J. Theoret. Biol. (1964) 7, 1–16

The Genetical Evolution of Social Behaviour. I

W. D. HAMILTON

The Galton Laboratory, University College, London, W.C.2

(*Received* 13 *May* 1963, *and in revised form* 24 *February* 1964)

A genetical mathematical model is described which allows for interactions between relatives on one another's fitness. Making use of Wright's Coefficient of Relationship as the measure of the proportion of replica genes in a relative, a quantity is found which incorporates the maximizing property of Darwinian fitness. This quantity is named "inclusive fitness". Species following the model should tend to evolve behaviour such that each organism appears to be attempting to maximize its inclusive fitness. This implies a limited restraint on selfish competitive behaviour and possibility of limited self-sacrifices.

Special cases of the model are used to show (a) that selection in the social situations newly covered tends to be slower than classical selection, (b) how in populations of rather non-dispersive organisms the model may apply to genes affecting dispersion, and (c) how it may apply approximately to competition between relatives, for example, within sibships. Some artificialities of the model are discussed.

1. Introduction

With very few exceptions, the only parts of the theory of natural selection which have been supported by mathematical models admit no possibility of the evolution of any characters which are on average to the disadvantage of the individuals possessing them. If natural selection followed the classical models exclusively, species would not show any behaviour more positively social than the coming together of the sexes and parental care.

Sacrifices involved in parental care are a possibility implicit in any model in which the definition of fitness is based, as it should be, on the number of adult offspring. In certain circumstances an individual may leave more adult offspring by expending care and materials on its offspring already born than by reserving them for its own survival and further fecundity. A gene causing its possessor to give parental care will then leave more replica genes in the next generation than an allele having the opposite tendency. The selective advantage may be seen to lie through benefits conferred indifferently on a set of relatives each of which has a half chance of carrying the gene in question.

From this point of view it is also seen, however, that there is nothing special about the parent-offspring relationship except its close degree and a certain fundamental asymmetry. The full-sib relationship is just as close. If an individual carries a certain gene the expectation that a random sib will carry a replica of it is again one-half. Similarly, the half-sib relationship is equivalent to that of grandparent and grandchild with the expectation of replica genes, or genes "identical by descent" as they are usually called, standing at one quarter; and so on.

Although it does not seem to have received very detailed attention the possibility of the evolution of characters benefitting descendants more remote than immediate offspring has often been noticed. Opportunities for benefitting relatives, remote or not, in the same or an adjacent generation (i.e. relatives like cousins and nephews) must be much more common than opportunities for benefitting grandchildren and further descendants. As a first step towards a general theory that would take into account all kinds of relatives this paper will describe a model which is particularly adapted to deal with interactions between relatives of the same generation. The model includes the classical model for "non-overlapping generations" as a special case. An excellent summary of the general properties of this classical model has been given by Kingman (1961b). It is quite beyond the author's power to give an equally extensive survey of the properties of the present model but certain approximate deterministic implications of biological interest will be pointed out.

As is already evident the essential idea which the model is going to use is quite simple. Thus although the following account is necessarily somewhat mathematical it is not surprising that eventually, allowing certain lapses from mathematical rigour, we are able to arrive at approximate principles which can also be expressed quite simply and in non-mathematical form. The most important principle, as it arises directly from the model, is outlined in the last section of this paper, but a fuller discussion together with some attempt to evaluate the theory as a whole in the light of biological evidence will be given in the sequel.

2. The Model

The model is restricted to the case of an organism which reproduces once and for all at the end of a fixed period. Survivorship and reproduction can both vary but it is only the consequent variations in their product, net reproduction, that are of concern here. All genotypic effects are conceived as increments and decrements to a basic unit of reproduction which, if possessed by all the individuals alike, would render the population both stationary and non-evolutionary. Thus the fitness $a^{\bullet}$ of an individual is treated as the sum

of his basic unit, the effect δa of his personal genotype and the total e° of effects on him due to his neighbours which will depend on their genotypes:

$$a^\bullet = 1+\delta a+e^\circ. \tag{1}$$

The index symbol $^\bullet$ in contrast to $^\circ$ will be used consistently to denote the inclusion of the personal effect δa in the aggregate in question. Thus equation (1) could be rewritten

$$a^\bullet = 1+e^\bullet.$$

In equation (1), however, the symbol $^\bullet$ also serves to distinguish this neighbour modulated kind of fitness from the part of it

$$a = 1+\delta a$$

which is equivalent to fitness in the classical sense of individual fitness.

The symbol δ preceding a letter will be used to indicate an effect or total of effects due to an individual treated as an addition to the basic unit, as typified in

$$a = 1+\delta a.$$

The neighbours of an individual are considered to be affected differently according to their relationship with him.

