}_2 = \hat{Y}_2,...,$ $\hat{X}_n = \hat{Y}_n$, and introduce a new mutant A_{n+1}, with frequencies X_{n+1} and Y_{n+1} in males and females small enough that quadratic terms may be neglected. Then from (4) and (5) we have

$$X'_{n+1} = (X_{n+1} + Y_{n+1}) \sum_{j=1}^{n} m_{n+1\,j}\hat{X}_j/2M + o(X_{n+1}, Y_{n+1}), \qquad (8)$$

$$Y'_{n+1} = (X_{n+1} + Y_{n+1}) \left(1 - \sum_{j=1}^{n} m_{n+1\,j}\hat{X}_j\right)\bigg/2(1 - M) + o(X_{n+1}, Y_{n+1}). \qquad (9)$$

Write $M_{n+1} = \sum_{j=1}^{n} m_{n+1\,j}\hat{X}_j$, the fraction of males with the mutant allele. Then, to the same order,

$$\frac{(X'_{n+1} + Y'_{n+1})}{X_{n+1} + Y_{n+1}} = \frac{M_{n+1}}{2M} + \frac{1 - M_{n+1}}{2(1 - M)}$$

$$= 1 + \frac{(1 - 2M)(M_{n+1} - M)}{2M(1 - M)}. \qquad (10)$$

The right-hand side of (10) is larger than unity if either $M < \frac{1}{2}$ and $M_{n+1} > M$ or $M > \frac{1}{2}$ and $M_{n+1} < M$. These are therefore the conditions for local instability of a symmetric equilibrium to invasion by A_{n+1}. Numerical iteration of (4) and (6) suggests that under these conditions the population will eventually achieve an equilibrium sex ratio involving $A_1, A_2,..., A_{n+1}$ which is closer to $1:1$ than at the original n-allele equilibrium.

The other equilibrium, characterized by $M = \frac{1}{2}$, may be investigated in the same way. X_{n+1} and Y_{n+1} have the same interpretation as before, but since the equilibrium value of M is $\frac{1}{2}$ the linearized versions of (4) and (5) are now

$$X'_{n+1} = X_{n+1} \sum_{j=1}^{n} m_{n+1\,j}\hat{Y}_j + Y_{n+1} \sum_{j=1}^{n} m_{n+1\,j}\hat{X}_j \qquad (11)$$

$$Y'_{n+1} = X_{n+1} \left(1 - \sum_{j=1}^{n} m_{n+1,j} \hat{Y}_j \right) + Y_{n+1} \left(1 - \sum_{j=1}^{n} m_{n+1,j} \hat{X}_j \right). \quad (12)$$

One eigenvalue of the matrix on the right of (11), (12) is unity and the other is $\sum_{j=1}^{n} m_{n+1,j}(\hat{Y}_j - \hat{X}_j)$ which is less than or equal to unity in absolute value, with equality only in degenerate cases. Numerical iteration of (4) and (6) from the neighborhood of an equilibrium with $M = \frac{1}{2}$ suggests that there is very fast return to an $(n + 1)$-allele surface characterized by $M = \frac{1}{2}$ so that the even sex-ratio is preserved from such a starting condition.

MODEL II: DIPLOID OFFSPRING SEX DETERMINED BY MOTHER'S AUTOSOMAL GENES

As for model I, X_{ii} and $2X_{ij}$ are the proportions of $A_i A_i$ and $A_i A_j$ in males with Y_{ii} and $2Y_{ij}$ the corresponding fractions in females. Also, as before,

$$X_i = \sum_j X_{ij}, \qquad Y_i = \sum_j Y_{ij}.$$

Now let m_{ij} be the fraction of males in the progeny of an $A_i A_j$ mother, and set

$$Z_i = \sum_j m_{ij} Y_{ij}. \qquad (13)$$

The fraction of males among the offspring produced by this generation is

$$M = \sum_{i=1}^{n} \sum_{j=1}^{n} m_{ij} Y_{ij} = \sum_{i=1}^{n} Z_i. \qquad (14)$$

Under the assumption of random mating, exactly $X_k/2$ of the offspring of a mother of genotype $A_i A_j$ will be of genotype $A_i A_k$ ($k = 1, 2,..., n$). Hence

$$X'_{ik} = \left(X_k \sum_{j=1}^{n} m_{ij} Y_{ij} + X_i \sum_{j=1}^{n} m_{kj} Y_{kj} \right) \Big/ 2M$$

$$= (X_k Z_i + X_i Z_k)/2M. \qquad (15)$$

In the same way

$$Y'_{ik} = (X_k Y_i + X_i Y_k - X_k Z_i - X_i Z_k)/2(1 - M)$$

$$= (X_k Y_i + X_i Y_k - 2M X'_{ik})/2(1 - M). \qquad (16)$$

Sum (15) and (16) over k to obtain

$$X_i' = X_i/2 + Z_i/2M \tag{17}$$

and

$$Y_i' = (Y_i + X_i - 2MX_i')/2(1 - M). \tag{18}$$

Hence at equilibrium

$$(1 - 2M) Y_i = (1 - 2M) X_i$$

and as before either $M = \frac{1}{2}$ at equilibrium or $X_i = Y_i$. In addition, from (17) at equilibrium

$$Z_i = MX_i,$$

and inserting this into (15) at equilibrium we have

$$X_{ik} = X_i X_k. \tag{19}$$

If $M \neq \frac{1}{2}$ at equilibrium, so that $X_i = Y_i$ then (16) becomes

$$Y_{ik} = X_i X_k = Y_i Y_k \tag{20}$$

at equilibrium, with $M = \sum_{i=1}^{n} \sum_{j=1}^{n} m_{ij} X_i X_j$.

Evolutionary Genetic Stability for Model II

Consider first any equilibrium at which $M \neq \frac{1}{2}$ and assume that a new allele A_{n+1} appears at low frequency. Set $\varepsilon_i = X_{in+1}$ and $\delta_i = Y_{in+1}$, half the frequency of heterozygotes $A_i A_{n+1}$ in males and females, respectively. Also set $\varepsilon = \sum_{i=1}^{n} \varepsilon_i$, and denote equilibrium values for the alleles and genotypes involving $A_1, A_2, ..., A_n$ by $\hat{X}_{ij}$, $\hat{Y}_{ij}$, etc. Then from (15) and (16) we have

$$\varepsilon' = \varepsilon/2 + \sum_{i=1}^{n} \delta_i m_{in+1}/2M \tag{21}$$

and

$$\delta_i' = \varepsilon(\hat{Y}_i - \hat{Z}_i)/2(1 - M) + \hat{X}_i \sum_{j=1}^{n} (1 - m_{j,n+1}) \delta_j/2(1 - M) \tag{22}$$

for $i = 1, 2, ..., n$ with quadratic terms form A_{n+1} neglected.

At a symmetric equilibrium from (17) to (20) the linear approximation in (21), (22) may be written

$$\begin{pmatrix} \varepsilon \\ \delta \end{pmatrix}' = \mathbf{A} \begin{pmatrix} \varepsilon \\ \delta \end{pmatrix}, \tag{23}$$

where $\delta = (\delta_{1,n+1},...,\delta_{n,n+1})$ written as a column, and

$$\mathbf{A} = \begin{bmatrix} 1/2 & m_{1n+1}/2M & \cdots & m_{nn+1}/2M \\ \hat{X}_1/2 & \dfrac{\hat{X}_1(1-m_{1n+1})}{2(1-M)} & \cdots & \dfrac{\hat{X}_1(1-m_{nn+1})}{2(1-M)} \\ \vdots & \vdots & & \vdots \\ \hat{X}_n/2 & \dfrac{\hat{X}_n(1-m_{1n+1})}{2(1-M)} & \cdots & \dfrac{\hat{X}_n(1-m_{nn+1})}{2(1-M)} \end{bmatrix}. \tag{24}$$

Now $\mathbf{A}$ is a strictly positive matrix and therefore has a unique strictly positive eigenvector associated with its largest eigenvalue. For such a right eigenvector to take the form $(\eta, \xi\hat{X}_1, \xi\hat{X}_2,..., \xi\hat{X}_n)$ the associated eigenvalue must solve the 2×2 system

$$\eta/2 + \xi M_{n+1}/2M = \lambda\eta$$

$$\eta/2 + \xi(1-M_{n+1})/2(1-M) = \lambda\xi. \tag{25}$$

The leading eigenvalue of this 2×2 system will therefore also be the leading eigenvalue of $\mathbf{A}$. A necessary and sufficient condition that this leading eigenvalue be greater than unity is that either $M < \frac{1}{2}$ and $M_{n+1} > M$ or $M > \frac{1}{2}$ and $M_{n+1} < M$, where $M_{n+1} = \sum_i \hat{X}_i m_{in+1}$. These are therefore the conditions for local instability of a symmetric equilibrium to invasion by A_{n+1}. Numerical iteration of (15) and (16) suggests that under these local instability conditions the population will eventually achieve an $(n + 1)$-allele equilibrium at which the sex ratio is closer to $1:1$ than at the original n-allele equilibrium.

Now consider the equilibrium $M = \frac{1}{2}$, and refer again to (21) and (22). Clearly, to this order of approximation,

$$\varepsilon' + \sum_{i=1}^{n} \delta_i = \varepsilon + \sum_{i=1}^{n} \delta_i,$$

so that there is an eigenvalue of unity for the linear approximation. Further manipulation of the local stability matrix in this case produces $n - 2$ zero eigenvalues while the remaining two are the roots of the quadratic

$$\lambda^2 - \lambda(\tfrac{1}{2} - M_{n+1}) + \sum_{i=1}^{n} (M_{n+1} - m_{in+1})(\hat{Y}_i - \hat{Z}_i).$$

These last two are less than unity in absolute value. As with model I, although we cannot answer the local stability question here, numerical iteration of (15) and (16) suggests that there is fast return to an $(n + 1)$-allele surface characterized by $M = \frac{1}{2}$ from such a starting condition.

MODEL III: XY SYSTEM WITH MOTHER CONTROLLING SEX DETERMINATION

Here females are XX, males XY, and m_{ij} is the proportion of males (i.e., haploids) in the progeny of an A_iA_j female. Again let $2Y_{ij}$ and Y_{ii} be the frequency of A_iA_j and A_iA_i females (diploids), respectively, and X_i the frequency of A_i males. Set $Y_i = \sum_j Y_{ij}$ and $M = \sum_i \sum_j Y_{ij}m_{ij}$. Then

$$X_i' = \sum_j Y_{ij}m_{ij}/M, \tag{26}$$

and

$$Y_{ij}' = \frac{1}{2(1-M)} \sum_k \left[(1-m_{ik}) Y_{ik}X_j + (1-m_{jk}) Y_{jk}X_i \right] \tag{27}$$

$$= \frac{1}{2(1-M)} \left[X_jY_i + X_iY_j - MX_i'X_j - MX_iX_j' \right], \tag{28}$$

using (26). Then clearly

$$Y_i' = \frac{1}{2(1-M)} \left[Y_i + X_i - M(X_i + X_i') \right]. \tag{29}$$

At equilibrium $X_i' = X_i$, $Y_i' = Y_i$ and (29) produces

$$(1-2M)X_i = (1-2M)Y_i.$$

Thus as in the previous models, at equilibrium either $M = \frac{1}{2}$ or $X_i = Y_i$.

