Bulletin of Mathematical Biology Vol. 52, No. 1/2, pp. 209–240, 1990.
Printed in Great Britain.

0092-8240/90$3.00+0.00
Pergamon Press plc
Society for Mathematical Biology

A MATHEMATICAL THEORY OF NATURAL AND ARTIFICIAL SELECTION—I*

■ J. B. S. HALDANE
Trinity College,
Cambridge, U.K.

Mathematical expressions are found for the effect of selection on simple Mendelian populations mating at random. Selection of a given intensity is most effective when amphimixis does not affect the character selected, e.g. in complete inbreeding or homogamy. Selection is very ineffective on autosomal recessive characters so long as they are rare.

1. Introduction. A satisfactory theory of natural selection must be quantitative. In order to establish the view that natural selection is capable of accounting for the known facts of evolution we must show not only that it can cause a species to change, but that it can cause it to change at a rate which will account for present and past transmutations. In any given case we must specify the following.

(1) The mode of inheritance of the character considered.
(2) The system of breeding in the group of organisms studied.
(3) The intensity of selection.
(4) Its incidence (e.g. on both sexes or only one).
(5) The rate at which the proportion of organisms showing the character increases or diminishes.

It should then be possible to obtain an equation connecting (3) and (5).

The principal work on the subject so far is that of Pearson (1908), Warren (1917), and Norton. Pearson's work was based on a pre-Mendelian theory of variation and heredity, which is certainly inapplicable to many, and perhaps to all characters. Warren has only considered selection of an extremely stringent character, whilst Norton's work is as yet only available in the table quoted by Punnett.

In this paper we shall only deal with the simplest possible cases. The character dealt with will be the effect of a single completely dominant Mendelian factor or its absence. The system of breeding considered will be random mating on the one hand or self-fertilization, budding, etc. on the other. Moreover we shall confine ourselves to organisms such as annual plants, and

---

* Reprinted from *Transactions of the Cambridge Philosophical Society*, Vol. 23, pp. 19–41 (1924) with the permission of Cambridge University Press.

many insects and fish in which different generations do not interbreed. Even so it will be found that in most cases we can only obtain rigorous solutions when selection is very rapid or very slow. At intermediate rates we should require to use functions of a hitherto unexplored type. Indeed the mathematical problems raised in the more complicated cases to be dealt with in subsequent papers seem to be as formidable as any in mathematical physics. The approximate solutions given in this paper are however of as great an order of accuracy as that of the data hitherto available.

It is not of course intended to suggest that all heredity is Mendelian, or all evolution by natural selection. On the other hand we know that besides non-Mendelain differences between species (e.g. in chromosome number) there are often Mendelian factor-differences. The former are important because they often lead to total or partial sterility in crosses, but their somatic expression is commonly less striking than that of a single factor-difference. Their behaviour in crosses is far from clear, but where crossing does not occur evolution takes place according to equations (1)–(3).

2. Specification on the Intensity of Selection. If a generation of zygotes immediately after fertilization consists of two phenotypes A and B in the ratio $pA:1B$, and the proportion which form fertile unions is $pA:(1-k)B$, we shall describe k as the coefficient of selection. Thus if $k=0.01$, a population of equal numbers of A and B would survive to form fertile unions in the proportion $100A:99B$, the A's thus having a slight advantage (k may be positive or negative; when it is small, selection is slow). When $k=1$ no B's reproduce, and when $k=-\infty$ no A's reproduce. It will be convenient to refer to these two case as "complete selection". They occur in artificial selection if the character is well marked.

If the character concerned affects fertility, or kills off during the breeding period, we can use just the same notation. in this case each B on the average leaves as many offspring as $(1-k)A$'s, e.g. if $k=0.01$ then 100 B's leave as many as 99 A's. The effect is clearly just the same as if one of the B's had died before breeding. It will be observed that no assumption is made as to the total number of the population. If this is limited by the environment, natural selection may cause it to increase or diminish. It will for example tend to increase if selection renders the organism smaller or fitter to cope with its environment in general. If on the other hand selection increases its size, or merely arms it in the struggle with other members of its species for food or mates, the population will tend to diminish or even to disappear.

Warren (1917) considered the case where the total population is fixed. He supposes that the parents produce l times their number of offspring, and that type A is p times as numerous as type B, but $1/m$ as likely to die. In this case it can be shown that:

$$k=\frac{(l-1)(m-1)(p+1)}{lm-l+p+1}.$$

Hence the advantage of one type over the other as measured by k is not independent of the composition of the population unless $m-1$ is very small, when $k=(l-1)(m-1)$ approximately. Hence when selection is slow—the most interesting case—the two schemes of selection lead to similar results. On the other hand the mathematical treatment of selection on our scheme is decidedly simpler.

3. Familial Selection. The above notation may easily be applied to the cases, such as Darwinian sexual selection, where one sex only is selected. There is however another type of selection which so far as I know has not been considered in any detail by former authors, but which must have been of considerable importance in evolution. So far we have assumed that the field of struggle for existence is the species as a whole, or at least those members of it living within a given area. But we have also to consider those cases where the struggle occurs between members of the same family. Such cases occur in many mammals, seed-plants, and nematodes, to mention no other groups. Here the size of the family is strictly limited by the food or space available for it, and more embryos are produced than can survive to enter into the struggle with members of other families. Thus in the mouse Ibsen and Steigleder (1917) have shown that some embryos of any litter perish *in utero*. Their deaths are certainly sometimes selective. In litters from the mating yellow × yellow one-quarter of the embryos die in the blastula stage, yet as Durham (1911) has shown, such litters are no smaller than the normal, because the death of the YY embryos allows others to survive which would normally have perished.

The above is a case of complete selection. Where the less viable type of embryo, instead of perishing inevitably, is merely at a slight disadvantage, it is clear that selection will only be effective, or at any rate will be much more effective, in the mixed litters. Thus let us consider 3 litters of 20 embryos each, the first consisting wholly of the stronger type, the second containing 10 strong and 10 weak, the third wholly of the weaker type. Suppose that in each case there is only enough food or space for 10 embryos, and that the strong type has an advantage over the weak such that, out of equal numbers, 50% more of the strong will survive, i.e. $k=\frac{1}{3}$. Then the survivors will be 10 strong from the first litter, 6 strong and 4 weak from the second, and 10 weak from the third, or 16 strong and 14 weak. If the competition had been free, as with pelagic larvae, the numbers would have been 18 strong and 12 weak. Clearly with familial selection the same advantage acts more slowly than with normal selection, since it is only effective in mixed families.

The "family" within which selection acts may have both parents in common, as in most mammals, or many different male parents, as in those plants whose pollen, but not seeds, is spread by the wind. In this case the seeds from any one plant will fall into the same area, and unless the plants are very closely packed, will compete with one another in the main. In rare cases familial sexual selection may occur. Thus in *Dinophilus* the rudimentary males fertilize their sisters before leaving the cocoon. Clearly so long as every female gets fertilized before hatching selection can only occur in the male sex between brothers, and must tend to make the males copulate at as early a date as possible.

The survival of many of the embryonic characters of viviparous animals and seed-plants must have been due to familial selection.

4. Selection in the Absence of Amphimixis. The simplest form of selection is uncomplicated either by amphimixis or dominance. It occurs in the following cases.

(1) Organisms which do not reproduce sexually, or are self-fertilizing.
(2) Species which do not cross, but compete for the same means of support.
(3) Organisms in which mating is always between brother and sister.
(4) Organisms like *Bryophyta* which are haploid during part of the life cycle, provided that selection of the character considered only occurs during the haploid phase.
(5) Heterogamous organisms in which the factor determining the character selected occurs in the gametes of one sex only. For example Renner (1917) has shown that *Oenothera muricata* transmits certain characters by the pollen only, others by the ovules only. Schmidt (1920) has found a character in *Lebistes* transmitted by males to males only, and Goldschmidt (1920) has postulated sex-factors in *Lymantria* transmitted only by females to females. As far as the characters in question are concerned there is no amphimixis, and these organisms behave as if they were asexual. Other species of *Oenothera* which are permanently heterozygous for other reasons would probably be selected in much the same way.

Let the n^{th} generation consist of types A and B in the ratio $u_n A : 1B$, and let the coefficient of selection be k, i.e. $(1-k)B$'s survive for every A. Then the survivors of the n^{th} generation, and hence the first numbers of the $(n+1)^{\text{th}}$, with be $u_n A : (1-k)B$. Therefore:

$$u_{n+1}=\frac{u_n}{1-k}, \tag{1}$$

and if u_0 be the original ratio $u_n=(1-k)^{-n}u_0$.

Now if we write y_n for the proportion of B's in the total population of the n^{th} generation:

$$y_n=\frac{1}{1+u_n}=\frac{1}{1+(1-k)^{-n}u_0}=\frac{y_0}{y_0+(1-k)^{-n}(1-y_0)},$$

or if we start with equal numbers of A and B, $y_0=\frac{1}{2}$, and:

$$y_n=\frac{1}{1+(1-k)^{-n}}. \tag{2}$$

If k is very small, i.e. selection slow, then approximately:

$$\text{or:}\quad \left.\begin{aligned} y_n&=\frac{1}{1+e^{kn}} \\ kn&=\log_e\left(\frac{1-y_n}{y_n}\right) \end{aligned}\right\}. \tag{3}$$

Hence the proportion of B's falls slowly at first, then rapidly for a short time, then slowly again, the rate being greatest when $y=\frac{1}{2}$. Before $y=\frac{1}{2}$, n is of course taken as negative. So long as k is small the time taken for any given change in the proportions varies inversely as k. The curve representing graphically the change of the population is symmetrical about its middle point, and is shown in Fig. 1 for the case where $k=0.001$, i.e. 999 B's survive for every 1000 A's. 9184

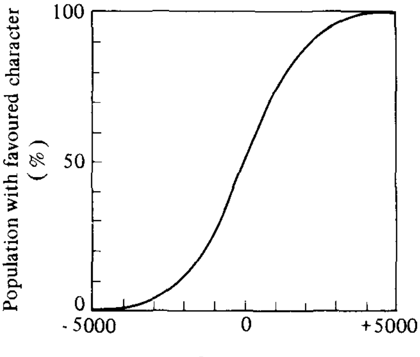

Figure 1. Effect of selection on a non-amphimictic character. $k=0.001$.

generations are needed for the proportion of A's to increase from 1 to 99%. Equation (3) gives an error of only four in this number.

As will be shown below, selection proceeds more slowly with all other systems of inheritance. In this case the speed must compensate to some extent for the failure to combine advantageous factors by amphimixis. Where occasional amphimixis occurs, as for example in wheat, conditions are very favourable for the evolution of advantageous combinations of variations.

5. Selection of a Simple Mendelian Character. Consider the case of a population which consists of zygotes containing two, one, or no "doses" of a completely dominant Mendelian factor A, mating is at random, and selection acts to an equal degree in both sexes upon the character produced by the factor. Pearson (1904) and Hardy (1908) have shown that in a population mating at random the square of the number of heterozygotes is equal to four times the product of the numbers of the two homozygous classes. Let $u_n A : 1a$ be the proportion of the two types of gametes produced by the $(n-1)^{\text{th}}$ generation. Then in the n^{th} generation the initial proportions of the three classes of zygotes are:

$$u_n^2 AA : 2u_n Aa : 1aa.$$

The proportion of recessives to the whole population is:

$$y_n = (1+u_n)^{-2}. \tag{4}$$

Now only $(1-k)$ of the recessives survive to breed, so that the survivors are in the proportions:

$$u_n^2 AA : 2u_n Aa : (1-k)aa.$$

The numbers of the next generation can be most easily calculated from the new gametic ratio u_{n+1}. This is immediately obvious in the case of aquatic organisms which shed their gametes into the water. If each zygote produces N gametes which conjugate, the numbers are clearly:

$$(Nu_n^2 + Nu_n)A, \text{ and } (Nu_n + N1 - k)a.$$

So the ratio:

$$u_{n+1} = \frac{u_n(1+u_n)}{1+u_n-k}. \tag{5}$$

It can easily be shown that this result follows from random mating, for matings will occur in the following proportions:

$AA \times AA$		$u_n^2 \times u_n^2$	or	u_n^4,
$AA \times Aa$	and reciprocally,	$2 \times u_n^2 \times 2u_n$	or	$4u_n^3$,

$AA \times aa$and reciprocally,	$2 \times u_n^2 \times (1-k)$	or	$2(1-k)u_n^2$,
$Aa \times Aa$	$2u_n \times 2u_n$	or	$4u_n^2$,
$Aa \times aa$ and reciprocally,	$2 \times 2u_n \times (1-k)$	or	$4(1-k)u_n$,
$aa \times aa$	$(1-k) \times (1-k)$	or	$(1-k)^2$.

Hence zygotes are formed in the following proportions:

AA	$u_n^4 + 2u_n^3 + u_n^2$	or	$u_n^2(1+u_n)^2$,
Aa	$2u_n^3 + 2(1-k)u_n^2 + 2u_n^2 + 2(1-k)u_n$	or	$2u_n(1+u_n)(1+u_n-k)$,
aa	$u_n^2 + 2(1-k)u_n + (1-k)^2$	or	$(1+u_n-k)^2$.

These ratios may be written:

$$\left[\frac{u_n(1+u_n)}{1+u_n-k}\right]^2 AA : \frac{2u_n(1+u_n)}{1+u_n-k} Aa : 1aa,$$

or:

$$u_{n+1}^2 AA : 2u_{n+1} Aa : 1aa,$$

where:

$$u_{n+1} = \frac{u_n(1+u_n)}{1+u_n-k},$$

as above. It is however simpler to obtain u_{n+1} directly from the ratio of A to a among the gametes of the population as a whole, and this will be done in our future calculations.

Now if we know the original proportion of recessives y_0, we start with a population:

$$u_0^2 AA : 2u_0 Aa : 1aa,$$

where:

$$u_0 = y_0^{-1/2} - 1,$$

and we can at once calculate:

$$u_1 = \frac{u_0(1+u_0)}{1+u_0-k},$$

and thence u_2 and so on, obtaining y_1, y_2, etc. from equation (4). Thus if we start with 25% of recessives, and $k=0.5$, i.e. the recessives are only half as viable as the dominants, then $u_0=1$, and:

$$u_1=\frac{1(1+1)}{1+1-\frac{1}{2}}=\frac{4}{3},$$

$$y_1=(1+\tfrac{4}{3})^{-2}=\tfrac{9}{49}=0.184, \text{ or } 18.4\%.$$

Similarly $y_2=13.75\%$, $y_3=10.9\%$, and so on. Starting from the same population, but with $k=-1$, so that the recessives are twice as viable as the dominants, we have $y_1=36\%$, $y_2=49.8\%$, $y_3=64.6\%$, $y_4=77.5\%$, $y_5=87.0\%$, and so on. If k is small this method becomes very tedious, but we can find a fairly accurate formula connecting y_n with n.

The case of complete selection is simple. If all the dominants are killed off or prevented from breeding we shall see the last of them in one generation, and $y_n=1$. Punnett (1917) and Hardy (1908) have solved the case where the recessives all die. Here $k=1$, and:

$$u_{n+1}=\frac{u_n(1+u_n)}{1+u_n-1}=1+u_n.$$

$$\therefore\ u_n=n+u_0;$$

$$\begin{aligned}\therefore\ y_n&=(n+1+u_0)^{-2}\\&=(n+y_0^{-1/2})^{-2}\\&=y_0(1+ny_0^{1/2})^{-2}.\end{aligned}\tag{6}$$

Thus if we start with a population containing $\frac{1}{4}$ recessives the second generation will contain $\frac{1}{9}$, the third $\frac{1}{16}$, the n^{th} $1/(n+1)^2$. Thus 999 generations will be needed to reduce the proportion to one in a million, and we need not wonder that recessive sports still occur in most of our domestic breeds of animals.

When selection is not very intense, we can proceed as follows:

$$u_{n+1}=\frac{u_n(1+u_n)}{1+u_n-k};$$

$$\therefore\ \Delta u_n\equiv u_{n+1}-u_n=\frac{ku_n}{1+u_n-k}.$$

When k is small we can neglect it in comparison with unity, and suppose that u_n increases continuously and not by steps, i.e. take $\Delta u_n = \mathrm{d}u_n/\mathrm{d}n$.

$$\therefore\ \frac{\mathrm{d}u_n}{\mathrm{d}n} = \frac{ku_n}{1+u_n}\ \text{approximately;}$$

$$\therefore\ kn = \int_{u_0}^{u_n} \frac{1+u}{u}\,\mathrm{d}u$$

$$= u_n - u_0 + \log_e\left(\frac{u_n}{u_0}\right). \tag{7}$$

If we start from or work towards a standard population containing 25% of recessives, and hence $u_0 = 1$, we have:

$$kn = u_n + \log_e u_n - 1. \tag{8}$$

This equation is accurate enough for any practical problem when $|k|$ is small, and as long as k lies between ± 0.1, i.e. neither phenotype has an advantage of more than 10%, it may be safely used. When $|k|$ is large the equation:

$$kn = u_n + (1-k)\log_e u_n - 1, \tag{9}$$

is fairly accurate for positive values of n. Thus when $k = \frac{1}{2}$, the error is always under 4%. For large values of $|k|$ and negative values of n the equation:

$$kn = u_n + \left(1 - \frac{k}{2}\right)\log_e u_n - 1, \tag{10}$$

gives results with a very small error. But for every case so far observed equation (8) gives results within the limits of observational error.

In the above equations we have only to make k negative to calculate the effects of a selection which favours recessives at the expense of dominants. For the same small intensity of selection the same time is clearly needed to produce a given change in the percentage of recessives whether dominants or recessives are favoured. Figure 2 shows graphically the rate of increase of dominants and recessives respectively when $k = \pm 0.001$, i.e. the favoured type has an advantage of one in a thousand, as in Fig. 1. In each case 16 582 generations are required to increase the proportion of the favoured type from 1 to 99%, but dominants increase more rapidly than recessives when they are few, more slowly when they are numerous. The change occurs most rapidly when y_n, the proportion of recessives, is 56.25%. When selection is 10 times as intense, the population will clearly change 10 times as fast, and so on.

In Table 1 the values of y_n calculated from equations (4) and (8) are given in

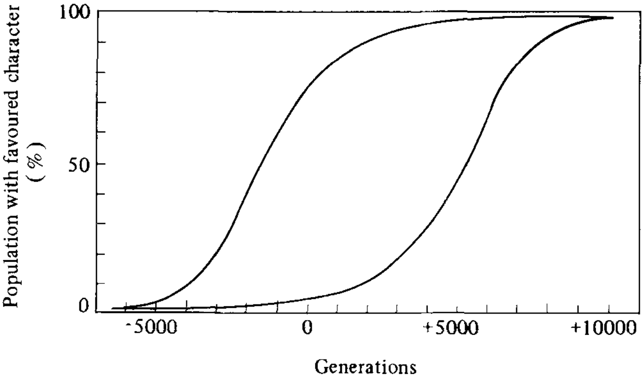

Figure 2. Effect of selection on an autosomal Mendelian character. $k=0.001$. Upper curve, dominants favoured; lower curve, recessives favoured.

terms of kn. In Table 2 kn is given in terms of y_n. The number of generations (forwards or backwards) is reckoned from a standard population containing 75% of dominants and 25% of recessives. A few examples will make the use of these tables clear.

(1) Detlefsen (1918) has shown that in a mixed population of mice about 95.9 without the factor G, causing light bellies and yellow-tipped hair, survive for every 100 with it. Hence $k=0.041$. It is required to find how many recessives will be left after 100 generations, starting from a population with 90% of recessives, and assuming that different generations do not interbreed.

From Table 2, when $y=0.9$, $kn=-3.863$, $\therefore n=-94.2$. So 94.2 generations of selection will bring the recessives down to 25%. The remaining 5.8 generations give $kn=0.238$, and from Table 1 by interpolation we find $y=0.224$, i.e. only 22.4% of recessives remain.

(2) In the same case how many generations are needed to reduce the number of recessives to 1%? $y_n=0.01$, hence, from Table 2, $kn=10.197$, $\therefore n=248.7$. So 248.7 generations after 25% is reached, or 343 in all, will be required.

(3) The dominant melanic form *doubledayaria* of the peppered moth *Amphidasys betularia* first appeared at Manchester in 1848. Some time before 1901 when Barrett described the case, it had completely ousted the recessive variety in Manchester. It is required to find the least intensity of natural selection which will account for this fact.

Assuming that there were not more than 1% of dominants in Manchester in 1848, nor less than 99% in 1898, we have, from Table 2, $kn=16.58$ as a minimum. But $n=50$, since this moth usually has one brood per year.

Table 1. Effect of slow selection on an autosomal Mendelian character

(1) *kn* (number of generations × *k*):						−1000	−100	−50	−20	−15	−10	−9
(2) % of recessives when dominants are favoured:						—	—	—	—	99.9998	99.975	99.933
(3) % of recessives when recessives are favoured:						0.0001	0.0105	0.0427	0.2773	0.4215	1.036	1.254
(1)	−8	−7	−6	−5	−4.5	−4	−3.5	−3	−2.5	−2	−1.5	−1
(2)	99.82	99.50	98.68	96.50	94.38	91.14	86.36	79.71	71.24	61.53	50.68	40.98
(3)	1.545	1.940	2.497	3.308	—	4.537	—	6.528	—	9.718	12.11	15.30
(1)	−0.5	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
(2)	32.05	25.0	19.53	15.30	12.11	9.718	—	6.528	—	4.537	—	3.308
(3)	19.53	25.0	32.05	40.98	50.68	61.53	71.24	79.71	86.36	91.14	94.38	96.50
(1)	6	7	8	9	10	15	20	50	100	1000		
(2)	2.497	1.940	1.545	1.254	1.036	0.4215	0.2773	0.0427	0.0105	0.0001		
(3)	98.68	99.50	99.82	99.933	99.975	99.9998	—	—	—	—		

Table 2. Effect of slow selection on an autosomal Mendelian character

(1) % of favoured type:					0.0001	0.001	0.01	0.05	0.1	0.2	0.5
(2) *kn* when dominants are favoured:					−15.51	−13.21	−10.90	−9.294	−8.600	−7.905	−6.996
(3) *kn* when recessives are favoured:					−1005	−320.0	−102.60	−45.50	−33.04	−23.42	−14.72
(1)	1	2	3	5	10	15	20	25	30	35	40
(2)	−6.286	−5.580	−5.161	−4.624	−3.863	−3.290	−2.979	−2.712	−2.439	−2.180	−1.964
(3)	−10.197	−6.875	−4.976	−3.717	−1.933	−1.041	−0.448	0	+0.366	+0.681	+0.962
(1)	45	50	55	60	65	70	75	80	85	90	95
(2)	−1.708	−1.467	−1.220	−0.962	−0.681	−0.366	0	+0.448	+1.041	+1.933	+3.717
(3)	+1.220	+1.467	+1.708	+1.964	+2.180	+2.439	+2.712	+2.979	+3.290	+3.863	+4.620
(1)	97	98	99	99.5	99.8	99.9	99.95	99.99	99.999	99.9999	
(2)	+4.976	+6.875	+10.197	+14.72	+23.42	+33.04	+45.50	+102.60	+320.0	+1005	
(3)	+5.161	+5.580	+6.286	+6.996	+7.905	+8.600	+9.294	+10.90	+13.21	+15.51	

Therefore $k=0.332$ at least, i.e. at least 3 dominants must survive for every 2 recessives, and probably more; or the fertility of the dominants must be 50% greater than that of the recessives. Direct calculation step by step from equation (5) shows that 48 generations are needed for the change if $k=0.3$. Hence the table is sufficiently accurate. After only 13 generations the dominants would be in a majority. It is perhaps instructive, in view of the fact that attempts have been made to explain such cases by epidemics of mutation due either to the environment or to unknown causes, to note that in such a case one recessive in every five would have to mutate to a dominant. Hence it would be impossible to obtain true breeding recessives as was done by Bate (1895). Another possible explanation would be a large excess of dominants begotten in mixed families, as occurs in human night-blindness according to Bateson. But this again does not agree with the facts, and the only probable explanation is the not very intense degree of natural selection postulated above.

6. Familial Selection of a Simple Mendelian Character. Consider the case of a factor A whose presence gives any embryo possessing it an advantage measured by k over those members of the same family which do not possess it. In this case the Pearson–Hardy law does not hold in the population. Each family may have both parents in common, as in mammals, or only the mother, as in cross-pollinated seed-plants. In the first case let the population consist of:

$$p_n AA : 2q_n Aa : r_n aa, \text{ where } p_n + 2q_n + r_n = 1.$$

Then in a mixed family where equality was to be expected the ratio of dominants to recessives will be $1:1-k$. But since the total is unaltered, the actual number of dominants will be to the expected as $2:2-k$, of recessives as $2-2k:2-k$, and similarly for a family where a $3:1$ ratio was to be expected. The n^{th} generation mating at random will therefore produce surviving offspring in the following proportions:

		AA	Aa	aa
From mating:	$AA \times AA$	p_n^2	0	0
	$AA \times Aa$	$2p_nq_n$	$2p_nq_n$	0
	$AA \times aa$	0	$2p_nr_n$	0
	$Aa \times Aa$	$\frac{4q_n^2}{4-k}$	$\frac{8q_n^2}{4-k}$	$\frac{(4-4k)q_n^2}{4-k}$
	$Aa \times aa$	0	$\frac{2q_nr_n}{2-k}$	$\frac{(2-2k)q_nr_n}{2-k}$
	$aa \times aa$	0	0	r_n^2

$$\left.\begin{aligned} \therefore\ [AA] &= p_{n+1} = (p_n+q_n)^2 + \frac{kq_n^2}{4-k} \\ \tfrac{1}{2}[Aa] &= q_{n+1} = (p_n+q_n)(q_n+r_n) + kq_n\left(\frac{q_n}{4-k}+\frac{r_n}{2-k}\right) \\ [aa] &= r_{n+1} = (q_n+r_n)^2 - kq_n\left(\frac{3q_n}{4-k}+\frac{2r_n}{2-k}\right) \end{aligned}\right\}. \qquad (11)$$

With complete selection, when $k=1$, we have $r_{n+1}=r_n^2$, so the proportion of recessives, starting from $\frac{1}{4}$, will be $\frac{1}{16}$, $\frac{1}{64}$, etc., in successive generations, provided of course that all-recessive families survive, as in *Oenothera.* So recessives are eliminated far more quickly than in the ordinary type of selection. Clearly however dominants are not eliminated at once when $k=-\infty$ (provided that they survive in all-dominant families), for:

$$p_{n+1} = p_n(1-r_n) = p_n p_{n-1}(2-p_{n-1}).$$

Starting from the standard population, successive proportions of recessives are 25%, 56.25%, 66.02%, 84.25%, etc.

In the more interesting case when k is small we can solve approximately, as follows. From equation (11) we see that $q_{n+1}^2 - p_{n+1}r_{n+1}$ and hence $q_n^2 - p_n r_n$ is a small quantity of the order kq_n^2, i.e. is less than k. Hence if we write $u_n = (p_n+q_n)/(q_n+r_n)$, then q_n only differs from $u_n/(1+u_n)^2$ by a small quantity of the order of k.

Now:

$$\begin{aligned} u_{n+1} &= \frac{p_{n+1}+q_{n+1}}{q_{n+1}+r_{n+1}} \\ &= \frac{p_n+q_n+kq_n\left(\dfrac{2q_n}{4-k}+\dfrac{r_n}{2-k}\right)}{q_n+r_n-kq_n\left(\dfrac{2q_n}{4-k}+\dfrac{r_n}{2-k}\right)} \\ &= \frac{p_n+q_n+\frac{1}{2}kq_n(q_n+r_n)}{q_n+r_n-\frac{1}{2}kq_n(q_n+r_n)} \text{ approximately} \\ &= \frac{u_n+\frac{1}{2}kq_n}{1-\frac{1}{2}kq_n} \\ &= u_n + \tfrac{1}{2}kq_n(1+u_n) \text{ approximately} \\ &= u_n + \frac{ku_n}{2(1+u_n)}. \end{aligned} \qquad (12)$$

Solving as for equation (5) we find:

$$\tfrac{1}{2}kn = u_n + \log_e u_n - 1. \qquad (13)$$

And since as above r_n (the proportion of recessives) $= (1+u_n)^{-2}$, it follows that the species changes its composition at half the rate at which it would change if selection worked on the species as a whole, and not within families only.