Genetically two related persons differ from two unrelated members of the population in their tendency to carry replica genes which they have both inherited from the one or more ancestors they have in common. If we consider an autosomal locus, not subject to selection, in relative B with respect to the same locus in the other relative A, it is apparent that there are just three possible conditions of this locus in B, namely that both, one only, or neither of his genes are identical by descent with genes in A. We denote the respective probabilities of these conditions by c_2, c_1 and c_0. They are independent of the locus considered; and since

$$c_2+c_1+c_0 = 1,$$

the relationship is completely specified by giving any two of them. Li & Sacks (1954) have described methods of calculating these probabilities adequate for any relationship that does not involve inbreeding. The mean number of genes per locus i.b.d. (as from now on we abbreviate the phrase "identical by descent") with genes at the same locus in A for a hypothetical population of relatives like B is clearly $2c_2+c_1$. One half of this number, $c_2+\frac{1}{2}c_1$, may therefore be called the expected fraction of genes i.b.d. in a relative. It can be shown that it is equal to Sewall Wright's Coefficient of Relationship r (in a non-inbred population). The standard methods of calculating r without obtaining the complete distribution can be found in Kempthorne (1957). Tables of

$$f = \tfrac{1}{2}r = \tfrac{1}{2}(c_2+\tfrac{1}{2}c_1) \quad \text{and} \quad F = c_2$$

for a large class of relationships can be found in Haldane & Jayakar (1962).

Strictly, a more complicated metric of relationship taking into account the parameters of selection is necessary for a locus undergoing selection, but the following account based on use of the above coefficients must give a good approximation to the truth when selection is slow and may be hoped to give some guidance even when it is not.

Consider now how the effects which an arbitrary individual distributes to the population can be summarized. For convenience and generality we will include at this stage certain effects (such as effects on parents' fitness) which must be zero under the restrictions of this particular model, and also others (such as effects on offspring) which although not necessarily zero we will not attempt to treat accurately in the subsequent analysis.

The effect of A on specified B can be a variate. In the present deterministic treatment, however, we are concerned only with the means of such variates. Thus the effect which we may write $(\delta a_{\text{father}})_{\text{A}}$ is really the expectation of the effect of A upon his father but for brevity we will refer to it as the effect on the father.

The full array of effects like $(\delta a_{\text{father}})_{\text{A}}$, $(\delta a_{\text{specified sister}})_{\text{A}}$, etc., we will denote

$$\{\delta a_{\text{rel.}}\}_{\text{A}}.$$

From this array we can construct the simpler array

$$\{\delta a_{r,\,c_2}\}_{\text{A}}$$

by adding together all effects to relatives who have the same values for the pair of coefficients (r, c_2). For example, the combined effect $\delta a_{\frac{1}{4},0}$ might contain effects actually occurring to grandparents, grandchildren, uncles, nephews and half-brothers. From what has been said above it is clear that as regards changes in autosomal gene-frequency by natural selection all the consequences of the full array are implied by this reduced array—at least, provided we ignore (a) the effect of previous generations of selection on the expected constitution of relatives, and (b) the one or more generations that must really occur before effects to children, nephews, grandchildren, etc., are manifested.

From this array we can construct a yet simpler array, or vector,

$$\{\delta a_r\}_{\text{A}},$$

by adding together all effects with common r. Thus $\delta a_{\frac{1}{4}}$ would bring together effects to the above-mentioned set of relatives and effects to double-first cousins, for whom the pair of coefficients is $(\frac{1}{4}, \frac{1}{16})$.

Corresponding to the effect which A causes to B there will be an effect of similar type on A. This will either come from B himself or from a person who stands to A in the same relationship as A stands to B. Thus corresponding to

an effect by A on his nephew there will be an effect on A by his uncle. The similarity between the effect which A dispenses and that which he receives is clearly an aspect of the problem of the correlation between relatives. Thus the term e° in equation (1) is not a constant for any given genotype of A since it will depend on the genotypes of neighbours and therefore on the gene-frequencies and the mating system.

Consider a single locus. Let the series of allelomorphs be $G_1, G_2, G_3, \ldots, G_n$, and their gene-frequencies $p_1, p_2, p_3, \ldots, p_n$. With the genotype G_iG_j associate the array $\{\delta a_{\text{rel.}}\}_{ij}$; within the limits of the above-mentioned approximations natural selection in the model is then defined.

If we were to follow the usual approach to the formulation of the progress due to natural selection in a generation, we should attempt to give formulae for the neighbour modulated fitnesses $a^\bullet_{ij}$. In order to formulate the expectation of that element of e°_{ij} which was due to the return effect of a relative B we would need to know the distribution of possible genotypes of B, and to obtain this we must use the double measure of B's relationship and the gene-frequencies just as in the problem of the correlation between relatives. Thus the formula for e°_{ij} will involve all the arrays $\{\delta a_{r,c_2}\}_{ij}$ and will be rather unwieldy (see Section 4).