Evolutionary Genetic Stability for Model III

Consider first the symmetric equilibrium with $X_i = Y_i = X_i' = Y_i'$. Assume that a new mutant allele A_{n+1} arises near this equilibrium with frequency ε in males. Replace the small variables Y_{in+1} by δ_i for $i = 1, 2,..., n$. From (26) and (27) $\delta\varepsilon'/\partial\varepsilon = 0$, $\partial\varepsilon'/\partial\delta_i = m_{in+1}/\hat{M}$. $\partial\delta_i'/\partial\varepsilon = \hat{X}_i/2$, $\partial\delta_i'/\partial\delta_j = (1 - m_{jn+1})\hat{X}_i/2(1-M)$. Hence the local stability matrix for the symmetric equilibrium $\mathbf{C}$ is

$$\mathbf{C} = \begin{bmatrix} 0 & m_{1,n+1}/\hat{M} & \cdots & m_{n,n+1}/\hat{M} \\ \hat{X}_1/2 & (1-m_{1n+1})\hat{X}_1/2(1-\hat{M}) & \cdots & (1-m_{nn+1})\hat{X}_1/2(1-\hat{M}) \\ \hat{X}_2/2 & (1-m_{1n+1})\hat{X}_2/2(1-\hat{M}) & \cdots & (1-m_{nn+1})\hat{X}_2/2(1-\hat{M}) \\ \vdots & \vdots & & \vdots \\ \hat{X}_n/2 & (1-m_{1n+1})\hat{X}_n/2(1-\hat{M}) & \cdots & (1-m_{nn+1})\hat{X}_n/2(1-\hat{M}) \end{bmatrix}.$$

By arguments similar to those of the previous sections the nonzero eigen-values of $\mathbf{C}$ are those of the 2×2 matrix $\mathbf{D}$:

$$\mathbf{D} = \begin{bmatrix} 0 & M_{n+1}/M \\ \frac{1}{2} & (1 - M_{n+1})/2(1 - M) \end{bmatrix}, \tag{30}$$

where $M_{n+1} = \sum_{l=1}^{n} m_{ln+1}\hat{X}_l$. Now $\mathbf{D}^2$ is a strictly positive matrix and hence, by the Frobenius theorem, the largest eigenvalue of the matrix is positive. This largest eigenvalue is greater than unity if and only if

$$(1 - 2M)(M - M_n) < 0. \tag{31}$$

Thus, as with models I and II we infer in the standard way that a symmetric n-allele equilibrium is unstable to invasion by A_{n+1} if either $M < \frac{1}{2}$ and $M_{n+1} < M$. Numerical iteration of (26) and (28) suggests that under these conditions the $(n + 1)$-allele equilibrium achieved has a more even sex ratio than $M: 1 - M$.

At the $M = \frac{1}{2}$ equilibrium the local stability to invasion by A_{n+1} is governed by an eigenvalue of unity with the other nonzero eigenvalues given by the roots of a quadratic which are less than unity in absolute value. As with the other models we expect that there is fast return to an $(n + 1)$-allele surface $M = \frac{1}{2}$ in this case.

DISCUSSION

There are three components to the equilibrium structure common to models I–III, and these are independent of the number of alleles, n, involved:

(i) There are two classes of equilibria; a symmetric class with allele frequencies equal in the sexes and unequal sex ratio, and another class at which the sex ratio is one to one.

(ii) Any n-allele symmetric equilibrium is unstable to allele A_{n+1} if the marginal sex ratio M_{n+1} induced by A_{n+1} satisfies the property that $M_{n+1} > M$ if $M < \frac{1}{2}$ or $M_{n+1} < M$ if $M > \frac{1}{2}$. This condition is satisfied if the marginal sex ratio is more even than $M: 1 - M$, but it may also be satisfied if, for example, $M_{n+1} > \frac{1}{2}$ with $M_{n+1} - \frac{1}{2} > \frac{1}{2} - M$. Even so, numerical iteration of the various recursions suggests that in all three models, when a symmetric n-allele equilibrium is unstable the $(n + 1)$-allele equilibrium ultimately reached has a more even sex ratio than $M: 1 - M$. It remains a conjecture that this is generally true in all three models. The proof promises to be difficult in view of the lack of monotonicity of the population sex ratio over time.

(iii)   At any n-allele equilibrium with one to one sex ratio the leading eigenvalue controlling the local stability of the equilibrium to invasion by a new allele is unity. Thus, local linear analysis is not informative for stability. There appears, however, from numerical work, to be an $(n + 1)$-allele surface of even sex ratio which is rapidly reached from a starting condition near an n-allele even sex ratio equilibrium.

These results produce the conjecture that there is a long-term evolutionary tendency to favor those alleles producing a more even sex ratio.

Suppose, more generally, that in a population with alleles $A_1, A_2,..., A_n$ affecting only the sex ratio there is a sex ratio $M^*: 1 - M^*$ with the following two properties:

I.   Any equilibrium of the population with alleles $A_1, A_2,..., A_n$ and a sex ratio $M: 1 - M$ other than $M^*: 1 - M^*$ is unstable to allele A_{n+1} if $M < M^*$ and $M_{n+1} > M$ or if $M > M^*$ and $M_{n=1} > M$, where M_{n+1} is the marginal sex ratio for A_{n+1} at the n-allele equilibrium.

II.   At any n-allele equilibrium with sex ratio $M^*: 1 - M^*$ the leading eigenvalue controlling the local stability of the equilibrium to invasion by a new allele is unity.

A sex ratio $M^*: 1 - M^*$ satisfying properties I and II is said to exhibit *evolutionary genetic stability*, EGS. In the models considered above the even sex ratio appears to have EGS. Properties I and II are not sufficient to guarantee that the population sex ratio should approach $M^*: 1 - M^*$; the latter would require global dynamic analysis of the genotype frequency recursions. Numerical work on the three models treated above suggests that there is long-term global evolution toward the even sex ratio.

It should be stressed that EGS invokes only the kinetics (and therefore only the parameters) of the genetic model. Fisher's original argument and many subsequent studies of the sex ratio (and other evolutionary problems) were made in terms of optimal population strategies. Many have used the concept of evolutionary stable strategy introduced by Maynard Smith and Price (1973). We suggest that since properties I and II involve only the underlying population genetic recursions, and require the introduction of no "payment" functions extraneous to the kinetic genotype frequency model, EGS arguments constitute a more direct population genetic approach, one which may prove to be useful for population genetic models of the evolution of other parameters.

REFERENCES

BODMER, W. F. AND EDWARDS, A. W. F. 1960. Natural selection and the sex ratio, *Ann. Human. Genet.* **24**, 239–244.

CHARLESWORTH, B. 1977. Population genetics, demography and the sex ratio, *in* "Measuring Selection in Natural Populations" (F. B. Christiansen and T. M. Fenchel, Eds.), pp. 345–363, Springer-Verlag, New York.

ESHEL, I. 1975. Selection on sex ratio and the evolution of sex determination, *Heredity* **34**, 351–361.

ESHEL, I. AND FELDMAN, M. W. 1982. On the evolution of sex determination and the sex ratio in haplodiploid populations, *Theor. Pop. Biol.* **21**,

FISHER, R. A. 1958. "The Genetical Theory of Natural Selection."

HAMILTON, W. D. 1967. Extraordinary sex ratios. *Science* **156**, 477–488.

NUR, U. 1974. The expected changes in the frequency of alleles affecting the sex ratio, *Theor. Pop. Biol.* **5**, 143–147.

SCUDO, F. M. 1946. Sex Population Genetics. *Ri. Sci.* **34**, II-B, 93–146.

THOMSON, G. J. AND FELDMAN, M. W. 1975. Population genetics of modifiers of meiotic drive. IV. On the evolution of sex-ratio distortion. *Theor. Pop. Biol.* **8**, 202–211.

UYENOYAMA, M. K. AND BENGTSSON, B. O. 1979. Towards a genetic theory for the evolution of the sex ratio, *Genetics* **93**, 721–736.

UYENOYAMA, M. K. AND BENGTSSON, B. O. 1981. Towards a genetic theory for the evolution of the sex ratio. II. Haplodiploid and Diploid models with sibling and parental control of the brood sex ratio and broodsize. *Theor. Pop. Biol.* **20**, 57–79.

MAYNARD SMITH, J. AND PRICE, G. R. 1973. The logic of animal conflicts. *Nature (London)* **246**, 15–18.

Evolutionarily Stable Strategies
and Viability Selection in Mendelian Populations

ILAN ESHEL*

*Department of Statistics,
Tel Aviv University,
Ramat-Aviv, Tel Aviv, Israel*

Received November 15, 1980

For various genetical structures, including haploid and diploid, one-locus n-alleles, and n-locus additive viability random mating models, natural selection resulting from intrapopulation conflicts between random individuals leads to exactly those genetical equilibria which determine a mixture of strategies evolutionarily stable according to the game theory definition of Maynard Smith and Price (1973).

1. INTRODUCTION

The concept of evolutionarily stable strategy (ESS) has been introduced by Maynard Smith and Price (1973) to characterize a strategy with the property that when adopted by a large enough majority in the population becomes the choice for each individual in that population, in the sense that it then maximizes the expected *fitness* or *survival* of its chooser. Once such a strategy has been established, it seems likely to be stable in a parthenogenetical population (Maynard Smith, 1974, 1976). A crucial question to be settled is whether natural selection, operating within the framework of known genetical structures does lead to the establishment and stabilization of ESSs, or strategies close to them (Lloyd, 1977; Maynard Smith, 1981). Concentrating on the simplest ESS problem of two alternative strategies in a population, this study is conducted under a wider assumption on the possible genetical basis for the choice of a strategy by a given individual. Indeed, a strategy in this sense may not necessarily be a behavioral pattern. It may be any phenotypic trait which affects the outcome of a conflict between individuals in the population.

For a class of genetical structures, namely locally adaptive ones, it is shown that all stable equilibria maintained by the system determine either fixation or mixture of strategies which are ESSs. The class of locally adaptive systems, for which this result holds, is shown to include all one-

* Present address: Departments of Mathematics and Biological Science, Stanford University, Stanford, Calif. 94305.