If each family has its mother only in common, but the fathers are a random sample of the population, we assume the n^{th} generation to consist of:

$$p_n AA : 2q_n Aa : r_n aa, \qquad \text{where } p_n + 2q_n + u_n = 1.$$

Let $u_n = \dfrac{p_n + q_n}{q_n + r_n}$, hence $p_n + q_n = \dfrac{u_n}{1+u_n}$, $q_n + r_n = \dfrac{1}{1+u_n}$.

Then families will be begotten as follows:

	AA	Aa	aa
From AA females:	$\dfrac{p_n u_n}{1+u_n}$	$\dfrac{p_n}{1+u_n}$	0
From Aa females:	$\dfrac{q_n u_n}{1+u_n}$	q_n	$\dfrac{q_n}{1+u_n}$
From aa females:	0	$\dfrac{r_n u_n}{1+u_n}$	$\dfrac{r_n}{1+u_n}$

After selection and replacement the proportions will be:

	AA	Aa	aa
From AA females:	$\dfrac{p_n u_n}{1+u_n}$	$\dfrac{p_n}{1+u_n}$	0
From Aa females:	$\dfrac{q_n u_n}{1+u_n-\frac{1}{2}k}$	$\dfrac{q_n(1+u_n)}{1+u_n-\frac{1}{2}k}$	$\dfrac{q_n(1-k)}{1+u_n-\frac{1}{2}k}$
From aa females:	0	$\dfrac{r_n u_n}{1+u_n-k}$	$\dfrac{r_n(1-k)}{1+u_n-k}$

With complete selection, where $k=1$, recessives are eliminated at once, provided families are large enough. Where $k=-\infty$, dominants are not eliminated at once if pure dominant families survive, since $p_{n+1} = p_n(1-p_n)/(1+p_n)$. Starting from the standard population, successive values of r_n are 25%, 75%, 87.5%, 99.7%, etc. When k is small were obtain approximate equations analogous to equation (11) whose solution is:

$$\tfrac{3}{4}kn = u_n + \log_e u_n - 1. \qquad (14)$$

Thus selection proceeds at $\frac{3}{4}$ of the rate given by equation (8).

7. Sex-Limited Characters and Unisexual Selection. We have next to deal with characters which only appear in one sex, for example milk yield or other secondary sexual characters; or on which selection at least is unisexual, as for example in Darwinian sexual selection. Let the $(n-1)^{\text{th}}$ generation form spermatozoa in the ratio $u_n A\colon 1a$, eggs in the ratio $v_n A\colon 1a$. Then the n^{th} generation consists of zygotes in the ratios:

$$u_n v_n AA\colon (u_n+v_n)Aa\colon 1aa,$$

$$\therefore\ y_n=(1+u_n)^{-1}(1+v_n)^{-1}. \tag{15}$$

If only $1-k$ recessive ♀ survives for every dominant ♀, whilst ♂'s are not affected by selection, we have:

$$\left.\begin{aligned} u_{n+1}&=\frac{2u_n v_n+u_n+v_n}{u_n+v_n+2}\\ v_{n+1}&=\frac{2u_n v_n+u_n+v_n}{u_n+v_n+2-2k}\end{aligned}\right\}. \tag{16}$$

With complete selection, when $k=-\infty$, and all dominants of one sex are weeded out, we have $v_n=0$, and $u_{n+1}=u_n/(2+u_n)$.

$$\therefore\ u_n=\left[2^{n-1}\left(1+\frac{1}{u_0}\right)-1\right]^{-1},$$

and:

$$y_n=1+2^{1-n}(y_0^{1/2}-1). \tag{17}$$

Hence the proportion of dominants is halved in every successive generation. When $k=1$, and all the recessives of one sex die childless, the proportions of recessives in successive generations, starting from the standard population, are 25%, 16.7%, 12.5%, 9.56%, 7.94%, and so on.

When k is small, since:

$$v_{n+1}-u_{n+1}=\frac{2k(2u_n v_n+u_n+v_n)}{(u_n+v_n+2)(u_n+v_n+2-2k)},$$

and:

$$\Delta u_n=\frac{(1+u_n)(v_n-u_n)}{u_n+v_n+2},$$

and hence the differences between u_n, u_{n+1}, v_n, v_{n+1} may be neglected in comparison with themselves:

$$\therefore\ v_n - u_n = \frac{ku_n}{1+u_n} \text{ approximately,}$$

and:

$$\Delta u_n = \frac{ku_n}{2(1+u_n)} \text{ approximately.}$$

$$\therefore\ \tfrac{1}{2}kn = u_n + \log_e u_n - 1, \tag{18}$$

and selection proceeds at half the rate given by equation (8), a result stated by Punnett.

8. Selection of an Alternatively Dominant Character. A few factors, such as that determining the presence or absence of horns in Dorset and Suffolk sheep, according to Wood (1905) are dominant in one sex, recessive in the other. Consider a factor dominant in the male sex, recessive in the female. Let the n^{th} generation be produced by:

$$\text{spermatozoa } u_n A : 1a, \qquad \text{eggs } v_n A : 1a.$$

It consists of:

$$\text{zygotes}: u_n v_n AA : (u_n + v_n)Aa : 1aa,$$

and the survivors after selection are in the ratios:

$$♂\quad u_n v_n AA : (u_n + v_n)Aa : (1-k)aa,$$

$$♀\quad \frac{u_n v_n}{1-k} AA : (u_n + v_n)Aa : 1aa,$$

$$\left.\begin{aligned} \therefore\ u_{n+1} &= \frac{2u_n v_n + u_n + v_n}{u_n + v_n + 2 - 2k} \\ v_{n+1} &= \frac{\dfrac{2}{1-k} u_n v_n + u_n + v_n}{u_n + v_n + 2} \end{aligned}\right\}. \tag{19}$$

Whilst:

$$y_n(\text{for males}) = (1+u_n)^{-1}(1+v_n)^{-1}. \tag{20}$$

With complete selection, when all members of the type dominant in the female sex are weeded out, $k=1$.

$$\therefore\ v_{n+1} = \infty, \text{ and } u_{n+1} = 1 + 2u_n, \text{ after the first generation.}$$

$$\therefore\ 1 + u_n = 2^{n-1}(1+u_1),$$

and if z_n be the proportion of the weeded out type occurring in the female sex:

$$\therefore \left.\begin{aligned} y_n&=0 \\ z_n&=2^{1-n}z_1 \end{aligned}\right\}. \tag{21}$$

So this type disappears in the male sex, and is halved in successive female generations. If $k=\infty$ the type recessive in the female sex disappears in that sex and is halved in successive male generations.

When k is small:

$$v_{n+1}-u_{n+1}=\frac{ku_n(u_n-1)}{1+u_n} \text{ approximately,}$$

and $\Delta u_n=\frac{1}{2}ku_n$.

$$\therefore\ kn=2\log_e u_n \tag{22}$$

if $u_0=1$, so selection occurs on the whole more rapidly than by equation (8) (see Table 5); y_n is given by equation (4).

9. Sex-Linked Characters Under No Selection. The events in an unselected population whose members differ with regard to a sex-linked factor have been considered by Jennings (1915) but can be treated more simply. We suppose the male to be heterozygous for sex, but the argument is the same where the female is heterozygous. Consider a fully dominant factor A, such that the female may be AA, Aa, or aa, the male Aa or aa (or in Morgan's notation, which will be adopted, A or a). As Jennings showed, a population with:

$$♀\text{'s } u^2AA:2uAa:1aa; \qquad ♂\text{'s } uA:1a,$$

is stable during random mating, and other populations approach it asymptotically. In any population let the eggs of the $(n-1)^{\text{th}}$ generation be $u_nA:1a$, the ♀-producing spermatozoa $v_nA:1a$. The the n^{th} generation will be: ♀'s $u_nv_nAA:(u_n+v_n)Aa:1aa$; ♂'s $u_nA:1a$.

$$\therefore \left.\begin{aligned} u_{n+1}&=\frac{2u_nv_n+u_n+v_n}{2+u_n+v_n} \\ v_{n+1}&=u_n \end{aligned}\right\}, \tag{23}$$

and if y_n be the proportion of recessive ♀'s, z_n of recessive ♂'s:

$$\left.\begin{aligned} y_n&=(1+u_n)^{-1}(1+v_n)^{-1} \\ z_n&=(1+u_n)^{-1} \end{aligned}\right\}. \tag{24}$$

$$\therefore\ z_{n+1}^{-1}=1+u_{n+1}=\frac{2(1+u_n)(1+u_{n-1})}{2+u_n-u_{n-1}}=\frac{2}{z_n+z_{n-1}}.$$

$$\therefore\quad 2z_n=z_{n-1}+z_{n-2}.$$

Solving as usual for recurring series, we have:

$$\left.\begin{array}{l} 3z_n=z_0+2z_1+(-\tfrac{1}{2})^n(2z_0-2z_1) \\ y_n=z_nz_{n-1} \end{array}\right\}. \tag{25}$$

$$\left.\begin{array}{l} \therefore\ z_\infty=\tfrac{1}{3}(z_0+2z_1)=\tfrac{1}{3}(z_n+2z_{n+1}) \\ y_\infty=z_\infty^2 \end{array}\right\}. \tag{26}$$

Hence from the proportion of males in two successive generations, or both sexes in one, we can calculate the final values. Successive values of y_n and z_n oscillate alternately above and below their final values, but converge rapidly towards them.

10. Bisexual Selection of a Sex-Linked Character. If the conditions are as above, except that in each generation one dominant survives for every $(1-k)$ recessives in each sex, then:

$$\left.\begin{array}{l} u_{n+1}=\dfrac{2u_nv_n+u_n+v_n}{u_n+v_n+2-2k} \\ v_{n+1}=\dfrac{u_n}{1-k} \end{array}\right\}. \tag{27}$$

and:

$$\left.\begin{array}{l} y_n=(1+u_n)^{-1}(1+v_n)^{-1} \\ z_n=(1+u_n)^{-1} \end{array}\right\}.$$

With complete selection if $k=-\infty$, and no dominants survive to breed, selection is complete in one generation. If $k=1$, and no recessives survive to breed:

$$u_{n+1}=1+2u_n,\qquad \text{and}\qquad v_n=\infty.$$

$$\therefore\ 1+u_n=2^n(1+u_0),$$

and:

$$\left.\begin{array}{l} z_n=2^{-n}z_0 \\ y_n=0 \end{array}\right\}. \tag{28}$$

So no recessive females are produced and the number of recessive males is halved in each generation. Selection is therefore vastly more effective than on an autosomal character. If colour-blind or haemophilic persons were prevented from breeding, these conditions could be almost abolished in a few generations, which is not the case with feeble-mindedness. If selection is slow we solve as for equation (16), and find approximately:

$$v_n - u_n = \frac{2ku_n^2}{3+3u_n},$$

$$\Delta u_n = \frac{ku_n(3+u_n)}{3+3u_n}.$$

So, reckoning generations to or from a standard population where $u_0 = 1$, and 50% of the males and 25% of the females are recessives:

$$kn = \log_e u_n + 2\log\left(\frac{3+u_n}{4}\right), \tag{29}$$

$$\left.\begin{aligned} y_n &= (1+u_n)^{-2} \\ z_n &= (1+u_n)^{-1} \end{aligned}\right\}. \tag{30}$$

Table 3 and Fig. 3 are calculated from these equations. Within the limits covered by the figure selection acts more rapidly on a sex-linked character in the homozygous sex than on an autosomal character. In the heterozygous sex selection occurs at about the same rate in the two cases. However, as appears from Table 5, sex-linked recessive characters increase far more rapidly in the early stages, and sex-linked dominants in the late stages of selection, the autosomal characters.

Table 3 is not quite accurate unless selection is very slow, the error being of the order of k. Thus when $k = 0.2$ the error in n is nearly 10%. Still even for these large values it furnishes a useful first approximation.

11. Bisexual Familial Selection of a Sex-Linked Character. Here we need only consider the case where the family within which selection occurs has both parents in common. Sex-linkage of the animal type is rare in plants, and families with many fathers per mother are rare in animals. Let the n^{th} generation be:

$$♀\ p_nAA : 2q_nAa : r_naa; \qquad ♂\ s_nA : t_na,$$

where $p_n + 2q_n + r_n = s_n + t_n = 1$. Let the dominants have an advatage of $1:1-k$ over the recessives in the mixed families. Then the $(n+1)^{\text{th}}$ generation occurs in the following proportions, after selection has operated:

Table 3. Effect of slow Selection in both sexes on a sex-linked character, dominants being favoured

(1) % of recessives of homozygous sex:						99.998	99.98	99.80	99.60	99.00	98.01
(2) % of recessives of heterozygous sex:						99.999	99.99	99.9	99.8	99.5	99
(3) *kn* (number of generations × *k*):						−12.09	−9.786	−7.481	−6.787	−5.866	−5.164
(1)	96.04	90.25	81	64	49	36	25	16	10	6.25	4
(2)	98	95	90	80	70	60	50	40	31.62	25	20
(3)	−4.454	−3.485	−2.700	−1.802	−1.156	−0.580	0	0.619	1.282	1.910	2.506
(1)	2	1	0.5	0.25	0.1	0.01	0.001	0.0001	0.0_51	0.0_71	
(2)	14.14	10	7.071	5	3.162	1	0.3162	0.1	0.01	0.001	
(3)	3.441	4.394	5.366	6.353	7.679	11.07	14.50	17.95	24.86	31.76	

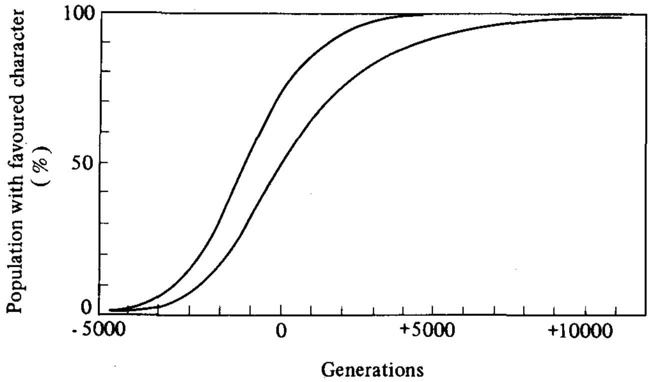

Figure 3. Effect of selection on a sex-linked character. $k=0.001$. Dominants favoured. Upper curve, homozygous sex; lower curve, heterozygous sex.

From mating	AA ♀	Aa ♀	aa ♀	A ♂	a ♂
$AA \times A$	$p_n s_n$	0	0	$p_n s_n$	0
$Aa \times A$	$q_n s_n$	$q_n s_n$	0	$\frac{2q_n s_n}{2-k}$	$\frac{2(1-k)q_n s_n}{2-k}$
$aa \times A$	0	$r_n s_n$	0	0	$r_n s_n$
$AA \times a$	0	$p_n t_n$	0	$p_n t_n$	0
$Aa \times a$	0	$\frac{2q_n t_n}{2-k}$	$\frac{2(1-k)q_n t_n}{2-k}$	$\frac{2q_n t_n}{2-k}$	$\frac{2(1-k)q_n t_n}{2-k}$
$aa \times a$	0	0	$r_n t_n$	0	$r_n t_n$

Hence, writing $k'=k/2-k$:

$$\left.\begin{aligned} p_{n+1} &= (p_n+q_n)s_n \\ 2q_{n+1} &= (p_n+q_n)t_n+(q_n+r_n)s_n+k'q_n t_n \\ r_{n+1} &= (q_n+r_n)t_n-k'q_n t_n \\ s_{n+1} &= p_n+q_n+k'q_n \\ t_{n+1} &= q_n+r_n-k'q_n \end{aligned}\right\}. \tag{31}$$

With complete selection, when the recessives are eliminated, $k'=k=1$, and:

$$r_{n+1}=r_n t_n,$$
$$t_{n+1}=r_n,$$
$$\therefore\ r_n=r_0^{\phi(n+1)}t_0^{\phi(n)},$$

where:

$$\phi(n)=\frac{2^{-n}}{\sqrt{5}}\left[(1+\sqrt{5})^n-(1-\sqrt{5})^n\right], \tag{32}$$

i.e. $\phi(n)$ is the n^{th} term of Fibonnacci's series 1, 2, 3, 5, 8, 13, 21, etc. So the recessives disappear very fast. When dominants are eliminated $k'=-1$, $k=-\infty$, and the equations are less tractable. The percentages of recessives in succedding generations, starting from a standard population, are:

♀	25	37.5	56.25	66.80	82.97	etc.
♂	50	75	75	89.06	93.16	etc.

When k is small we solve as in equation (11), and find:

$$\Delta u_n=\frac{k'u(2+u)}{3+3u}\text{ approximately.}$$

$$\left.\begin{aligned} \therefore\ kn&=3\log_e\left[\frac{u_n(2+u_n)}{3}\right] \\ r_n&=(1+u_n)^{-2} \\ t_n&=(1+u_n)^{-1} \end{aligned}\right\}, \tag{33}$$

starting from the standard population, and p_n, q_n have very nearly the values for a population in equilibrium. Selection therefore proceeds much as in racial selection but at from a half to a third of the rate. Some figures are given in Table 5.

12. Selection of a Sex-Linked Character in the Homozygous Sex Only. Several sex-linked factors are known which have a much more marked effect on the homozygous than the heterozygous sex. Thus in *Drosophila melanogaster* "fused" females are sterile, males fertile; whilst the character "dot" occurs in 8% of the genetically recessive females, but only 0.8% of the males (Morgan and Bridges, 1916). But the chief evolutionary importance of this type of selection must have been in the Hymenoptera and other groups where the males are haploid and all amphimictic inheritance sex-linked. The characters of the diploid females and neuters are generally more important (especially in the social species) than those of the males. On the other hand it must be remembered that for a few drone characters selection must be very intense, and largely familial. Using the usual notation:

$$\left.\begin{aligned} u_{n+1}&=\frac{2u_nv_n+u_n+v_n}{u_n+v_n+2-2k} \\ v_{n+1}&=u_n \end{aligned}\right\}. \tag{34}$$

With complete selection, if all dominants are eliminated and $k = -\infty$, all dominants disappear in two generations. If all recessives are eliminated $k = 1$, and starting with a standard population the percentages of recessives in successive generations are:

♂ (heterozygous sex)	50	33.3	30	23.2	21.4	18.6
♀ (homozygous sex)	25	16.7	10	6.96	5.14	3.95.

So elimination is vastly slower than when selection occurs in both sexes [equation (29)]. When k is small we solve as in equation (27), and find:

$$3\Delta u_n = \frac{2ku_n}{1+u_n} \text{ approximately.}$$

$$\therefore\ \tfrac{2}{3}kn = u_n + \log_e u_n - 1, \tag{35}$$

$$\left.\begin{aligned} y_n &= (1+u_n)^{-2} \\ z_n &= (1+u_n)^{-1} \end{aligned}\right\}.$$

So selection of the homozygotes proceeds as in Fig. 2 and Tables 1 and 2, but 1.5 times as many generations are needed for a given change. The heterozygous sex changes rather more slowly.

13. Selection of a Sex-Linked Character in the Heterozygous Sex Only. In certain cases sex-linked factors appear only or mainly in the heterozygous sex. Thus in *Drosophila melanogaster* "eosin" eye-colour is far more marked in the male than the female, and the sex-linked fertility factor L_2 postulated by Pearl (1912) in poultry can only show in the female sex. If selection is limited to the heterozgous sex:

$$\left.\begin{aligned} u_{n+1} &= \frac{2u_n v_n + u_n + v_n}{u_n + v_n + 2} \\ v_{n+1} &= \frac{u_n}{1-k} \end{aligned}\right\}. \tag{36}$$

With complete selection, if all dominants are eliminated, $k = -\infty$, and:

$$u_{n+1} = \frac{u_n}{2+u_n} \text{ (after the second generation),}$$

$$v_n = 0.$$

$$\therefore\ u_n = \left[2^{n-1}\left(1+\frac{1}{u_1}\right)-1\right]^{-1}, \text{ but } u_1 = u_0;$$

$$\therefore\ y_n = z_n = 1 - 2^{1-n} z_0. \tag{37}$$

So the number of dominants is halved in each generation after the second. If recessives are eliminated, $k=1$, and:

$$u_{n+1}=1+2u_n \text{ (after the second generation),}$$

$$v_{n+1}=\infty.$$

$$\therefore\; u_n=2^{n-1}(1+u_1)-1;$$

$$\therefore\; \left.\begin{aligned} y_n&=0 \\ z_n&=2^{1-n}z_0 \end{aligned}\right\}, \tag{38}$$

the proportion of recessives being halved in each generation. If selection is slow

$$\Delta u_n=\tfrac{1}{3}ku_n \text{ approximately;}$$

$$\therefore\; kn=3\log u_n, \tag{39}$$

if $u_0=1$; and y_n, z_n are given by equation (30). Hence selection in the heterozygous sex proceeds as in Fig. 1, but at one-third of the pace, whilst selection in the homozygous sex is slightly faster.

14. Certation, or Gametic Selection of an Autosomal Character. The work of Renner (1917) and Heribert-Nilsson (1920) shows that gametes or gametophytes may be selected according to what factors they carry. The field of such selection may be wide, as in wind pollination, but is more often restricted, and mainly familial, i.e. among the gametes of the same individual. Except in homosporous plants the intensity must be different in gametes of different genders, and we shall here only consider the case where one is affected. Let the n^{th} generation be formed from gametes carrying $u_nA\colon 1a$, this proportion being reduced by selection in one gender to $u_nA\colon(1-k)a$, the selection being general and not familial. Then the n^{th} generation will be $u_n^2AA\colon(2-k)u_nAa\colon(1-k)aa$.

$$\therefore\; u_{n+1}=\frac{u_n(2u_n+2-k)}{(2-k)u_n+2-2k}. \tag{40}$$

With complete selection, if all dominant-carrying gametes are eliminated, $k=-\infty$, and:

$$u_{n+1}=\frac{u_n}{2+u_n};$$

$$\therefore\; y_n=\frac{1}{1+u_n}=1-2^{1-n}(1-y_1). \tag{41}$$

So the proportion of dominants is halved in each generation. If recessivecarriers are eliminated, no recessive zygotes appear, and the proportion of heterozygotes is halved in each generation. If selection is slow:

$$\begin{aligned} \Delta u_n &= \tfrac{1}{2}ku_n \text{ approximately;} \\ \therefore\ kn &= 2\log_e u_n, \qquad (42) \\ y_n &= (1+u_n)^{-2}. \end{aligned}$$

If the gametes are replaced in heterozygous organisms, as must happen in a large batch of pollen-grains or spermatozoa from the same source, let the n^{th} zygotic generation be formed from unselected gametes (say megagametes) $u_nA:1a$, and selected gametes (say microgametes) $v_nA:1a$, so its proportions are $u_nv_nAA:(u_n+v_n)Aa:1aa$.

$$\left.\begin{aligned} \therefore\ u_{n+1} &= \frac{2u_nv_n+u_n+v_n}{u_n+v_n+2} \\ v_{n+1} &= \frac{2u_nv_n+u_n+v_n+k'(u_n+v_n)}{u_n+v_n+2-k'(u_n+v_n)} \end{aligned}\right\}, \qquad (43)$$

where $k'=k/(2-k)$, as in equation (31). With complete selection (when for example there is a very great disparity between growth-rates of pollen-tubes, though both types are viable), if dominant gametes are eliminated, $k'=-1$, and the percentages of recessive zygotes in successive generations, starting from a standard population, are:

25, 37.5, 54.69, 71.48, 84.16, 91.83, etc.

If recessive gametes are eliminated, $k'=1$, and the percentages of recessive zygotes in successive generations are:

25, 12.5, 4.56, 1.14, 0.17, 0.014, etc.

When selection is slow, $\Delta = \tfrac{1}{4}ku_n$ approximately.

$$\therefore\ kn = 4\log_e u_n, \qquad (44)$$

if $u_0=1$, so selection proceeds at half the rate given by equation (42), y_n being given by equation (4).

15. Gametic Selection of a Sex-Linked Character. This is not known to occur, and at all complete gametic selection is very unlikely in animals, so we need only consider slow selection. Let selection occur among the gametes of the homozygous sex, with no replacement within heterozygous organisms. Let the n^{th} generation be formed from eggs in the ratio $u_nA:1a$ before, or $u_nA:(1-k)a$

after selection, and ♀-producing spermatozoa in the ratio $v_nA : 1a$. Then the n^{th} generation is:

$$♀\ u_nv_nAA : (u_n + v_n - kv_n)Aa : (1-k)aa; \qquad ♂\ u_nA : (1-k)a.$$

$$\left.\begin{aligned} \therefore\ u_{n+1} &= \frac{2u_nv_n + u_n + v_n - kv_n}{u_n + v_n + 2 - kv_n - 2k} \\ v_{n+1} &= \frac{u_n}{1-k} \end{aligned}\right\}; \tag{45}$$

$$\therefore\ \Delta u_n = \tfrac{2}{3}ku_n \text{ approximately.}$$

$$\therefore\ \tfrac{2}{3}kn = \log_e u_n, \tag{46}$$

whilst y_n and z_n are given by equation (30), so selection proceeds twice as fast as in equation (39). In the more important case of familial selection (replacement in heterozygous individuals), if $k' = k/(2-k)$, then:

$$\left.\begin{aligned} u_{n+1} &= \frac{2u_nv_n + u_n + v_n + k'(u_n + v_n)}{u_n + v_n + 2 - k'(u_n + v_n)} \\ v_{n+1} &= u_n, \end{aligned}\right\}, \tag{47}$$

u_n being here the gametic ratio after selection.

$$\therefore\ \Delta u_n = \tfrac{2}{3}k'u_n = \tfrac{1}{3}k'v_n;$$

$$\therefore\ \tfrac{1}{3}kn = \log_e u_n, \tag{48}$$

so selection proceeds as in equation (39).

If selection occurs among the gametes of the heterozygous sex there is clearly no effect if they are replaced, whilst otherwise the effects are the same as those of zygotic selection, and are given by equation (39).

16. Comparative Results of Complete (Artificial) Selection. The results of complete selection in the more important cases are summarized in Table 4. In every case the field of selection considered is the whole population. Complete familial selection occasionally occurs through natural causes, but never through human agency. Column 3 gives the sex to which the numbers in columns 4 and 5 refer. Selection is supposed to begin on a population in equilibrium containing equal numbers of dominants and recessives of the sex considered. It is worth noting that in the case of sex-linked characters, and autosomal recessives when selection is gametic, individuals of types which have wholly disappeared reappear if selection ceases. With many types of heredity

Table 4. Effects of complete selection

Character eliminated	Type of selection	Sex	% After 5 generations from 50%	% After 10 generations from 50%	Equation
Non-amphimietic	Any	Both	0	0	—
Autosomal dominant	Bisexual	Both	0	0	—
Autosomal recessive	Bisexual	Both	2.44	0.768	6
Autosomal dominant	Unisexual	Both	1.83	0.0572	17
Autosomal recessive	Unisexual	Both	8.88	3.27	—
Sex-linked dominant	Bisexual	Homozygous	0	0	—
		Heterozygous	0	0	—
Sex-linked recessive	Bisexual	Homozygous	0	0	28
		Heterozygous	1.56	0.0484	28
Sex-linked dominant	In homozygous sex	Homozygous	0	0	—
		Heterozygous	0	0	—
Sex-linked recessive	In homozygous sex	Homozygous	5.34	1.74	—
		Heterozygous	18.5	13.28	—
Sex-linked dominant	In heterozygous sex	Homozygous	1.83	0.0572	37
		Heterozygous	3.125	0.0977	37
Sex-linked recessive	In heterozygous sex	Homozygous	0	0	—
		Heterozygous	3.125	0.0977	38
Autosomal dominant	Gametic unisexual	Both	1.83	0.0572	41
Autosomal recessive	Gametic unisexual	Both	0	0	—

dominants are eliminated in one or two generations, and where this is not the case they generally decrease more rapidly than recessives.

17. Applications to Slow Selection. With the exception of equation (2) the equations found for the rate of slow selection are not rigorously accurate. n is in general a higher transcendental function of u, but of what nature is not clear. It will be shown later that the finite difference equations found in this paper are special cases of integral equations which may possibly prove more tractable. The values for kn found in terms of u all have inexactitudes of the order k^2n. Thus if one type has an advantage of 1%, the number of generations required for a given change can also be found within about 1%.

Table 5 shows the effect of slow selection in the various cases considered. The third column gives the sex to which the subsequent figures apply. Selection is throughout supposed to give the favoured type an advantage of $\frac{1}{1000}$, i.e. 1000 of this type survive for 999 of the other. If the advantage is $\frac{1}{100}$, one-tenth of the number of generations is required for a given change, and so on, but when selection is very rapid the numbers are somewhat inaccurate.