An alternative approach, however, shows that the arrays $\{\delta a_r\}_{ij}$ are sufficient to define the selective effects. Every effect on reproduction which is due to A can be thought of as made up of two parts: an effect on the reproduction of genes i.b.d. with genes in A, and an effect on the reproduction of unrelated genes. Since the coefficient r measures the expected fraction of genes i.b.d. in a relative, for any particular degree of relationship this breakdown may be written quantitatively:

$$(\delta a_{\text{rel.}})_A = r(\delta a_{\text{rel.}})_A + (1-r)(\delta a_{\text{rel.}})_A.$$

The total of effects on reproduction which are due to A may be treated similarly:

$$\sum_{\text{rel.}} (\delta a_{\text{rel.}})_A = \sum_{\text{rel.}} r(\delta a_{\text{rel.}})_A + \sum_{\text{rel.}} (1-r)(\delta a_{\text{rel.}})_A,$$

or

$$\sum_{r} (\delta a_r)_A = \sum_{r} r(\delta a_r)_A + \sum_{r} (1-r)(\delta a_r)_A,$$

which we rewrite briefly as

$$\delta T^\bullet_A = \delta R^\bullet_A + \delta S_A,$$

where $\delta R^\bullet_A$ is accordingly the total effect on genes i.b.d. in relatives of A, and δS_A is the total effect on their other genes. The reason for the omission of an index symbol from the last term is that here there is, in effect, no question of whether or not the self-effect is to be in the summation, for if it is included it has to be multiplied by zero. If index symbols were used

we should have $\delta S^{\bullet}_{A} = \delta S^{\circ}_{A}$, whatever the subscript; it therefore seems more explicit to omit them throughout.

If, therefore, all effects are accounted to the individuals that cause them, of the total effect $\delta T^{\bullet}_{ij}$ due to an individual of genotype G_iG_j a part $\delta R^{\bullet}_{ij}$ will involve a specific contribution to the gene-pool by this genotype, while the remaining part δS_{ij} will involve an unspecific contribution consisting of genes in the ratio in which the gene-pool already possesses them. It is clear that it is the matrix of effects $\delta R^{\bullet}_{ij}$ which determines the direction of selection progress in gene-frequencies; δS_{ij} only influences its magnitude. In view of this importance of the $\delta R^{\bullet}_{ij}$ it is convenient to give some name to the concept with which they are associated.

In accordance with our convention let

$$R^{\bullet}_{ij} = 1 + \delta R^{\bullet}_{ij};$$

then $R^{\bullet}_{ij}$ will be called the *inclusive fitness*, $\delta R^{\bullet}_{ij}$ the *inclusive fitness effect* and δS_{ij} the *diluting effect*, of the genotype G_iG_j.

Let

$$T^{\bullet}_{ij} = 1 + \delta T^{\bullet}_{ij}.$$

So far our discussion is valid for non-random mating but from now on for simplicity we assume that it is random. Using a prime to distinguish the new gene-frequencies after one generation of selection we have

$$p'_i = \frac{\sum_j p_i p_j R^{\bullet}_{ij} + p_i \sum_{j,k} p_j p_k \delta S_{jk}}{\sum_{j,k} p_j p_k T^{\bullet}_{jk}} = p_i \frac{\sum_j p_j R^{\bullet}_{ij} + \sum_{j,k} p_j p_k \delta S_{jk}}{\sum_{j,k} p_j p_k T^{\bullet}_{jk}}.$$

The terms of this expression are clearly of the nature of averages over a part (genotypes containing G_i, homozygotes G_iG_i counted twice) and the whole of the existing set of genotypes in the population. Thus using a well known subscript notation we may rewrite the equation term by term as

$$p'_i = p_i \frac{R^{\bullet}_{i.} + \delta S_{..}}{T^{\bullet}_{..}}$$

$$\therefore \quad p'_i - p_i = \Delta p_i = \frac{p_i}{T^{\bullet}_{..}} (R^{\bullet}_{i.} + \delta S_{..} - T^{\bullet}_{..})$$

or

$$\Delta p_i = \frac{p_i}{R^{\bullet}_{..} + \delta S^{\bullet}_{..}} (R^{\bullet}_{i.} - R^{\bullet}_{..}). \tag{2}$$

This form clearly differentiates the roles of the $R^{\bullet}_{ij}$ and $\delta S^{\bullet}_{ij}$ in selective progress and shows the appropriateness of calling the latter diluting effects.

For comparison with the account of the classical case given by Moran (1962), equation (2) may be put in the form

$$\Delta p_i = \frac{p_i}{T^{\bullet}_{\cdot\cdot}}\left(\frac{1}{2}\frac{\partial R^{\bullet}_{\cdot\cdot}}{\partial p_i} - R^{\bullet}_{\cdot\cdot}\right)$$

where $\partial/\partial p_i$ denotes the usual partial derivative, written $\mathrm{d}/\mathrm{d}p_i$ by Moran.

Whether the selective effect is reckoned by means of the $a^{\bullet}_{ij}$ or according to the method above, the denominator expression must take in all effects occurring during the generation. Hence $a^{\bullet}_{\cdot\cdot} = T^{\bullet}_{\cdot\cdot}$.

As might be expected from the greater generality of the present model the extension of the theorem of the increase of mean fitness (Scheuer & Mandel, 1959; Mulholland & Smith, 1959; a much shorter proof by Kingman, 1961a) presents certain difficulties. However, from the above equations it is clear that the quantity that will tend to maximize, if any, is $R^{\bullet}_{\cdot\cdot}$, the mean inclusive fitness. The following brief discussion uses Kingman's approach.