204

locus n-alleles, random mating viability models of a diploid population, but also other structures of genetic transmission of strategies. With some restriction on the intensity of the selection forces within this class of genetic structures it is shown further that any ESS corresponds to some stable equilibrium.

2. EVOLUTIONARY STABILITY AND CONVERGENCE OF GENOTYPE FREQUENCIES

Consider a large population in which any individual can choose one out of n strategies, say $S_1,..., S_n$. Let $x_j \geqslant 0$ be the proportion of individuals choosing the strategy S_j before selection ($j = 1,..., n$); and assume that the result of a given encounter confers an incremental or decremental fitness (say, viability) of v_{ij} to an individual following the strategy S_i, whose opponent practises the strategy $S_j(i, j = 1, 2,..., n)$. Thus, $V = \|v_{ij}\|$ can be construed as a population game matrix. Assuming random encounter between individuals in the population (for different assumptions see Cavalli-Sforza and Eshel, 1982), the expected change in fitness of an individual with the strategy S_i due to a random encounter is $\sum_i x_j V_{ij}$. The expected fitness of an individual, choosing the strategy S_i, is, therefore, proportional to

$$u_i(x) = 1 + \theta \sum_j V_{ij} x_j = \sum_j a_{ij} x_j = (Ax)_i, \qquad (2.1)$$

where $\theta > 0$ is up to a scale factor of the expected number of encounters for individuals per generation and we use the abbreviation $a_{ij} = 1 + \theta V_{ij}$.

If an individual employs a mixed strategy $y = (y_1,..., y_n)$, where y_i is the probability of him choosing the strategy S_i while population strategy is x his expected fitness is

$$V(x, y) = \sum_i y_i u_i(x) = \sum_{ij} y_i a_{ij} x_j = yAx. \qquad (2.2)$$

Within this framework, the formal definition of ESS suggested by Maynard Smith (1974) and further refined by Bishop and Cannings (1976) is as follows:
A strategy y is an ESS if for any strategy $x \neq y$,

$$V(y, y) \geqslant V(x, y) \qquad (2.3)$$

and if (2.3) holds as an equality, then

$$V(y, x) > V(x, x). \qquad (2.4)$$

Condition (2.3) guarantees that the strategy x will not be advantageous

when adopted by a single deviant individual. Condition (2.4) guarantees that if the strategy x is not disadvantageous at that stage, it will become disadvantageous when slightly accumulated. A version of the Nash theorem for symmetric games guarantees the existence of a strategy x satisfying (2.3) (namely, a symmetric *Nash solution* of the population game). There is no guarantee that an ESS exists for a given population game. In this work we see, however, that an ESS always exists in a 2×2 population game.

We now assume that the choice of a given strategy by an individual in a population depends, at least in a probabilistic way, on the genotype of the individual. We concentrate on the simplest situation where $k = 2$, i.e., each individual has exactly two possible strategies, S_1 and S_2, but there are m genotypes, $B_1, B_2,..., B_m$. Let p_i be the frequency of the genotype B_i among juveniles at a given generation and let h_i be the probability that an individual of genotype B_i will choose the strategy S_1.

The frequency of individuals choosing the strategy S_1 in the population will then be

$$x = x(\mathbf{p}) = \sum_{i=1}^{n} p_i h_i. \tag{2.5}$$

For convenience we speak of population strategy $0 \leqslant x \leqslant 1$ when we mean that fraction x of the population adopts strategy S_1 and the remaining fraction $1 - x$ uses strategy S_2.

The average fitness $u_1(x)$ of an individual choosing strategy S_1 where a proportion x in the population adopts strategy S_1 in this population is given by (2.1) and the fitness of the genotype B_i in this situation is

$$w_i = w_i(x) = h_i u_1(x) + (1 - h_i) u_2(x). \tag{2.6}$$

Assuming mating according to some given rule, appropriate segregation (depending on the structure of genotypes $B_1,..., B_m$) and natural selection with the frequency-dependent coefficient (2.6), one can calculate the frequencies $p_1',..., p_n'$ of the relevant genotypes in the next generation.

For example, in a one-locus random mating diploid model with n alleles, if h_{ij} is the probability that an individual of genotype $A_i A_j$ will choose the strategy S_1, $p_1,..., p_n$ are the allelic frequencies at a given generation, then by the Hardy–Weinberg law,

$$x = \sum_{ij}' p_i p_j h_{ij}. \tag{2.7}$$

(Note that both the p_i's and x are measured before selection.) After selection and random mating we, therefore, have

$$p_i' = \frac{p_i \sum_j p_j w_{ij}(x)}{\sum_{kj} p_k p_j w_{kj}(x)}, \tag{2.8}$$

where, following (2.6), the $w_{ij}(x)$ are given by $w_{ij}(x) = h_{ij} u_1(x) + (1 - h_{ij}) u_2(x)$. At this point, a traditional population geneticist would focus on the stable fixed points $\hat{\mathbf{p}}$ of transformation (2.8). An animal-conflict-oriented population biologist would emphasize the population strategy

$$\hat{x} = \hat{x}(\mathbf{p}) = \sum_{i=1}^{n} h_i \, p_i \qquad (2.9)$$

coresponding to each stable equilibrium. Sometimes, information about the population strategy $\hat{x} = x(\hat{p})$ is much easier to obtain than direct information about $\hat{p}$. Moreover, the population strategies $\hat{x}(\hat{p})$, corresponding to the stable genetical equilibria, do not depend on the specific genetical system except for the obvious restriction on the range of possible strategies allowed by the system, namely,

$$\lambda \leqslant \hat{x} \leqslant \mu \qquad (0 \leqslant \lambda < \mu \leqslant 1) \qquad (2.10)$$

where

$$\lambda = \min_{\mathbf{p} \in \Delta^n} \sum_{ij} p_i h_{ij} p_j \qquad (2.11)$$

$$\mu = \max_{\mathbf{p} \in \Delta^n} \sum_{ij} p_i h_{ij} p_j \qquad (2.12)$$

and Δ^n is the space of probability n-vectors.

A relevant ESS value should therefore, be defined in terms of the game matrix $\| V_{ij} \|$ and restriction (2.10) on the choice of strategies.

Following Bishop and Cannings, we have

DEFINITION. A strategy $\lambda \leqslant x \leqslant \mu$ is an ESS of the population game $\| V_{ij} \|$ with the restriction $\lambda \leqslant y \leqslant \mu$ on the choice of a strategy y (namely, an ESS of the *restricted game*) if for any $y \neq x$ with that *restriction*, conditions (2.3)–(2.4) hold.

Our objective now is to determine a structure under which the population strategies (2.9), determined by the stable genetical equilibria will be the ESSs of the restricted population game.

3. LOCALLY ADAPTIVE SYSTEMS

In order to emphasize the main concepts and methods it is convenient to describe the dynamics of the two-strategy population game in a general setting which includes the examples of (2.8).

Consider a 2×2 population game $\|V_{ij}\|$. To avoid degeneracies, we assume

$$V_{11} - V_{12} - V_{21} + V_{22} \neq 0. \tag{3.1}$$

Generally, consider that the proportion x of individuals in the population choosing the strategy S_1 may depend on a number of parameters $p_1,...,p_n$, with the vector $\mathbf{p} = (p_1,...,p_n)$ belonging to some compact set (e.g., $\mathbf{p}$ are probability vectors belonging to the simplex).

In a formal way

$$x = \phi(\mathbf{p}). \tag{3.2}$$

For technical convenience, we assume that ϕ has all continuous derivatives and that these derivatives are not all identically zero on any open set of ϕ.

Finally, it is assumed that the parameters P_i $(i = 1,...,n)$ change over successive generations to the new values p_i' determined by the parameters $\mathbf{p}$ of the first generation and the values $u_1(x)$ and $u_2(x)$, the game matrix corresponding to the population strategy $x(\mathbf{p})$ (cf. (2.1)): Formally,

$$\begin{aligned} p_i' &= f_i(\mathbf{p}, u_1, u_2) = f_i(\mathbf{p}, u_1(\phi(\mathbf{p})), u_2(\phi(\mathbf{p}))] \\ &= F_i(p), \qquad \text{say} \quad (i = 1, 2,..., n). \end{aligned} \tag{3.3}$$

The frequency of individuals choosing the strategy S_1 in the next generation is, therefore,

$$x' = \phi(\mathbf{p}'). \tag{3.4}$$

We are interested in a structure for which iteration of (3.3) converges to a stable vector $\hat{\mathbf{p}}$, the induced population strategy $\phi(\hat{p})$ of which is always an ESS of the restricted game. We show that this structure includes the one-locus n-allele random mating diploid system as a special case.

DEFINITION. The system (3.2)–(3.3) associated with the game matrix $\|a_{ij}\|$ and possibly with the restriction $\lambda \leqslant x \leqslant \mu$ on the choice of strategies is called a locally adaptive system with respect to the (restricted) game if for all $\mathbf{p} \in \Gamma$ i.e., $\mathbf{p}$ is a strictly positive frequency vector with $\lambda < x = \phi(\mathbf{p}) < \mu$,

$$\begin{aligned} x' > x \qquad &\text{if} \quad u_1(x) > u_2(x) \\ x' < x \qquad &\text{if} \quad u_1(x) < u_2(x) \end{aligned} \tag{3.5}$$

or equivalently

$$\text{sign}\{\phi(\mathbf{p}') - \phi(\mathbf{p})\} = \text{sign}\{u_1(\phi(\mathbf{p})) - u_2(\phi(\mathbf{p}))\}. \tag{3.6}$$

In this case, the transformation (3.3) is also said to be locally adaptive. This means that at any step the change $p \to p'$ in the population parameters results in an increase in the frequency of individuals practising the strategy that was advantageous at the time of the change. Note, however, that in the context of frequency-dependent selection, imposed by the viability game $\|a_{ij}\|$ on a given genetical structure, the condition (3.5) for local adaptivity of the system does not imply an increase in the average fitness $xu_1(x) + (1 - x) u_2(x)$ of the population, since $u_1(x)$ and $u_2(x)$ also change with x.