It will at once be seen that selection is most rapid when amphimixis is avoided by any of the means cited above. Moreover selection is ineffective on recessive characters when these are rare, except in the case of sex-linked factors, when selection is effective in the heterozygous sex, and in gametic selection. It seems therefore very doubtful whether natural selection in random-mating organisms can cause the spread of autosomal recessive characters unless they are extraordinarily valuable to their possessors. Such characters appear far more frequently than dominant mutations, but in their early stages are selected infinitely more slowly. It is thus intelligible that none of the melanic varieties of Lepidoptera which are known to have spread should be recessive.

There are at least four ways out of this impasse.

(1) In a species which adopts self-fertilization or very close inbreeding advantageous autosomal recessive characters can spread rapidly. Thus supposing that in each of two otherwise similar species, one of which is mainly self-fertilizing, an advantageous recessive mutation occurs, it will spread far more quickly in the self-fertilizing species, and this species will tend to replace the other. This fact may well account for the widespread presence of self-fertilization and close inbreeding, in spite of the fact that they seem often to be physiologically harmful, and must certainly check the combination of useful variations which have arisen independently.

(2) Recessives may be helped to spread by assortative fertilization. This may take place in the following ways.

Psychological isolation. Thus Pearson and Lee (1903) have shown that a tall man is more likely to marry a tall woman than a short woman if presented with equal numbers of each. Of course the recessives must not

Table 5. Generations required for a given change with various types of slow selection ($k = 0.001$)

Dominant factor favoured	Type of selection	Sex	0.001–1%	1–50%	50–99%	99–99.999%	Equations
Non-amphimictic	Bisexual racial	Both	6 921	4 592	4 592	6 921	2
Autosomal	Bisexual racial	Both	6 920	4 819	11 664	309 780	4, 8
Autosomal	Unisexual racial Bisexual familial	Both	13 841	9 638	23 328	619 560	4, 13, 18
Autosomal	Bisexual familial*	Both	9 227	6 425	15 522	413 040	4, 14
Autosomal†	Bisexual racial	♂	13 831	8 819	6 157	7 112	4, 22
Sex-linked	Bisexual racial	Homozygous	6 916	4 668	5 593	10 106	29, 30
		Heterozygous	6 928	5 164	11 070	20 693	
Sex-linked	Bisexual familial	Homozygous	20 753	13 785	13 785	20 753	33
		Heterozygous	20 768	14 987	24 332	41 450	
Sex-linked	Racial of homozygous sex	Homozygous	10 380	7 228	17 496	464 670	30, 35
		Heterozygous	10 392	8 378	153 893	149 860 377	
Sex-linked	Racial of heterozygous sex	Homozygous	20 746	13 228	9 236	10 668	30, 39
		Heterozygous	20 753	13 785	13 785	20 753	
Autosomal	Unisexual gametic	Both	13 831	8 819	6 157	7 112	4, 42
Autosomal	Unisexual gametic‡	Both	27 661	17 638	12 314	14 224	4, 44
Sex-linked	Gametic of homozygous sex	Homozygous	10 373	6 619	4 618	5 334	30, 46
		Heterozygous	10 377	6 892	6 892	10 377	
Sex-linked	Gametic of homozygous sex‡	Homozygous	20 746	13 228	9 236	10 668	30, 48
		Heterozygous	20 753	13 785	13 785	20 753	

*The families have only one parent in common.

†Dominant in ♂, recessive in ♀.

‡In heterozygous individuals gametes are replaced (as zygotes *in familial* selection).

The effect of selection on recessive characters may be found by inverting the order of the four numerical columns. Thus 309 780 generations are needed for an autosomal recessive to increase from 0.001 to 1%, 11 664 generations to increase from 1 to 50%, and so on.

be so repulsive to the dominants as to escape mating altogether at first. In plants psychological isolation may be due to the psychology of the insect or other pollinating organism. Thus a mutant plant with a new colour, scent, or shape may be isolated because it attracts a different insect from the type plant.

Anatomical isolation. Pearl (1907) and Crozier and Snyder (1922) have shown that in *Paramoecium* and *Gammarus* there is a strong tendency for organisms of like size to mate. This will be effective provided mutations are not so great as to leave the first mutants unmated.

Temporal isolation. If the recessive factor causes (or is very closely coupled with a factor causing) a change in the breeding or flowering time, this will serve as an effective barrier against crossing.

Spatial isolation. If the recessive has a different habitat, e.g. a different range of soil or temperature conditions to which it is adapted, some of its individuals will be spatially isolated from the dominants.

Selective fertilization. If the results of Jones (1920) are due to this cause, as seems almost certain, we have here a *vera causa*, though it must be remembered that he did not work with single factor-differences. He found that when either of two races of maize is fertilized with a mixture of pollen the proportion of hybrids was less than was to be expected from random fertilization. This does not seem to have been due to inviability of the hybrids, which were more vigorous and fertile than the parent races. Clearly if the hybrid zygotes are inviable or sterile the rarer form of the species will be weeded out whether it is dominant or recessive, weak or vigorous. But if there is selective fertilization due for example to increased activity of pollen-tubes in tissue of the type which produced them, the increase of the rare form, especially if it is recessive, will be facilitated.

All these types of isolation, then, will favour the replacement of a type species by a recessive mutant. May it not be that in many cases mutual infertility is the cause and not the effect of specific differences? A new mutant form arises within a species. If it crosses freely with the type we call it a variety, and a moderately advantageous recessive variety will only spread very slowly indeed. But if it does not cross freely we call it a new species, and it is much more likely to establish itself. Possibly then interspecific sterility is partly to be explained by its having a selective value.

(3) The increase of recessives is greatly facilitated, as will be shown later, by incomplete dominance. Thus if there is only one recessive in a million, and the recessives have an advantage of 0.001, their rate of increase will be speeded up elevenfold if the heterozygotes have an advantage of 0.00001 over the pure dominants.

(4) If heterozygotes have any advantage as such this will tend to favour any new factors so long as they are rare. But no "stimulus of heterozygosis" has yet been demonstrated in cases of single factor-differences.

Whether the isolation of small communities, or what comes to much the same thing, great immobility of individuals at all stages of their lives, will help or hinder the spread of a new recessive type in the species as a whole is a nice question. It will certainly slow the spread of a dominant.

At first sight the selection of dominant factors would not seem to be a probable cause of the origin of species rather than new varieties. But it must be remembered that dominant mutations are very often lethal in the homozygous condition. Under certain circumstances, to be discussed later, their selection may lead to the establishment of a system of balanced lethals, and a probable change in the chronosome number.

The theory so far developed gives a quantitative account of the spread of a new advantageous type within a population under certain simple conditions, and demonstrates that inbreeding, homogamy, and inter-varietal sterility may sometimes be of selective value, and therefore preserved by natural selection. It is proposed in later papers to discuss the selection of semi-dominant, multiple, linked, and lethal factors, partial inbreeding and homogamy, overlapping generations, and other complications.

LITERATURE

Barrett. 1901. *Lepidoptera of the British Islands*, Vol. 7, p. 130.
Bate. 1895. *Ent. Rev.*, 27.
Bateson. *Mendel's Principles of Heredity*, p. 221.
Crozier and Snyder. 1922. *Proc. Soc. Exp. Biol. Med.* **19**, 327.
Detlefsen. 1915. *Genetics* **3**, 573.
Durham. 1911. *J. Genetics* **1**, 107.
Goldschmidt. 1920. *Z. indukt. Abstamm.-u. Verblehre* **23**, 1.
Hardy. 1908. *Science* **28**, 49.
Heribert-Nilsson. 1920. *Hereditas* **1**, 41.
Ibsen and Steigleder. 1917. *Am. Nat.* **51**, 740.
Jennings. 1915. *Genetics* **1**, 53.
Jones. 1920. *Biol. Bull.* **38**, 251.
Morgan and Bridges. 1916. *Carn. Inst. Wash. Pub.*, 237.
Pearl. 1907. *Biometrika* **5**, 213.
Pearl. 1912. *Am Nat.* **46**, 130.
Pearson. 1904. *Phil. Trans. R. Soc.* **A203**, 53.
Pearson. 1908. *Proc. R. Soc.* **54**, 72.
Pearson and Lee. 1903. *Biometrika* **2**, 371.
Punnett. *Mimicry in Butterflies*, 154.
Punnett. 1917. *J. Hered.* **8**, 464.
Renner. 1917. *Z. indukt. Abstamm.-u. Verblehre* **18**, 121.
Schimidt. 1920. *C. r. Trav. Lab. Carlsberg*, 14.
Warren. 1917. *Genetics* **2**, 305.
Wood. 1905. *J. Agric. Science* **1**, 364.

A MATHEMATICAL THEORY OF NATURAL AND ARTIFICIAL SELECTION. PART II

THE INFLUENCE OF PARTIAL SELF-FERTILISATION, INBREEDING, ASSORTATIVE MATING, AND SELECTIVE FERTILISATION ON THE COMPOSITION OF MENDELIAN POPULATIONS, AND ON NATURAL SELECTION.

By J. B. S. HALDANE, M.A.
Trinity College, Cambridge.

(*Received* 1 August 1924.)

In the first paper (1) of this series expressions were found for the effect of natural selection of small and constant intensity on Mendelian populations whose generations do not overlap; either during random mating, or when all zygotes are self-fertilised. An intermediate condition as regards mating may arise when there is a tendency to self-fertilisation, to mating between relatives, or to unions between similar but not necessarily related zygotes or gametes. We consider a population whose mth generation consists of $p_m AA : 2q_m Aa : r_m aa$, where A is a completely dominant Mendelian factor, and $p_m + 2q_m + r_m = 1$. When such a population is subjected to any system of mating it falls rapidly or instantly into a new equilibrium. During this process it will be shown that the gametic ratio $u_m = \dfrac{p_m + q_m}{q_m + r_m}$ is unaltered. When equilibrium is reached under the given mating system we find p, q, r in terms of u.

We now suppose selection to take place at such a rate that $(1 - k)$ recessives survive for every dominant, and so slowly that the population is always very nearly in equilibrium under the mating system. If this condition were not fulfilled we should have to investigate the problem by the method of Lotka (2), which in this case presents considerable difficulties. During selection we have

$$u_{n+1} = \frac{p_n + q_n}{q_n + r_n - kr_n},$$

$$\therefore \Delta u_n = \frac{kr_n u_n (1 + u_n)}{1 - kr_n (1 + u_n)}.$$

Since k is small, and $r_n (1 + u_n) = \dfrac{r_n}{q_n + r_n}$ and is therefore less than unity,

$$\therefore \frac{du}{dn} = \Delta u_n = kr_n u_n (1 + u_n) \text{ approximately.}$$

$$\therefore kn = \int_1^{u_n} \frac{du}{ru(1+u)} \qquad \ldots(1),$$

putting $u_0 = 1$, as in Part I. This can be evaluated as r is a known function of u. Under random mating when recessives are few $r_n = \dfrac{1}{(1 + kn)^2}$ approximately, so selection is very slow. It will be shown that with some systems of mating successive small values of r_n approximate to a geometrical series, so that selection is vastly more rapid.

PARTIAL SELF-FERTILISATION.

Let a proportion l of the population be self-fertilised, $(1 - l)$ mated at random, where l may have any value from 0 to 1 inclusive.

$$\begin{aligned} \therefore\ p_{m+1} &= l\left(p_m + \tfrac{1}{2}q_m\right) + (1 - l)\,(p_m + q_m)^2, \\ q_{m+1} &= \tfrac{1}{2}lq_m + (1 - l)\,(p_m + q_m)\,(q_m + r_m), \\ r_{m+1} &= l\left(\tfrac{1}{2}q_m + r_m\right) + (1 - l)\,(q_m + r_m)^2. \end{aligned}$$

Clearly $u_{m+1} = u_m$, and

$$q_m = \frac{2(1 - l)\,u}{(2 - l)(1 + u)^2} + \left(\frac{l}{2}\right)^m \left[q_0 - \frac{2(1 - l)\,u}{(2 - l)(1 + u)^2}\right].$$

So there is a rapid approach to equilibrium, when

$$\left.\begin{aligned} p &= \frac{u\,(l + 2u - lu)}{(2 - l)(1 + u)^2} \\ q &= \frac{2(1 - l)\,u}{(2 - l)(1 + u)^2} \\ r &= \frac{2 - l + lu}{(2 - l)(1 + u)^2} \end{aligned}\right\} \qquad \ldots(2\cdot 1).$$

During selection

$$\left.\begin{aligned} kn &= \int_1^{u_n} \frac{(2 - l)(1 + u)\,du}{u\,(2 - l + lu)} \\ &= \log_e u_n + \frac{2}{l}\log_e\left(\frac{2 - l + lu_n}{2}\right) \\ r_n &= \frac{2 - l + lu_n}{(2 - l)(1 + u_n)^2} \end{aligned}\right\} \qquad \ldots(2\cdot 2),$$

unless $l = 0$, when $kn = u_n + \log_e u_n - 1$.

When recessives are sufficiently few

$$(2 - l)\,r_n = \frac{l}{u_n} = le^{-\frac{lkn}{2 + l}} \text{ approximately,}$$

so $\dfrac{r_n}{r_{n+1}} = 1 + \dfrac{lk}{2 + l}$ approximately, and selection is rapid.

PARTIAL INBREEDING.

Let a proportion l of the population be mated to whole brothers or sisters, $(1 - l)$ mated at random. Let matings occur in the following proportions:

Mating	Proportion	Producing offspring	Matings of inbred offspring
$AA \times AA$	α_m	$\alpha_m AA$	$l\alpha_m (AA \times AA)$
$AA \times Aa$	$4\beta_m$	$2\beta_m (AA + Aa)$	$l\beta_m (AA \times AA + 2AA \times Aa + Aa \times Aa)$
$AA \times aa$	$2\gamma_m$	$2\gamma_m Aa$	$2l\gamma_m (Aa \times Aa)$
$Aa \times Aa$	$16\delta_m$	$4\delta_m (AA + 2Aa + aa)$	$l\delta_m (AA \times AA + 4AA \times Aa + 2AA \times aa + 4Aa \times Aa + 4Aa \times aa + aa \times aa)$
$Aa \times aa$	$4\epsilon_m$	$2\epsilon_m (Aa + aa)$	$l\epsilon_m (Aa \times Aa + 2Aa \times aa + aa \times aa)$
$aa \times aa$	ζ_m	$\zeta_m aa$	$l\zeta_m (aa \times aa)$

Where
$$\alpha_m + 4\beta_m + 2\gamma_m + 16\delta_m + 4\epsilon_m + \zeta_m = 1,$$
$$\therefore\ p_m = \alpha_m + 2\beta_m + \gamma_m,\ p_{m+1} = \alpha_m + 2\beta_m + 4\delta_m,$$
$$q_m = \beta_m + 8\delta_m + \epsilon_m,\ q_{m+1} = \beta_m + \gamma_m + 4\delta_m + \epsilon_m,$$
$$r_m = \gamma_m + 2\epsilon_m + \zeta_m,\ r_{m+1} = 4\delta_m + 2\epsilon_m + \zeta_m,$$
$$\alpha_{m+1} = (1-l)\,{p_{m+1}}^2 + l\,(\alpha_m + \beta_m + \delta_m),$$
$$4\beta_{m+1} = 4\,(1-l)\,p_{m+1}q_{m+1} + 2l\,(\beta_m + 2\delta_m),$$
$$2\gamma_{m+1} = 2\,(1-l)\,p_{m+1}r_{m+1} + 2l\delta_m,$$
$$16\delta_{m+1} = 4\,(1-l)\,{q_{m+1}}^2 + l\,(\beta_m + 2\gamma_m + 4\delta_m + \epsilon_m),$$
$$4\epsilon_{m+1} = 4\,(1-l)\,q_{m+1}r_{m+1} + 2l\,(2\delta_m + \epsilon_m),$$
$$\zeta_{m+1} = (1-l)\,{r_{m+1}}^2 + l\,(\delta_m + \epsilon_m + \zeta_m),$$
$$\therefore\ u_{m+1} = u_m.$$

When equilibrium is reached we can suppress suffixes in the above, and find $\gamma = 4\delta$.

$$\therefore\ (1-l)\,pr = (4-l)\,\delta = (1-l)\,q^2 + \frac{lq}{4}.$$

But
$$pr = \left(\frac{u}{1+u} - q\right)\left(\frac{1}{1+u} - q\right).$$
$$\therefore\ lq = 4\,(1-l)\left(\frac{u}{(1+u)^2} - q\right).$$
$$\left.\begin{aligned} \therefore\ p &= \frac{l + (4-3l)\,u}{(4-3l)\,(1+u)^2} \\ q &= \frac{4\,(1-l)\,u}{(4-3l)\,(1+u)^2} \\ r &= \frac{4-3l+lu}{(4-3l)\,(1+u)^2} \end{aligned}\right\} \qquad \ldots(3{\cdot}1),$$
$$\left.\begin{aligned} \therefore\ kn &= \int_1^{u_n} \frac{(4-3l)\,(1+u)\,du}{u\,(4-3l+lu)} \\ &= \log_e u_n + \frac{4\,(1-l)}{l}\log_e\left(\frac{4-3l+lu_n}{4-2l}\right) \\ r_n &= \frac{4-3l+lu_n}{(4-3l)\,(1+u_n)^2} \end{aligned}\right\} \qquad \ldots(3{\cdot}2),$$
unless $l = 0$, when $kn = u_n + \log_e u_n - 1$.

When recessives are very few,
$$(4-3l)\,r_n = \frac{l}{u_n} = le^{-\frac{lkn}{3l-4}} \text{ approximately,}$$
so $\dfrac{r_n}{r_{n+1}} = 1 + \dfrac{lk}{4-3l}$ approximately, and selection is rapid.

PARTIAL ASSORTATIVE MATING.

We consider a population containing a proportion r of recessives, the sexes being in equal numbers and mating so conducted that while each zygote is mated once and only once in a given period, the probability of a recessive mating with a given recessive is greater than that of its mating with a given dominant, and similarly for dominants. Let θ be the proportion of dominant × recessive and recessive × dominant matings, then that of matings between two dominants is $1 - r - \theta$, between two recessives $r - \theta$,

$$\therefore\ (r - \theta)(1 - r - \theta) = (1 + \lambda)\,\theta^2,$$

where λ is positive. In general λ is a function of r, but since $\dfrac{\lambda}{1 + 2\lambda}$ is the coefficent of association as defined by Yule (3), between the phenotypic characters of spouses, and such coefficients are found to be valuable even when the proportions of the different classes vary greatly, it is probable that λ varies rather little with changes in the population. In a case of human assortative mating given by Yule $\lambda = 0{\cdot}18$.

$$\theta = \frac{\sqrt{1 + 4\lambda r(1 - r)} - 1}{2\lambda},$$

$$\therefore\ p_{m+1} = \frac{(p_m + q_m)^2}{1 - r_m} - \theta_m \left(\frac{p_m + q_m}{1 - r_m}\right)^2,$$

$$q_{m+1} = \frac{q_m(p_m + q_m)}{1 - r_m} + \theta_m \left(\frac{p_m + q_m}{1 - r_m}\right)^2,$$

$$r_{m+1} = r_m + \frac{q_m^2}{(1 - r_m)} - \theta_m \left(\frac{p_m + q_m}{1 - r_m}\right)^2.$$

$\therefore\ u_{m+1} = u_m$; and, at equilibrium,

$$\left.\begin{aligned} p &= \frac{u}{1 + u} - q \\ \lambda(1 + u)^4 q^4 &+ u^2(1 + u)^2 q - u^3 = 0 \\ r &= \frac{1}{1 + u} - q \end{aligned}\right\} \qquad \ldots(4{\cdot}1).$$

During selection

$$\frac{du_n}{dn} = kr_n u_n(1 + u_n),$$

$$\therefore\ \lambda\left(1 - \frac{1}{ku_n}\frac{du_n}{dn}\right)^4 - \frac{u_n(1 + u_n)}{k}\frac{du_n}{dx} + u_n^2 = 0.$$

$$\left.\begin{aligned} \therefore\ kn &= \int_1^{u_n} \frac{du}{u - uf(\lambda, u)} \\ r_n &= \frac{1 - f(\lambda, u_n)}{1 + u_n} \end{aligned}\right\} \qquad \ldots(4{\cdot}2),$$

where $f(\lambda, u)$ is the real positive root of

$$\lambda x^4 + u^2(1 + u)x - u^3 = 0.$$

Clearly

$$\frac{u}{1 + u} \geqslant f(\lambda, u) \geqslant 0,$$

$$\therefore\ |u_n + \log_e u_n - 1| \geqslant |kn| \geqslant |\log_e u_n|$$

and

$$(1 + u_n)^{-1} \geqslant r_n \geqslant (1 + u_n)^{-2}.$$

Hence selection proceeds at a rate intermediate between those of equations (1·2) and (2·3) of Part I. When recessives are few, so that u_n^3 is large compared with λ, $f(\lambda, u) = \frac{u}{1+u}$, approximately, and selection proceeds according to equation (2·3) of Part I. Hence the effect of partial assortative mating in speeding up selection is unimportant.

SELECTIVE FERTILISATION.

If λ has the same meaning as above, except that it applies to unions between gametes and not zygotes, as in Jones' (4) case, where λ was generally less that 100, though in one experiment it exceeded 10,000, equilibrium is reached in one generation, and $pr = (1+\lambda)q^2$,

$$\left.\begin{aligned} \therefore p &= \frac{2\lambda + (1+2\lambda)u - \sqrt{(1+u)^2 + 4\lambda u}}{2\lambda(1+u)} \\ q &= \frac{\sqrt{(1+u)^2+4\lambda u} - 1 - u}{2\lambda(1+u)} \\ r &= \frac{1 + 2\lambda + u - \sqrt{(1+u)^2 + 4\lambda u}}{2\lambda(1+u)} \end{aligned}\right\} \qquad \ldots(5·1).$$

During selection,

$$\begin{aligned} kn &= \int_1^{u_n} \frac{2\lambda\, du}{u\left[1+2\lambda+u-\sqrt{(1+u)^2+4\lambda u}\right]} \\ &= \int_1^{u_n} \frac{1+2\lambda+u+\sqrt{(1+u)^2+4\lambda u}}{2(1+\lambda)u}\, du, \end{aligned}$$

and, if λ be constant,

$$\left.\begin{aligned} kn = \log_e u_n + \frac{1}{2+2\lambda}\Big[u_n - 1 + \sqrt{(1+u_n)^2+4\lambda u_n} - 2\sqrt{1+\lambda} \\ + (1+2\lambda)\log_e\big(1+2\lambda+u_n \\ + \sqrt{(1+u_n)^2+4\lambda u_n}\big) - \log_e\big(1+u_n \\ + 2\lambda u_n + \sqrt{(1+u_n)^2+4\lambda u_n}\big) \\ - 2\lambda\log_e 2\left(1+\lambda+\sqrt{1+\lambda}\right)\Big] \end{aligned}\right\} \qquad \ldots(5·2),$$

$$r_n = \frac{1+2\lambda+u_n-\sqrt{(1+u_n)^2+4\lambda u_n}}{2\lambda(1+u_n)}.$$

Here again selection occurs at a rate intermediate between that of equations (1·2) and (2·3) of Part I, and when recessives are few $r_n = \frac{1+\lambda}{u_n^2} = \frac{1}{(1+\lambda)k^2n^2}$ approximately, so selection is only very slightly more rapid than during random mating.

DISCUSSION.

Effects similar to those produced by partial brother-sister mating may be expected from less drastic types of inbreeding, *e.g.* mating of cousins. Such moderate degrees of inbreeding must occur in any population where neither zygotes nor gametes of both genders are very mobile. When recessives are sufficiently rare any cause which promotes inbreeding, even of distant relatives, will enormously increase their number for a given gametic ratio, and will make u_n tend to vary as r_n^{-1} rather than $r_n^{-\frac{1}{2}}$, making u an exponential function of kn instead of being proportional to it. Assortative mating will have little effect. Thus, if recessives number one in a million, and if only one mating in a hundred is between whole brothers and sisters, more than one recessive in 400 will mate with another recessive. To attain a like result by assortative mating a recessive must be more than 2500 times as likely to mate with a recessive as a dominant. This would imply such obstacles to mating with a dominant that the first recessive to appear could never mate at all. Probabilities of this order may, however, occur in selective fertilisation. Hence inbreeding or self-fertilisation appears to be necessary in the early stages of selection of a recessive character if this process is to be fast enough to be an effective cause of evolution. They cannot be replaced by moderate degrees of selective mating or fertilisation.

SUMMARY.

Expressions (2·1), (3·1), (4·1), (5·1) are found for the composition of Mendelian populations subjected to partial self-fertilisation, inbreeding, assortative mating, or selective fertilisation, and equations (2·2), (3·2), (4·2), (5·2) derived for the effect of selection on such populations. The effect of selection is greatly increased by inbreeding and self-fertilisation.

REFERENCES.

(1) Haldane (1924). *Trans. Camb. Phil. Soc.* **23**, 19.
(2) Lotka (1921). *Proc. Nat. Acad. Sci.* **7**, 168.
(3) Yule (1919). *An Introduction to the Theory of Statistics*, p. 38.
(4) Jones (1920). *Biol. Bull.* **38**, 251.

A mathematical theory of natural and artificial selection. Part III. By Mr J. B. S. HALDANE, Trinity College.

[*Received* 9 July, *read* 26 July 1926.]

In this part the cases of a single but incompletely dominant factor, and of several interacting factors are considered. Mating is supposed to be at random, populations to be very large, and generations not to overlap. The notation is, so far as possible, that of Part I (1).

Selection of an incompletely dominant autosomal character.

Let the nth generation be formed from female gametes in the ratio $u_nA : 1a$, male gametes in the ratio $v_nA : 1a$. The nth generation is therefore in the proportions $u_nv_nAA : (u_n + v_n)Aa : 1aa$. Let the ratios after selection has occurred be:

$$\begin{aligned} ♂\quad & u_nv_nAA : (1 - K_m)(u_n + v_n)Aa : (1 - k_m)aa, \\ ♀\quad & u_nv_nAA : (1 - K_f)(u_n + v_n)Aa : (1 - k_f)aa, \end{aligned}$$

where K_m, K_f, k_m, k_f are small.

$$\therefore \quad u_{n+1} = \frac{2u_nv_n + (1 - K_f)(u_n + v_n)}{(1 - K_f)(u_n + v_n) + 2 - 2k_f}.$$

Hence, since $\frac{u_n - v_n}{u_n}$ is clearly small,

$$\Delta u_n = \frac{v_n - u_n}{2} + \frac{u_n(K_fu_n - K_f + k_f)}{1 + u_n}, \text{ approximately,}$$

and $$\Delta v_n = \frac{u_n - v_n}{2} + \frac{u_n(K_mu_n - K_m + k_m)}{1 + u_n}, \text{ approximately.}$$

Δu_n and Δv_n can be shewn to differ by a small quantity of the second order.

$$\therefore \quad \Delta u_n = \frac{u_n(Ku_n - K + k)}{1 + u_n} \quad \ldots\ldots\ldots\ldots\ldots (1{\cdot}0)$$

where $$K = \tfrac{1}{2}(K_f + K_m); \; k = \tfrac{1}{2}(k_f + k_m).$$

Equilibrium can only occur when $\Delta u_n = 0$, i.e. u_n tends either to zero, infinity, or to $1 - \frac{k}{K}$. Hence for equilibrium to be possible $\frac{k}{K} < 1$. If K be positive, i.e. heterozygotes are at a disadvantage compared with pure dominants, then $\Delta u_n \gtrless 0$ according as $u_n \gtrless 1 - \frac{k}{K}$. Hence the equilibrium is unstable if it exists. If K

be negative the equilibrium is stable if it exists. We have thus three cases to consider. In each

$$\frac{du_n}{dn} = \frac{u_n (Ku_n - K + k)}{1 + u_n} \text{ approximately,}$$

and the proportion of recessives $y_n = (1 + u_n)^{-2}$.

(*a*) No equilibrium, $\frac{k}{K} > 1$.

$$\therefore \quad (k - K) n = \log_e u_n + \frac{k - 2K}{K} \log_e \left(\frac{Ku_n - K + k}{k}\right) \quad \ldots (1{\cdot}1)$$

making the usual convention that $u_0 = 1$.

Hence the values of u_n lie between two geometrical series, and selection is therefore vastly more efficacious on recessives than when dominance is complete, as in equations 2·4 and 4·3 of Part I.

