

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 $0 \leq x \leq 1$ 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 $x \leq r$, 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

* This work has been supported in part by NIH Grants (1 RO1 HD 12731-01), 10452 and 20816.

[†] Present address: Department of Mathematics, Stanford University, Stanford, CA 94305.

[‡] Present address: Department of Genetics, University of California, Berkeley, CA 94720.

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 , $0 \leq r \leq 1$.

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 $0 \leq \sigma \leq 1$.

With these assumptions, the fitness of individual A is

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

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

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

(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:*

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

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 $x \geq 0$ 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

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

(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 $(\partial/\partial x)\lambda(y, x) \leq 0$ for all values of x and from Proposition 1 it follows that

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

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

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

$$\begin{aligned} \frac{\partial \Omega(x, y)}{\partial x} = & \frac{-(\partial/\partial x)\lambda(y, x)}{\lambda[\lambda(x, y) + \lambda(y, x)]^2} \{(1+r)(1-x)\lambda \\ & - (1+\sigma)(1-r)[\lambda(x, y) + x\lambda(y, x)]\}. \end{aligned} \quad (3.2)$$

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

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

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 $-1 < x^*(y) \leq 1$.*

Proof. By applying Proposition 1 we have

$$\begin{aligned} \frac{dB_y(x)}{dx} &= -(1+r)\lambda + (1+\sigma)(1-r) \left[\frac{\partial}{\partial x} \lambda(x, y) + \lambda(y, x) + x \frac{\partial}{\partial x} \lambda(y, x) \right] \\ &= -(1+r)\lambda - (1+\sigma)(1-r) \lambda(y, x) < 0 \end{aligned} \quad (3.3)$$

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

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

therefore

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

whereas

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

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 $0 \leq r \leq 1$, $x^*(y)$ 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:

$$\begin{aligned} \frac{\partial}{\partial y} B_y(x) &= -(1 + \sigma)(1 - r) \left[\frac{\partial}{\partial y} \lambda(x, y) + x \frac{\partial}{\partial y} \lambda(y, x) \right] \\ &= -\frac{\partial}{\partial y} \lambda(x, y)(1 + \sigma)(1 - r)(1 - xy) \geq 0 \end{aligned}$$

for all $-1 \leq x \leq 1$. 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. ■

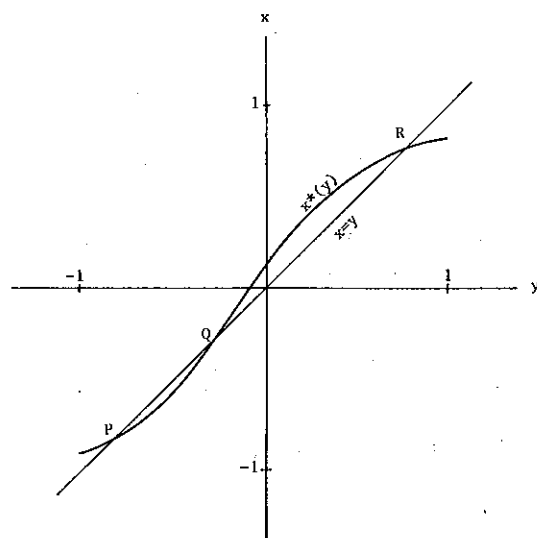


FIG. 1. The function $x^*(y)$, $-1 \leq y \leq 1$. The intersection points of $x^*(y)$ with the line $x = y$ represent ESSs. The points P and R are CSSs, whereas Q is an ESS which is not continuously stable. For each y , $\Omega(x, y)$ (the inclusive fitness of A) is a decreasing function of x for $x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

(See Fig. 1.)

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

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

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))

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

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

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

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:

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

and by substituting

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

we obtain

$$\begin{aligned} \left. \frac{\partial h(x, y)}{\partial x} \right|_{x=y} &= \left. \frac{-(\partial \lambda(y, x)/\partial x)}{2\lambda^2(y, y)\lambda} \right|_{x=y} \{ \tfrac{3}{2}(1 - y)\lambda - \tfrac{1}{2}(1 + \sigma)(1 + y)\lambda(y, y) \} \\ &= -\frac{1}{2\lambda^2(y, y)\lambda} \left. \frac{\partial \lambda(y, x)}{\partial x} \right|_{x=y} \phi_{r=1/2}(y), \end{aligned}$$

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

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

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 $\phi_{1/2}(y) < 0$.

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 , $0 < x < 1$, $\lim_{k \rightarrow \infty} \lambda(x, x) = 0$. Hence $\lim_{k \rightarrow \infty} \phi(x) = (1+r)(1-x)\lambda > 0$. 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

We wish to thank Professor Reinhardt Selten for acquainting us with the correspondence principle of Samuelson.

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.