The mean inclusive fitness in the succeeding generation is given by

$$R^{\bullet\prime}_{\cdot\cdot} = \sum_{i,j} p'_i p'_j R^{\bullet}_{ij} = \frac{1}{T^{\bullet 2}_{\cdot\cdot}} \sum_{i,j} p_i p_j R^{\bullet}_{ij}(R^{\bullet}_{i\cdot} + \delta S_{\cdot\cdot})(R^{\bullet}_{\cdot j} + \delta S_{\cdot\cdot}).$$

$$\therefore \quad R^{\bullet\prime}_{\cdot\cdot} - R^{\bullet}_{\cdot\cdot} = \Delta R^{\bullet}_{\cdot\cdot} = \frac{1}{T^{\bullet 2}_{\cdot\cdot}}\Big\{\sum_{i,j} p_i p_j R^{\bullet}_{ij} R^{\bullet}_{i\cdot} R^{\bullet}_{\cdot j} + 2\delta S_{\cdot\cdot} \sum_{i,j} p_i p_j R^{\bullet}_{ij} R^{\bullet}_{i\cdot} + \\ + R^{\bullet}_{\cdot\cdot}\delta S^2_{\cdot\cdot} - R^{\bullet}_{\cdot\cdot} T^{\bullet 2}_{\cdot\cdot}\Big\}.$$

Substituting $R^{\bullet}_{\cdot\cdot} + \delta S_{\cdot\cdot}$ for $T^{\bullet}_{\cdot\cdot}$ in the numerator expression, expanding and rearranging:

$$\Delta R^{\bullet} = \frac{1}{T^{\bullet 2}_{\cdot\cdot}}\Big\{\Big(\sum_{i,j} p_i p_j R^{\bullet}_{ij} R^{\bullet}_{i\cdot} R^{\bullet}_{\cdot j} - R^{\bullet 3}_{\cdot\cdot}\Big) + \\ + 2\delta S_{\cdot\cdot}\Big(\sum_{i,j} p_i p_j R^{\bullet}_{ij} R^{\bullet}_{i\cdot} - R^{\bullet 2}_{\cdot\cdot}\Big)\Big\}.$$

We have $(\ ) \geqslant 0$ in both cases. The first is the proven inequality of the classical model. The second follows from

$$\sum_{i,j} p_i p_j R^{\bullet}_{ij} R^{\bullet}_{i\cdot} = \sum_i p_i R^{\bullet 2}_{i\cdot} \geqslant \Big(\sum_i p_i R^{\bullet}_{i\cdot}\Big)^2 = R^{\bullet 2}_{\cdot\cdot}.$$

Thus a sufficient condition for $\Delta R^{\bullet}_{\cdot\cdot} \geqslant 0$ is $\delta S_{\cdot\cdot} \geqslant 0$. That $\Delta R^{\bullet}_{\cdot\cdot} \geqslant 0$ for positive dilution is almost obvious if we compare the actual selective changes with those which would occur if $\{R^{\bullet}_{ij}\}$ were the fitness matrix in the classical model.

It follows that $R^{\bullet}_{\cdot\cdot}$ certainly maximizes (in the sense of reaching a local maximum of $R^{\bullet}_{\cdot\cdot}$) if it never occurs in the course of selective changes that $\delta S_{\cdot\cdot} < 0$. Thus $R^{\bullet}_{\cdot\cdot}$ certainly maximizes if all $\delta S_{ij} \geqslant 0$ and therefore also if all $(\delta a_{\text{rel.}})_{ij} \geqslant 0$. It still does so even if some or all δa_{ij} are negative, for, as we have seen δS_{ij} is independant of δa_{ij}.

Here then we have discovered a quantity, inclusive fitness, which under the conditions of the model tends to maximize in much the same way that fitness tends to maximize in the simpler classical model. For an important class of genetic effects where the individual is supposed to dispense benefits to his neighbours, we have formally proved that the average inclusive fitness in the population will always increase. For cases where individuals may dispense harm to their neighbours we merely know, roughly speaking, that the change in gene frequency in each generation is aimed somewhere in the direction of a local maximum of average inclusive fitness,† but may, for all the present analysis has told us, overshoot it in such a way as to produce a lower value.

As to the nature of inclusive fitness it may perhaps help to clarify the notion if we now give a slightly different verbal presentation. Inclusive fitness may be imagined as the personal fitness which an individual actually expresses in its production of adult offspring as it becomes after it has been first stripped and then augmented in a certain way. It is stripped of all components which can be considered as due to the individual's social environment, leaving the fitness which he would express if not exposed to any of the harms or benefits of that environment. This quantity is then augmented by certain fractions of the quantities of harm and benefit which the individual himself causes to the fitnesses of his neighbours. The fractions in question are simply the coefficients of relationship appropriate to the neighbours whom he affects: unity for clonal individuals, one-half for sibs, one-quarter for half-sibs, one-eighth for cousins, ... and finally zero for all neighbours whose relationship can be considered negligibly small.