(*b*) Stable equilibrium, $k > K$, $0 > K$.

$$\therefore \quad (K - k) n = \log_e \left(\frac{u_n}{u_0}\right) + \frac{2K - k}{K} \log_e \left(\frac{Ku_n - K + k}{Ku_0 - K + k}\right) \quad \ldots (1{\cdot}2).$$

We must take $u_0 \gtrless 1 - \frac{k}{K}$ according as $u_n \gtrless 1 - \frac{k}{K}$.

Here again successive values of u_n lie between two geometrical series, so that the population proceeds fairly rapidly towards equilibrium. As Fisher (2) has pointed out, such cases probably occur in nature in connexion with factors governing size, where the heterozygote is at an advantage as compared with either type of homozygote.

(*c*) Unstable equilibrium, $K > 0$, $K > k$.

The population proceeds towards homozygosis in one direction or the other. This case can hardly occur in nature, as any mutants, either in an AA or an aa population, would be weeded out while still few in number.

Selection of an incompletely dominant sex-linked character.

The female sex is throughout supposed to be homogametic; if the male is homogametic the argument is the same, *mutatis mutandis*. Let the nth generation be formed from ova in the ratio $u_n A : 1a$, female-producing spermatozoa in the ratio $v_n A : 1a$. Let the ratios of the nth generation after selection be:

$$♀ \quad u_n v_n AA : (1 - K)(u_n + v_n) Aa : (1 - k)\, aa,$$
$$♂ \quad u_n A : (1 - k')\, a,$$

where K, k and k' are small.

$$\therefore \quad u_{n+1} = \frac{2u_n v_n + (1-K)(u_n + v_n)}{(1-K)(u_n + v_n) + 2 - 2k},$$

$$v_{n+1} = \frac{u_n}{1-k'}.$$

$$\therefore \quad \Delta u_n = \Delta v_n = \frac{u_n}{3+3u_n}[(2K+k')u_n - 2K + 2k + k'],$$

approximately, ...(2·0)

and $u_n = v_n$, approximately.

Hence u_n tends to zero, infinity, or $\frac{2K-2k-k'}{2K+k'}$. Equilibrium is possible if $\frac{k+k'}{2K+k'} < \frac{1}{2}$. It is stable if $2K+k'$ be negative, unstable if this quantity be positive. In each case

$$\frac{du_n}{dn} = \frac{u_n(2Ku_n + k'u_n - 2K + 2k + k')}{3(1+u_n)}, \text{ approximately},$$

and the proportion of recessive females is $(1+u_n)^{-2}$, of recessive males $(1+u_n)^{-1}$. Three cases occur.

(*a*) No equilibrium, $\frac{k+k'}{2K+k'} > \frac{1}{2}$.

$$\therefore \quad \frac{2k+k'-2K}{3}n = \log_e u_n + \frac{2k-4K}{2K+k'}\log_e\left(\frac{2Ku_n+k'u_n-2K+2k+k'}{2k+2k'}\right)$$

......(2·1),

putting $u_0 = 1$. Selection therefore proceeds much as according to equation 7·2 of Part I.

(*b*) Stable equilibrium, $0 > 2K+k'$, $2k+k' > 2K$.

$$\therefore \quad \frac{2K-2k-k'}{3}n = \log_e\left(\frac{u_n}{u_0}\right) + \frac{2k-4K}{2K+k'}\log_e\left(\frac{2Ku_n+k'u_n-2K+2k+k'}{2Ku_0+k'u_0-2K+2k+k'}\right)$$

......(2·2)

where $u_0 \gtrless u_\infty$ according as $u_n \gtrless u_\infty$.

The results of Robertson (3) suggest that milk-yield in cattle depends on one or more sex-linked factors which act most effectively when heterozygous, besides autosomal factors. If so human effort in this case has given K a negative value, while k and k' are nearly zero. Hence an equilibrium should be reached.

(*c*) Unstable equilibrium, $2K+k' > 0$, $2K > 2k+k'$.

The population proceeds in one direction or the other to homozygosis. This case can hardly occur in nature.

Multiple factors.

Many cases exist in nature where several factors are needed to ensure the appearance of a character. Thus in wheat Nilsson-Ehle (4) found that any one of three dominant factors will produce redness, that is to say a white plant must be a triple recessive. On the other hand Saunders (5) found that in Matthiola three dominant factors are needed for slight hoariness, four for complete hoariness, so that a hoary plant is a multiple dominant. In other cases the effects of factors are merely additive, and selection will act on each independently of the others. It will be shown later that linkage between factors, unless very strong, is unimportant. We shall therefore at first consider unlinked factors, and shall confine ourselves to the case of complete dominance.

Selection of a multiple autosomal recessive character.

Let $A_1, A_2, \ldots A_r, \ldots A_m$ be m autosomal dominant factors, each of which produces the same effect, so that the multiple recessive alone competes with the other genotypes. Let $1-k$ of this type survive for every one of the others. Let y_n be the proportion of multiple recessives in the nth generation, formed from gametes in the proportion ${}_ru_n A_r : 1 a_r$, and similarly for the other factor pairs. Then the nth generation consists of zygotes in the ratios:

$$ {}_ru_n{}^2 A_r A_r : 2\,{}_ru_n A_r a_r : 1 a_r a_r, \text{ etc.} $$

$$ \therefore \quad y_n = \prod_{r=1}^{m} (1 + {}_ru_n)^{-2}. $$

Of the $a_r a_r$ zygotes only $y_n (1 + {}_ru_n)^2$ are multiple recessives. Hence

$$ {}_ru_{n+1} = \frac{{}_ru_n (1 + {}_ru_n)}{1 + {}_ru_n - k y_n (1 + {}_ru_n)^2}. $$

$\therefore \quad \Delta_r u_n = k y_n\, {}_ru_n (1 + {}_ru_n)$, approximately, if k be small.

Putting $x = kn$ we have approximately:

$$ \left. \begin{array}{l} \dfrac{d\,{}_1u_n}{dx} = y_n\, {}_1u_n (1 + {}_1u_n) \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ \dfrac{d\,{}_ru_n}{dx} = y_n\, {}_ru_n (1 + {}_ru_n) \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ \dots\dots\dots\dots\dots\dots\dots\dots \end{array} \right\}. $$

To eliminate the u's from these m equations put $s_r = \dfrac{{}_ru_n}{1 + {}_ru_n}$.

$$ \therefore \quad \frac{ds_r}{dx} = y_n s_r. \qquad \therefore \quad \log s_r = \int y_n dx + \log a_r. $$

$\therefore\ s_r = a_r s$, where a_r is an integration constant independent of n and given by the initial state of the population.

$$\therefore\ \left.\begin{aligned} y_n &= \prod_{r=1}^{m} (1 - a_r s)^2 \\ kn = x &= \int \frac{ds}{s y_n} \end{aligned}\right\} \qquad \ldots\ldots\ldots\ldots\ldots(3{\cdot}0).$$

The latter equation is integrable, and the elimination of s gives the required relation between y_n and kn.

$$\frac{dy_n}{dx} = -\,2y_n^2 \sum_{r=0}^{m} {}_r u_n,$$

whereas if only one factor is concerned,

$$\frac{dy_n}{dx} = -\,2y_n^2 u_n.$$

Now comparing these rates for equal values of y_n in the two cases, we note that since $y_n^{-\frac{1}{2}} = 1 + u_n = \prod_{r=1}^{m} (1 + {}_r u_n)\ \therefore\ u_n > \sum_{r=1}^{m} {}_r u_n$. Hence selection is slower than in the case of a character determined by one factor only. When however dominants are very rare, or when one a_r greatly exceeds the rest, i.e. one recessive factor is far commoner than the others, selection proceeds at about the same rate in the two cases. It is slowest when all the a_r's are equal.

Selection of a multiple sex-linked recessive character.

If $A_1, A_2, \ldots A_r, \ldots A_m$ are sex-linked (the female being homogametic) the nth generation formed from eggs in the ratios ${}_r u_n A : 1a$, etc., and female-producing spermatozoa in the ratios ${}_r v_n A : 1a$, etc., while z_n is the proportion of multiple recessive males, y_n of such females, and k is the coefficient of selection.

$$\therefore\ {}_r u_{n+1} = \frac{2\,{}_r u_n\,{}_r v_n + {}_r u_n + {}_r v_n}{{}_r u_n + {}_r v_n + 2 - 2k y_n (1 + {}_r u_n)}.$$

$${}_r v_{n+1} = \frac{{}_r u_n}{1 - k z_n (1 + {}_r u_n)}.$$

$$y_n = \prod_{r=1}^{m} (1 + {}_r u_n)^{-1} (1 + {}_r v_n)^{-1}.$$

$$z_n = \prod_{r=1}^{m} (1 + {}_r u_n)^{-1}.$$

$\therefore$ Approximately ${}_r u_n = {}_r v_n,\ y_n = z_n^2$.

$$\therefore\ 3\Delta\,{}_r u_n = k\,{}_r u_n (1 + {}_r u_n)(2z_n^2 + z_n).$$

As above, putting $kn = x$, $\frac{{}_ru_n}{1 + {}_ru_n} = a_r s$, we have a_r constant, and

$$\left.\begin{aligned} z_n &= \prod_{r=1}^{m} (1 - a_r s) \\ kn &= x = 3 \int \frac{ds}{s z_n (2z_n + 1)} \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots (4{\cdot}0).$$

This again is soluble in finite terms by the elimination of s.

$$\frac{dz_n}{dx} = -\tfrac{1}{3} z_n^2 (2z_n + 1) \sum_{r=1}^{m} {}_ru_n,$$

while in the single factor case

$$\frac{dz_n}{dx} = -\tfrac{1}{3} z_n^2 (2z_n + 1) u_n, \text{ where } u_n = z_n^{-1} - 1.$$

Comparing these rates for equal values of z_n, we find as above $u_n > \Sigma\, {}_ru_n$. Hence selection proceeds more slowly with many factors than with one. When, however, dominants are very rare or one a_r much larger than the rest, selection proceeds as with one factor.

Selection of a multiple autosomal dominant character.

When each of m autosomal dominant factors is needed to produce a character, we find, using the same notation as above except that y_n is the proportion of dominants,

$$y_n = \prod_{r=1}^{m} [1 - (1 + {}_ru_n)^{-2}],$$

$$\frac{d\,{}_ru_n}{dx} = \frac{y_n (1 + {}_ru_n)}{2 + {}_ru_n}, \text{ with } m - 1 \text{ similar equations.}$$

$$\therefore \quad (1 + {}_ru_n)\, e^{1 + {}_ru_n} = e^{\int y_n dx} = a_r s.$$

Hence the problem can be reduced to the elimination of s between:

$$\left.\begin{aligned} y_n &= \prod_{r=1}^{m} [1 - \{\phi (a_r s)\}^{-2}] \\ kn &= x = \int \frac{ds}{sy} \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots (5{\cdot}0)$$

where ϕ is defined by the equation $t = \phi(t)\, e^{\phi(t)}$.

Numerical integration would be possible for known values of a_r,

$$\frac{dy_n}{dx} = 2y_n^2 \sum_{r=1}^{m} \frac{1}{{}_ru_n (2 + {}_ru_n)^2},$$

while in the single factor case

$$\frac{dy_n}{dx} = \frac{2y_n^2}{u_n (2 + u_n)^2}.$$

Now when one a_r is very much smaller than the rest these two rates are nearly equal for equal values of y_n. When all the a_r's are equal,

$$\frac{dy_n}{dx} = my_n^2 \left[\left(1 - y_n^{\frac{1}{m}}\right)^{-\frac{1}{2}} + 1\right]^{-2} \left[\left(1 - y_n^{\frac{1}{m}}\right)^{-2} - 1\right]^{-1}.$$

The ratio of this rate to the rate with a single factor (putting $t^m = y_n$) is

$$\frac{mt^{m-1}(1 + \sqrt{1 - t^m})(1 - t)^{\frac{3}{2}}}{(1 + \sqrt{1 - t})(1 - t^m)^{\frac{3}{2}}}.$$

When t is small this tends to the small mt^{m-1}; when t is nearly unity, to $m^{-\frac{1}{2}}$ which is also small. The ratio when all the a_r's are equal is, by Purkiss' theorem, the minimum value. Hence it would seem that in general natural selection acts more slowly on a multiple dominant than a single dominant. The case of a multiple sex-linked dominant and various more complicated cases present still greater difficulties to analysis, though of course individual cases could always be solved numerically.

Linkage.

Consider two autosomal factors A, B, linked with such intensity that the cross-over value is $100l$ in the female, $100l'$ in the male sex. Let the nth generation be formed from:—

eggs $p_n AB : q_n Ab : r_n aB : s_n ab$,
spermatozoa $p_n' AB : q_n' Ab : r_n' aB : s_n' ab$,

where $p_n + q_n + r_n + s_n = p_n' + q_n' + r_n' + s_n' = 1$.

The nth generation therefore consists of:—

$$p_n p_n' ABAB : (p_n q_n' + p_n' q_n) ABAb : (p_n r_n' + p_n' r_n) ABaB$$
$$: q_n q_n' AbAb : (p_n s_n' + p_n' s_n) AB.ab : (q_n r_n' + q_n' r_n) Ab.aB$$
$$: r_n r_n' aB.aB : (q_n s_n' + q_n' s_n) Abab : (r_n s_n' + r_n' s_n) aBab$$
$$: s_n s_n' abab.$$

If no selection occurs they produce gametes in the proportions:

$$\begin{aligned} 2p_{n+1} &= p_n + p_n' + l(q_n r_n' + q_n' r_n - p_n s_n' - p_n' s_n) \\ 2q_{n+1} &= q_n + q_n' - l(q_n r_n' + q_n' r_n - p_n s_n' - p_n' s_n) \\ 2r_{n+1} &= r_n + r_n' - l(q_n r_n' + q_n' r_n - p_n s_n' - p_n' s_n) \\ 2s_{n+1} &= s_n + s_n' + l(q_n r_n' + q_n' r_n - p_n s_n' - p_n' s_n), \end{aligned}$$

whilst the values of p_{n+1}', etc., are given by the same expressions with l' substituted for l. Hence after one generation

$$\frac{p_n + q_n}{r_n + s_n} \text{ and } \frac{p_n' + q_n'}{r_n' + s_n'}$$

have the same constant value u, while

$$\frac{p_n + r_n}{q_n + s_n} = \frac{p_n' + r_n'}{q_n' + s_n'} = v.$$

We may therefore write:

$$p_n = \frac{uv}{(1+u)(1+v)} + x_n;\quad q_n = \frac{u}{(1+u)(1+v)} - x_n;$$

$$r_n = \frac{v}{(1+u)(1+v)} - x_n;\quad s_n = \frac{1}{(1+u)(1+v)} + x_n.$$

$$p_n' = \frac{uv}{(1+u)(1+v)} + x_n';\quad q_n' = \frac{u}{(1+u)(1+v)} - x_n';$$

$$r_n' = \frac{v}{(1+u)(1+v)} - x_n';\quad s_n' = \frac{1}{(1+u)(1+v)} + x_n'.$$

$$\therefore\quad q_n r_n' + q_n' r_n - p_n s_n' - p_n' s_n = -2(x_n + x_n').$$

$$\therefore\quad 2x_{n+1} = (1-2l)(x_n + x_n');\ 2x_{n+1}' = (1-2l')(x_n + x_n').$$

Hence if $x_0 + x_0' = c$,

$$\therefore\quad \left.\begin{aligned} x_n &= (\tfrac{1}{2} - l)(1-l-l')^{n-1}c \\ x_n' &= (\tfrac{1}{2} - l')(1-l-l')^{n-1}c \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots(6\cdot0).$$

Hence the proportions of the various types of gamete approach asymptotically those which would be reached in one generation without linkage, the ratio of successive differences from the final values being $1-l-l'$. Hence if either l or l' is larger than k the effects of linkage are unimportant. A similar proof holds for a pair of sex-linked factors.

Selection in a tetraploid organism.

In a tetraploid race which is stable, i.e. yields only diploid gametes, five types of zygote and three of gamete exist. Gregory (6) and Blakeslee, Belling and Farnham (7) have shown that zygotes produce gametes as follows:

Zygotes	Gametes
$AAAA$	AA
$AAAa$	$1AA : 1Aa$
$AAaa$	$1AA : 4Aa : 1aa$
$Aaaa$	$1Aa : 1aa$
$aaaa$	aa

Gregory thought that $AAaa$ gave $1AA : 2Aa : 1aa$, but his results, as well as theory, suggest the above ratio. As in Part II we first consider tetraploidy without selection, and then the process of selection in a population which would be in equilibrium but for that selection. Let the mth generation be formed from gametes in the ratios $p_m AA : 2q_m Aa : r_m aa$, where $p_m + 2q_m + r_m = 1$, and $u_m = \dfrac{p_m + q_m}{q_m + r_m}$. They form zygotes in the ratios:

$$p_m^2 AAAA : 4p_m q_m AAAa : (4q_m^2 + 2p_m r_m) AAaa$$
$$: 4q_m r_m Aaaa : r_m^2 aaaa.$$

$$\begin{aligned} \therefore \quad p_{m+1} &= p_m + \tfrac{2}{3}(q_m^2 - p_m r_m) \\ q_{m+1} &= q_m - \tfrac{2}{3}(q_m^2 - p_m r_m) \\ r_{m+1} &= r_m + \tfrac{2}{3}(q_m^2 - p_m r_m). \end{aligned}$$

Hence $u_{m+1} = u_m = u$, and when equilibrium is reached $q_\infty^2 = p_\infty r_\infty$.

Hence $p_\infty = \dfrac{u^2}{(1+u)^2}$, $q_\infty = \dfrac{u}{(1+u)^2}$, $r_\infty = \dfrac{1}{(1+u)^2}$,
and the population in equilibrium is in the ratios:

$$\frac{u^4}{(1+u)^4} AAAA : \frac{4u^3}{(1+u)^4} AAAa : \frac{6u^2}{(1+u)^4} AAaa$$

$$: \frac{4u}{(1+u)^4} Aaaa : \frac{1}{(1+u)^4} aaaa.$$

Putting $\theta_m = q_m^2 - p_m r_m$, we find $\theta_{m+1} = \tfrac{1}{3}\theta_m$, $\therefore \theta_m = 3^{-m}\theta_0$.

$$\begin{aligned} \therefore \quad p_m &= p_{m-1} + \theta_{m-1} \\ &= p_0 + \sum_{r=0}^{m-1} \theta_r \\ &= p_0 + \tfrac{3}{2}(1 - 3^{-m})\theta_0 \\ &= p_\infty - \frac{3^{1-m}}{2}\theta_0 \quad \text{..........................(7·0).} \end{aligned}$$

Hence the ratios of the different classes converge very rapidly to their final values. Under selection of a population which has reached such an equilibrium, if A is completely dominant,

$$\begin{aligned} u_{n+1} &= \frac{p_n + q_n}{q_n + r_n - k r_n^2} \\ &= \frac{u_n}{1 - k(1+u_n)^{-3}}, \end{aligned}$$

$\therefore \quad \dfrac{du_n}{dn} = \Delta u_n = \dfrac{k u_n}{(1+u_n)^3}$, approximately, if k be small.

$\therefore$ if $u_0 = 1$, $kn = \log_e u_n + 3u_n + \tfrac{3}{2}u_n^2 + \tfrac{1}{3}u_n^3 - 4\tfrac{5}{6}$ ...(7·1),

Hence when dominants are few u_n changes at the same rate as in a diploid organism; when they are many, much more slowly. To compare the change in the number y_n of recessives we find

$kn = \log_e(1 - y_n^{\frac{1}{4}}) - \tfrac{1}{4}\log_e y_n + y_n^{-\frac{1}{4}} + \tfrac{1}{2}y_n^{-\frac{1}{2}} + \tfrac{1}{3}y_n^{-\frac{3}{4}} - 6\tfrac{2}{3}$...(7·2),

$$\therefore \quad \frac{dy_n}{dn} = -4ky_n^2(y_n^{-\frac{1}{4}} - 1),$$

while in a diploid population

$$\frac{dy_n}{dn} = -2ky_n^2(y_n^{-\frac{1}{2}} - 1).$$

Hence here too the rate is always slower in the tetraploids, though not much so when recessives are few.

If dominance is incomplete, as is usual in tetraploid organisms, and after selection the zygotes are in the ratios:

$$u_n^4 AAAA : 4(1-k_1)u_n^3 AAAa : 6(1-k_2)u_n^2 AAaa$$
$$: 4(1-k_3)u_n Aaaa : (1-k_4) aaaa,$$

$$\therefore \quad \Delta u_n = \frac{k_1 u_n^4 + 3(k_2-k_1)u_n^3 + 3(k_3-k_2)u_n^2 + (k_4-k_3)u_n}{(1+u_n)^3},$$

approximately, if the coefficients are small. The possible equilibria, if any, are given by the roots of

$$k_1 u_\infty^3 + 3(k_2-k_1)u_\infty^2 + 3(k_3-k_2)u_\infty + k_4 - k_3 = 0.$$

The various possible cases, and their stability, could easily be investigated. If the advantage of the various genotypes increases or decreases with the number of dominant factors they contain, so that $k_4 > k_3 > k_2 > k_1 > 0$, or $0 > k_1 > k_2 > k_3 > k_4$, no equilibrium is possible,

$$\therefore \quad n = \int \frac{(1+u_n)^3 du_n}{k_1 u_n^4 + 3(k_2-k_1)u_n^3 + 3(k_3-k_2)u_n^2 + (k_4-k_3)u_n}. \quad \ldots(7{\cdot}3).$$

If $k_1 = 0$ this contains a term proportional to u_n or u_n^2. If $k_1 \neq 0$ all the terms are logarithmic and selection is always rapid. But $AAAa$ is more likely to resemble $AAAA$ than Aa to resemble AA. Hence polyploidy diminishes the probability of a rapid selection in populations where recessives are few. Since stable polyploidy is only known in hermaphrodite plants there is no need to discuss cases of sex-linkage or different intensities of selection in the two sexes. The theory can readily be extended to the higher forms of polyploidy.

SUMMARY.

Expressions are found for the changes caused by slow selection in populations whose characters are determined by incompletely dominant, multiple, or polyploid factors, and for the equilibria attained in certain of these cases.

REFERENCES.

1. J. B. S. Haldane. *Trans. Camb. Phil. Soc.*, vol. 22, p. 19, 1924.
2. R. A. Fisher. *Proc. Roy. Soc. Edin.*, vol. 42, p. 321, 1922.
3. E. Robertson. *Journ. Genetics*, vol. 11, p. 79, 1921.
4. H. Nilsson-Ehle. *Lund's Univ. Arsskrift.*, p. 69, 1909.
5. E. R. Saunders. *Journ. Genetics*, vol. 10, p. 149, 1920.
6. Gregory. *Proc. Roy. Soc.*, B, vol. 87, p. 484, 1914.
7. Blakeslee, Belling and Farnham. *Science*, vol. 52, p. 388, 1920.

A mathematical theory of Natural and Artificial Selection. Part IV. By Mr J. B. S. Haldane, Trinity College.

[*Received* 11 November, *read* 22 November 1926.]

In such organisms as annual plants, in which successive generations do not overlap, the composition of the $n+1$th generation can be calculated from that of the nth, and the resulting finite difference equation investigated. Where generations overlap we may obtain a similar relation between the compositions of the population at times t and t', but the finite difference equation is now represented by an integral equation. This fact was first pointed out in 1910 by Mr H. T. J. Norton of Trinity College. At a much later date I arrived at the same conclusion, and Mr Norton showed me his results in 1922, stating that he would publish them shortly. He has been prevented from doing so by illness, and, although I believe that all the results here given were reached by me independently, there can be no question that Mr Norton had obtained many of them previously, and had treated the problem rigorously, which I have not done.

The only case considered here is the very simple one in which the intensity of selection is independent of the size of the population. A preliminary lemma will first be discussed.

The growth of a population.

If the death-rate and birth-rate of a population are not functions of its density, its number at any time may be calculated as follows:

Let $N(t)$ be the number at time t. Only the female sex need be considered if the sexes are separate.

$S(x)$ be the probability of an individual surviving to the age x.

$U(t)\,\delta t$ be the number of individuals produced between times t and $t+\delta t$.

$K(x)\,\delta x$ be the probability of an individual between the ages x and $x+\delta x$ producing one (female) offspring. All individuals of this age, both alive and dead, are considered, so that if $P(x)$ be the corresponding function for living individuals only, $K(x)=S(x)\,P(x)$. Then

$$\left.\begin{aligned} N(t) &= \int_0^\infty U(t-x)\,S(x)\,dx \\ U(t) &= \int_0^\infty U(t-x).K(x)\,dx \end{aligned}\right\} \qquad \text{............(1·0).}$$

Instead of infinity any upper limit exceeding the maximum life of the organism may be taken. Equation (1·0) has been considered by Herglotz*. Let $U(t) = ce^{zt}$. Then

$$\int_0^\infty e^{-zx} K(x)\, dx = 1, \text{ or } \int_0^a e^{-zx} K(x) = 1 \quad \text{.........(1·1)},$$

where a is sufficiently large. Since $K(x)$ is always real and zero or positive, the above integral is a monotone function of z when z is real, and can have any real positive value. Hence it has one and only one real root for z, say α_0. The complex roots clearly occur in pairs $\alpha_r \pm \iota\beta_r$. Then

$$\int_0^\infty e^{-\alpha_r x} K(x)\, dx > \int_0^\infty e^{-(\alpha_r + \iota\beta_r)x} K(x)\, dx$$
$$= 1.$$

Therefore $\qquad \alpha_r < \alpha_0.$

If any two functions of t are solutions of (1·0) so is their sum. Therefore

$$\left.\begin{aligned} U(t) &= \alpha_0 e^{\alpha_0 t} + \sum_{r=1}^{\infty} a_r e^{\alpha_r t} \cos \beta_r (t - b_r) \\ N(t) &= c_0 e^{\alpha_0 t} + \sum_{r=1}^{\infty} c_r e^{\alpha_r t} \cos \beta_r (t - d_r) \end{aligned}\right\} \quad \text{......(1·2)}.$$

In general there will be an infinite number of terms. The values of a_r and b_r, and hence of c_r and d_r, depend on the initial conditions. Where multiple roots occur there will be periodic terms including powers of t as factors. Since $\alpha_r < \alpha_0$, all the periodic terms become negligible compared with the first after the lapse of a sufficient time. That is to say, oscillations of the population about an exponential function of the time are either damped or at least increase less rapidly than the population itself. In particular, if $\int_0^\infty K(x)dx = 1$, so that the population is in equilibrium, oscillations are damped and the equilibrium is stable. We are therefore justified in neglecting periodic terms in the solution of equations which only differ by small terms from (1·0), and which occur in the subsequent analysis. It is proposed to discuss the stability of the equilibrium when $S(x)$ and $K(x)$ depend on the number of the population in a subsequent paper.

Incidentally, if $C(x)\,\delta x$ be the probability of any member of the population being between the ages x and $x + \delta x$, then

$$C(x) = \frac{U(t-x)\, S(x)}{N(t)}.$$

---

* Herglotz, *Math. Ann.* 65, p. 87.

Hence, when oscillations have died down,

$$C(x) = \frac{e^{-a_0 x} S(x)}{\int_0^\infty e^{-a_0 x} S(x)\, dx}$$

This constitutes a new proof of Lotka's* theorem on the stability of the normal age distribution.

Selection of an autosomal factor.

Consider a population consisting, at time t, of $D(t)$ female zygotes possessing a dominant factor A, $R(t)$ female recessives. The sex-ratio at birth is taken as fixed.

Let $F(t)\,\delta t$ be the number of fertile A ova produced between times t and $t+\delta t$.
$f(t)\,\delta t$ be the number of fertile a ova produced between times t and $t+\delta t$.
$M(t)\,\delta t$ be the number of A spermatozoa produced between times t and $t+\delta t$.
$m(t)\,\delta t$ be the number of a spermatozoa produced between times t and $t+\delta t$.
$S(x)$ be the probability of a female dominant reaching the age x.
$s(x)$ be the probability of a female recessive reaching the age x.
$K(x)\,\delta x$ be the probability of a female dominant (alive or dead, as above) producing a female offspring between the ages x and $x+\delta x$.
$[K(x)-k(x)]\,\delta x$ be the same probability for a female recessive.
$L(x)\,\delta x$ be the same probability for a male dominant.
$[L(x)-l(x)]\,\delta x$ be the same probability for a male recessive.

$$S=\int_0^\infty S(x)\,dx,\quad s=\int_0^\infty s(x)\,dx.$$

$$K=\int_0^\infty K(x)\,dx,\quad K'=\int_0^\infty xK(x)\,dx,\quad k=\int_0^\infty k(x)\,dx.$$

$$L=\int_0^\infty L(x)\,dx,\quad L'=\int_0^\infty xL(x)\,dx,\quad l=\int_0^\infty l(x)\,dx.$$

In general the functions $S(x)$, $s(x)$, $K(x)$, etc., will not be functions of age alone, but of $D(t)$, $R(t)$, etc. We make the assumption however that selection and population growth are proceeding so slowly that $k(x)$ and $l(x)$, and $K-1$ are small, and $S(x)$, etc., do not vary appreciably in the course of a generation.