From arguments of continuity, it follows that for any adaptive system, inequality (3.5) holds, at least in a weak sense, at the boundaries $x = \lambda$ or $x = \mu$, and also

$$x' = x \qquad \text{if} \quad u_1(x) = u_2(x). \tag{3.7}$$

EXAMPLE 1. *A one-locus haploid model.* $p_1,...,p_n$ are the frequencies of the types $A_1,..., A_n$ in the population. h_i is the probability that an individual of type A_i $(i = 1, 2,..., n)$ will choose the strategy S_1. This is the only difference between types, thus we assume $h_i \neq h_j$ for all $i \neq j$. In this example, for any probability vector $\mathbf{p} \in S^n$

$$x = \phi(\mathbf{p}) = \sum_{i=1}^{n} p_i h_i, \tag{3.8}$$

which is obviously differentiable and, with the h_i being different from one another, nonconstant on an open set. We further set

$$\lambda = \min_{i=1,...,n} h_i, \qquad \mu = \max_{i=1,..., n} h_i, \tag{3.9}$$

and

$$f_i(\mathbf{p}, u_1, u_2) = \frac{p_i w_i}{\sum_j p_j w_j} \tag{3.10}$$

where

$$w_j = w_j(x) = h_j u_1(x) + (1 - h_j) u_2(x). \tag{3.11}$$

From (3.8) and (3.10) we obtain

$$x' = \phi(\mathbf{p}) = \frac{\sum h_i w_i p_i}{\sum w_i p_i}. \tag{3.12}$$

Hence, employing (3.8),

$$x' - x = \phi(\mathbf{p}') - \phi(\mathbf{p}) = \frac{\sum w_i h_i p_i - \sum w_i p_i \sum h_i p_i}{\sum w_i p_i}.$$

With (3.11), we have

$$x' - x = \left(\sum p_i w_i\right)^{-1} (u_1 - u_2)\left[\sum_i p_i h^2 - \left(\sum_i p_i h_i\right)^2\right], \qquad (3.13)$$

But $\sum p_i h_i^2 > (\sum p_i h_i)^2$ unless the $h_i - s$ are identical, and condition (3.5) for local adaptiveness of the haploid system with respect to any game of individual survival is satisfied.

EXAMPLE 2.   *A one-locus, random mating diploid model.* $(p_i,...,p_n) \in S^n$ are the frequencies of alleles in the population. As shown in the previous section

$$x = \phi(\mathbf{p}) = \sum_{ij} p_i\, p_j h_{ij}. \qquad (3.14)$$

λ and μ are given by (2.11) and (2.12), respectively,

$$p_i' = \frac{p_i \sum_i p_j w_{ij}(x)}{\sum_{jk} p_j\, p_k w_{jk}(x)} \qquad (3.15)$$

where

$$w_{ij} = h_{ij} u_1 + (1 - h_{ij})\, u_2 = u_2 + (u_1 - u_2)\, h_{ij}. \qquad (3.16)$$

But for any given symmetric matrix $\|w_{ij}\|$ of positive values and p_i' $(i = 1, 2,..., n)$ as determined in (3.15) we know (Kingman, 1961) that

$$\sum_{ij} p_i'\, p_j'\, w_{ij} \geqslant \sum_{ij} p_i\, p_j w_{ij} \qquad (3.17)$$

with equality only if $\mathbf{p}' = \mathbf{p}$.
  Inserting (3.16) into (3.17), one obtains

$$u_2 + (u_1 - u_2)\sum_{ij} p_i'\, p_j'\, h_{ij} \geqslant u_2 + (u_1 - u_2)\sum_{ij} p_i\, p_j h_{ij}$$

or

$$(u_1 - u_2)(x' - x) \geqslant 0 \qquad (3.18)$$

with equality only at equilibrium. This proves condition (3.5) for local adaptivity in respect to games of individual viability.

EXAMPLE 3.   *A multilocus, multiallele nonepistatic additive viability model of a randomly mated diploid population.* $B_i,...,B_m$ are the various genotypes allowed by the model; $p_1,...,p_m$ are their relative frequencies in a

given generation. $h_1,..., h_n$ are the probabilities of their choosing the first strategy. $x = \sum p_i h_i$ is, therefore, the population strategy and

$$w_i = w_i(x) = u_2 + (u_1 - u_2) h_i$$

is the viability of the genotype B_i. If the probability h_i is a sum of probabilistic effects, each being determined at a different locus, then the viability $w_i(x)$, determined by h_i and the population strategy, is, at any given generation, additive as well. In this case, we know (Ewens 1969) that with recombination, ransom mating, and selection, the new frequencies p_i' will satisfy the inequality

$$\sum p_i' w_i \geqslant \sum p_i w_i$$

with equality only at equilibrium (the w_i's are those of the previous generation). Again, this inequality is equivalent to

$$(u_1 - u_2)\left(\sum p_i' h_i - \sum p_i h_i\right) = (u_1 - u_2)(x' - x) > 0. \qquad (3.19)$$

Hence, the multilocus, nonepistatic additive model is shown to be locally adaptive with respect to any give of individual survival.

4. Stable Fixed Points of an Adaptive System and ESS

THEOREM 1.   *If* $\hat{p} \in \text{Int } \Gamma$ *(i.e., if* $p_i > 0$, $i = 1, 2,..., n$) *is a stable polymorphic equilibrium of the locally adaptive transformation* (3.3), *then* $y = \phi(p)$ *is an ESS of the game.*

In order to prove this theorem we need the following lemmas about ESS.

LEMMA 1.   *A strategy y satisfy* $\lambda \leqslant y \leqslant \mu$ *(i.e., a mixture* $(y, 1 - y)$*) is an ESS of the population game if for all* $x \neq y$ *(within the restrictions* $\lambda \leqslant x \leqslant \mu$ *of the game)*

$$(y - x)[u_1(y) - u_2(y)] \geqslant 0 \qquad (4.1)$$

and in the case of equality in (4.1)

$$(y - x)[u_1(x) - u_2(x)] > 0. \qquad (4.2)$$

LEMMA 2. (i) *If the edge strategy* $y = \lambda$ *is an* ESS, *then at least for some* $\varepsilon > 0$ *and for all* $\lambda < x < \lambda + \varepsilon$,

$$u_1(x) < u_2(x). \tag{4.3}$$

(ii) *If the edge strategy is not an* ESS, *then for all* $\varepsilon > 0$ *there exist at least one* $\lambda < x < \lambda + \varepsilon$.

$$u_1(x) > u_2(x). \tag{4.4}$$

LEMMA 3. (i) *If* $\lambda < y < \mu$ *is an* ESS, *then* $u_1(y) = u_2(y)$ *and* (4.2) *holds.*

(ii) *If* $\lambda < y < \mu$ *is not an* ESS, *then either* $u_1(y) \neq u_2(y)$ *or* $u_1(y) = u_2(y)$ *and*

$$(y - x)[u_1(x) - u_2(x)] < 0 \qquad \text{for all} \quad x \neq y. \tag{4.5}$$

LEMMA 4. *For any* $\lambda \leqslant y \leqslant \mu$: (i) *If* $u_1(y) = u_2(y)$ *and if for all* $0 < |x - y| < \varepsilon$ ($\varepsilon > 0$ *small enough*), $(x - y)[u_1(x) - u_2(x)] < 0$, *then* y *is an* ESS.

(ii) *If* y *is an* ESS, *then* $(x - y)[u_1(x) - u_2(x)] < 0$ *for all* $0 < |x - y| < \varepsilon$.

The proofs of these lemmas are given in the Appendix.

Proof of Theorem 1. That $\hat{\mathbf{p}}$ is a stable equilibrium of the transformation F, means that for some vicinity S of $\hat{\mathbf{p}}$, $\mathbf{p}^{(n)} = \mathbf{F}^{(n)}(\mathbf{p}^{(0)}) \rightarrow \hat{\mathbf{p}}$ for all $\mathbf{p}^0 \in S$. From the continuity of ϕ, $x_n = \phi(\mathbf{p}^{(n)}) \rightarrow y = \phi(\hat{\mathbf{p}})$. But from the assumption that ϕ is not constant on any open set we know that for some $p^{(0)} \in S$, $x_0 = \phi(\mathbf{p}^0) \neq y$. For any $\varepsilon > 0$, there is, therefore, n such that $|x_{n+1} - y| < |x_n - y| < \varepsilon$. If $x_n > y$, then $x_{n+1} < x_n$ and, from (3.5), we therefore know that $u_1(x_n) < u_2(x_n)$. If $x_n < y$, then $x_{n+1} > x_n$ and therefore $u_1(x_n) > u_2(x_n)$. In both cases, $(x_n - y)[u_1(x_n) - u_2(x_n)] < 0$. Moreover, since $y = \phi(\hat{\mathbf{p}}) = \phi(\mathbf{F}(\hat{p}))$, it follows from (3.5) that $u_1(y) = u_2(y)$. Hence, it follows from Lemma 4 that y is an ESS. ∎

Next is the opposite question: Are all ESSs of the population game stable with respect to the transformation determined by the genetical structure?

DEFINITION. A population strategy x ($\lambda \leqslant x \leqslant \mu$) is said to be stably maintained by the transformation (3.3) if, when the initial population state ϕ determines a population strategy $\phi(\mathbf{P})$ close to x, then the iteration of (3.3) results with convergence of the population strategy to x. In a formal way:

A population strategy $\lambda \leqslant x \leqslant \mu$ is said to be stably maintained by the transformation (3.3) if for small enough $\varepsilon > 0$ and for all $\mathbf{p} \in \Gamma$ for which $|\phi(\mathbf{p}) - x| < \varepsilon$

$$|\phi(\mathbf{F}^{(n)}(\mathbf{p})) - x| \to 0. \tag{4.6}$$

DEFINITION. The transformation $\mathbf{F}$ is said to maintain a protected mixture of strategies if neither of the edge strategies is stably maintained by it.

THEOREM 2.   (i) *If an edge strategy of a restricted game is an* ESS, *then it is stably maintained by all locally adaptive transformations corresponding to the game.*

(ii) *If a substantially mixed ESS $\lambda < y < \mu$ exists, then a protected mixture of strategies is maintained by any locally adaptive transformation $\mathbf{F}$ corresponding to the population game. Moreover, in this case $\mathbf{F}$ either stably maintains the mixed strategy y or it allows for initially increasing fluctuations around it.*

Proof. (i) Let $y = \lambda$ be an ESS. From Lemma 2 we know that for $\lambda < x < \lambda + \varepsilon$, $u_1(x) < u_2(x)$. Hence, if $\phi(\mathbf{p}) = x$, (3.5) implies that

$$x' = \phi(\mathbf{p}') = \phi(\mathbf{F}(\mathbf{p})) < x. \tag{4.7}$$

Convergence of the sequence $\phi(\mathbf{F}^{(n)}(\mathbf{p}))$ to λ is immediately implied by (4.7) and the continuity of ϕ and $\mathbf{F}$.

(ii) Suppose $\lambda < y < \mu$ is an ESS. From Lemma 3 it follows that

$$(y - \lambda)[u_1(\lambda) - u_2(\lambda)] > 0.$$

Hence, $u_1(\lambda) > u_2(\lambda)$ and for x close enough to λ, say $0 < |x - \lambda| < \varepsilon$, $u_1(x) > u_2(x)$.