Actually, in the preceding mathematical account we were not concerned with the inclusive fitness of individuals as described here but rather with certain averages of them which we call the inclusive fitnesses of types. But the idea of the inclusive fitness of an individual is nevertheless a useful one. Just as in the sense of classical selection we may consider whether a given character expressed in an individual is adaptive in the sense of being in the interest of his personal fitness or not, so in the present sense of selection we may consider whether the character or trait of behaviour is or is not adaptive in the sense of being in the interest of his inclusive fitness.

3. Three Special Cases

Equation (2) may be written

$$\Delta p_i = p_i \frac{\delta R_{i\cdot}^{\bullet} - \delta R_{\cdot\cdot}^{\bullet}}{1 + \delta T_{\cdot\cdot}^{\bullet}}. \tag{3}$$

† That is, it is aimed "uphill": that it need not be at all directly towards the local maximum is well shown in the classical example illustrated by Mulholland & Smith (1959).

Now $\delta T^{\bullet}_{ij} = \sum_r (\delta a_r)_{ij}$ is the sum and $\delta R^{\bullet} = \sum_r r(\delta a_r)_{ij}$ is the first moment about $r = 0$ of the array of effects $\{\delta a_{\text{rel.}}\}_{ij}$ cause by the genotype G_iG_j; it appears that these two parameters are sufficient to fix the progress of the system under natural selection within our general approximation.

Let

$$r^{\bullet}_{ij} = \frac{\delta R^{\bullet}_{ij}}{\delta T^{\bullet}_{ij}}, \qquad (\delta T^{\bullet}_{ij} \neq 0); \tag{4}$$

and let

$$r^{\circ}_{ij} = \frac{\delta R^{\circ}_{ij}}{\delta T^{\circ}_{ij}}, \qquad (\delta T^{\circ}_{ij} \neq 0). \tag{5}$$

These quantities can be regarded as average relationships or as the first moments of reduced arrays, similar to the first moments of probability distributions.

We now consider three special cases which serve to bring out certain important features of selection in the model.

(a) The sums $\delta T^{\bullet}_{ij}$ differ between genotypes, the reduced first moment $r^{\bullet}$ being common to all. If all higher moments are equal between genotypes, that is, if all arrays are of the same "shape", this corresponds to the case where a stereotyped social action is performed with differing intensity or frequency according to genotype.

Whether or not this is so, we may, from equation (4), substitute $r^{\bullet}\delta T^{\bullet}_{ij}$ for $\delta R^{\bullet}_{ij}$ in equation (3) and have

$$\Delta p_i = p_i r^{\bullet} \frac{\delta T^{\bullet}_{i.} - \delta T^{\bullet}_{..}}{1 + \delta T^{\bullet}_{..}}.$$

Comparing this with the corresponding equation of the classical model,

$$\Delta p_i = p_i \frac{\delta a_{i.} - \delta a_{..}}{1 + \delta a_{..}}. \tag{6}$$

we see that placing genotypic effects on a relative of degree $r^{\bullet}$ instead of reserving them for personal fitness results in a slowing of selection progress according to the fractional factor $r^{\bullet}$.

If, for example, the advantages conferred by a "classical" gene to its carriers are such that the gene spreads at a certain rate the present result tells us that in exactly similar circumstances another gene which conferred similar advantages to the sibs of the carriers would progress at exactly half this rate.

In trying to imagine a realistic situation to fit this sort of case some concern may be felt about the occasions where through the probabilistic nature of things the gene-carrier happens not to have a sib, or not to have one suitably placed to receive the benefit. Such possibilities and their frequencies of reali-

zation must, however, all be taken into account as the effects $(\delta a_{\text{sibs}})_A$, etc., are being evaluated for the model, very much as if in a classical case allowance were being made for some degree of failure of penetrance of a gene.

(b) The reduced first moments $r_{ij}^{\bullet}$ differ between genotypes, the sum $\delta T^{\bullet}$ being common to all. From equation (4), substituting $r_{ij}^{\bullet}\delta T^{\bullet}$ for $\delta R_{ij}^{\bullet}$ in equation (3) we have

$$\Delta p_i = p_i \frac{\delta T^{\bullet}}{T^{\bullet}}(r_{i.}^{\bullet} - r_{..}^{\bullet}).$$

But it is more interesting to assume δa is also common to all genotypes. If so it follows that we can replace $^{\bullet}$ by $^{\circ}$ in the numerator expression of equation (3). Then, from equation (5), substituting $r_{ij}^{\circ}\delta T^{\circ}$ for δR_{ij}°, we have

$$\Delta p_i = p_i \frac{\delta T^{\circ}}{T^{\bullet}}(r_{i.}^{\circ} - r_{..}^{\circ}).$$

Hence, if a giving-trait is in question (δT° positive), genes which restrict giving to the nearest relative ($r_{i.}^{\circ}$ greatest) tend to be favoured; if a taking-trait (δT° negative), genes which cause taking from the most distant relatives tend to be favoured.

If all higher reduced moments about $r = r_{ij}^{\circ}$ are equal between genotypes it is implied that the genotype merely determines whereabouts in the field of relationship that centres on an individual a stereotyped array of effects is placed.