---

* Lotka, *Proc. Nat. Ac. Sci.* 8, p. 339, 1922.

If mating be at random, the rates of production of the three female phenotypes at time t are

$$AA,\ \frac{F(t)\,M(t)}{M(t)+m(t)};\quad Aa,\ \frac{F(t)\,m(t)+f(t)\,M(t)}{M(t)+m(t)};\quad aa,\ \frac{f(t)\,m(t)}{M(t)+m(t)}.$$

The group aged x at time t was hatched or born at time $t-x$. Therefore

$$\left.\begin{aligned}
F(t)&=\tfrac{1}{2}\int_0^\infty \frac{2F(t-x)\,M(t-x)+F(t-x)\,m(t-x)+f(t-x)\,M(t-x)}{M(t-x)+m(t-x)}\,K(x)\,dx\\
f(t)&=\tfrac{1}{2}\int_0^\infty \frac{F(t-x)\,m(t-x)+f(t-x)\,M(t-x)+2f(t-x)\,m(t-x)}{M(t-x)+m(t-x)}\,K(x)\,dx\\
&\qquad\qquad-\int_0^\infty \frac{f(t-x)\,m(t-x)\,k(x)\,dx}{M(t-x)+m(t-x)}\\
M(t)&=\tfrac{1}{2}\int_0^\infty \frac{2F(t-x)\,M(t-x)+F(t-x)\,m(t-x)+f(t-x)\,M(t-x)}{M(t-x)+m(t-x)}\,L(x)\,dx\\
m(t)&=\tfrac{1}{2}\int_0^\infty \frac{F(t-x)\,m(t-x)+f(t-x)\,M(t-x)+2f(t-x)\,m(t-x)}{M(t-x)+m(t-x)}\,L(x)\,dx\\
&\qquad\qquad-\int_0^\infty \frac{f(t-x)\,m(t-x)\,l(x)\,dx}{M(t-x)+m(t-x)}
\end{aligned}\right\}\ldots(2{\cdot}0).$$

Since selection and population growth are slow, we may put $F(t-x)=F(t)-xF'(t)$, etc., $M(t)=\lambda F(t)$, $m(t)=\lambda f(t)$, all to the first order of small quantities. Hence, to this degree of approximation,

$$\begin{aligned}
F(t)&=\tfrac{1}{2}KF(t)+\frac{\tfrac{1}{2}K[F(t)+f(t)]\,M(t)}{M(t)+m(t)}-\tfrac{1}{2}K'F'(t)-\frac{\tfrac{1}{2}K'[F'(t)+f'(t)]\,M(t)}{M(t)+m(t)}\\
&\qquad-\frac{\tfrac{1}{2}K'[F(t)+f(t)]\,M'(t)\,m(t)-M(t)\,m'(t)}{[M(t)+m(t)]^2}.\\
&=\tfrac{1}{2}KF(t)+\frac{\tfrac{1}{2}K[F(t)+f(t)]\,M(t)}{M(t)+m(t)}-K'F'(t).
\end{aligned}$$

Similarly

$$\begin{aligned}
f(t)&=\tfrac{1}{2}Kf(t)+\frac{\tfrac{1}{2}K[F(t)+f(t)]\,m(t)}{M(t)+m(t)}-K'f(t)-\frac{k[f(t)]^2}{F(t)+f(t)}\\
M(t)&=\tfrac{1}{2}LF(t)+\frac{\tfrac{1}{2}L[F(t)+f(t)]\,M(t)}{M(t)+m(t)}-L'F'(t)\\
m(t)&=\tfrac{1}{2}Lf(t)+\frac{\tfrac{1}{2}L[F(t)+f(t)]\,m(t)}{M(t)+m(t)}-L'f'(t)-\frac{l[f(t)]^2}{F(t)+f(t)},
\end{aligned}$$

all approximately. Therefore

$$\begin{aligned}
\frac{M(t)}{m(t)}&=\frac{\Big[2LF(t)+Lf(t)-2L'F(t)\Big]\dfrac{M(t)}{m(t)}+LF(t)-2L'F'(t)}{\Big[LF(t)-2L'f'(t)-\dfrac{2l\,\{f(t)\}^2}{F(t)+f(t)}\Big]\dfrac{M(t)}{m(t)}+LF(t)+2Lf(t)-2L'f'(t)-\dfrac{2l\,[f(t)]^2}{F(t)+f(t)}}\\
&=\frac{L[2F(t)+f(t)]\dfrac{M(t)}{m(t)}+LF(t)-\dfrac{2L'F'(t)\,[(F(t)+f(t)]}{f(t)}}{Lf(t)\dfrac{M(t)}{m(t)}+LF(t)+2Lf(t)-\dfrac{2L'f'(t)\,[F(t)+f(t)]}{f(t)}-2lf(t)}
\end{aligned}$$

approximately,

$$=\frac{F(t)}{f(t)}+\frac{2L'[F(t)f'(t)-F'(t)f(t)]}{L[F(t)+f(t)]}+\frac{lF(t)[f(t)]^2}{L[F(t)+f(t)]^2},$$

approximately, by solving the quadratic. Therefore

$$\left.\begin{aligned}(K-1)F(t)-K'F'(t)+\frac{L'[F(t)f'(t)-F'(t)f(t)]}{L[F(t)+f(t)]}+\frac{lF(t)[f(t)]^2}{L[F(t)+f(t)]^2}=0\\(K-1)f(t)-K'f'(t)-\frac{L'[F(t)f'(t)-F'(t)f(t)]}{L[F(t)+f(t)]}-\frac{lF(t)[f(t)]^2}{L[F(t)+f(t)]^2}-\frac{k[f(t)]^2}{F(t)+f(t)}=0\end{aligned}\right\}, \qquad \ldots\ldots\ldots\ldots\ldots(2{\cdot}1).$$

If $$u(t)=F(t)/f(t),$$

$$\frac{d}{dt}u(t)=\frac{(l+kL)\,u(t)}{(L'+K'L)[1+u(t)]} \qquad \ldots\ldots\ldots\ldots(2{\cdot}2).$$

This is equivalent to equation (2·1) of Part I*, $\frac{l+kL}{L'+K'L}$ being the coefficient of selection. In general this quantity is not independent of t, hence the equation cannot be integrated, but if its upper and lower limits are known, the march of the composition of the population can be roughly calculated from equation (2·3) of Part I. If, however, the population is very nearly in equilibrium, and either dominants or recessives are very rare, more accurate results are possible.

When recessives are rare, $F(x)$ is large and equal to a constant F, $D(x)$ being also large and equal to a constant N. $K=1$, and $F'(t)$ is negligible, while $f(t)$ is small. Therefore

$$K'f'(t)+\frac{L'}{L}f'(t)+\left(\frac{l+kL}{L}\right)\frac{[f(t)]^2}{F}=0,$$

and $$f(t)=\frac{(L'+K'L)F}{(l+kL)(t-t_0)},$$

where t_0 is an integration constant. But

$$N=D(t)=\int_0^\infty F(t-x)S(x)\,dx=FS,$$

$$R(t)=\int_0^\infty\frac{[f(t-x)]^2 s(x)\,dx}{F(t-x)+f(t-x)}=\frac{(L'+K'L)^2Fs}{(l+kL)^2(t-t_0)^2}.$$

$$R(t)=\frac{(L'+K'L)^2sN}{(l+kL)^2S(t-t_0)^2} \qquad \ldots\ldots\ldots\ldots\ldots(2{\cdot}3).$$

Hence selection proceeds at the same rate as when generations are separate, with a selection coefficient equal to

$$\left(\frac{l+kL}{L'+K'L}\right)\left(\frac{S}{s}\right)^{\frac{1}{2}}.$$

---

* **Haldane**, *Trans. Camb. Phil. Soc.* **23**, p. 19, 1924.

When dominants are rare, $f(x)$ is large and equal to a constant f, $K = 1 + k$, and $F(x)$ is small. Therefore

$$\left(k + \frac{l}{L}\right) F(t) + \left(K' + \frac{L'}{L}\right) F'(t) = 0,$$

$$F(t) = e^{\frac{(l+kL)(c-t)}{L'+K'L}},$$

where c is a constant of integration. Therefore

$$D(t) = \int_0^\infty 2F(t-x)\, S(x)\, dx = 2SF(t)$$

$$= e^{\frac{(l+kL)(t_0-t)}{L'+K'L}} \qquad \text{..............................(2·4)},$$

where t_0 is an arbitrary constant. Selection therefore proceeds as when generations are separate, but with a selection coefficient

$$\frac{l + kL}{L' + K'L}.$$

When the death rates and fertility rates are the same in the two sexes, or in a hermaphrodite species, we have in general,

$$\frac{d}{dt} u(t) = \frac{ku(t)}{K'[1 + u(t)]} \qquad \text{..................(2·5)},$$

when recessives are rare,

$$R(t) = \frac{K'^2 sN}{k^2 S (t - t_0)^2} \qquad \text{..................(2·6)},$$

when dominants are rare,

$$D(t) = e^{\frac{k(t_0-t)}{K'}} \qquad \text{....................(2·7)}.$$

Now k has approximately the same meaning as when generations are separate, provided $K = 1$. Hence the two cases become comparable if we choose our unit of time, or "generation," so as to make K' or $\int_0^\infty xK(x)\,dx = 1$, as in the calculations of Dublin and Lotka* on the rate of increase of a population. In each case, if functions analogous to $S(x)$, $s(x)$ are known for the males, their numbers can be calculated.

---

* Dublin and Lotka, *Journ. Amer. Stat. Assoc.* **1925**, p. 306.

Selection of a sex-linked factor.

Here, using the same notation as above, we find

$$\left.\begin{aligned} &(K-1)\,F(t)-K'F'(t)+\frac{L'[F(t)f'(t)]-F'(t)\,f(t)]+lF(t)\,f(t)}{2L\,[F(t)+f(t)]}=0\\ &(K-1)\,f(t)-K'f'(t)-\frac{L'\,[F(t)\,f'(t)-F'(t)\,f(t)]+lF(t)\,f(t)}{2L\,[F(t)+f(t)]}-\frac{k\,[f(t)]^2}{F(t)+f(t)}=0 \end{aligned}\right\}\quad\ldots(3{\cdot}1).$$

In general, if $u=F(t)/f(t)$,

$$(L'+2K'L)\frac{d}{dt}u\,(t)=lu\,(t)+\frac{2kLu\,(t)}{1+u\,(t)}\quad\ldots\ldots(3{\cdot}2),$$

an equation analogous to (2·0) of Part III*.

When recessives are rare,

$$(L'+2K'L)\,f'(t)+lf(t)+\frac{2kL\,[f(t)]^2}{F}=0,$$

where F is the birth-rate of dominants. Three cases occur:

(*a*) If kL is negligible compared with lN, which will be the case if selection is of the same order of intensity in the two sexes, or more intense among males, then

$$f(t)=e^{\frac{l(t_0-t)}{L'+2K'L}}\quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3{\cdot}31).$$

The number of recessive males is proportional to $f(t)$, of recessive females to its square.

(*b*) If kL is of the same order of magnitude as lF, then

$$\frac{2kLf(t)}{2kLf(t)+lF}=e^{\frac{l(t_0-t)}{L'+2K'L}}\quad\ldots\ldots\ldots\ldots\ldots(3{\cdot}32).$$

Hence if V, v are quantities corresponding to S, s for the male sex, the proportion of recessive males is

$$\frac{lv}{2kLV\left(e^{\frac{l(t-t_0)}{L'+2K'L}}-1\right)},$$

of recessive females

$$\frac{l^2s}{4k^2L^2S\left(e^{\frac{l(t-t_0)}{L'+2K'L}}-1\right)^2}.$$

---

* Haldane, *Proc. Camb. Phil. Soc.* 23, p. 363, 1926.

(*c*) If kL is much larger than lF, i.e. selection is confined to females, then

$$f(t) = \frac{(2K'L + L')\,F}{2kL\,(t - t_0)} \quad \ldots\ldots\ldots\ldots\ldots (3{\cdot}33).$$

Hence the proportion of recessive males is

$$\frac{(L' + 2K'L)\,v}{2kLV\,(t - t_0)},$$

of recessive females

$$\frac{(L' + 2K'L)^2 s}{4k^2 L^2 S\,(t - t_0)^2}.$$

When dominants are rare,

$$F(t) = e^{\frac{(l + 2kL)\,(t - t_0)}{L' + 2K'L}} \quad \ldots\ldots\ldots\ldots\ldots\ldots (3{\cdot}4).$$

The number of male dominants is proportional to $F(t)$, that of females being double that of males.

When the intensity of selection is equal in both sexes, these equations simplify to

$$\frac{d}{dt}u(t) = \frac{ku(t)\,[3 + u(t)]}{3K'\,[1 + u(t)]} \quad \ldots\ldots\ldots\ldots\ldots (3{\cdot}5),$$

$$f(t) = e^{\frac{k(t_0 - t)}{3K'}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3{\cdot}6),$$

$$F(t) = e^{\frac{k(t_0 - t)}{K'}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3{\cdot}7),$$

analogous to equation (7·2) of Part I.

DISCUSSION.

The most satisfactory table of $K(x)$ known to me is that given by Dublin and Lotka[*] for certain American women. Here the population is growing, and $K = 1{\cdot}17$, while $\frac{K'}{K}$, the length of a "generation," is 28·45 years. No satisfactory values of k are known in the present state of genetics, though the data on mice discussed in Part I suggest that here $k = {\cdot}04$ approximately. In man mating is highly assortative for age, and the above formulae cannot be applied. Moreover, a change in the coefficient of correlation between the ages of spouses would undoubtedly affect the values of $K(x)$, etc., if other conditions remained equal. Thus old men would beget more children if they were more likely to have young

---

[*] Dublin and Lotka, *loc. cit.*

wives. It is thus impossible to calculate the effect of this correlation on selection. But a consideration of the extreme case when the age of the wife fixes that of the husband makes it clear that selection must follow equations of the type here arrived at, with changes in the parameters only.

SUMMARY.

Expressions are found for the progress of slow selection in a Mendelian population where generations overlap. The changes are very similar to those which occur when generations are separate.

A Mathematical Theory of Natural and Artificial Selection, Part V: Selection and Mutation. By Mr J. B. S. HALDANE, Trinity College.

[*Received* 21 May, *read* 25 July, 1927.]

New factors arise in a species by the process of mutation. The frequency of mutation is generally small, but it seems probable that it can sometimes be increased by changes in the environment (1, 2). On the whole mutants recessive to the normal type occur more commonly than dominants. The frequency of a given type of mutation varies, but for some factors in Drosophila it must be less than 10^{-6}, and is much less in some human cases. We shall first consider initial conditions, when only a few of the new type exist as the result of a single mutation; and then the course of events in a population where the new factor is present in such numbers as to be in no danger of extinction by mere bad luck. In the first section the treatment of Fisher (3) is followed.

In a large population let p_r be the chance that a factor present in a zygote at a given stage in the life-cycle will appear in r of its children in the next generation. If the individual considered is homozygous, this is the chance of leaving r children, if mutation is neglected. Let $\sum_{r=0}^{\infty} p_r x^r = f(x)$. Therefore $f(1) = 1$, $f(0) = p_0$, the probability of the factor disappearing, while $f'(1) = \sum_{r=0}^{\infty} rp_r$, *i.e.*, the probable number of individuals possessing the factor in the next generation. The probability of m individuals bearing one each of the factors considered leaving r descendants is clearly the coefficient of x^r in $[f(x)]^m$, if we neglect the possibility of a mating between two such individuals, which we may legitimately do if m is small compared with the total number of the population. If then the probability of the factor being present in r zygotes of the nth generation be the coefficient of x^r in $F(x)$, the corresponding probability in the $(n+1)$th generation is the same coefficient in $F[f(x)]$. Hence if a single factor appears in one zygote, the probability of its presence in r zygotes after n generations is the coefficient of x^r in $S_f^n(x)$, *i.e.* $f(f(f..f(x)...))$, the operation being repeated n times. The probability of its disappearance is therefore $\operatorname*{Lt}_{n \to \infty} S_f^n(0)$. By Koenigs' theorem (4) this is the root of $x = f(x)$ in the neighbourhood of zero.

Now in the case of a dominant factor appearing in a population in equilibrium, and conferring an advantage measured by k, as in Part I (5), $f'(1) = 1 + k$. Since $f'(x)$ and $f''(x)$ are positive when x is positive, and $f(0)$ is positive, $x = f(x)$ has two and only two real positive roots, one equal to unity, the other lying between 0 and 1, but near the latter value if k be small. Hence any advantageous dominant factor which has once appeared has a finite chance of survival, however large the total population may be.

If a large number of offspring is possible, as in most organisms, the series p_n approximates to a Poisson series, provided that adult organisms are counted, and since $f'(1) = 1 + k, f(x) = e^{(1+k)(x-1)}$. Hence the probability of extinction $1 - y$ is given by

$$1 - y = e^{-(1+k)y}.$$

Hence $$(1 + k)y = -\log(1 - y) \quad\ldots\ldots\ldots\ldots\ldots(1 \cdot 0),$$

and $$k = \frac{y}{2} + \frac{y^2}{3} + \frac{y^3}{4} + \ldots,$$

and if k be small, $y = 2k$ approximately. Hence an advantageous dominant gene has a probability $2k$ of survival after only a single appearance in an adult zygote, and if in the whole history of a species it appears more than $\frac{\log_e 2}{2k}$ times it will probably spread through the species. But, however large k may be, the factor may be extinguished after a single appearance. Thus if $k = 1$, so that the new type probably leaves twice as many offspring as the normal, the probability of its extinction is still ·203. If in any generation there are m dominant individuals the probability of extinction is reduced to y^m, where y is the smaller positive root of $x = f(x)$. When k is small this reduces to $(1 - 2k)^m$. Hence if in any generation more than $\frac{\log_e 2}{2k}$ adult dominants exist, the factor will probably spread through the whole population.

On the other hand a recessive factor whose phenotype is advantageous has a quite negligible advantage in a random mating population provided that the number of its bearers is small compared with the square root of the total population. This is best seen by considering the case of a hermaphrodite: in a dioecious organism the argument, though similar, is more complicated. Let N be the fixed number of the population, and z_n the number of heterozygotes plus double the number of recessives for the factor A in the nth generation. It therefore produces gametes in the ratio $(2N - z_n)A : z_n a$. If now the recessives have a small advantage measured by k, the probabilities of production of each genotype in the next generation are

$$(2N - z_n)^2 AA : 2z_n(2N - z_n)Aa : (1 + k)z_n^2 aa.$$

Hence if, as above, $f(x)$ be the function defining the probable number of offspring of a dominant, so that $f'(1) = 1$, the probability of r heterozygotes in the $(n + 1)$th generation is the coefficient of r in

$$[f(x)]^{\frac{2Nz_n(2N - z_n)}{4N^2 + kz_n^2}},$$

that of r recessives the same coefficient in

$$[f(x)]^{\frac{N(1 + k)z_n^2}{4N^2 + kz_n^2}}.$$

Hence the probability of z_{n+1} in the next generation is the coefficient of $x^{z_{n+1}}$ in

$$[f(x)]^{\frac{2Nz_n(2N + kz_n)}{4N^2 + kz_n^2}},$$

or, approximately, if z_n be small compared with N, in

$$[f(x)]^{z_n\left(1 + \frac{kz_n}{2N}\right)}.$$

The corresponding expression for a dominant factor is

$$[f(x)]^{z_n(1+k)}.$$

Hence provided that z_n is small the probability of escaping extinction is much smaller than k. I have been unable to evaluate it exactly, but it seems from a comparison with the case of a dominant factor, that the value of z_n such that the factor is as likely to survive as to be extinguished, is of the order of $\left(\frac{N}{k}\right)^{\frac{1}{2}}$, *i.e.* generally $> N^{\frac{1}{2}}$. So if N is sufficiently large the probability of a single mutation leading to the establishment of a recessive factor is negligible.

When the population is wholly self-fertilized or inbred by brother-sister mating, on the other hand, a recessive factor has almost as good a chance of survival as a dominant. With partial self-fertilization or inbreeding it can be shown by methods similar to those of Part II (6) that an advantageous recessive factor has a finite chance of establishment after one appearance, however large be the population.

If mutation occurs with a finite frequency any advantageous or not too disadvantageous factor will certainly be established. Consider a random mating population in which, in each generation, a proportion p of the A genes mutate to a, a proportion q of the a genes to A, and the coefficient of selection is k. Let u_n be the gametic ratio of the nth generation. But for mutation we should have

$$u_{n+1} = \frac{u_n(u_n + 1)}{u_n + 1 - k};$$

allowing for mutation

$$u_{n+1} = \frac{(1-p)(u_n^2 + u_n) + q(u_n + 1 - k)}{(1-q)(u_n + 1 - k) + p(u_n^2 + u_n)}.$$

$$\text{Hence } \Delta u_n = \frac{ku_n^2 - pu_n(u_n+1)^2 + q(u_n+1)(u_n+1-k)}{u_n + 1 - k + pu_n(u_n+1) - q(u_n+1-k)} \quad \ldots(2{\cdot}0)$$

$$= \frac{ku_n}{u_n+1} - pu_n(u_n+1) + q(u_n+1) \quad \ldots\ldots\ldots\ldots\ldots(2{\cdot}1)$$

approximately, if p, q, and k are small, as is generally the case. It is clear that u_n must lie between $\frac{1-p}{p}$ and $\frac{q}{1-q}$, *i.e.* between $\frac{1}{p}$ and q approximately, and that when near these values it alters rapidly. But as p and q may be less than 10^{-6} these limits are very wide. The population is in equilibrium when

$$pu^3 + (2p - q)u^2 + (p - 2q - k + kq)u - q + kq = 0.$$

There is always one real positive root since p and q are positive and less than unity. If k be positive there is only one such root, defining a stable condition towards which the population tends when dominants have the advantage. If k or q be large compared with p this root approximates to $\left(\frac{k}{p}\right)^{\frac{1}{2}}$ or $\frac{q}{p}$ as the case may be, *i.e.* recessives nearly disappear. If p be of the same order of magnitude as the larger of k and q, u has a moderate value and the population is dimorphic. If p be much larger than k or q, u is small and approximates to $\frac{q}{p}$, *i.e.* dominants are rare.

If k be negative all the roots are positive if they are real, provided $q > 2p$ and $-k(1-q) > 2q - p$. They are real if $\frac{\Delta}{3k}$, *i.e.*

$$4(p+q)^3 + [-27p^2 + 18p(p+q)(1-q) + (p+q)^2(1-q)^2]k + 4p(1-q)^3k^2$$

is positive, that is to say, when q is small, if

$$4pk^2 + (-8p^2 + 20pq + q^2)k + 4(p+q)^2$$

is positive. All these three conditions can rarely be fulfilled, but such cases may presumably occur. Thus if $p = {\cdot}000{,}001$, $q = {\cdot}0004$, $k = -{\cdot}008$; $u^3 - 398u^2 + 7197{\cdot}8u - 403{\cdot}2 = 0$. Therefore $u = {\cdot}057$, $18{\cdot}93$, or $379{\cdot}0$, giving $89{\cdot}5\,{}^\circ/_\circ$, $0{\cdot}252\,{}^\circ/_\circ$, or ${\cdot}000{,}693\,{}^\circ/_\circ$ of recessives. In such a case the middle root defines an unstable equilibrium, the other two equilibria being stable. Thus the above considered population would be stable with only about seven recessives per million, the small tendency of dominant genes to mutate to recessive being balanced by reverse mutation. But if a group containing more than one recessive gene in twenty were isolated from it, selection would be effective, and it would pass into a condition where only 10·5 °/₀ were dominants, this number being kept up by mutation.

Usually when k is negative there is only one real root. If p or $-k$ be large compared with q, it is small and approximates to $\frac{q}{p}$ or $\frac{-q}{k}$ as the case may be, so that dominants are rare. If q be of the same order of magnitude as the larger of p and $-k$, the root has a moderate value and the population is dimorphic. If q be larger than p or $-k$, u is large and approximates to $\frac{q}{p}$, so that recessives are few.

The rate of approach to equilibrium is given by

$$\frac{du_n}{dn} = \frac{ku_n}{u_n+1} - pu_n(u_n+1) + q(u_n+1) \quad \ldots\ldots(2\cdot2),$$

provided that the constants are small. The exact expression for n in terms of u_n depends on the nature of the roots and the side from which an equilibrium is being approached, but it always contains logarithmic terms. Hence the numbers of the rarer type of the population in succeeding generations always lie between two geometric series until equilibrium is nearly reached. That is to say, the march of events is comparatively rapid.

In a self-fertilizing population we can similarly show that

$$\Delta u_n = ku_n - pu_n(u_n+1) + q(u_n+1) \quad \ldots\ldots\ldots(2\cdot3).$$

Only one equilibrium is possible, and the course of events can readily be calculated in any given case. Similarly for a sex-linked factor

$$\Delta u_n = \frac{ku_n(u_n+3)}{3(u_n+1)} - pu_n(u_n+1) + q(u_n+1) \ldots\ldots(2\cdot4).$$

In this case if k be negative, three equilibria are sometimes found, and selection is more effective than in the autosomal case when recessives are rare.

To sum up, if selection acts against mutation, it is ineffective provided that the rate of mutation is greater than the coefficient of selection. Moreover, mutation is quite effective where selection is not, namely in causing an increase of recessives where these are rare. It is also more effective than selection in weeding out rare recessives provided that it is not balanced by back mutation of dominants. Mutation therefore determines the course of evolution as regards factors of negligible advantage or disadvantage to the species. It can only lead to results of importance when its frequency becomes large.

Addendum. Equilibrium and selection in Sciara and similar animals.

In Part I of this series all the then known types of single-factor Mendelian inheritance were discussed. Since then Metz (7) has discovered a new type in Sciara which is here treated on the lines of Part I. Gametogenesis is normal in the female, but spermatozoa are formed from maternal chromatin only. Hence there are two types of heterozygous male, which may be symbolized by $A(a)$ and $(A)a$ according as the A is received from the mother or father. They yield A and a spermatozoa respectively, the other genotypes behaving normally.

In the absence of selection let eggs and spermatozoa be produced by the mth generation in the proportions $u_m A : 1a$ and $v_m A : 1a$, respectively. The next generation is therefore:

$$♀\ u_m v_m AA : (u_m + v_m) Aa : 1aa.$$

$$♂\ u_m v_m AA : u_m A(a) : v_m (A)a : 1aa.$$

Hence

$$\left.\begin{aligned} u_{m+1} &= \frac{2u_m v_m + u_m + v_m}{u_m + v_m + 2} \\ v_{m+1} &= u_m \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots(3{\cdot}0),$$

which is the same as equation (6·0) of Part I (5). Hence, as in the above equation, we find, if y_m be the proportion of recessives in the mth generation,

$$y_m = y_\infty - \left(\frac{-1}{2}\right)^m c^{\frac{1}{2}} y_\infty^{\frac{1}{2}} + \left(\frac{-1}{2}\right)^{2m-1} c \quad \ldots\ldots(3{\cdot}1),$$

where c is a constant depending on the initial conditions. Hence equilibrium is rapidly approached, the values in successive generations being alternately greater and less than the final value.

If selection occurs with a coefficient k in ♀s, l in ♂s, then

$$\left.\begin{aligned} u_{n+1} &= \frac{2u_n v_n + u_n + v_n}{u_n + v_n + 2 - 2k} \\ v_{n+1} &= \frac{u_n (v_n + 1)}{v_n + 1 - l} \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots(3{\cdot}2).$$

If the population is nearly in equilibrium apart from selection and k and l are small, so that u_n and v_n are nearly equal,

$$\Delta u_n = \frac{v_n - u_n}{2} + \frac{k u_n}{u_n + 1},$$

$$\Delta v_n = u_n - v_n + \frac{l u_n}{u_n + 1}, \text{ both approximately.}$$

Hence
$$\Delta u_n = \frac{2k+l}{3}\frac{u_n}{u_n+1},$$

and
$$\frac{2k+l}{3}n = u_n - u_0 + \log_e\left(\frac{u_n}{u_0}\right) \quad \text{.........(3·3)},$$

approximately. Selection therefore occurs much as with a normally inherited autosomal factor.