From this and from the local adaptivity property (3.3) of $\mathbf{F}$ it follows that for any $\mathbf{p} \in \Gamma$ with $0 < |\phi(\mathbf{p}) - \lambda| < \varepsilon$, $\phi(\mathbf{F}(\mathbf{p})) > \phi(\mathbf{p})$; thus $|\phi(\mathbf{F}(\mathbf{p})) - \lambda| > |\phi(\mathbf{p}) - \lambda|$ and the edge strategy λ is not stably maintained. In the same way, $y = \mu$ is also not stably maintained by $\mathbf{F}$; thus, $\mathbf{F}$ maintains a protected polymorphism.

Moreover, Lemma 3 and (3.5) also implies that if $\lambda < y < \mu$ is an ESS, then for any $\mathbf{p} \in \Gamma$,

$$\mathrm{sign}(y - x) = \mathrm{sign}(x' - x). \tag{4.8}$$

This means that x' is either closer to y than x or, in the case of very

strong selection forces, it is further apart on the other side of y. Thus, the locally adaptive system either stably maintains the ESS value y or it allows fluctuations around it. ∎

5. DISCUSSION

The Darwinian theory of evolution concerns long-run changes in phenotypes in a population due to natural selection. However, long lasting phenotypic changes may reflect changes in the distribution of genotypes in a population. And quite generally, it is impossible to understand and predict the effect of natural selection on a given trait without a knowledge of the mechanism, genetical or other, by which this trait is transmitted from generation to generation. This partly explains the fact that changes in genotypic frequencies became the focus of modern studies in population biology. Another reason, quite attractive to quantitatively oriented scientists, is the possibility of drawing rigorous results about the changes in genotype frequencies once the selection forces operating on a given genetical structure are known.

Unfortunately, the genetical basis for evolutionary changes in phenotypic traits which are of a direct interest to students of natural history is rarely, if ever, known. Even less so are the exact selection forces operating on them (e.g., Lewontin, 1974).

An alternative approach to the study of this sort of trait attempts to avoid complications stemming from the specific nature of one genetical structure or another. Instead, intuitively understood criteria of phenotype optimization are suggested, with the basic asumption that despite technical counterexamples, accumulated through years of population genetic research, the basic Darwinian relation between adaptation and natural selection must lead to some sort of local optimization, at least as a workable approximation. Thus, instead of dealing with many technical, unmeasurable, and presumably insignificant details, it is preferable to ignore them in order to obtain simple qualitative results which are at least easy to interpret. The crucial question is under what circumstances the technical details being ignored are, indeed, insignificant.

This can be settled only by a comparison between results, obtained by the intuitive model with those achieved under a sufficiently general family of rigorous models. It should be kept in mind, though, that the very concept of a "sufficiently general" family of rigorous models cannot, in itself, be determined in a rigorous way; and no family of models is general enough as to provide us with more than a sample information about the validity of a given criteriona in the much wider context of the theory of evolution.

One of the two perhaps most important examples involves the concept of

evolutionarily stable strategies (ESS), suggested by Maynard Smith and Price (1973). The question here is, what is the relevance of results drawn by methods of game theory to the actual laws of natural selection in population genetics theory. At least in one case, namely, that of selection for optimal spread of seeds, interesting results drawn by pure methods of ESS (Hamilton and May, 1977) were fully verified by a rigorous analysis of both haploid and diploid populations (Motro, 1982a, b). In other cases, such as parent offspring conflict (Trivers, 1974) the result of a rigorous analysis of genetical models seems less in agreement with the ESS model of local optimality (e.g., Feldman and Eshel, 1982).

The present work is an attempt to develop a theoretical basis for comparison between ESSs of a given population game and the strategies which are determined by exact genetical structures.

The selective value of a strategy (or phenotype) affecting the outcome of intrapopulation conflicts is, by definition, frequency dependent. When conflicts occur with random encounters between individuals in the population (though not necessarily only then), the selective value of a strategy is additively frequency dependent. In this study we concentrate on the case of two alternative strategies (phenotypes), the choice (or manifestation) of each depending, at least statistically, on the individual's genotype, or maybe on some other inherited character of it.

For those genetical structures which obey Fisher's fundamental law of natural selection (including all one locus random mating viability models) it is shown that natural selection due to intrapopulation conflicts between random individuals stabilizes only those genetical equilibria which determine evolutionarily stable mixtures of strategies according to the game theory definition of Maynard Smith and Price. Moreover, with some further assumption, precluding fluctuations, any ESS of the appropriate population game is stably mainained by the genetical structure.

It is now well established, however, that Fisher's fundamental law, connecting adaptation and natural selection in the intuitively Darwinian way is mathematically false for almost all multilocus genetical systems (e.g., Karlin, 1975). These cast serious limitations on our ability to predict any precise adaptive pattern, at all (cf. Maynard Smith, 1978; Lewontin 1974. It is, therefore, not the intention of this study to claim that exact ESS values are, indeed, predicted to be found in natural populations. Instead it is shown that the game theory argument of ESS is mathematically equivalent in result to rigorous analysis of the specific models studied here and, thus, can legitimately replace them.

APPENDIX

Proof of Lemma 1. Y is an ESS if and only if for all $x \neq y$, $V(y, y) \geqslant V(x, y)$ and if in the case of equality $V(y, x) > V(x, x)$.

In the case of two competing strategies, condition (4.1) is written as

$$yu_1(y) + (1 - y)u_2(y) \geqslant xu_1(y) + (1 - x)u_2(y)$$

and (4.2) becomes

$$yu_1(x) + (1 - y)u_2(x) \geqslant xu_1(x) + (1 - x)u_2(x).$$

Thus, (4.1) and (4.2) immediately follow.

Proof of Lemma 2. (i) Suppose $y = \lambda$ is an ESS. Condition (4.1) can then be written as

$$u_1(\lambda) - u_2(\lambda) \leqslant 0.$$

If $u_1(\lambda) - u_2(\lambda) < 0$, it follows from the continuity of u_1 and u_2 that (4.3) holds for at least $\lambda < x < \lambda + \varepsilon$. If, on the other hand, $u_1(\lambda) - u_2(\lambda) = 0$, (4.2) means that $u_1(x) - u_2(x) < 0$ for all $\lambda < x \leqslant \mu$.

(ii) Suppose $y = \lambda$ is not an ESS. In this case, either (4.1) is not true, or it holds as an equality and (4.4) results from the continuity of u_1 and u_2, If (4.1) holds as an equality, (4.2) means that for all $\lambda < x \leqslant \mu$, $u_1(x) - u_2(x) \geqslant 0$. But in this case,

$$u_1(x) - u_2(x) = u_1(\lambda) - u_2(\lambda) + (a_{11} - a_{12} - a_{21} + a_{22})(x - \lambda)$$

$$= \theta(V_{11} - V_{12} - V_{21} - V_{22})(x - \lambda) \geqslant 0$$

and from (3.1) it follows that $V_{11} - V_{12} - V_{21} + V_{22} > 0$ (it can be neither zero nor negative); hence (4.4) must hold for all $x > \lambda$.

Proof of Lemma 3. (i) is immediate from (4.1) and (4.2). Also, if y is not an ESS and $u_1(y) = u_2(y)$, (4.6) holds at least as a weak inequality. As in the proof of Lemma 1, the sharp version of (4.6) is implied from (3.1).

ACKNOWLEDGMENTS

REFERENCES

BISHOP, D. T., AND CANNINGS, C. 1976. Models of animal conflict *Advan. Appl. Prob.* **8**, 616–621.

CAVALLI-SFORZA, L. L., AND ESHEL, I. 1982. Assortment of encounters and evolution of cooperativeness, *Proc. Nat. Acad. Sci. USA* **79**, 1331–1335.

DAWKINS, R. 1980. Goof dytzyrgy or evolutionarily stable strategy? *In* "Sociobiology: Beyond Nature/Nuture?" (G. W. Barlow and J. Silvergerg, Eds.) (AAS Selected Symp. Ser.). Westview Press, Boulder, Colo.

ESHEL, I. 1981a. A remark on evolutionary stability in respect to more than one minority strategy. Manuscript.

ESHEL, I. 1981b. Evolutionary stability and continuous stability. Manuscript.

ESHEL, I. 1981c. On a coevolutionary instability of a fully mixed Nash solution of a 2×2 two player games. Manuscript.

EWENS, W. J. 1969. With additive fitnesses, the mean fitness increases, *Nature (London)* **221**, 1076.

FELDMAN, M. W., AND ESHEL, I. 1982. On the theory of parent offspring conflict—A two locus genetical model, *Amer. Natur.* **119**, 285–292.

HAMILTON, W. D. 1964. The genetical evolution of social behavior, I, II, *J. Theor. Biol.* **7**, 1–521.

HAMILTON, W. D. 1980. Selfish and spiteful behavior in an evolutionary model, *Nature (London)* **228**, 1218–1220.

HAMILTON, W. D., AND MAY, R. 1977. Dispersal in stable habitats, *Nature (London)* **269**, 578–581.

KARLIN, S. 1975. General two locus selection models: Some objectives, results, and interpretations, *Theor. Pop. Biol.* **7**, 364–398.

KINGMAN, J. F. C. 1961. A mathematical problem in populatio genetics, *Proc. Cambridge Phil. Soc.* **57**, 574–582.

LEWONTIN, R. C. 1974. "The Genetic Basis for Evolutionary Change," Columbia Univ. Press, New York/London.

LLOYD, D. G. 1977. Genetic and phenotypic models of natural selection, *J. Theor. Biol.* **69**, 543–560.

MAYNARD SMITH, J. 1972. Game theory and the evolution of fighting, *in* "On Evolution" (J. Maynard Smith, Ed.), pp. 8–20, Edinburgh Univ. Press, Edinburgh.

MAYNARD SMITH, J. 1974. The theory of games and the evolution of animal conflict, *J. Theor. Biol.* **47**, 209–221.

MAYNARD SMITH, J. 1976. Evolution and the theory of games, *Amer. Sci* **64**, 41–45.

MAYNARD SMITH, J. 1978. Optimization theory in evolution, *An. Rev. Ecol. Systemat.* **9**, 31–56.

MAYNARD SMITH, J. 1980. Models of the evolution of altruism, *Theor. Pop. Biol.* **18**, 151–159.

MAYNARD SMITH. J. 1981. Will sexual population evolve to an ESS? *Amer. Natur.* **117**, 1015–1018.

MOTRO, U. 1982a. Optimal rates of dispersion and migration in biological populations. I. Haploid models, *Theor. Pop. Biol.* **21**,

MOTRO, U. 1982b. Optimal rates of dispersion and migrations in biological populations. II. Diploid models, *Theor. Pop. Biol.* **21**,

TRIVERS, R. L. 1974. Parent offspring conflict, *Amer. Zool.* **14**, 249–264.