With many natural populations it must happen that an individual forms the centre of an actual local concentration of his relatives which is due to a general inability or disinclination of the organisms to move far from their places of birth. In such a population, which we may provisionally term "viscous", the present form of selection may apply fairly accurately to genes which affect vagrancy. It follows from the statements of the last paragraph but one that over a range of different species we would expect to find giving-traits commonest and most highly developed in the species with the most viscous populations whereas uninhibited competition should characterize species with the most freely mixing populations.

In the viscous population, however, the assumption of random mating is very unlikely to hold perfectly, so that these indications are of a rough qualitative nature only.

(c) $\delta T_{ij}^{\bullet} = 0$ for all genotypes.

$$\therefore \quad \delta T_{ij}^{\circ} = -\delta a_{ij}$$

for all genotypes, and from equation (5)

$$\delta R^{\circ}_{ij} = -\delta a_{ij} r^{\circ}_{ij}.$$

Then, from equation (3), we have

$$\begin{aligned}\Delta p_i = p_i(\delta R^{\bullet}_{i.} - \delta R^{\bullet}_{..}) &= p_i\{(\delta a_{i.} + \delta R^{\circ}_{i.}) - (\delta a_{..} + \delta R^{\circ}_{..})\} \\ &= p_i\{\delta a_{i.}(1 - r^{\circ}_{i.}) - \delta a_{..}(1 - r^{\circ}_{..})\}.\end{aligned}$$

Such cases may be described as involving transfers of reproductive potential. They are especially relevant to competition, in which the individual can be considered as endeavouring to transfer prerequisites of survival and reproduction from his competitors to himself. In particular, if $r^{\circ}_{ij} = r^{\circ}$ for all genotypes we have

$$\Delta p_i = p_i(1 - r^{\circ})(\delta a_{i.} - \delta a_{..}).$$

Comparing this to the corresponding equation of the classical model (equation (6)) we see that there is a reduction in the rate of progress when transfers are from a relative.

It is relevant to note that Haldane (1923) in his first paper on the mathematical theory of selection pointed out the special circumstances of competition in the cases of mammalian embryos in a single uterus and of seeds both while still being nourished by a single parent plant and after their germination if they were not very thoroughly dispersed. He gave a numerical example of competition between sibs showing that the progress of gene-frequency would be slower than normal.

In such situations as this, however, where the population may be considered as subdivided into more or less standard-sized batches each of which is allotted a local standard-sized pool of reproductive potential (which in Haldane's case would consist almost entirely of prerequisites for pre-adult survival), there is, in addition to a small correcting term which we mention in the short general discussion of competition in the next section, an extra overall slowing in selection progress. This may be thought of as due to the wasting of the powers of the more fit and the protection of the less fit when these types chance to occur positively assorted (beyond any mere effect of relationship) in a locality; its importance may be judged from the fact that it ranges from zero when the batches are indefinitely large to a halving of the rate of progress for competition in pairs.

4. Artificialities of the Model

When any of the effects is negative the restrictions laid upon the model hitherto do not preclude certain situations which are clearly impossible

from the biological point of view. It is clearly absurd if for any possible set of gene-frequencies any $a^{\bullet}_{ij}$ turns out negative; and even if the magnitude of δa_{ij} is sufficient to make $a^{\bullet}_{ij}$ positive while $1+e^{\circ}_{ij}$ is negative the situation is still highly artificial, since it implies the possibility of a sort of overdraft on the basic unit of an individual which has to be made good from his own takings. If we call this situation "improbable" we may specify two restrictions: a weaker, $e^{\circ}_{ij} > -1$, which precludes "improbable" situations; and a stronger, $e^{\bullet}_{ij} > -1$, which precludes even the impossible situations, both being required over the whole range of possible gene-frequencies as well as the whole range of genotypes.

As has been pointed out, a formula for $e^{\bullet}_{ij}$ can only be given if we have the arrays of effects according to a double coefficient of relationship. Choosing the double coefficient (c_2, c_1) such a formula is

$$e^{\bullet}_{ij} = \sum_{c_2, c_1}{}^{\bullet}[c_2 \operatorname{Dev}(\delta a_{c_2, c_1})_{ij} + \tfrac{1}{2}c_1\{\operatorname{Dev}(\delta a_{c_2, c_1})_{i.} + \operatorname{Dev}(\delta a_{c_2, c_1})_{.j}\}] + \delta T^{\circ}_{..}$$

where

$$\operatorname{Dev}(\delta a_{c_2, c_1})_{ij} = (\delta a_{c_2, c_1})_{ij} - (\delta a_{c_2, c_1})_{..} \quad \text{etc.}$$

Similarly

$$e^{\circ}_{ij} = \sum{}^{\circ}[''] + \delta T^{\circ}_{..},$$

the self-effect $(\delta a_{1,\,0})_{ij}$ being in this case omitted from the summations.

The following discussion is in terms of the stronger restriction but the argument holds also for the weaker; we need only replace $^{\bullet}$ by $^{\circ}$ throughout.