REFERENCES.

1. HARRISON and GARRETT. *Proc. Roy. Soc.* B, 99, p. 241, 1926.
2. GAGER and BLAKESLEE. *Proc. Nat. Ac. Sci.* 13, p. 75, 1927.
3. FISHER. *Proc. Roy. Soc. Edin.* 42, p. 321, 1922.
4. KOENIGS. *Darb. Bull.* (2) 7, p. 340, 1883.
5. HALDANE. *Trans. Camb. Phil. Soc.* 23, p. 19, 1924.
6. HALDANE. *Biol. Proc. Camb. Phil. Soc.* 1, p. 158, 1924.
7. METZ. *Proc. Nat. Ac. Sci.* 12, p. 690, 1926.

A mathematical theory of natural and artificial selection. (*Part VI, Isolation.*) By Mr J. B. S. HALDANE, Trinity College.

[*Received* 12 February, *read* 24 February, 1930.]

It is generally believed that isolation has played an important part in evolution. If an organism is to evolve so as to adapt itself to a special type of environment, e.g. a cave or a desert, it must not be swamped in each generation by migrants from the original habitat.

We consider a series of cases, in each of which a new form is favoured in a limited area, the coefficient of selection being k. In each generation a number of migrants of the original type, equal to the whole population of the limited area multiplied by a constant l, migrate into it. k and l are taken to be small and of course positive. It is required to find the relationship between k and l if selection is to take place, and what equilibrium is reached, if any. We shall consider ten different cases.

1. *No amphimixis.* This is analogous to the cases considered on p. 21 of Part I of this series*, where several reasons are discussed which may lead to a failure of amphimixis. We may consider two types, A and B, which do not interbreed, A being the normal type, B that favoured in the area considered. Let u_n be the ratio of $A:B$ in the nth generation. The proportion of A is transformed in one generation, as the result of selection and immigration, from $\frac{u_n}{u_n+1}$ to $\frac{u_n(1-k)}{u_n+1}+l$. Hence

$$u_{n+1}=(1-k)\,u_n+l\,(u_n+1),\quad\text{and}\quad \Delta u_n=l\,(u_n+1)-ku_n.$$

Hence equilibrium is possible when $u_\infty=\frac{l}{k-l}$, a number which is positive only if $k>l$. If $k<l$, Δu_n is always positive, and the type B disappears. If $k>l$, $\Delta u_n=(k-l)(u_\infty-u_n)$. Hence the equilibrium is stable, and whatever be the value of u_0, that is to say the initial state of the population, the final state is $lA:(k-l)\,B$. Hence if A is to disappear nearly completely, k must greatly exceed l.

2. *Dominants favoured, recessives immigrate.* Let the nth generation produce gametes in the ratio $uA:1a$, where aa is the immigrant recessive type. Then the composition of the population after selection and immigration is

$$\frac{u_n^2}{(u_n+1)^2}AA:\frac{2u_n}{(u_n+1)^2}Aa:\left[\frac{1-k}{(u_n+1)^2}+l\right]aa.$$

---

* *Trans. Camb. Phil. Soc.* 23 (1924), 19–41.

Thus

$$u_{n+1} = \frac{u_n(u_n+1)}{u_n+1-k+l(u_n+1)^2} \text{ and } \Delta u_n = \frac{u_n}{u_n+1}\left[k-l(u_n+1)^2\right],$$

approximately.

Hence $u_\infty = \sqrt{\frac{k}{l}} - 1$, which is positive if $k > l$.

If $k < l$, Δu_n is always negative, so that dominants disappear, and selection is ineffective. If $k > l$ the equilibrium is stable, the final ratio being $k - l$ dominants to l recessives.

3. *Recessives favoured, dominants immigrate.* With the same convention as above, except that AA is the immigrant type, the transformed population is

$$\left[\frac{u_n^2}{(u_n+1)^2} + l\right] AA : \frac{2u_n}{(u_n+1)^2} Aa : \frac{1+k}{(u_n+1)^2} aa.$$

Thus

$$u_{n+1} = \frac{u_n^2 + u_n + l(u_n+1)^2}{u_n+1+k}, \text{ and } \Delta u_n = \frac{l(u_n+1)^2 - ku_n}{u_n+1},$$

approximately.

Hence $$u_\infty^2 + (2 - k/l)\,u_\infty + 1 = 0,$$

i.e. $$u_\infty = \{k - 2l \pm \sqrt{k(k-4l)}\}/2l.$$

If $k < 4l$, these roots are complex, Δu_n is always positive, and recessives disappear. If $k > 4l$, Δu_n is negative when u_n lies between the two values of u_∞, otherwise it is positive. Hence the larger root represents an unstable equilibrium. If

$$u_0 > \{k - 2l + \sqrt{k(k-4l)}\}/2l,$$

i.e. if initially there are too few recessives, recessives ultimately disappear in the face of natural selection. If u_0 be less than this value, the final state is given by the stable equilibrium

$$u_\infty = \{k - 2l - \sqrt{k(k-4l)}\}/2l,$$

and the final proportion of recessives is

$$\{k - 2l + \sqrt{k(k-4l)}\}/2k,$$

a number lying between unity and $\frac{1}{2}$. When we compare this case with the last, it is clear that selection is much less effective. Not only must it be relatively four times more intense to produce any permanent result, but, if k is not much greater than $4l$, a chance fluctuation may push the population past the point of unstable equilibrium, and the recessives be finally eliminated.

4. *No dominance.* Consider a population as above, where, after selection and immigration, the proportions are

$$\frac{u_n^2}{(u_n+1)^2}AA : \frac{2(1-K)u_n}{(u_n+1)^2}Aa : \left[\frac{1-k}{(u_n+1)^2}+l\right]aa,$$

K, as well as k, being small and positive. If K were negative, the population would come into equilibrium in the absence of immigration. The most important case is that in which $K<k$. We have

$$u_{n+1} \doteq \frac{u_n^2+(1-K)u_n}{(1-K)u_n+1-k+l(u_n+1)^2},$$

so that

$$\Delta u_n = \frac{Ku_n(u_n-1)+ku_n-lu_n(u_n+1)^2}{u_n+1}, \text{ approximately.}$$

Hence

$$K(u_\infty-1)+k-l(u_\infty+1)^2=0, \text{ or } u_\infty^2+\left(2-\frac{K}{l}\right)u_\infty+1+\frac{K-k}{l}=0,$$

i.e. $$u_\infty=\{K-2l\pm\sqrt{4kl-8Kl+K^2}\}/2l.$$

The roots are real if $4kl+K^2>8Kl$. This condition is obviously fulfilled if $k>2K$. But both are negative, i.e. no equilibrium is possible, if $2l>K$, and $l+4K>4k$. Hence for an equilibrium to be possible $K>2l$, or $4k>4K+l$. If only the former is true one root is positive, and one negative. The positive root represents a stable equilibrium, and this is reached whatever the initial composition. If $K>2l$, and $4k>4K+l$, both roots are positive, the smaller representing an unstable equilibrium, so that u_0 must exceed the smaller root for selection to be effective.

In any case, when equilibrium is reached, the proportion of recessives is

$$\frac{K^2+2kl-4Kl-K\sqrt{4kl+K^2-8Kl}}{2(k-2K)^2}.$$

If $k=2K$, i.e. the heterozygotes are exactly intermediate, this expression is equal to l^2/K^2.

5. *Sex-linked dominants favoured, recessives immigrate.* Let the nth generation produce eggs in the ratio $u_nA:1a$, spermatozoa in the ratio $v_nA:1a$. After selection and migration, the population is in the proportions

$$\frac{u_nv_n}{(u_n+1)(v_n+1)}AA♀ : \frac{u_n+v_n}{(u_n+1)(v_n+1)}Aa♀$$
$$: \left[\frac{1-k}{(u_n+1)(v_n+1)}+l\right]aa♀,$$

and $$\frac{u_n}{u_n+1}A♂ : \left[\frac{1-k}{u_n+1}\right]a♂.$$

Thus $$u_{n+1}=\frac{2u_n v_n+u_n+v_n}{u_n+v_n+2-2k+2l(u_n+1)(v_n+1)},$$

$$v_{n+1}=\frac{u_n}{1-k+l(u_n+1)}.$$

Hence, since u_n-v_n is a small quantity,

$$\Delta u_n=\frac{v_n-u_n}{2}+\frac{ku_n}{u_n+1}-lu_n(u_n+1), \text{ approximately,}$$

and $$\Delta v_n=u_n-v_n+ku_n-lu_n(u_n+1), \text{ approximately.}$$

Thus $$\Delta u_n=\frac{ku_n(u_n+3)}{3(u_n+1)}-lu_n(u_n+1), \text{ approximately,}$$

and $3l(u_\infty+1)^2=k(u_\infty+3)$, i.e. $u_\infty=\{k-6l\pm\sqrt{k(k+24l)}\}/6l$.

The roots are always real, but neither is positive unless $k>l$, and one is always negative. Hence, if $k<l$, selection is ineffective, whilst, if $k>l$, the population, whatever its initial composition, reaches a stable equilibrium in which the proportion of recessive males is $\frac{1}{4}(\sqrt{1+24l/k}-1)$, that of recessive females being the square of this quantity. Clearly they may assume any value between 0 and 1.

6. *Sex-linked recessives favoured, dominants immigrate.* Adopting the same convention as above, we find

$$u_{n+1}=\frac{2u_n v_n+u_n+v_n+2l(u_n+1)(v_n+1)}{u_n+v_n+2+2k}, \quad v_{n+1}=\frac{u_n+l(u_n+1)}{1+k}.$$

Thus $$\Delta u_n=\frac{v_n-u_n}{2}-\frac{ku_n}{u_n+1}+l(u_n+1), \text{ approximately,}$$

$$\Delta v_n=u_n-v_n-ku_n+l(u_n+1), \text{ approximately,}$$

and $$\Delta u_n=l(u_n+1)-\frac{ku_n(u_n+3)}{3(u_n+1)}, \text{ approximately.}$$

Thus $$3l(u_\infty+1)^2=ku_\infty(u_\infty+3),$$

i.e. $$u_\infty=\{3k-6l\pm\sqrt{3k(3k-8l)}\}/(6l-2k).$$

The roots are real if $k>8l/3$. If $k<8l/3$ selection is ineffective. If $3l>k>8l/3$ the larger root represents an unstable equilibrium, the smaller a stable. So selection is ineffective if the original population contains too many dominants. If however $k>3l$, one root is negative, the positive root represents a stable equilibrium, and selection is effective whatever the original composition. When stable equilibrium is reached the proportion of recessive males is $\frac{1}{4}\{1+\sqrt{9-24l/k}\}$. The proportion may thus have any value between $\frac{1}{2}$ and 1. The proportion of females is the square of this number.

In the remaining four cases we first consider a character determined by two genes, and then generalize the result to apply to one determined by m. In the former case we can clearly represent any population by a point in a plane. If our coordinates are u and v, the ratios of the genes $A:a$ and $B:b$, we require one quadrant of an infinite plane. If we plot $\frac{1}{u+1}$ and $\frac{1}{v+1}$, or $\frac{1}{(u+1)^2}$ and $\frac{1}{(v+1)^2}$, which correspond to the actual proportions of genotypes, all populations can be represented within a finite square. Under the influence of natural selection the representative point takes up a series of positions, which, if selection is slow, are very close together, and lie on a definite trajectory passing to a point of stable equilibrium. In the case of m genes, the populations are represented by points in m-dimensional space, through each of which, in general, passes one and only one trajectory.

7. *Double or single dominants (i.e. all genotypes save aa bb) favoured, double recessives immigrate.* Consider a population whose nth generation produces gametes in the ratios $u_n A : 1a$, $v_n B : 1b$. The proportion $(u_n+1)^{-2}(v_n+1)^{-2}$ of double recessives is reduced by $k(u_n+1)^{-2}(v_n+1)^{-2}$ as the result of selection. Hence the proportion $(u_n+1)^{-2}$ of the genotype aa is reduced by

$$[1-k(v_n+1)^{-2}].$$

At the same time it is increased by l as the result of immigration.

Thus
$$u_{n+1}=\frac{u_n^2+u_n}{u_n+1-\dfrac{k}{(v_n+1)^2}+l(u_n+1)^2},$$

whence

$$\Delta u_n=\frac{ku_n}{(u_n+1)(v_n+1)^2}-lu_n(u_n+1),\text{ approximately,}$$

$$\Delta v_n=\frac{kv_n}{(u_n+1)^2(v_n+1)}-lv_n(v_n+1),\text{ approximately.}$$

Thus
$$(u_\infty+1)^2(v_\infty+1)^2=k/l.$$

Any of the singly infinite number of populations in which the proportion of double recessives is l/k is therefore in equilibrium, provided that $k>l$, and the equilibrium is stable.

We have $$\frac{du_n}{dv_n} = \frac{\Delta u_n}{\Delta v_n}, \text{ approximately}$$
$$= \frac{u_n(u_n+1)}{v_n(v_n+1)},$$

so that $$\frac{u_n}{u_n+1}\,\frac{u_0+1}{u_0} = \frac{v_n}{v_n+1}\,\frac{v_0+1}{v_0}.$$

If therefore we put $x_n = 1/(u_n+1)$, $y_n = 1/(v_n+1)$, x_n and y_n are the proportions of the genes a and b respectively, lying between 0 and 1, and $\frac{1-x_n}{1-x_0} = \frac{1-y_n}{1-y_0}$. The trajectories are thus all straight lines passing through the point (1, 1) and all end on the segment of the hyperbola $xy = (l/k)^{\frac{1}{2}}$ included within the square containing the representative points.

In the case of m genes, let z_n be the proportion $\Pi\,({}_ru_n+1)^{-2}$ of multiple recessives, where ${}_ru_n$ is the ratio of $A_r : a_r$.

Thus $${}_ru_{n+1} = \frac{{}_ru_n}{1+({}_ru_n+1)(l-kz_n)},$$
$$\Delta\,{}_ru_n = {}_ru_n\,({}_ru_n+1)(kz_n-l).$$

Hence the trajectories are straight lines passing through
$$(1, 1, 1, \ldots, 1)$$
and ending on the $(m-1)$-dimensional manifold
$$x_1x_2x_3\ldots x_m = (l/k)^{\frac{1}{2}},$$
where x_1, x_2, etc. are the proportions of the gene a_1, a_2, etc. in the population. The final proportion of recessives is again k/l.

8. *Double recessives favoured, double dominants immigrate.* With a population constituted as in the last case,
$$u_{n+1} = \frac{u_n^2+u_n+l(u_n+1)^2}{u_n+1+\dfrac{k}{(v_n+1)^2}},$$
so that
$$\Delta u_n = l(u_n+1) - \frac{ku_n}{(u_n+1)(v_n+1)^2}, \text{ approximately,}$$
$$\Delta v_n = l\,(v_n+1) - \frac{kv_n}{(u_n+1)^2(v_n+1)}, \text{ approximately}$$

Thus $$l\,(u_\infty+1)^2(v_\infty+1)^2 = ku_\infty = kv_\infty,$$
or, putting $$u_\infty+1 = v_\infty+1 = p,$$
$$lp^4 - kp + k = 0.$$

This has two real positive roots if $k > 256l/27$. Otherwise the roots are complex and selection is ineffective. When the roots are real the larger represents an unstable equilibrium, the smaller a stable. Plotting v against u, each representative point lies on a trajectory leading either to the point of stable equilibrium or to $u = v = \infty$. The two families of trajectories are separated by a curve passing through the point of unstable equilibrium, symmetrical about the line $u = v$, and having $u = 0$, $v = 0$ for asymptotes. At equilibrium the proportion of double recessives is necessarily greater than $\frac{81}{256}$.

In the case of m genes,

$${}_ru_{n+1} = \frac{{}_ru_n + l\,({}_ru_n + 1)}{1 + k\,({}_ru_n + 1)\,y_n},$$

so that

$$\Delta\,{}_ru_n = ({}_ru_n + 1)\,(l - k\,{}_ru_n\,y_n).$$

Putting ${}_ru_\infty + 1 = p$, we have

$$lp^{2m} - kp + k = 0.$$

This equation has two real positive roots if, and only if,

$$\frac{k}{l} > \frac{2m^{2m}}{(2m-1)^{2m-1}}.$$

Again the larger represents an unstable, the smaller a stable equilibrium. The families of trajectories either pass to the latter or to infinity, being separated by a $(m-1)$-dimensional manifold passing through the point of unstable equilibrium. At equilibrium the proportion of double recessives necessarily exceeds $\dfrac{(2m-1)^{2m}}{2m^{2m}}$.

We now pass from these cases, characteristic of allopolyploids such as wheat, where a character may be determined by any one of a number of genes, to cases where, as in many diploids, all of several dominants are needed to determine it.

9. *Double, but not single, dominants selected, double recessives immigrate.*

In a population composed as above the favoured individuals possess at least one A and one B. Now a proportion $[1 - (v_n + 1)^{-2}]$ of the AA and Aa zygotes are BB or Bb. Hence the effect of selection is to increase the numbers of AA and Aa by the factor $[1 + k - k\,(v_n + 1)^{-2}]$. Thus

$$u_{n+1} = \frac{u_n(u_n + 1)\,[1 + k\,(v_n^2 + 2v_n)\,(v_n + 1)^{-2}]}{u_n\,[1 + k\,(v_n^2 + 2v_n)\,(v_n + 1)^{-2}] + 1 + l\,(u_n + 1)^2},$$

so that

$$\Delta u_n = \frac{ku_n\,(v_n^2 + 2v_n)}{(u_n + 1)\,(v_n + 1)^2} - lu_n\,(u_n + 1), \text{ approximately,}$$

$$\Delta v_n = \frac{kv_n\,(u_n^2 + 2u_n)}{(v_n + 1)\,(u_n + 1)^2} - lv_n\,(v_n + 1), \text{ approximately.}$$

Thus

$$u_\infty = v_\infty = 0, \text{ or } l(u_\infty+1)^2(v_\infty+1)^2 = k(u_\infty{}^2+2u_\infty) = k(v_\infty{}^2+2v_\infty).$$

Hence $u_\infty = v_\infty$. Putting $(u_\infty+1)^2 = q$, we have

$$lq^2 - kq + k = 0,$$

$$q = \{k \pm \sqrt{k^2-4kl}\}/2l.$$

The roots are real if $k > 4l$. Otherwise selection is ineffective, and dominants disappear. If the roots are real, both exceed unity, and therefore represent equilibria, the larger representing a stable equilibrium, the smaller an unstable. The proportion of double dominants in the population in stable equilibrium is

$$\{k - 2l + \sqrt{k^2-4kl}\}/2k,$$

which necessarily exceeds $\frac{1}{4}$.

As before, the trajectories form two families, one passing to the point of stable equilibrium, the other through the point (0, 0) when u and v are plotted. They are separated by a curve passing through the point of unstable equilibrium.

Generalising for the case of m genes A_r, the ratio $A_r : a_r$ in the nth generation being ${}_ru_n$, the proportion of multiple dominants

$$y_n = \prod_{r=1}^{m} [1 - ({}_ru_n+1)^{-2}].$$

Of these $\dfrac{{}_ru_n}{{}_ru_n+2}$ are A_rA_r, $\dfrac{2}{{}_ru_n+2}$ are A_ra_r. Thus

$${}_ru_{n+1} = \frac{\dfrac{{}_ru_n{}^2 + {}_ru_n}{({}_ru_n+1)^2} + \dfrac{k({}_ru_n+1)\,y_n}{{}_ru_n+2}}{\dfrac{{}_ru_n+1}{({}_ru_n+1)^2} + \dfrac{ky_n}{({}_ru_n+2)} + l},$$

and

$$\Delta\,{}_ru_n = ({}_ru_n+1)\left(\frac{ky_n}{{}_ru_n+2} - l\,{}_ru_n\right),$$

whence

$$ky_\infty = l\,{}_ru_\infty({}_ru_\infty+2).$$

Hence ${}_ru_\infty$ is independent of the value of r. Putting

$$1 - (u_\infty+1)^{-2} = z^{-1},$$

we have $y_\infty = z^{-m}$, where $lz^m - kz + k = 0$.

This has two real positive roots if, and only if,

$$\frac{k}{l} > \frac{m^m}{(m-1)^{m-1}}.$$

The larger root of z corresponds to small values of u_∞ and defines an unstable equilibrium, the smaller a stable equilibrium. At the stable equilibrium $z < m/(m-1)$, so that $y_\infty > \{(m-1)/m\}^m > \frac{1}{4}$. The trajectories in m-dimensional space again form two distinct families.

10. *Double, but not single, dominants selected against, double dominants immigrate.* The effect of selection is as in the last case, the sign of k being changed. Thus

$$u_{n+1} = \frac{u_n(u_n+1)\{1-k(v_n^2+2v_n)(v_n+1)^{-2}\}+l(u_n+1)^2}{u_n\{1-k(v_n^2+2v_n)(v_n+1)^{-2}\}+1},$$

and $\Delta u_n = l(u_n+1) - \dfrac{ku_n(v_n^2+2v_n)}{(u_n+1)(v_n+1)^2}$, approximately,

$$\Delta v_n = l(v_n+1) - \frac{kv_n(u_n^2+2u_n)}{(v_n+1)(u_n+1)^2}, \text{ approximately.}$$

Thus $l(u_\infty+1)^2(v\ \ +1)^2 = ku_\infty v_\infty(v_\infty+2) = ku_\infty v_\infty(u_\infty+2)$.

Hence $u_\infty = v_\infty$, being a root of

$$u_\infty^4 + (4-k/l)u_\infty^3 + (6-2k/l)u_\infty^2 + 4u + 1 = 0.$$

This equation always has two real negative roots. It has also two real positive roots if and only if $k > 2^{-10}.7^{-1}(\sqrt{17}+1)^6(\sqrt{17}-3)l$, or about $3{\cdot}57\,l$. Otherwise selection is ineffective. If k exceeds this value the situation is similar to that of case 8, the larger root representing an unstable equilibrium, the smaller a stable. At stable equilibrium the proportion of double dominants is less than $\left(\dfrac{23+\sqrt{17}}{32}\right)^2$, or 71·8 °/ₒ; that of double recessives exceeds $\left(\dfrac{\sqrt{17}-1}{8}\right)^4$, or 2·32 °/ₒ. As before, the trajectories in the u, v plane fall into two families, one passing to infinity, the other to the point of stable equilibrium.

In the case of m factors,

$${}_ru_{n+1} = \frac{\dfrac{{}_ru_n}{{}_ru_n+1} - \dfrac{k({}_ru_n+1)y_n}{{}_ru_n+2} + l}{\dfrac{1}{{}_ru_n+1} - \dfrac{ky_n}{{}_ru_n+2}},$$

so that $\Delta\,{}_ru_n = ({}_ru_n+1)\left(l - \dfrac{ky_n}{{}_ru_n+2}\right)$.

At equilibrium

$$u_\infty + 2 = (k/l)[1-(u_\infty+1)^{-2}]^m,$$

or, if $u_\infty + 1 = p$,

$$p + 1 = (k/l)(1-p^{-2})^m.$$

This has two real positive roots if

$$\frac{k}{l} > \frac{\sqrt{8m+1}+3}{2}\left(\frac{4m+1+\sqrt{8m+1}}{4m}\right)^m\left(\frac{4m-1-\sqrt{8m+1}}{4(m-1)}\right)^m.$$

As above, the larger root represents an unstable equilibrium, the smaller a stable, and the trajectories fall into two families. At equilibrium the proportion of double dominants is less than

$$\left(\frac{4m-1+\sqrt{8m+1}}{4m}\right)^m\left(\frac{4m+1-\sqrt{8m+1}}{4m}\right)^m,$$

that of double recessives exceeds

$$\left(\frac{\sqrt{8m+1}-1}{4m}\right)^m.$$

It will be seen that in every case k/l must exceed a certain critical value unless the selected type is to be completely swamped by immigration. When this value is exceeded it is further necessary, in some cases, that the proportion of the selected type in the initial population should exceed a certain fraction. Now in nature the value of k/l must fluctuate widely round an average. Occasionally a flood must sweep large numbers of normal-eyed aquatic animals into a cave where the majority are blind, and so on. And where one of the equilibria is unstable, this may lead to the disappearance of the selected type. It would thus seem that in partially isolated communities selection is most likely to be effective when it favours a dominant or sex-linked character.

The problem can readily be generalized in several ways. Thus the effects of partial inbreeding might be considered. As however the results of complete inbreeding (case 1) are not very different from those of random mating, the general character of the solution will be unaltered.

One group of cases of a practical character will be considered, namely the result of breeding plants which are partially cross-pollinated, but where seed is only gathered from those of a desirable type. For the sake of simplicity it is assumed that rogues (plants of undesirable types) are destroyed before they produce pollen. We also assume complete self-sterility, so that all fertilization is by pollen of other plants. The average proportion of pollen from the wild population surrounding the group under selection is denoted by l, which in this case need not be a small quantity.

If the cultivated plant be recessive to the wild population, whether for one or many factors, it is obvious that the proportion of rogues in each generation is l. If it is dominant to the wild type with respect to a single factor A, suppose that, after the removal of rogues, gametes are formed in the ratio $u_n A : 1\,a$. After admixture of foreign pollen, the ratio of pollen grains will be $(1-l)\,u_n A : (1+lu_n)\,a$. Hence zygotes are formed in the ratios

$$(1-l)\,u_n^2 AA : (2-l+lu_n)\,Aa : (1+lu_n)\,aa.$$

The AA and Aa zygotes contribute to the next generation, so that

$$u_{n+1} = \frac{2(1-l)\,u_n^2 + (2-l+lu_n)\,u_n}{(2-l+lu_n)\,u_n} = \frac{(2-l)(u_n+1)}{2-l+lu_n},$$

and $$\Delta u_n = \frac{2-l-lu_n^2}{2-l+lu_n},$$

whence $$u_\infty = \sqrt{(2/l-1)}.$$

This represents a stable equilibrium, approached fairly rapidly from both sides. At equilibrium the proportion of rogues is

$$\frac{1+lu_\infty}{(u_\infty+1)^2}, \text{ or } \frac{l}{2}.$$

The expression in the case when the selected type is dominant to the wild with regard to several factors is more complicated, but such a case is too improbable to be worth considering. Provided that it is recessive for even one factor, all other factors are automatically eliminated in one generation.

A Mathematical Theory of Natural and Artificial Selection. Part VII. Selection intensity as a function of mortality rate. By Mr J. B. S. HALDANE, Trinity College.

[*Received* 12 November, *read* 8 December 1930.]

The assumption is often made that when competition is extremely intense at any stage in a life cycle, natural selection is bound to be intense also. This assumption will be examined quantitatively and it will be shown that the intensity of selection may diminish and become negative at high rates of elimination, while at its best its increase is extremely slow.

The intensity of competition is measured by the ratio, z, of organisms eliminated, to survivors. This may be small, e.g. $z = 0{\cdot}1$ or less for the period between birth and maturity in civilised human societies. It may exceed 10^6 in marine organisms producing many million eggs per year, or spermatozoa of which 10^9 are ejaculated at a time. But in few cases can it exceed 10^{12}.

Confining ourselves for the moment to a population consisting of two types A and B, the intensity of selection is measured by the coefficient of selection k, where the ratio of A to B is increased $1 + k$ times as the result of selection. k is taken to be small throughout the argument.

Consider a character whose measure x is normally distributed, according to Gauss' law, in the A and B groups, the standard deviation being the same in each, and the differences between the means and the standard deviations being small in comparison with the standard deviations. For example, Johansen (1926) found the mean breadths of 8·091 and 8·152 mm. and standard deviations of ·400 and ·405 mm. in two lines, BB and GG, of beans, the difference being clearly significant in the first case, doubtfully so in the second. If all individuals in which the variate x falls below a certain value are eliminated by selection, we can readily calculate the proportion of the whole population eliminated, and the proportion of A to B among the survivors.

Conditions are not grossly dissimilar under natural selection. We may imagine a variate, to be called viability, which is normally distributed and such that only those individuals possessing more than a certain viability survive. The large size of its standard deviation compared to the difference of the mean values would signify the relatively large part played by chance in natural selection. The best studied case is that of pollen-tube growth, described by Buchholz and Blakeslee (1929). Here those tubes which arrive first at the ovules are selected. The distribution of growth rates is definitely skew, but the skewness is not likely to affect the general character of the result if the two types compared are sufficiently similar. Where viability depends on a greater variety of accidental causes, as is generally the case, the distribution is likely to be more normal.