Kin Selection and Strong Evolutionary Stability
of Mutual Help*

Ilan Eshel[†] and Uzi Motro[‡]

Department of Statistics, Tel Aviv University, Tel Aviv, Israel

Received August 1980

1. Introduction

The theory of kin selection, as developed by Hamilton (e.g., 1964, 1970) stems from the fact that in a sexually reproducing population, the genotype of a progeny of an individual is not necessarily identical to the genotype of either parent. Thus natural selection cannot operate through the preservation of the most fit type. Instead, it can only be expressed in terms of changes in gene frequencies. Such changes, as suggested by Hamilton, are likely to be in favor of those genes which, by their effect on their carriers, act to increase the expected number of their copies in the population of the next generation. Thus, in Hamilton's terminology, natural selection is expected to favor genes which increase their carrier's *inclusive fitness* (e.g., Hamilton, 1964).

More specifically, it has been maintained by Hamilton that if, in order to save a kin of relatedness r (see Wright, 1922), a risk of $0 \leqslant x \leqslant 1$ is needed, then, by taking this risk, an individual will add a value of $r - x$ to its inclusive fitness. Taking such a risk will, therefore, be selected for if and only if $x \leqslant r$, and the value r is expected to be the maximal risk accepted by an individual in a population in order to save a relative of relatedness r.

This prediction is, however, not always in a satisfactory agreement with empirical observations, a discrepancy that provoked some attacks on the very theory of kin selection (e.g., Zahavi, 1981). For example, parents' help to their offspring is almost always more generous than offspring's help to their parents or to their sibs, even though the relatedness in both cases is the same. Moreover, even on a theoretical level, the prediction mentioned above cannot possibly be true, for example, on an isolated island, overpopulated by

420

a highly inbred population, where only those individuals which manage to exterminate their neighbors, even though of close relatedness, will leave their genes to the next generation. (See Harpending, 1979, for a more detailed treatment.)

More detailed study of kin selection, thus, concentrates on two parallel lines. On one hand, a rigorous model-study of various sorts of sib-to-sib altruism (e.g., Cavalli-Sforza and Feldman, 1978, Uyenoyama and Feldman, 1980, Boorman and Levitt, 1980 and references therein) reveals some of the difficulties concerning selection of such complicated traits as altruism under the restriction of a genetical structure, even in the simplest cases. For example, negative correlation is shown to be created between alleles carried by the same individual when alleles are measured by their effect on the altruistic action (Feldman and Eshel, unpublished). Such negative correlation is shown to be favorable for the non-altruist, even though its effect is negligible near fixation.

On the other hand, deviations from the classic prediction of kin selection are likely to occur in nature due to inevitable complications of the altruistic act itself which can differ from one situation to another (e.g., Cohen and Eshel, 1976). In a previous study (Eshel and Cohen, 1976) an attempt has been made to incorporate three crucial factors into the classic model of kin selection:

(i)   Differences in potential fertility makes it advantageous to take a higher risk in order to save potentially fertile (e.g., younger) kin.

(ii)   Competition among kins is likely to reduce, eliminate or even negate a loss in inclusive fitness due to death of a relative.

(iii)   As long as mutual help of some level is established among kins, a death of one individual results in a loss of its potential help to other members of its family and thereby to a further reduction in their inclusive fitness.

The purpose of this work is to quantitatively study the combined effect of relatedness, competition and mutual dependence on the relations between two relatives having the same potential fitness. Most results are drawn from a model of a two-player game with the inclusive fitness as the evolutionary relevant payoff function. We consider the maximal risk (MR) accepted by one individual in order to save a relative of a given relatedness r and, adopting the definition of Maynard Smith and Price (1973), we calculate the value of the MR which is an evolutionarily stable strategy (ESS) in the population. In some cases there exist more than one positive value of maximal risk that can be established in the population as an ESS, and there might be also some negative values, corresponding to evolutionarily stable

strategies of spitefulness towards kin. (See also Hamilton 1970, Eshel and Cohen, 1976.)

It is shown, however, that the dynamic process of natural selection, if operating to increase the inclusive fitness, can lead to the establishment of some, but not other, so-called evolutionarily stable strategies. A stronger, and for this purpose more natural, definition of evolutionary stability is, therefore, suggested in Section 4, and the stability of the various ESS values of maximal risk are checked according to this definition.

In a special case of mutual help between sibs, the results achieved by the game theory model are validated (see Section 5) by a direct Fisherian analysis of changes in gene frequencies.

2. THE GAME THEORY MODEL

Let A and B be two individuals with a symmetric degree of relationship r, $0 \leqslant r \leqslant 1$.

Let A be ready to help B up to the maximal risk (MR) x. By this we mean that if the help of A can increase the survival probability of B by p, then A will donate this help to B unless by doing so it would decrease its own survival probability by more than px. It is convenient to extend this notion of maximal risk (MR) to include also negative values. Thus, by saying that the MR of A to B is $x < 0$, we mean that if A can inflict damage on B, he will do so unless the risk he takes is greater than the proportion $-x$ of the damage inflicted on B, damage and risk being measured in terms of survival probability. We assume, likewise, that B is ready to help A up to the MR y $(-\infty < y < \infty)$.

We concentrate on a time-continuous model in which each individual has an infinitesimal probability of dying during any infinitesimal time interval dt. We naturally assume that this probability depends both on the individual's willingness to take a risk in order to help its relative (or harm him) (i.e., its MR value x) and on the willingness of its relative to take risk in order to help (or harm) him (i.e., the relative's MR value y), provided both individuals are alive at the beginning of the time interval in question. Denote this probability by $\lambda(x, y)\, dt$. The probability of B dying during the same time interval is, indeed, $\lambda(y, x)\, dt$. The death probability of each of them in the absence of the other is $\lambda(0, 0)\, dt = \lambda\, dt$, say. From the definition of y as the MR of B we see that $\lambda(x, y)$ is a decreasing function of y.

It follows that the length of time during which both individuals are alive is distributed exponentially with a parameter $\lambda(x, y) + \lambda(y, x)$ and an expectation $1/(\lambda(x, y) + \lambda(y, x))$.

The distribution of the remaining life span of each individual, after the

death of its relative, is exponential with a parameter λ and expectation $1/\lambda$. The probability that A is the first one to die is $\lambda(x, y)/(\lambda(x, y) + \lambda(y, x))$.

We finally assume that the number of offspring produced by each individual is proportional to the length of time it lives in the presence of its relative (and, perhaps, competitor) plus $1 + \sigma$ times the length of time it lives after the death of this relative (provided it outlives it). σ is the *degree of competition* between the two relatives and it is natural (though not necessary) to assume $0 \leqslant \sigma \leqslant 1$.

With these assumptions, the fitness of individual A is

$$\omega(x, y) = \frac{1}{\lambda(x, y) + \lambda(y, x)} + \frac{\lambda(y, x)}{\lambda(x, y) + \lambda(y, x)} \cdot \frac{1 + \sigma}{\lambda}. \qquad (2.1)$$

Obviously, the Fisherian fitness of B is $\omega(y, x)$.

Hence, the inclusive fitness of A, defined as the expected number of copies of its genes to be represented in the next generation (Eshel and Cohen, 1976) turns out to be

$$\Omega(x, y) = \omega(x, y) + r\omega(y, x) + c. \qquad (2.2)$$

(The same equality, though with a different c, we get with Hamilton's definition of inclusive fitness as "own offspring $+ r$ times the additional offspring B has because of A's help.")

Our first objective is to calculate the optimal MR readiness of A to help B when B's MR readiness to help A is given. Optimality is to be understood in terms of maximization of the inclusive fitness. Then we shall calculate evolutionarily stable strategies for mutual maximum risk.

3. ANALYSIS OF THE MODEL AND EVOLUTIONARILY STABLE STRATEGIES

We first prove the following useful proposition, resulting from the definition of a maximal risk strategy.

PROPOSITION 1. *For each x and y the following relation exists:*

$$x \frac{\partial}{\partial x} \lambda(y, x) = -\frac{\partial}{\partial x} \lambda(x, y). \qquad (3.1)$$

Proof. $\lambda(x, y)\, dt$ and $\lambda(y, x)\, dt$ are the death probabilities of A and B, respectively, during a time interval of length dt at the beginning of which both are alive. Increasing the maximal risk x of A by dx results in decreasing the death probability of B by $\lambda(y, x)\, dt - \lambda(y, x + dx)\, dt$. But for $x \geqslant 0$ this

is due to help, given by A at a relative cost to himself no greater than $x + dx$ and no smaller than x (immediate from the definition of maximal risk), cost being measured in terms of death probability. Hence, the increase in the death probability of A, resulting from his readiness to take high risks, cannot be either larger than $x + dx$ times the decrease in the death probability of B or smaller than x times that decrease. Thus

$$x \leqslant \frac{\lambda(y, x)\, dt - \lambda(y, x + dx)\, dt}{\lambda(x + dx, y)\, dt - \lambda(x, y)\, dt} \leqslant x + dx$$

(for $dx < 0$ the inequality signs are reversed). By letting dx tend to zero, we get the required result.

By similar arguments we show the validity of (3.1) also for $x < 0$. ∎

Note that the infinitesimal term $(\partial/\partial x)\,\lambda(y, x)\, dt\, dx$ measures the probability that within a time interval of length dt, one individual will have an opportunity to save (or harm) the other with a relative risk between x and $x + dx$ to himself. Obviously the derivative $(\partial/\partial x)\,\lambda(y, x)$ (and, therefore, $(\partial/\partial y)\,\lambda(y, x)$) will be higher in populations wherein some intrinsic social structure increases the probability of one individual helping or harming another. Thus, except for a constant $\lambda = \lambda(0, 0)$, the function $\lambda(x, y)$ is determined by the structure of the ecological interaction between individuals in the population and will be referred to as the *interaction function*.

From the definition of $\lambda(x, y)$ we have $(\partial/\partial x)\,\lambda(y, x) \leqslant 0$ for all values of x and from Proposition 1 it follows that

$$\frac{\partial}{\partial x}\lambda(x, y) \geqslant 0 \qquad \text{for} \quad x \geqslant 0,$$

$$\frac{\partial}{\partial x}\lambda(x, y) \leqslant 0 \qquad \text{for} \quad x \leqslant 0.$$

Indeed, it is the absolute value $|x|$ that determines the maximal risk taken in order to help (if $x > 0$) or harm a relative.

We now return to Eq. (2.2) and find, for each y, a strategy $x^* = x^*(y)$ that will maximize the inclusive fitness of A.

By differentiating (2.2) with respect to x and by applying (3.1), we have

$$\frac{\partial \Omega(x, y)}{\partial x} = \frac{-(\partial/\partial x)\,\lambda(y, x)}{\lambda[\lambda(x, y) + \lambda(y, x)]^2}\{(1 + r)(1 - x)\lambda$$

$$- (1 + \sigma)(1 - r)[\lambda(x, y) + x\lambda(y, x)]\}. \tag{3.2}$$

We shall denote the expression in braces by $B_y(x)$. Because

$$\frac{-(\partial/\partial x)\,\lambda(y,x)}{\lambda[\lambda(x,y)+\lambda(y,x)]^2} \geqslant 0$$

for every x and y, the sign of $\partial\Omega(x,y)/\partial x$ is identical to the sign of $B_y(x)$.