If there are no dominance deviations, i.e. if

$$(\delta a_{\text{rel.}})_{ij} = \tfrac{1}{2}\{(\delta a_{\text{rel.}})_{ii} + (\delta a_{\text{rel.}})_{jj}\} \quad \text{for all } ij \text{ and rel.},$$

it follows that each ij deviation is the sum of the $i.$ and the $j.$ deviations. In this case we have

$$e^{\bullet}_{ij} = \sum{}^{\bullet} r \operatorname{Dev}(\delta a_r)_{ij} + \delta T^{\bullet}_{..}.$$

Since we must have $e^{\bullet}_{..} = \delta T^{\bullet}_{..}$, it is obvious that some of the deviations must be negative.

Therefore $\delta T^{\bullet}_{..} > -1$ is a necessary condition for $e^{\bullet}_{ij} > -1$. This is, in fact, obvious when we consider that $\delta T^{\bullet}_{..} = -1$ would mean that the aggregate of individual takings was just sufficient to eat up all basic units exactly. Considering that the present use of the coefficients of relationships is only valid when selection is slow, there seems little point in attempting to derive mathematically sufficient conditions for the restriction to hold;

intuitively however it would seem that if we exclude over- and under-dominance it should be sufficient to have no homozygote with a net taking greater than unity.

Even if we could ignore the breakdown of our use of the coefficient of relationship it is clear enough that if $\delta T^{\bullet}_{\cdot\cdot}$ approaches anywhere near -1 the model is highly artificial and implies a population in a state of catastrophic decline. This does not mean, of course, that mutations causing large selfish effects cannot receive positive selection; it means that their expression must moderate with increasing gene-frequency in a way that is inconsistent with our model. The "killer" trait of *Paramoecium* might be regarded as an example of a selfish trait with potentially large effects, but with its only partially genetic mode of inheritance and inevitable density dependance it obviously requires a selection model tailored to the case, and the same is doubtless true of most "social" traits which are as extreme as this.

Really the class of model situations with negative neighbour effects which are artificial according to a strict interpretation of the assumptions must be much wider than the class which we have chosen to call "improbable". The model assumes that the magnitude of an effect does not depend either on the genotype of the effectee or on his current state with respect to the prerequisites of fitness at the time when the effect is caused. Where taking-traits are concerned it is just possible to imagine that this is true of some kinds of surreptitious theft but in general it is more reasonable to suppose that following some sort of an encounter the limited prerequisite is divided in the ratio of the competitive abilities. Provided competitive differentials are small however, the model will not be far from the truth; the correcting term that should be added to the expression for Δp_i can be shown to be small to the third order. With giving-traits it is more reasonable to suppose that if it is the nature of the prerequisite to be transferable the individual can give away whatever fraction of his own property that his instincts incline him to. The model was designed to illuminate altruistic behaviour; the classes of selfish and competitive behaviour which it can also usefully illuminate are more restricted, especially where selective differentials are potentially large.

For loci under selection the only relatives to which our metric of relationship is strictly applicable are ancestors. Thus the chance that an arbitrary parent carries a gene picked in an offspring is $\frac{1}{2}$, the chance that an arbitrary grandparent carries it is $\frac{1}{4}$, and so on. As regards descendants, it seems intuitively plausible that for a gene which is making steady progress in gene-frequency the true expectation of genes i.b.d. in a n-th generation descendant will exceed $\frac{1}{2}^n$, and similarly that for a gene that is steadily declining in frequency the reverse will hold. Since the path of genetic connection with a

simple same-generation relative like a half-sib includes an "ascending part" and a "descending part" it is tempting to imagine that the ascending part can be treated with multipliers of exactly $\frac{1}{2}$ and the descending part by multipliers consistently more or less than $\frac{1}{2}$ according to which type of selection is in progress. However, a more rigorous attack on the problem shows that it is more difficult than the corresponding one for simple descendants, where the formulation of the factor which actually replaces $\frac{1}{2}$ is quite easy at least in the case of classical selection, and the author has so far failed to reach any definite general conclusions as to the nature and extent of the error in the foregoing account which his use of the ordinary coefficients of relationship has actually involved.

Finally, it must be pointed out that the model is not applicable to the selection of new mutations. Sibs might or might not carry the mutation depending on the point in the germ-line of the parent at which it had occurred, but for relatives in general a definite number of generations must pass before the coefficients give the true—or, under selection, the approximate—expectations of replicas. This point is favourable to the establishment of taking-traits and slightly against giving-traits. A mutation can, however, be expected to overcome any such slight initial barrier before it has recurred many times.

5. The Model Limits to the Evolution of Altruistic and Selfish Behaviour

With classical selection a genotype may be regarded as positively selected if its fitness is above the average and as counter-selected if it is below. The environment usually forces the average fitness $a..$ towards unity; thus for an arbitrary genotype the sign of δa_{ij} is an indication of the kind of selection. In the present case although it is $T^{\bullet}_{..}$ and not $R^{\bullet}_{..}$ that is forced towards unity, the analogous indication is given by the inclusive fitness effect $\delta R^{\bullet}_{ij}$, for the remaining part, the diluting effect δS_{ij}, of the total genotypic effect $\delta T^{\bullet}_{ij}$ has no influence on the kind of selection. In other words the kind of selection may be considered determined by whether the inclusive fitness of a genotype is above or below average.