Without loss of generality we can put the initial numbers of A and B equal, and take the mean value of x as zero and its standard deviation as unity. We suppose the mean value of x for the A type to be λ and its standard deviation to be $1+\mu$, the corresponding values for B being $-\lambda$ and $1-\mu$. The ratio of the frequency of any value of x in the population to that in a strictly normal population is $1-(1+x^2)(\lambda+\mu x)^2+$ higher powers of λ and μ. Hence the mixed population is normal to the second order of small quantities provided that λx and μx^2 are small. x will rarely exceed 7 even in a population of 10^{12}.

Then provided that the population is numerous compared with both the numbers surviving and eliminated, the survivors will be those members for which $x > X$, X being given by

$$\frac{1}{\sqrt{2\pi}}\int_X^\infty e^{-\frac{1}{2}x^2}\,dx=\frac{1}{z+1}\,.$$

the proportion of the A type exceeding this value is

$$\begin{aligned}\frac{1}{\sqrt{2\pi}\,(1+\mu)}\int_X^\infty e^{\frac{-(x-\lambda)^2}{2(1+\mu)^2}}\,dx&=\frac{1}{\sqrt{2\pi}}\int_{(X-\lambda)/(1+\mu)}^\infty e^{-\frac{1}{2}t^2}\,dt\\&=\frac{1}{\sqrt{2\pi}}\int_X^\infty e^{-\frac{1}{2}t^2}\,dt+\frac{1}{\sqrt{2\pi}}\int_{(X-\lambda)/(1+\mu)}^X e^{-\frac{1}{2}t^2}\,dt\\&=\frac{1}{z+1}+\frac{(\lambda+\mu X)}{\sqrt{2\pi}}e^{-\frac{1}{2}X^2},\text{ approximately,}\end{aligned}$$

so that

$$k=\frac{2(\lambda+\mu X)(z+1)}{\sqrt{2\pi}}e^{-\frac{1}{2}X^2},$$

the value of X being found as above.

First consider the case when $\mu=0$, i.e. the standard deviations are equal. The value of $q=k/2\lambda$ is plotted against $\log_{10} z$ in the figure (calculated from Pearson's (1924) tables). When $z=1$, $q=(2/\pi)^{\frac{1}{2}}=\cdot 798$. When z is large, we may put

$$\frac{1}{1+z}=\frac{e^{-\frac{1}{2}X^2}}{\sqrt{2\pi}\,X}\text{ approximately,}$$

whence

$$q=\sqrt{\log_e\frac{z^2}{2\pi}}\text{ approximately.}$$

So the intensity of selection only increases extremely slowly with z. Thus q is only doubled when z increases from 1 to about 6·4, or from 10 to 1800, and only increased 9 times over the whole range from 1 to 10^{12}. On the other hand when z is small, q approximates to $z\sqrt{-\log_e 2\pi z^2}$, a small quantity of the order of z, and is roughly proportional to z over small ranges. For example, when $z = 10^{-4}$, $q = \cdot 0004$, and when $z = 10^{-2}$, $q = \cdot 03$. To sum up, the efficiency of selection increases very rapidly with z until about 80 °/₀ of the population is eliminated, and thereafter very slowly.

The only experimental data known to me are those of Correns (1918) who measured the sex-ratio of *Melandrium* when pollinated with mixtures of male-producing and female-producing pollen, and used numbers of pollen-grains either less than that of ovules, so that $z = 0$, or greater, in various proportions. When z approximated to 6, k was 0·195; when z was about 142, k rose to ·710. The value of k thus increased only 3·6 times while z increased 24 times. According to Fig. 1 the increase of q should be only 1·8 times. But the values of both z and k are very uncertain, thus the value ·195 of k has a standard error of ·07. Figures well within the limit of experimental error would give complete agreement with the theory. Moreover μ is probably not zero nor is the distribution of growth rates normal. Certainly, however, k does not increase anything like proportionally to z, even when λ has the somewhat large value of 0·1, which as we shall see later will tend to exaggerate the rate of increase of k with z.

When μ is not zero the case is rather more complicated. If λ and μ have the same sign, i.e. the type with the largest mean has also the largest standard deviation, selection favours them unless X is negative and less than $-\lambda/\mu$. In this case the group of lower average viability will be favoured when competition is very slight, but their selective advantage will be extremely small at best. For example, Johansen's bean line *GG* had a mean breadth of 8·152 mm. with standard deviation ·415, while the corresponding figures for the line *MM* were 7·976 and ·348. Hence $\lambda = \cdot 101$ and $\mu = \cdot 076$. Selection for greater breadth would favour line *MM* slightly when z was less than 0·1, while for higher values *GG* would be considerably favoured, k being ·159 when $z = 1$.

In many cases the coefficient of variation of the two groups is approximately equal, i.e. $\mu = \lambda$. In four of Johansen's pure lines the coefficient of variation for length only varied between 5·0 °/₀ and 4·4 °/₀. In a family described on p. 136 two slightly impure genotypes had coefficients of 7·0 °/₀ and 6·8 °/₀, the heterozygote a coefficient of 6·8 °/₀. It thus seems likely that in a large number of cases λ and μ will be very nearly equal. When this is the case q is small and negative for low mortalities, attaining a minimum value of $-0\cdot 070$ when $z = \cdot 066$, i.e. with a mortality of 6·2 °/₀, vanishes when $z = \cdot 1886$, i.e. with a mortality of 15·9 °/₀, and then increases, being ·798 when $z = 1$ (50 °/₀ mortality), 23·4 when $z = 10^6$, and 57·6 when $z = 10^{12}$. For large values of z, k varies as $\log z$.

Even though λ and μ are not quite equal, in a very large number, perhaps the majority of cases, λ/μ will lie between ·5 and 2, and the direction of selection will be reversed at a mortality of between 31 °/₀ and 2·3 °/₀. It is of interest to note that during the last fifty

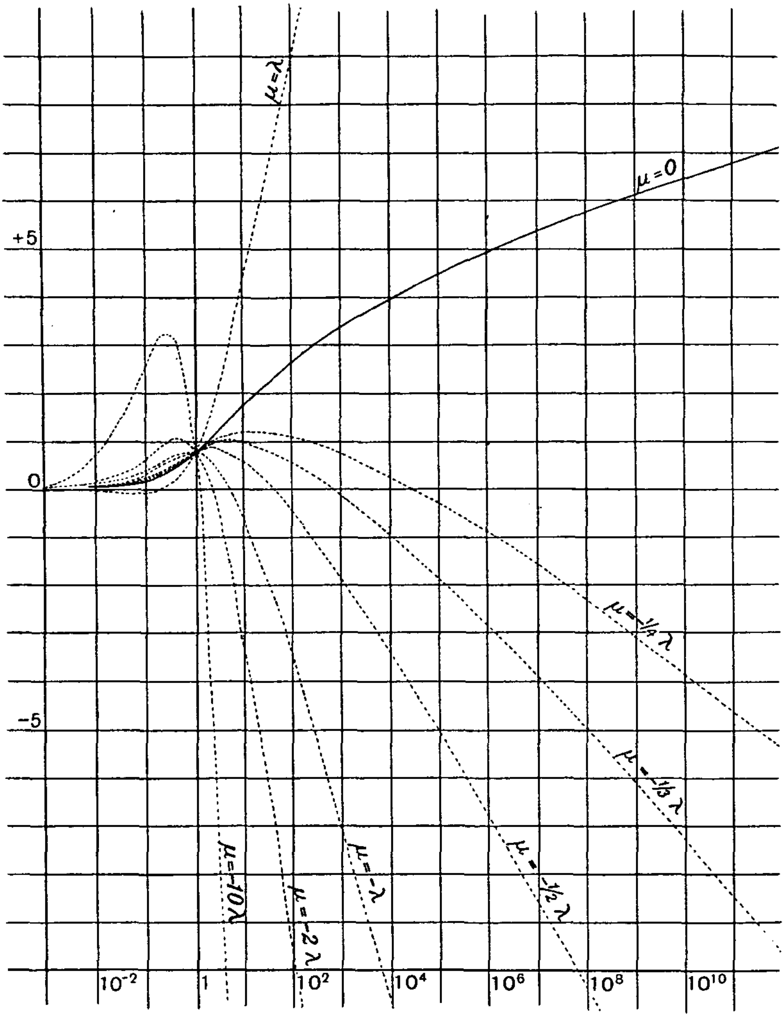

Fig. 1. Abscissa: $\text{Log}_{10} z$. Ordinate: q, which measures the intensity of selection.

years infantile mortality in most civilised countries has fallen from well above the critical value of 15·9 °/₀ to well below it. It seems probable, therefore, that the direction of selection for certain genes has been reversed.

If λ and μ have opposite signs, the group of highest average viability will be favoured until X exceeds $-\lambda/\mu$, i.e. until

$$z > \frac{\sqrt{2\pi}}{\int_{-\lambda/\mu}^{\infty} e^{-\frac{1}{2}x^2}\,dx} - 1.$$

If $-\lambda/\mu$ is fairly large, say greater than 3, this becomes approximately

$$z > \sqrt{2\pi}\frac{\lambda}{\mu} e^{\lambda^2/\mu^2}.$$

In practice however since z rarely exceeds 10^{12}, the direction of selection is not likely to be reversed if $-\mu < \lambda/7$.

In the figure q is plotted against z when

$$\mu = 0,\ \lambda,\ -10\lambda,\ -2\lambda,\ -\tfrac{1}{2}\lambda,\ -\tfrac{1}{3}\lambda,\ -\tfrac{1}{4}\lambda.$$

The maximum intensity of selection is reached when z is greater or less than unity according as $-\mu/\lambda$ is less or greater than unity.

But wherever it is not zero the results of slight and intense competition are in opposite directions, although the required competition may sometimes be too intense or the selection too slight to be of practical importance. This is in full accordance with the views of Bidder, who points out that, where "cataclasms" occasionally destroy the vast majority of a species, characters which are useless or worse under normal conditions may be selected. He specially mentions the case of a violent or erratic response of an animal by migration or otherwise to unfavourable environments, which would be likely to lower the average viability, but increase its dispersion.

It is easy to extend the above arguments to a population consisting of many genotypes. To take one example, suppose that $\mu = 0$, but λ is normally distributed with a standard deviation σ. Then the new frequency of any value of λ will be given by

$$df = \frac{1 + q\lambda}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\lambda^2/\sigma^2}\,d\lambda.$$

The new mean value of λ will therefore be

$$\frac{\int_{-\infty}^{\infty} \lambda(1 + q\lambda)\, e^{-\frac{1}{2}\lambda^2/\sigma^2}\,d\lambda}{\int_{-\infty}^{\infty} (1 + q\lambda)\, e^{-\frac{1}{2}\lambda^2/\sigma^2}\,d\lambda}$$

$$= \frac{q}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} \lambda^2 e^{-\frac{1}{2}\lambda^2/\sigma^2}\,d\lambda, \text{ approximately,}$$

$$= q\sigma^2.$$

All the results here given apply only to the results of a single act of selection. The way in which the population will change depends on the way in which the mean viability and its dispersion are inherited, and on the system of mating. The effects of these have been considered in former papers of this series.

The theory can readily be extended to cover cases where λ and μ are no longer small, but the results are no longer elegant or simple. In particular the proportion of types in the original population must be taken into account. When the difference of the means is large compared with the standard deviations, *A* being more viable than *B*, it is convenient to take κ as the intensity of selection, where u is the ratio of *A* to *B* before, and ue^{κ} after selection. κ and k are of course equal when both are small. It is clear that

$$\kappa = \log_e \left(\frac{1+z}{1+uz}\right),$$

approximately, so that the intensity of selection is proportional to z when this is small, but becomes very large when $z = 1/u$. Such intense selection occasionally occurs in nature, for example between normal types and semi-lethal mutants, but its results as between competing types of organism would be very rapid, and it is not of much interest in a study of evolution. In general when λ is not small, the value of q for any value of z will be increased.

REFERENCES.

G. P. Bidder (1930). "The importance of cataclasms in evolution." *Nature*, 125, p. 783.

J. T. Buchholz and A. F. Blakeslee (1929). "Pollen-tube growth in crosses between balanced chromosomal types of *Datura Stramonium*." *Genetics*, 14, p. 538.

C. Correns. "Fortsetzung der Versuche zur experimentellen Verschiebung des Geschlechtsverhältnisses." *Sitzber. K. Preuss. Akad. Wiss.* 1918, p. 1175.

W. Johansen (1926). *Elemente der exakten Erblichkeitslehre*, p. 171.

K. Pearson (1924). *Tables for Statisticians and Biometricians.*

A Mathematical Theory of Natural Selection. Part VIII. Metastable Populations. By Mr J. B. S. HALDANE, Trinity College.

[*Received* 20 November, *read* 8 December 1930.]

Almost every species is, to a first approximation, in genetic equilibrium; that is to say no very drastic changes are occurring rapidly in its composition. It is a necessary condition for equilibrium that all new genes which arise at all frequently by mutation should be disadvantageous, otherwise they will spread through the population. Now each of two or more genes may be disadvantageous, but all together may be advantageous. An example of such balance has been given by Gonsalez(1). He found that, in purple-eyed *Drosophila melanogaster*, arc wing or axillary speck (each due to a recessive gene) shortened life, but the two together lengthened it.

Consider the case of two dominant genes A, B, where the relative chances of producing offspring by the four phenotypes are as follows: AB 1, aaB $1-k_1$, Abb $1-k_2$, $aabb$ $1+K$. k_1 and k_2 are small and positive. K is small, and if negative its absolute value is less than k_1 or k_2.

Consider a random mating population where in the nth generation the genic ratios are $u_n A : 1a$; $v_n B : 1b$.

Then

$$u_{n+1} = \frac{(u_n^2 + u_n)\{1 - k_2(1+v_n)^{-2}\}}{u_n\{1 - k_2(1+v_n)^{-2}\} + 1 + \{K - k_1(v_n^2 + 2v_n)\}(1+v_n)^{-2}},$$

whence $$\Delta u_n = \frac{u_n\{k_1(1+v_n)^2 - K - k_1 - k_2\}}{(1+u_n)(1+v_n)^2}, \text{ approximately.}$$

So, taking a generation as the unit of time,

$$\frac{du}{dt} = \frac{u\{k_1(1+v)^2 - K - k_1 - k_2\}}{(1+u)(1+v)^2}, \text{ approximately.}$$

Let $x = 1/(1+u)$ (the proportion of recessive genes) and $y = 1/(1+v)$, so that $1 > x > 0$, $1 > y > 0$.

Then $$\frac{dx}{dt} = x^2(1-x)[(K + k_1 + k_2)y^2 - k_1].$$

Similarly $$\frac{dy}{dt} = y^2(1-y)[(K + k_1 + k_2)x^2 - k_2].$$

Clearly $x=0$, $y=0$; and $x=1$, $y=1$ are the only stable equilibria, though Fisher(2) appears to regard a mixed population as stable in such a case. Putting

$$\frac{k_1}{K+k_1+k_2}=a^2, \quad \frac{k_2}{K+k_1+k_2}=b^2,$$

we have
$$\frac{dy}{dx}=\frac{y^2(1-y)(x^2-a^2)}{x^2(1-x)(y^2-b^2)}.$$

So
$$\int_{y_0}^{y}\frac{(s^2-b^2)\,ds}{s^2(1-s)}=\int_{x_0}^{x}\frac{(s^2-a^2)\,ds}{s^2(1-s)},$$

whence
$$f(y, b)-f(x, a)=c=f(y_0, b)-f(x_0, a),$$

where x_0, y_0 represent the initial conditions, and

$$f(x, a)=a^2/x-a^2\log x+(a^2-1)\log(1-x).$$

Each value of $f(y_0, b)-f(x_0, a)$ determines a trajectory passing to $(0, 0)$ or $(1, 1)$, which represent populations composed entirely of double dominants or double recessives respectively. The minimum value of $f(x, a)$ occurs when $x=a$ and is

$$a-a^2\log a+(a^2-1)\log(1-a),$$

and $f(x, a)$ is always real and positive, becoming infinite when $x=0$ or 1. If $c>f(b, b)-f(a, a)$, there are two values of x corresponding to each value of y, but some values of x are excluded. Hence the trajectories fall into four families divided by the two branches of the curve whose equation is

$$f(y, b)-f(x, a)=f(b, b)-f(a, a).$$

This consists of two trajectories running from $(0, 1)$ and $(1, 0)$ to (a, b) and two from (a, b) to $(0, 0)$ and $(1, 1)$. These are represented by the dotted lines in Fig. 1, where $a=\frac{1}{2}$, $b=\frac{1}{4}$. The former divides the whole area into two portions. Populations in the one tend to the values $x=y=0$, in the other to the values $x=y=1$. Some examples of trajectories are given. It is clear that a population consisting mainly of *AABB* or *aabb* tends, as the result of selection, to return to those compositions. If the signs of K, k_1, and k_2 be changed, the same trajectories will be described in the reverse direction.

If the original population is *AABB*, the factors *A* and *B* will generally have a small tendency to mutate to *a* and *b* respectively. Let p_1 and p_2 be the probabilities that *A* will mutate to *a* and *B* to *b* in the course of a generation. These appear to be generally small numbers of the order of 10^{-6} or less. The population is in equilibrium when $x=p_1/k_1$, $y=p_2/k_2$ (Haldane(3)). In general x will be much smaller than a, and y than b, but from time to time chance fluctuations may isolate a population where this is no longer the case. Its representative point will lie in the area whose stable type is *aabb*, and the whole population will be transformed into this type, apart from rare exceptions due to back mutation. In such a population modifying factors will be selected in such a way as to increase the viability of the *aabb* type, i.e. the value of K. But even so it may be expected to be swamped by hybridisation on coming into contact with the original *AABB* population, unless one of two things has happened.

aabb may possess or develop characters which render mating with *AABB* rare. For example, it may have a different flowering time if a plant, or a different psychology if an animal. In this case the species will divide into two. Or chromosome changes may occur to cause close linkage of A and B when the populations are crossed. Thus if the loci of A and B are in the same chromosome an inversion of the portion containing them will lead to their behaving as a single factor on crossing. In this case if K is positive the whole species will be transformed into the type *aabb*. A species which is liable to transformations of this kind may be called metastable. Possibly metastability is quite a general phenomenon, but it is only rarely that the circumstances arise which favour a change of the type considered.

In a population which is mainly self-fertilised, conditions are probably more favourable. Were self-fertilisation universal, the proportion of *aaBB* zygotes, when mutation and selection were in equilibrium, would be $\frac{1}{2}p_1/k_1$. So that of *aabb* would be $p_1p_2/(4k_1k_2)$ or less. This is presumably a small number, probably of the order 10^{-9}, and when such individuals occur, they will generally be wiped out by chance. But their probability of spreading through the population, though small, will be finite, and roughly equal to $2K$ (Haldane (3)). Hence, within a geologically short period we may expect evolution to occur in such cases.

The theory may be extended in two different ways. We may consider m genes. In this case any population can be represented by a point in m-dimensional space, all populations being represented by the points of a regular orthotope, or hypercube. Each of the 2^m apices of this figure represents a homozygous population. Clearly the condition for stability of any such population is that no change in a *single* factor should yield a more viable type. In other words, no adjacent apices can both represent stable populations. The maximum number of stable populations is thus 2^{m-1}, represented by the vertices of the polytope arising from the omission of alternate vertices of the regular orthotope. This is not regular but only semi-regular if $m > 3$. In general the numbers of stable genotypes will be much smaller than this, and may not exceed 1.

If there is more than one stable population the orthotope is divided into two or more regions analogous to the two areas of Fig. 1. A population in any given region tends to the same point of stable equilibrium. The regions are separated by a variety (surface or hyper-surface) of $m-1$ dimensions. If we take as our variables x_1, x_2, x_3, etc. not the proportions of recessive genes, but their squares, i.e. the proportion of recessive zygotes, we have

$$\frac{dx_1}{dt}=x_1^{\frac{3}{2}}(1-x_1^{\frac{1}{2}})f_1(x_2, x_3, x_4, \ldots),$$

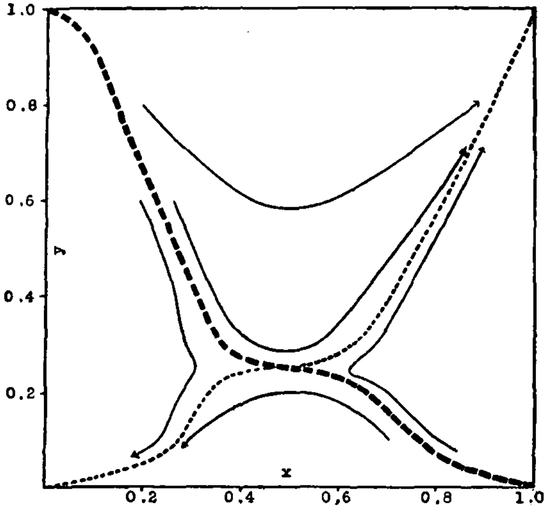

Fig. 1. Abscissa and ordinate. Proportions of genes *a* and *b* in a population. Trajectories of points representing populations are represented by continuous lines, and boundaries between families of trajectories by dotted lines.

where $f_1(x_2, x_3, x_4, \ldots)$ is linear in each of x_2, x_3, etc. and has 2^{m-1} constant coefficients; and $m-1$ similar equations. The $(m-1)$-dimensional space defined by x_2, x_3, etc. is thus divided into two regions, in one of which x_1 increases with time, whilst it diminishes in the other. These are not necessarily connected, as is obvious in the case where there are only three variables, and $f_1(x_2, x_3)$ may define a hyperbola which divides the unit square into three regions, in two of which dx_1/dt has the same sign. Hence in the course of a trajectory dx_1/dt may change sign several times. I have been unable to obtain the general equation for the trajectories or for the boundaries of the regions in which they lie.

So far we have only considered cases of complete dominance. If the heterozygotes are exactly intermediate in viability between the corresponding homozygous types, we have, in the terminology of the case first considered,

$$\frac{dx}{dt} = \tfrac{1}{2}x(1-x)\{(K+k_1+k_2)y-(K+k_1)\},$$

$$\frac{dy}{dt} = \tfrac{1}{2}y(1-y)\{(K+k_1+k_2)y-(K+k_2)\}.$$

Thus

$$\frac{dy}{dx} = \frac{y(1-y)(x-a)}{x(1-x)(y-b)}, \quad \text{where} \quad a = \frac{K+k_2}{K+k_1+k_2}, \quad b = \frac{K+k_1}{K+k_1+k_2}.$$

Hence
$$\frac{x^a(1-x)^{1-a}}{x_0{}^a(1-x_0)^{1-a}} = \frac{y^b(1-y)^{1-b}}{y_0{}^b(1-y_0)^{1-b}}.$$

By an argument similar to that used above we can show that the trajectories fall into two families, separated by one branch of the curve whose equation is

$$\left(\frac{x}{a}\right)^a\left(\frac{1-x}{1-a}\right)^{1-a} = \left(\frac{y}{b}\right)^b\left(\frac{1-y}{1-b}\right)^{1-b}.$$

The general case where heterozygotes are of any arbitrary viability is rather complicated. But where a heterozygote has a greater viability than any genotype differing from it in respect of a single gene only, there will be a stable population including some of these heterozygotes. Thus if *aabb* has a viability $1+K$, *AABb* of $1+K_2$, all other genotypes having unit viability,

$$\frac{dx}{dt} = xy(1-x)\{K_1xy - K_2(1-x)(1-y)\},$$

$$\frac{dy}{dt} = y\{K_1x^2y - K_2(1-x)^2(2y-1)\}.$$

The stable equilibria are at $x=1$, $y=1$ and $x=0$, $y=\tfrac{1}{2}$. But I have not been able to integrate these equations, since the variables are not readily separable. Nevertheless it is clear that the trajectories fall into two groups bounded by a curve passing through $(0, 1)$ and $\left(\frac{K_2}{K_1+2K_2}, \tfrac{1}{2}\right)$.

In the case of m genes, if heterozygotes have an advantage as such there may be points of stable equilibrium anywhere in the m-dimensional space, but it seems fairly clear that their number cannot exceed 2^{m-1}.

It is suggested that in many cases related species represent stable types such as I have described, and that the process of species formation may be a rupture of the metastable equilibrium. Clearly such a rupture will be specially likely where small communities are isolated. I have to thank Mr C. H. Waddington for calculating and drawing the figure.

REFERENCES.

(1) B. M. Gonsalez, *Am. Nat.*, 57, p. 289 (1923).

(2) R. A. Fisher, *The Genetical Theory of Natural Selection*, p. 102 (1930).

(3) J. B. S. Haldane, *Proc. Camb. Phil. Soc.*, 23, p. 838 (1927).

A MATHEMATICAL THEORY OF NATURAL AND ARTIFICIAL SELECTION

PART X. SOME THEOREMS ON ARTIFICIAL SELECTION*

J. B. S. HALDANE
University of California, Berkeley, California

Received October 25, 1933

The former papers of this series have been mainly occupied with cases where no genotype was completely eliminated, though some were fitter than others. The main earlier work dealing with the case where certain genotypes were wholly prevented from breeding, is that of ZIRKLE (1926). He considers selection in random mating populations where the characters are determined by a number of unlinked genes, all present in equal proportions. This case has assumed a special importance in experimental work where a cross is made between pure lines. Thus LITTLE and his colleagues have investigated the susceptibility of mice to tumor transplantation. A tumor can be transplanted into any individual carrying each of k dominant genes, where k varies between 2 and 12 or more (CLOUDMAN 1932). There are only two phenotypes, susceptible and immune. If either of these is bred from in any generation to the exclusion of the other it is clear that in any generation the ratio of dominant to recessive genes will be the same in all the k loci, apart from differences due to the smallness of the sample. This is so however the population is derived, whether it is an F_2, a backcross or some later derivative. In consequence, as ZIRKLE pointed out, we can fix our attention on one of the k gene pairs, and any statement made about it is true for each other pair.

HALDANE (1926) has already dealt with the case where all the genes concerned are not present in equal numbers. This case, though important for evolution and eugenics, is relatively intractable. ZIRKLE'S analysis, though accurate and valuable, can be enormously simplified. Further he confines himself to random mating populations, and a geneticist desiring to fix a character would probably employ inbreeding. Hence certain theorems on inbreeding combined with selection will be proved. Again there is a close analogy between the populations considered and autopolyploid populations.

In each case we shall consider a population derived from an F_1 obtained by mating two pure lines, though many results can be extended to other populations. These latter are marked by an asterisk. The results are

---

* Part of the cost of the mathematical composition in this article is paid by the GALTON AND MENDEL MEMORIAL FUND.

given for very large populations, but they may be regarded as giving probabilities for smaller populations, provided that these are not so small as to entail appreciable inbreeding where mating is supposed to be at random. We shall suppose that all genes are autosomal and unlinked. By the expression "multiple dominant" is meant a zygote containing at least one of each of the k dominant genes, by "multiple recessive" a zygote containing none of them.

SELECTION OF MULTIPLE DOMINANTS; RANDOM MATING

This case is very simple for the following reason. A zygote is eliminated if it is recessive for any gene. The probability of this being the case is the same whatever other genes are present. Each of the k genes *A*, *B*, *C*,— must be present in a multiple dominant. Since mating is at random the genotypes in F_n are in the ratios

$$u_n^2 AA : 2u_n Aa : 1aa$$

and so for the other genes. The effect of eliminating all recessives for any gene will be to remove all *aa* zygotes, and some of the *AA* and *Aa*. These latter are eliminated in equal proportions, as *A* is not linked with any of the other genes concerned. The survivors are in the ratio

$$u_n AA : 2Aa,$$

giving gametes in the ratio

$$(u_n+1)A : 1a$$

whence

$$u_{n+1} = u_n + 1$$

just as when one gene only is concerned, and

$$u_n = u_0 + n$$

so that the proportion of multiple dominants in F_n is

$$D_n = \left[1 - \frac{1}{(u_0+n+1)^2}\right]^k. \tag{1.1*}$$

But in F_2, $u_2 = 1$, whence $u_n = n-1$, and

$$D_n = (1 - n^{-2})^k. \tag{1.2}$$

This is equivalent to ZIRKLE's formula of p. 562 but simpler. For example if $k=8$, $D_2 = (1-\frac{1}{4})^8$ or .1001, $D_5 = (1-1/25)^8$ or .7218.