PROPOSITION 2. *For every y between -1 and 1, the equation $B_y(x) = 0$ has a unique solution $x^*(y)$ with $-1 < x^*(y) \leqslant 1$.*

Proof. By applying Proposition 1 we have

$$\frac{dB_y(x)}{dx} = -(1+r)\lambda + (1+\sigma)(1-r)\left[\frac{\partial}{\partial x}\lambda(x,y) + \lambda(y,x) + x\frac{\partial}{\partial x}\lambda(y,x)\right]$$

$$= -(1+r)\lambda - (1+\sigma)(1-r)\,\lambda(y,x) < 0 \qquad\qquad (3.3)$$

and the equation $B_y(x) = 0$ has at most one solution. But, applying Proposition 1 again, we also have

$$\frac{d}{dy}\,[\lambda(-1,y) - \lambda(y,-1)] = (1+y)\frac{d}{dy}\lambda(-1,y) \leqslant 0,$$

therefore

$$
\begin{aligned}
B_y(-1) &= 2(1+r)\lambda - (1+\sigma)(1-r)[\lambda(-1,y) - \lambda(y,-1)] \\
&\geqslant 2(1+r)\lambda - (1+\sigma)(1-r)[\lambda(-1,-1) - \lambda(-1,-1)] \\
&= 2(1+r)\lambda > 0,
\end{aligned}
$$

whereas

$$B_y(1) = -(1+\sigma)(1-r)[\lambda(1,y) + \lambda(y,1)] \leqslant 0$$

and the equation $B_y(x) = 0$ has exactly one solution at $(-1, 1]$.

PROPOSITION 3. *For every y, $\Omega(x,y)$ has a unique global maximum in x at the point $x = x^*(y)$, and this maximum is strict.*

Proof. For $x < x^*$, $B_y(x) > 0$ and so $(\partial/\partial x)\,\Omega(x,y) > 0$, i.e., $\Omega(x,y)$ is increasing in x for $x < x^*$. In the same way we show that $\Omega(x,y)$ is decreasing in x for $x > x^*$. ∎

Next we shall find the evolutionarily stable strategies (ESS) of mutual help or harm (and, indeed, show their existence). Using the definition of Maynard Smith and Price (1973), a strategy is an ESS if a population of individuals adopting that strategy cannot be "invaded" by an initially rare mutant

adopting an alternate strategy. (See also Maynard Smith, 1974 and 1976; Maynard Smith and Parker, 1976; Bishop and Cannings, 1976.) Equivalently, $\hat{x}$ is an ESS if $\Omega(\hat{x}, \hat{x}) > \Omega(x, \hat{x})$ for $x \neq \hat{x}$, or if there exists a strategy such that $\Omega(\hat{x}, \hat{x}) = \Omega(x, \hat{x})$, then $\Omega(\hat{x}, x) > \Omega(x, x)$.

PROPOSITION 4. *For all $0 \leqslant r \leqslant 1$, $x^*(y)$ is a non-decreasing function of y on the interval $[-1, 1]$ (it is a constant 1 for $r = 1$), i.e., the higher the MR readiness y of B to help A, the higher is the optimal MR readiness of A to help B.*

Proof. Employing again Proposition 1 we get:

$$\frac{\partial}{\partial y} B_y(x) = -(1 + \sigma)(1 - r) \left[\frac{\partial}{\partial y} \lambda(x, y) + x \frac{\partial}{\partial y} \lambda(y, x) \right]$$

$$= -\frac{\partial}{\partial y} \lambda(x, y)(1 + \sigma)(1 - r)(1 - xy) \geqslant 0$$

for all $-1 \leqslant x \leqslant 1$. From (3.3) we also know that $(\partial/\partial x) B_y(x) < 0$. But $x^*(y)$ is the unique solution of $B_y(x) = 0$, thus Proposition 4 follows from the implicit function theorem. ∎
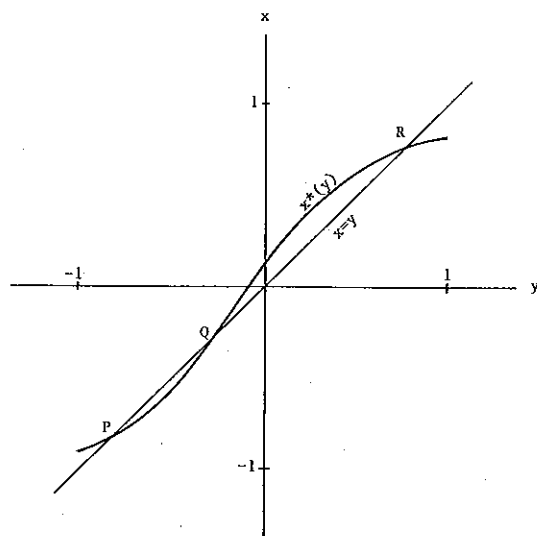

FIG. 1. The function $x^*(y)$, $-1 \leqslant y \leqslant 1$. The intersection points of $x^*(y)$ with the line $x = y$ represent ESSs. The points P and R are CSSs, whereas Q is an ESS which is not continuously stable. For each y, $\Omega(x, y)$ (the inclusive fitness of A) is a decreasing function of x for $x > x^*(y)$ and an increasing function of x for $x < x^*(y)$.

PROPOSITION 5.   *The curve* $x = x^*(y)$ *intersects the main diagonal* $x = y$ *from above at least once on the interval* $-1 \leqslant y \leqslant 1$.

(See Fig. 1.)

Proof. $B_{-1}(-1) = 2(1 + r)\lambda > 0$ and   therefore $x^*(-1) > -1$. $B_1(1) = -2(1 + \sigma)(1 - r)\lambda(1, 1) \leqslant 0$, with a strict inequality for $r < 1$, therefore $x^*(1) \leqslant 1$, with a strict inequality if $r < 1$.   ∎

As we see (Appendix), there might be more than one intersection of the curve $x = x^*(y)$ and the main diagonal.

PROPOSITION 6.   *The intersections of the curve* $x = x^*(y)$ *and the main diagonal are the ESSs for the maximal risk.*

Proof.   Let $(\hat{x}, \hat{x})$ be such an intersection point. Then $\hat{x} = x^*(\hat{x})$ and Proposition 3 states that for all $x \neq \hat{x}$, $\Omega(\hat{x}, \hat{x}) > \Omega(x, \hat{x})$. Thus $\hat{x}$ is an ESS.
On the other hand, let $\hat{x}$ be an ESS. Then $\Omega(\hat{x}, \hat{x}) > \Omega(x, \hat{x})$ and it follows again from Proposition 3 that $\hat{x} = x^*(\hat{x})$.   ∎

COROLLARY.   *The ESSs of the model are the solution of the equation* $B_x(x) = 0$, *i.e.,*

$$\phi(x) = (1 + r)(1 - x)\lambda - (1 + \sigma)(1 - r)(1 + x)\lambda(x, x) = 0. \qquad (3.4)$$

4. WEAK AND STRONG ESS

The ESSs of the model, being the intersections of the curve $x = x^*(y)$ and the main diagonal, can be divided into two kinds:

(i)   points where $x^*(y)$ intersects $x = y$ from above;

(ii)   points where $x^*(y)$ intersects $x = y$ from below or is tangential to it.

We see that the ESSs of the first kind represent a stronger stability than the stability represented by the ESSs of the second kind.

DEFINITION.   An ESS will be called a *continuously stable strategy* (CSS) if, whenever the entire population has a strategy which is close enough to it, there will be a selective advantage to some individual strategies which are closer to the CSS.

This definition is, indeed, meaningful only if there is a continuum of pure strategies.
Continuously stable strategies are the only class that represents a possible

dynamic selection process which eventually leads to the establishment of a CSS in the population. Note, however, that if for some historical reason a consensus on an ESS is established in a population, this ESS, even if not CSS, will be immune to invading mutant strategies.

PROPOSITION 7. *The CSSs of the model are only those ESSs in which the curve $x = x^*(y)$ intersects $x = y$ from above.*

Proof. Immediately follows from Proposition 4. For if $x^*(y)$ intersects $x = y$ from above at $x = \hat{x}$, then for $\varepsilon > 0$ sufficiently small and for $\hat{x} - \varepsilon < y < \hat{x}$, $y < x^*(y) < x^*(\hat{x}) = \hat{x}$. For $\hat{x} + \varepsilon > y > \hat{x}$, $y > x^*(y) > x^*(\hat{x}) = \hat{x}$. In both cases there is a selective advantage to $x^*(y)$ over y, and $x^*(y)$ is closer to $\hat{x}$ than y.

If, on the other hand, $x^*(y)$ intersects $x = y$ from below, then for any strategy y in the vicinity of the ESS $\hat{x}$ there is a preferable strategy $x^*(y)$ which is further from $\hat{x}$ than y. If $x^*(y)$ is tangential to $x = y$, then the latter statement holds for one side of the ESS.

Notes and Remarks

(1) In the case studied here of two equal relatives, there always exists an ESS which is continuously stable (an immediate result from Proposition 5).

(2) $\phi(0) = [2r - \sigma(1 - r)]\lambda$. Thus, if $\sigma < 2r/(1 - r)$ then $\phi(0) > 0$. This implies that the indifferent behavior $x = 0$ is in the domain of attraction of some altruistic CSS $\hat{x} > 0$. If $\sigma > 2r/(1 - r)$ then $x = 0$ is in the domain of attraction of a spitefulness CSS $\hat{x} < 0$. (If $\sigma = 2r/(1 - r)$ then $x = 0$ is an ESS.)

(3) The simplest case of competition occurs when the resources of the population are limited and equally shared by the offspring of all surviving individuals. In this case, $\sigma = 1/(n - 1)$, where n is the population size at a given time, and the sufficient condition for the existence of a CSS of altruistic behavior toward a relative of relatedness r turns out to be

$$(2n - 1)r > 1.$$

(4) If there is no competition ($\sigma = 0$), then (provided $r > 0$) there is a CSS which is greater than zero. In this case, for $x = r$ we have $\phi(r) = (1 - r^2)[\lambda - \lambda(r, r)]$. But $\lambda(x, x)$ is a decreasing function of x, since (using (3.1))

$$\frac{d}{dx}\lambda(x, x) = (1 - x)\frac{\partial}{\partial y}\lambda(x, y)\bigg|_{x=y} < 0. \qquad (4.1)$$

Thus $\lambda = \lambda(0, 0)$ is greater than $\lambda(r, r)$ and so $\phi(x = r) > 0$. This implies the existence of a CSS which is greater than r.