We proceed, therefore, to consider certain elementary criteria which determine the sign of the inclusive fitness effect. The argument applies to any genotype and subscripts can be left out.

Let

$$\delta T^{\circ} = k\,\delta a. \tag{7}$$

According to the signs of δa and δT° we have four types of behaviour as set out in the following diagram:

			Neighbours: gain; δT° +ve	Neighbours: lose; δT° −ve
Individual	gains;	δa +ve	k +ve *Selected*	k −ve Selfish behaviour ?
	loses;	δa −ve	k −ve Altruistic behaviour ?	k +ve *Counter-selected*

The classes for which k is negative are of the greatest interest, since for these it is less obvious what will happen under selection. Also, if we regard fitness as like a substance and tending to be conserved, which must be the case in so far as it depends on the possession of material prerequisites of survival and reproduction, k −ve is the more likely situation. Perfect conservation occurs if $k = -1$. Then $\delta T^{\bullet} = 0$ and $T^{\circ} = 1$: the gene-pool maintains constant "volume" from generation to generation. This case has been discussed in Case (c) of section 3. In general the value of k indicates the nature of the departure from conservation. For instance, in the case of an altruistic action $|k|$ might be called the ratio of gain involved in the action: if its value is two, two units of fitness are received by neighbours for every one lost by an altruist. In the case of a selfish action, $|k|$ might be called the ratio of diminution: if its value is again two, two units of fitness are lost by neighbours for one unit gained by the taker.

The alarm call of a bird probably involves a small extra risk to the individual making it by rendering it more noticeable to the approaching predator but the consequent reduction of risk to a nearby bird previously unaware of danger must be much greater.† We need not discuss here just how risks are to be reckoned in terms of fitness: for the present illustration it is reasonable to guess that for the generality of alarm calls k is negative but $|k| > 1$. How large must $|k|$ be for the benefit to others to outweigh the risk to self in terms of inclusive fitness?

† The alarm call often warns more than one nearby bird of course—hundreds in the case of a flock—but since the predator would hardly succeed in surprising more than one in any case the total number warned must be comparatively unimportant.

$$\begin{aligned}\delta R^{\bullet} &= \delta R^{\circ} + \delta a \\ &= r^{\circ}\,\delta T^{\circ} + \delta a && \text{from (5)} \\ &= \delta a(kr^{\circ} + 1) && \text{from (7).}\end{aligned}$$

Thus of actions which are detrimental to individual fitness (δa −ve) only those for which $-k > \frac{1}{r^{\circ}}$ will be beneficial to inclusive fitness ($\delta R^{\bullet}$ +ve).

This means that for a hereditary tendency to perform an action of this kind to evolve the benefit to a sib must average at least twice the loss to the individual, the benefit to a half-sib must be at least four times the loss, to a cousin eight times and so on. To express the matter more vividly, in the world of our model organisms, whose behaviour is determined strictly by genotype, we expect to find that no one is prepared to sacrifice his life for any single person but that everyone will sacrifice it when he can thereby save more than two brothers, or four half-brothers, or eight first cousins . . . Although according to the model a tendency to simple altruistic transfers ($k = -1$) will never be evolved by natural selection, such a tendency would, in fact, receive zero counter-selection when it concerned transfers between clonal individuals. Conversely selfish transfers are always selected except when from clonal individuals.

As regards selfish traits in general (δa +ve, k −ve) the condition for a benefit to inclusive fitness is $-k < \frac{1}{r^{\circ}}$. Behaviour that involves taking too much from close relatives will not evolve. In the model world of genetically controlled behaviour we expect to find that sibs deprive one another of reproductive prerequisites provided they can themselves make use of at least one half of what they take; individuals deprive half-sibs of four units of reproductive potential if they can get personal use of at least one of them; and so on. Clearly from a gene's point of view it is worthwhile to deprive a large number of distant relatives in order to extract a small reproductive advantage.

REFERENCES

HALDANE, J. B. S. (1923). *Trans. Camb. phil. Soc.* **23,** 19.
HALDANE, J. B. S. & JAYAKAR, S. D. (1962). *J. Genet.* **58,** 81.
KEMPTHORNE, O. (1957). "An Introduction to Genetical Statistics". New York: John Wiley & Sons, Inc.
KINGMAN, J. F. C. (1961a). *Quart. J. Math.* **12,** 78.
KINGMAN, J. F. C. (1961b). *Proc. Camb. phil. Soc.* **57,** 574.
LI, C. C. & SACKS, L. (1954). *Biometrics,* **10,** 347.
MORAN, P. A. P. (1962). *In* "The Statistical Processes of Evolutionary Theory", p. 54. Oxford: Clarendon Press.
MULHOLLAND, H. P. & SMITH, C. A. B. (1959). *Amer. math. Mon.* **66,** 673.
SCHEUER, P. A. G. & MANDEL, S. P. H. (1959). *Heredity,* **31,** 519.