(5) If $r = 0$ (but $\sigma > 0$), i.e., the two individuals are not related at all, then $\phi(0) < 0$. Hence there is a CSS which has a negative value.

(6) ϕ is a decreasing function of σ, implying a decrease in the values of the continuously stable strategies with the increase in the degree of competition. Also, ϕ is an increasing function of r, thus the values of the CSSs increase with the increase of the degree of relatedness. These statements, which are also intuitively logical, are not valid for the ESSs of the second kind. On the contrary, the values of such strategies (whenever they exist) increase with the increase in the degree of competition and decrease with increasing relatedness, in contradiction to common sense.

(7) If both individuals have the same strategy, the inclusive fitness of each is

$$\Omega(x, x) = \frac{1+r}{2} \left(\frac{1}{\lambda(x, x)} + \frac{1+\sigma}{\lambda} \right). \tag{4.2}$$

Since $\lambda(x, x)$ is a decreasing function of x, it follows that $\Omega(x, x)$ is maximal if $x = 1$. But, unless $r = 1$, this point is unstable—for values of x close enough to 1, $\Omega(x, 1) > \Omega(1, 1)$. Subsequently the population fixes on a CSS which has a value smaller than 1. Thus we see that the possibility of exploitation by the relative leads to the establishment of a CSS for which the inclusive fitness does not attain its maximal value.

(8) An example showing the possible existence of several stable strategies is presented in the Appendix.

5. An Individual Selection Example

In the following example we shall see that direct Fisherian selection will ultimately bring the population to the same degree of mutual help as predicted by the game theory model of the previous sections. (For comparison, see Eshel, 1981.)

We assume that the maximal risk is genetically determined by a single locus with two alleles: the dominant allele A implies a MR of the amount x, and y is the amount implied by the recessive allele a. The allele A is rare, and the relative frequency in the population of the heterozygote type is ε. Assuming random mating, it is easy to see that the relative frequency of AA is $o(\varepsilon)$. For simplicity, we shall further assume that in each brood there are always two offspring. (There can be more than one brood for an individual, but only brothers of the same brood can recognize each other.) We shall

assume as before (Section 2) that the Fisherian fitness of an individual whose MR is x and that of its brother is y, is

$$\omega(x, y) = \frac{1 + ((1 + \sigma)/\lambda)\, \lambda(y, x)}{\lambda(x, y) + \lambda(y, x)}.$$

Mating type	Frequency	Progeny		
		Both Aa	One Aa, one aa	Both aa
$Aa \times aa$	$2\varepsilon + o(\varepsilon)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
$aa \times aa$	$1 - 2\varepsilon + o(\varepsilon)$	—	—	1
other than above	$o(\varepsilon)$			

Hence, the relative frequency of Aa in the next generation is

$$\frac{\varepsilon[\omega(x, x) + \omega(x, y)]}{2\omega(y, y)} + o(\varepsilon)$$

and natural selection will favor the rare allele A if and only if

$$h(x, y) \equiv \omega(x, x) + \omega(x, y) - 2\omega(y, y) > 0$$

We concentrate on small changes in the MR, i.e., on values of x which are close enough to y. For $x = y$, $h(y, y) = 0$ while:

$$\frac{\partial h(x, y)}{\partial x}\bigg|_{x=y} = \frac{\partial \omega(x, x)}{\partial x} + \frac{\partial \omega(x, y)}{\partial x}\bigg|_{x=y},$$

and by substituting

$$\omega(x, y) = \frac{1 + ((1 + \sigma)/\lambda)\, \lambda(y, x)}{\lambda(x, y) + \lambda(y, x)}$$

we obtain

$$\frac{\partial h(x, y)}{\partial x}\bigg|_{x=y} = \frac{-(\partial \lambda(y, x)/\partial x)}{2\lambda^2(y, y)\lambda}\bigg|_{x=y} \{\tfrac{3}{2}(1 - y)\lambda - \tfrac{1}{2}(1 + \sigma)(1 + y)\, \lambda(y, y)\}$$

$$= -\frac{1}{2\lambda^2(y, y)\lambda}\frac{\partial \lambda(y, x)}{\partial x}\bigg|_{x=y} \phi_{r=1/2}(y),$$

where $\phi(y) = \phi_r(y)$ is defined in (3.4) above. But $\partial\lambda(y, x)/\partial x < 0$ and we have

$$\text{sign}\left\{\frac{\partial h(x, y)}{\partial x}\bigg|_{x=y}\right\} = \text{sign}\{\phi_{1/2}(y)\}.$$

Hence, if $\phi_{1/2}(y) > 0$, then there exists an interval around y in which $h(x, y)$ increases as a function of x. Thus, selection will act, at least at the outset, in favor of an allele which increases, in a limited amount, the tendency of its possessor to help its brother and against an allele which decreases such a tendency, and vice versa if $\phi_{r=1/2}(y) < 0$.

Hence, natural selection, by means of small changes in MR, tends to establish in the population a degree of mutual help (in this case between brothers) which is equal to the value at which $\phi(x)$ intersects the x axis from above, i.e., individual selection, in our example, will establish a degree of mutual help which is one of the CSSs obtained earlier by considerations of maximization of the inclusive fitness.

6. DISCUSSION

In this paper we have considered the problem of risk taking by an individual in order to save its relative. The combined effect of relatedness, competition and mutual dependence between individuals in the same population has been incorporated in a two-player game model, with the inclusive fitness as the payoff function.

We have found that for each strategy of maximal risk (MR) adopted by one of the players, a single MR strategy exists for the other player which maximizes the latter's inclusive fitness. Also, some of the admissible strategies are evolutionarily stable, i.e., if a sufficiently large proportion of the population adopts it, there is no "mutant" strategy that would yield a larger inclusive fitness.

Moreover, in regard to stability, the ESSs in our model can be classified into two categories—some of the ESSs (the ESSs of the first kind) possess a stronger form of stability than is exhibited by the formal definition of an ESS. Thus, by introducing the notion of continuous stability of strategies, (i.e., an ESS is continuously stable (CSS) if, whenever the entire population has a strategy which is not an ESS, there are strategies closer to the CSS which endow any individual adopting them with a selective advantage over the entire population), we have shown that the stable strategies of the first kind (and only those) are CSSs.

Except for singular cases (in which the curve $\phi(x)$ is tangent to the x axis), all the ESSs of our model exhibit the principle of small perturbations (Karlin

and McGregor, 1972), i.e., sufficiently small displacements in the parameters involved will not eliminate the ESS, but rather slightly change its position. However, only the CSSs are stable according to Samuelson's correspondence principle (see Samuelson, 1947). That is to say, if the population had maintained an ESS and if, in the course of time, the values of the parameters have been slightly changed (thus causing a small displacement of the ESS), evolution will tend to restore the ESS only in the case of a CSS. If the population had maintained an ESS of the second kind (a stable strategy which is not continuous), the population will subsequently be in the convergence region of another, continuous, stable strategy (and not of the near, new located ESS).

APPENDIX: An Example Showing the Possible Existence of ESSs Which Are Not CSS

Let us assume

$$\lambda(x, y) = \frac{\lambda}{k} [1 - (1 + kx)e^{-kx} + ke^{-ky}].$$

It is easy to see that this function fulfills the requirements for $\lambda(x, y)$. Let us further assume $\sigma > 2r/(1 - r)$. Thus, $\phi(0) < 0$ and there exists a CSS which is smaller than zero.

For any x, $0 < x < 1$, $\lim_{k \to \infty} \lambda(x, x) = 0$. Hence $\lim_{k \to \infty} \phi(x) = (1 + r)(1 - x)\lambda > 0$. Thus, for a large enough value of k, there exist two more solutions to the equation $\phi(x) = 0$ (which are both positive). The larger of these solutions is continuously stable, while the smaller is an ESS which is not continuous.

Remark. The possible existence of more than one Nash solution to an inclusive fitness game with mutual help has already been noticed by Eshel and Cohen (1976). However, the problem of stability was not studied there. For the case of mutual help between relatives with unequal fertility, the reader is referred to Motro, 1981.

ACKNOWLEDGMENT

REFERENCES

BISHOP, D. T. AND CANNINGS, C. 1976. Models of animal conflict, *Advances in Appl. Prob.* 8, 616–621.

BOORMAN, S. A. AND LEVITT, P. R. 1980. "The Genetics of Altruism," Academic Press, New York.

CAVALLI-SFORZA, L. L. AND FELDMAN, M. W. 1978. Darwinian selection and altruism, *Theor. Pop. Biol.* 14, 268–280.

COHEN, D. AND ESHEL, I. 1976. On the founder effect and the evolution of altruistic traits. *Theor. Pop. Biol.* 10, 276–302.

ESHEL, I. 1981. Evolutionarily stable strategies and natural selection in Mendelian populations. *Theor. Pop. Biol.* In press.

ESHEL, I. AND COHEN, D. 1976. Altruism, competition and kin selection in populations, *in* "Population Genetics and Ecology" (S. Karlin and E. Nevo, Eds.), Academic Press, New York.

HAMILTON, W. D. 1964. The genetical evolution of social behavior, *J. Theor. Biol.* 7, I, 1–16, and II, 17–52.

HAMILTON, W. D. 1970. Selfish and spiteful behavior in an evolutionary model, *Nature* 228, 1218–1220.

HARPENDING, H. C. 1979. The population genetics of interactions, *Amer. Natur.* 113, 622–630.

KARLIN, S. AND McGREGOR, J. 1972. Polymorphisms for genetic and ecological systems with weak coupling, *Theor. Pop. Biol.* 3, 210–238.

MAYNARD SMITH, J. 1974. The theory of games and the evolution of animal conflicts, *J. Theor. Biol.* 47, 209–221.

MAYNARD SMITH, J. 1976. Evolution and the theory of games, *Amer. Sci.* 64, 41–45.

MAYNARD SMITH, J. AND PARKER, G. A. 1976. The logic of asymmetric contests, *Anim. Behav.* 24, 159–175.

MAYNARD SMITH, J. AND PRICE, G. R. 1973. The logic of animal conflict, *Nature* 246, 15–18.

MOTRO, U. 1981. Evolutionarily stable strategies of mutual help between relatives with unequal fertility, to appear.

SAMUELSON, P. A. 1947. "Foundations of Economic Analysis," Harvard Univ. Press, Cambridge, Mass.

UYENOYAMA, M. AND FELDMAN, M. W. 1980. On relatedness and adaptive topography in kin selection, to appear.

WRIGHT, S. 1922. Coefficients of inbreeding and relationship, *Amer. Natur.* 56, 330–338.

ZAHAVI, A. 1981. Competitive altruism and its implication to the theory of kin selection, to appear.