RELATION BETWEEN GAMETES OF SUCCESSIVE GENERATIONS OF A RANDOM MATING POPULATION, WITHOUT SELECTION

Before solving the next two problems an expression must be found for this relationship. Let F_n be formed from gametes such that the proportion of *each* type of gamete carrying r dominants is g_r. There are ${}_kC_r$ types of gamete carrying r dominants, each present in equal numbers. Hence

$$\sum_{r=0}^{k} {}_kC_r g_r = 1.$$

Let Z(i, j) be the proportion in F_n of *each* type of zygote homozygous for i dominant and j recessive genes. Clearly $k \geqq i+j \geqq 0$, and there are $\frac{k!}{i!j!(k-i-j)!}$ or ${}_kC_i\,{}_{k-i}C_j$ different types of such zygote, so that

$$\sum_{i=0}^{k} {}_kC_i \sum_{j=0}^{k-i} {}_{k-i}C_j\, Z(i, j) = 1.$$

It also follows that such a zygote can be made up in ${}_{k-i-j}C_r$ ways from gametes carrying (i+r) and (k−j−r) dominants, so that

$$Z(i, j) = \sum_{r=0}^{k-i-j} {}_{k-i-j}C_r\, g_{i+r}\, g_{k-j-r}. \tag{2.1*}$$

Such a zygote produces gametes containing from i to k−j (inclusive) dominant genes in proportions which are terms in the expansion of $\left(\frac{1+1}{2}\right)^{k-i-j}$.

That is to say it contributes to one type of gamete carrying m dominants a proportion $2^{i+j-k}\,{}_{k-i-j}C_{m-i}$ of its gametes. So if $g_m{}^1$ be the proportion of each type of gamete carrying m dominants produced by F_n, we have, since there are ${}_kC_i\,{}_{k-i}C_j$ types of zygote in the proportion Z(i, j) and ${}_kC_m$ types of gamete in the proportion $g_m{}^1$,

$$\begin{aligned} g_m{}^1 &= \frac{1}{{}_kC_m} \sum_{i=0}^{m} \sum_{j=0}^{k-m} 2^{i+j-k}\,{}_{k-i-j}C_{m-i}\,{}_kC_i\,{}_{k-i}C_j\, Z(i, j). \\ &= 2^{-k} \sum_{i=0}^{m} 2^i\,{}_mC_i \sum_{j=0}^{k-m} 2^j\,{}_{k-m}C_j\, Z(i, j) \qquad (2.2^*) \\ &= 2^{-k} \sum_{i=0}^{m} 2^i\,{}_mC_i \sum_{j=0}^{k-m} 2^j\,{}_{k-m}C_j \sum_{r=0}^{k-i-j} {}_{k-i-j}C_r\, g_{i+r}\, g_{k-j-r}. \end{aligned}$$

This expression can readily be altered if any classes of zygote are removed by selection.

ELIMINATION OF MULTIPLE DOMINANTS, RANDOM MATING

The classes of zygotes eliminated are recessive for none of the genes, that is to say in the expression for Z(i, j), j = 0. Hence the proportion of dominants in F_n is

$$D_n = \sum_{i=0}^{k} {}_kC_i \, Z(i, 0) \qquad (3.1^*)$$

and the gametes of the survivors, which form F_{n+1}, are given by the expression

$$g_m{}^1 = \frac{2^{-k} \sum\limits_{i=0}^{m} 2^i \, {}_mC_i \sum\limits_{j=1}^{k-m} 2j \, {}_{k-m}C_i \, Z(i, j)}{1 - D_n}. \qquad (3.2^*)$$

Further in the gametes of F_1, every $g_r = 2^{-k}$, and in all later generations $g_k = 0$.

For purposes of calculation it is desirable to write out the necessary formulae for each Z(i, j) from equation (2.1). These number $\frac{1}{2}(k+1)(k+2)$. They can then be substituted in equations (3.1) and (3.2). For example if $k = 6$ we obtain 28 such expressions as

$$Z(2, 0) = 2g_2g_6 + 8g_3g_5 + 6g_4{}^2.$$

Whence

$$D_n = 2g_6 - g_6{}^2 + 6g_5(2g_1 + 10g_2 + 20g_3 + 20g_4 + 5g_5) + 30g_4(g_2 + 4g_3 + 3g_4) + 20g_3{}^2$$

$$g_0{}^1 = \frac{1}{8(1 - D_n)} [g_0(8g_0 + 48g_1 + 60g_2 + 40g_3 + 15g_4 + 3g_5) + 15g_1(4g_1 + 8g_2 + 4g_3 + g_4) + 15g_2(3g_2 + 2g_3)]$$

$$g_1{}^1 = \frac{1}{16(1 - D_n)} [g_0(16g_1 + 40g_2 + 40g_3 + 20g_4 + 5g_5) + g_1(56g_1 + 200g_2 + 160g_3 + 65g_4 + 10g_5) + 10g_2(14g_2 + 17g_3 + 4g_4) + 30g_3{}^2]$$

$$g_2{}^1 = \frac{1}{4(1 - D_n)} [g_0(2g_2 + 4g_3 + 3g_4 + g_5) + g_1(2g_1 + 20g_2 + 28g_3 + 17g_4 + 4g_5) + g_2(29g_2 + 62g_3 + 28g_4 + 4g_5) + 12g_3(2g_3 + g_4)]$$

$$g_3{}^1 = \frac{1}{16(1 - D_n)} [g_0(4g_3 + 6g_4 + 3g_5) + 3g_1(4g_2 + 16g_5 + 17g_4 + 6g_5) + 6g_2(7g_2 + 31g_3 + 24g_4 + 6g_5) + 2g_3(71g_3 + 78g_4 + 12g_5) + 24g_4{}^2]$$

$$g_4{}^1 = \frac{1}{8(1 - D_n)} [g_0(g_4 + g_5) + g_1(4g_3 + 13g_4 + 8g_5) + g_2(3g_2 + 34g_3 + 56g_4 + 24g_5) + 8g_3(6g_3 + 13g_4 + 4g_5) + 8g_4(5g_4 + 2g_5)]$$

$$g_5{}^1 = \frac{1}{16(1 - D_n)} [g_0g_5 + 5g_1(g_4 + 2g_5) + 10g_2(g_3 + 4g_4 + 4g_5) + 10g_3(3g_3 + 12g_4 + 8g_5) + 80g_4(g_4 + g_5) + 16g_5{}^2].$$

Hence $D_2 = \frac{729}{4096} = .1780$, and, among the gametes forming F_3

$$g_0 = \frac{63}{3367},\quad g_1 = \frac{62}{3367},\quad g_2 = \frac{60}{3367},\quad g_3 = \frac{56}{3367},\quad g_4 = \frac{48}{3367},\quad g_5 = \frac{32}{3367},\quad g_6 = 0$$

whence $D_3 = \frac{1,416,512}{11,336,689} = .1249$, and D_4 and subsequent terms can easily be calculated.

ZIRKLE has given the requisite expressions for $k = 2, 3, 4$, and carried out the calculations. In the case of $k = 2$ the composition of each generation may be represented by a single parameter. BENNETT (1924) has solved the equation

$$S_{n+1} = S_n + 1 - \frac{1}{S_n}, \tag{3.3}$$

which arises if we represent the gametes forming F_n by

$$\frac{1}{S_n+1}Ab,\quad \frac{1}{S_n+1}aB,\quad \frac{S_n-1}{S_n+1}ab,$$

(where $S_3 = 5/2$).
Here

$$D_n = \frac{2}{(S_n+1)^2},$$

and

$$n + C = S_n + \log(S_n - 1) + \frac{1}{2S_n} - \frac{1}{3S_n^2} - \frac{5}{36S_n^3} - \frac{13}{240S_n^4} - \frac{193}{1800S_n^5} - \frac{947}{7560S_n^6} - \cdots\cdots \tag{3.4}$$

where $C = .64018855$——. An alternative solution is as follows:—

Let the gametes forming F_n be:—

$$\frac{x_n-1}{2x_n}Ab,\quad \frac{x_n-1}{2x_n}aB,\quad \frac{1}{x_n}ab.$$

Then

$$D_n = \frac{(x_n-1)^2}{2x_n^2},\qquad x_{n+1} = \frac{x_n}{2} + 1 - \frac{1}{2x_n}$$

or

$$\Delta x_n = x_{n+1} - x_n = -\frac{(x_n-1)^2}{2x_n}.$$

This is a particular case of the equation $\Delta x_n = \frac{-k(x_n-1)^2}{x_n}$ solved by

HALDANE (1932) in another connection. It was shown that for $x > 1$, as in this case, a very approximate solution is

$$n = C + \frac{1}{k(x_n - 1)} + \frac{\log x_n}{\log (1 - k)} + \frac{1 - k}{k} \log_e \left(\frac{x_n}{x_n - 1} \right).$$

Here $k = \frac{1}{2}$, so

$$n = C + \frac{2}{x_n - 1} - \frac{\log x_n}{\log 2} + \log_e \left(\frac{x_n}{x_n - 1} \right)$$

$$= C^1 + \sqrt{\frac{2}{D_n}} + \frac{\log (1 - \sqrt{2D_n})}{\log 2} - \frac{1}{2} \log_e D_n. \qquad (3.5^*)$$

It was shown that this equation has an error of about 1 percent in the neighbourhood of $x_n = 2$, and is much more accurate for smaller values. It may be used to solve such problems as the following: "How many generations of selection are needed to reduce the proportion of double dominants to 1 percent?" In F_3, the first generation for which the equation (3.5) holds, we substitute $n = 3$, $D_n = 2/9$, and find $C^1 = .833$. Substituting this value, and $D_n = .01$, we find $n = 12.005$. Hence the value is very nearly reached in F_{12}.

ELIMINATION OF A MULTIPLE RECESSIVE, RANDOM MATING

Under any type of mating, selection of a multiple recessive is of course complete in one generation. To derive the equations for the case where the single genotype recessive for all k genes is eliminated we have only to note that the proportion of multiple recessives $Z(0, k)$ of equation (2.2) is $R_n = g_0^2$ and the relation between gametes of successive generations is

$$\left.\begin{aligned} (1 - R_n) g_m^1 &= 2^{-k} \sum_{i=0}^{m} 2^i \, {}_mC_i \sum_{j=0}^{k-m} 2^j \, {}_{k-m}C_j \, Z(i, j), \text{ when } m \neq 0, \text{ and} \\ (1 - R_n) g_0^1 &= 2^{-k} \sum_{j=0}^{k-1} 2^j \, {}_kC_j \, Z(0, j). \end{aligned}\right\} \qquad (4.1^*)$$

ZIRKLE has tabulated values of R_n for $k = 2$, 3, and 4; and higher values of k are not known to occur. If $k = 2$, and $p_n = g_2$, $q_n = g_1$, $r_n = g_0$ we have, as he points out

$$p_{n+1} = \frac{p_n + \frac{1}{2}(q_n^2 - p_n r_n)}{1 - r_n^2}$$

$$q_{n+1} = \frac{q_n - \frac{1}{2}(q_n^2 - p_n r_n)}{1 - r_n^2}$$

$$r_{n+1} = \frac{r_n - r_n^2 + \frac{1}{2}(q_n^2 - p_n r_n)}{1 - r_n^2}.$$

Putting $q_n+r_n=z_n$, we find

$$z_{n+1}=\frac{z_n-r_n^2}{1-r_n^2}$$

$$r_{n+1}=\frac{\frac{1}{2}(r_n+z_n^2)-r_n^2}{1-r_n^2} \tag{4.2*}$$

$$R_{n+1}=r_n^2.$$

Whence calculation is easy, since $z_2=\frac{1}{2}$, $r_2=\frac{1}{4}$. For large values of n, r_n approximates to z_n^2, and Δz_n to $-z_n^4$, whence $n=\text{constant}+\frac{1}{3}z^{-3}$, approximately, and R_n approximates to $(3n+c)^{-4/3}$.

SELECTION OF DOMINANTS IN AN AUTOTETRAPLOID, RANDOM MATING

This case is included here owing to its close similarity to the last. It is assumed that double reduction (HALDANE 1930) does not occur. Let the gametes forming F_n be in the ratios $p_nAA:2q_nAa:r_naa$, where $p_n+2q_n+r_n=1$. The recessives, which are eliminated, occur in F_n in the proportion $R_n=r_n^2$, and:—

$$p_{n+1}=\frac{p_n+\frac{2}{3}(q_n^2-p_nr_n)}{1-r_n^2}$$

$$q_{n+1}=\frac{q_n-\frac{2}{3}(q_n^2-p_nr_n)}{1-r_n^2}$$

$$r_{n+1}=\frac{r_n-r_n^2+\frac{2}{3}(q_n^2-p_nr_n)}{1-r_n^2}.$$

Putting $z_n=q_n+r_n$, we have

$$z_{n+1}=\frac{z_n-r_n^2}{1-r_n^2}$$

$$r_{n+1}=\frac{\frac{1}{3}(r_n+2z_n^2)-r_n^2}{1-r_n^2} \tag{5.1*}$$

$$R_n=r_n^2.$$

As above, when n is large, R_n approximates to $(3n+c)^{-4/3}$. Starting with an F_1 between homozygotes, we find the following percentage values of R_n in successive generations:—0, 2.7, 4.0, 3.671, 6.432, 2.623, 2.003, 1.644. . . . It will be seen that the values oscillate at first, as they do when there is no selection, reaching a maximum in F_5.

SELECTION OF MULTIPLE DOMINANTS: SELF-FERTILIZATION

A self-fertilized population consists of lines of one individual per generation. Selection does not alter the character of these lines, but only their relative proportions. If there is no selection F_{n-1} consists of zygotes in the ratios $(2^{n-1}-2)\,AA:4\,Aa:(2^{n-1}-2)aa$, and similarly for the other genes. Now (if we confine our attention to A and a) all the aa zygotes are eliminated. But as all their descendants in later generations are also aa, the composition of F_n is unaltered if we postpone our selection till F_{n-1}. The parents of F_n, after selection, are thus in the ratio

$$(2^{n-1}-2)AA:4Aa,$$

and F_n consists of $(2^{n-1}-1)AA:2Aa:1aa$ or $\frac{2^{n-1}+1}{2^{n-1}+2}$ AA and Aa. Hence the proportion of multiple dominants in F_n is

$$D_n \doteq \left(1-\frac{1}{2^{n-1}+2}\right)^k. \tag{5.1}$$

Comparing this with $(1-n^{-2})^k$ of equation (1.2) we see that from F_3 to F_6 inclusive the multiple dominants are fewer than in the case of random mating, but from F_7 onwards more numerous. This is because the rapid elimination of heterozygotes causes a large number of recessives to appear in the first few generations. The final population consists of homozygous multiple dominants.

ELIMINATION OF MULTIPLE DOMINANTS; SELF-FERTILIZATION

This case is extremely simple. Elimination is complete in one generation. For clearly no multiple dominant can appear save in the progeny of a zygote carrying all the requisite genes, that is, itself a multiple dominant. But the end result is quite different from that in the case of random mating. In the latter case all dominant genes are ultimately eliminated in a large population, and most in a small one. With self-fertilization no selective elimination occurs after F_2, and the average number of genes for which a homozygote in the final population is dominant is $\frac{2k(4^{k-1}-3^{k-1})}{4^k-3^k}$, or $\frac{k}{2}\left(1-\frac{3^{k-1}}{4^k-3^k}\right)$. The multiple dominant phenotype will thus reappear if crossbreeding commences in this heterogeneous population.

ELIMINATION OF MULTIPLE RECESSIVES; SELF-FERTILIZATION

This case is of some practical importance in allopolyploid plants such as wheat. As in the last case but one we can imagine selection postponed either until F_{n-1} or F_n without altering the final result. The complete P_{n-1} would have consisted of

$$(\tfrac{1}{2}-2^{1-n})AA,\ 2^{2-n}Aa,\ (\tfrac{1}{2}-2^{1-n})aa,$$

and so on for the other $k-1$ genes, the total multiple recessives removed up to F_{n-1} being therefore $(\tfrac{1}{2}-2^{1-nk})$. The additional multiple recessives removed from F_n would therefore be $(\tfrac{1}{2}-2^{-n})^k-(\tfrac{1}{2}-2^{1-n})^k$ taken from a population of $1-(\tfrac{1}{2}-2^{1-n})^k$ of the original total. Thus the proportion of recessives appearing in F_n is

$$\begin{aligned} R_n &= \frac{(\tfrac{1}{2}-2^{-n})^k-(\tfrac{1}{2}-2^{1-n})^k}{1-(\tfrac{1}{2}-2^{1-n})^k} \\ &= \frac{(2^{n-1}-1)^k-(2^{n-1}-2)^k}{2^{nk}-(2^{n-1}-2)^k}. \end{aligned} \tag{6.1}$$

For large values of n this approximates to $2^{1-n-k}k$. Table 1 gives the value of $100\ R_n$, the percentage of multiple recessives, for F_2 to F_{10} for $k=2$ and 3, the practically important cases.

TABLE 1

Percentages of multiple recessives in F_n, recessives eliminated by selfing.

n	2	3	4	5	6	7	8	9	10
$k=2,\ 100\,R_n$	6.25	$8.\dot{3}$	$5.\dot{9}\dot{0}$	3.502	1.909	0.9968	0.5095	0.2435	0.1295
$k=3,\ 100\,R_n$	1.5625	3.770	3.273	2.102	1.187	0.6374	0.3112	0.1649	0.08308

As in the last case, selection does not lead to a homogeneous population. It ceases in any line as soon as any single recessive gene is eliminated. The average number of dominant genes in a homozygote of the final population is $\dfrac{k}{2-2^{1-k}}$, instead of k in the case of random mating. But of course the elimination of the multiple recessives is greatly speeded up by selfing.

SELECTION OF DOMINANTS IN AN AUTOTETRAPLOID: SELF-FERTILIZATION

HALDANE (1930) has shown that in a self-fertilized autotetraploid F_n, where F_1 is a hybrid between two homozygotes, the proportion of recessives is

$$r_n=\frac{1}{2}\left[1-\frac{7}{5}\left(\frac{5}{6}\right)^n+\left(\frac{1}{6}\right)^n\right].$$

Hence, as above, the proportion appearing in F_n when recessives are eliminated in each generation is

$$R_n = \frac{r_n - r_{n-1}}{1 - r_{n-1}}$$

Hence

$$R_n = \frac{5(7.5^{n-3}-1)}{6(6^{n-1}+7.5^{n-3}-1)} \tag{7.1}$$

and when n is very large it approximates to $\frac{7.5^{n-2}}{6^n}$ or $\frac{1}{6}$ of the total proportion of heterozygotes in an unselected F_{n-1}.

SELECTION OF A SINGLE DOMINANT, BROTHER-SISTER MATING

Before we can proceed to consider the selection of a multiple dominant, this problem, which so far as I know has never been fully treated, must be solved. Since *aa* zygotes are not allowed to breed, there are only three types of mating. Let these occur among the parents of F_n in the proportions $x_n AA \times AA$, $y_n AA \times Aa$ (and reciprocally) and $z_n Aa \times Aa$, where $x_n + y_n + z_n = 1$. Then the proportion of dominants in F_n is $d_n = 1 - \frac{1}{4}z_n$. Only 3/4 of the offspring of the $Aa \times Aa$ matings are allowed to breed. Hence the contribution of these matings to the next generation is

$$\frac{3}{4}\left(\frac{1}{9}AA \times AA + \frac{4}{9}AA \times Aa + \frac{4}{9}Aa \times Aa\right)$$

or

$$\frac{1}{12}AA \times AA,\ \frac{1}{3}AA \times Aa,\ \frac{1}{3}Aa \times Aa.$$

Hence:—

$$\begin{aligned} x_{n+1} &= \frac{x_n + \frac{1}{4}y_n + \frac{1}{12}z_n}{d_n} \\ y_{n+1} &= \frac{\frac{1}{2}y_n + \frac{1}{3}z_n}{d_n} \\ z_{n+1} &= \frac{\frac{1}{4}y_n + \frac{1}{3}z_n}{d_n}\,. \end{aligned} \tag{8.1*}$$

Putting $z_n = 4 - 4d_n$ we find

$$y_{n+1} = \frac{\frac{1}{2}y_n + \frac{4}{3} - \frac{4}{3}d_n}{d_n}$$

and

$$4-4d_{n+1}=\frac{\frac{1}{4}y_n+\frac{4}{3}-\frac{4}{3}d_n}{d_n}$$

Hence $$\frac{3}{4}y_{n+1}=\frac{\frac{3}{8}y_n+1}{d_n}-1$$

and $$\tfrac{3}{4}y_n=16d_n-12d_{n+1}d_n-4$$

so that $$12d_{n+2}d_{n+1}d_n-22d_{n+1}d_n+11d_n-1=0 \qquad (8.2^*)$$

If $d_n=\frac{p_{n+1}}{p_n}$, we find

$$12p_{n+3}-22p_{n+2}+11p_{n+1}-p_n=0$$

p_n is therefore the sum of the nth terms of three geometric series whose common ratios are the roots of $12\lambda^3-22\lambda^2+11\lambda-1=0$, or 1 and $\frac{1}{5\pm\sqrt{13}}$.

Since $z_2=1$, we find

$$p_{n+2}=2+\frac{1+\frac{1}{\sqrt{13}}}{(5-\sqrt{13})^n}+\frac{1-\frac{1}{\sqrt{13}}}{(5+\sqrt{13})^n} \qquad (8.3)$$

If r_n and h_n are the proportions of recessives and heterozygotes respectively in F_n, then $r_n=1-d_n$, so from equation (8.2)

$$r_{n+2}=\frac{1+10r_{n+1}-\frac{1}{1-r_n}}{12(1-r_{n+1})} \qquad (8.4^*)$$

from which r_n is easily calculated, since $r_2=\frac{1}{4}$, $r_3=\frac{1}{9}$. Since $r_n=\frac{p_n-p_{n+1}}{p_n}$, then from equation (8.3),

$$r_{n+3}=\frac{\left(2-\frac{7}{\sqrt{13}}\right)(5-\sqrt{13})^{-n}+\left(2+\frac{7}{\sqrt{13}}\right)(5+\sqrt{13})^{-n}}{2+\left(1+\frac{1}{\sqrt{13}}\right)(5-\sqrt{13})^{-2}+\left(1-\frac{1}{\sqrt{13}}\right)(5+\sqrt{13})^{-n}}\,. \qquad (8.5)$$

Hence the values of r_n ultimately approximate to $\left(1-\frac{7}{2\sqrt{13}}\right)(5-\sqrt{13})^{3-n}$, a geometric series whose common ratio is .717, thus diminishing more rapidly than in the case of random mating.

$$\begin{aligned} h_n &= \tfrac{1}{2}(y_n+z_n) \\ &= \tfrac{2}{3}(13d_n-12d_{n+1}d_n-1) \qquad (8.6^*) \\ 8 &= r_{n+1}(1-r_n)-\tfrac{2}{3}r_n \end{aligned}$$

whence it is readily calculated. When n is large the value of $\frac{h_n}{r_n}$ approximates to $\frac{8+2\sqrt{13}}{3}$, or 5.07. Hence h_n approximates to a geometric series whose common ratio is .717, and it diminishes more rapidly than in the case of brother-sister mating without selection, where the common ratio is $\frac{1}{\sqrt{5}-1}$, or .809. Table 2 gives the values of r_n and h_n for the first 12 generations, calculated from equations (8.4) and (8.6). They are compared with the corresponding values in the case of random mating. It will be seen that after F_3 the proportion of homozygous dominants is increased by inbreeding, that of heterozygotes diminished. However the proportion of recessives between F_4 and F_{11} inclusive is higher when inbreeding is practiced than when mating is at random. It may be remarked that there is no advantage in beginning brother-sister mating before F_3, since the F_2 from two pure lines are all as closely related to one another genetically as if they were sibs.

In practice the elimination of recessive genes could of course be increased still further either by a refusal to breed from individuals with recessive sibs, or better, by test matings with recessives. But the consideration of this somewhat artificial case is necessary if we are to solve the next problem.

TABLE 2

n	RANDOM MATING		BROTHER-SISTER MATING	
	RECESSIVES	HETEROZYGOTES	RECESSIVES (r_n)	HETEROZYGOTES (h_n)
1	0	1.00	0	1.00
2	.25	.50	.25	.50
3	$.\dot{1}$	$.\dot{4}$	$.\dot{1}$	$.\dot{4}$
4	.0625	.375	$.07291\dot{6}$	$.3541\dot{6}$
5	.0400	.3200	.054307	.27341
6	$.0\dot{3}6$	$.2\dot{7}$	.040924	$.20\dot{7}2\dot{6}$
7	.020408	.2449	.030569	.15497
8	.015625	.21875	.022609	.11464
9	.012346	.19753	.016555	.084109
10	.0100	.1800	.012097	.061335
11	.008265	.165289	.0087854	.045115
12	$.0069\dot{4}$	$.152\dot{7}$	.0063392	.032127

SELECTION OF MULTIPLE DOMINANTS, BROTHER-SISTER MATING

The situation is very similar to that in the case of random mating. All the zygotes recessive for any one of the k genes are eliminated at once in each generation. The process of the last paragraph thus takes place for each gene separately. And these processes are independent. So to find the proportion of multiple dominants in F_n we have only to put

$$D_n = (1 - r_n)^k \tag{9.1*}$$

where r_n is given by equation (8.5) or table 2. Thus in F_7 of an inbred population a fraction .96943 carry any given one of the dominant genes, and $.96943^k$ carry all of them. A fraction $1 - h_n - r_n$ or .81446 of F_7 is homozygous for any one dominant gene, a fraction $(1 - h_n - r_n)^k$ or $.81446^k$ for all of them. Table 3 gives the results for $k = 10$, with random mating results for comparison. It will be seen that up to F_{12} inbreeding slightly slows down the actual appearance of dominants, but greatly increases their genetic purity. Thus in F_{10} 53 percent of the dominants are homozygous in the case of inbreeding, and only 13 percent in that of random mating.

Again there is no need to practice brother-sister mating before F_3, and it would be practicable to begin it in F_3 even if as many as 10 genes were concerned. The process of selection could of course be speeded up if families containing any recessives were rejected, which would be practicable after about F_5.

TABLE 3

Composition of F_n when only multiple dominants for 10 genes are bred from.

n	RANDOM MATING		BROTHER-SISTER MATING	
	TOTAL 10-ple DOMINANTS	HOMOZYGOUS 10-ple DOMINANTS	TOTAL 10-ple DOMINANTS	HOMOZYGOUS 10-ple DOMINANTS
1	1.0	0	1.0	0
2	.0563	9.5×10^{-7}	.0563	9.5×10^{-7}
3	.3080	.00030	.3080	.00030
4	.5244	.00317	.4690	.00381
5	.6648	.01152	.5722	.01886
6	.7720	.02578	.6585	.05769
7	.8137	.04582	.7331	.1284
8	.8543	.06921	.7956	.2285
9	.8832	.09483	.8453	.3460
10	.9043	.1215	.8854	.4665
11	.9203	.1486	.9155	.5746
12	.9327	.1755	.9385	.6754

ELIMINATION OF MULTIPLE DOMINANTS: BROTHER-SISTER MATING

The population in F_2 and later consists of pairs of mated zygotes, apart from the multiple dominants, which are eliminated. Thus mating pairs fall into four classes.

1. Pairs giving only multiple dominants.
2. Pairs giving multiple dominants and other genotypes among their immediate offspring, but no multiple dominants later.
3. Pairs giving multiple dominants and other genotypes both among their immediate progeny and in later generations.
4. Pairs never giving multiple dominants.

Clearly it is only class 3 which contribute multiple dominants to remote generations.

A consideration of the case when $k=2$ will illustrate the principles involved. Nine-sixteenths of F_2 are double dominants, so $D_2=9/16$. The five genotypes which are the parents of F_3 occur in the proportions:

$$\frac{1}{7}AAbb,\ \frac{1}{7}aaBB,\ \frac{2}{7}Aabb,\ \frac{2}{7}aaBb,\ \frac{1}{7}aabb.$$

The matings occur with the following frequencies:—

Class 1. 2/49 $AAbb \times aaBB$

Class 2. 4/49 $AAbb \times aaBb$, 4/49 $aaBB \times Aabb$

Class 3. 8/49 $Aabb \times aaBb$

Class 4. 31/49 other matings, for example, 1/49 $AAbb \times AAbb$, 4/49 $Aabb \times aabb$.

That is to say 2/49 give all double dominants, 8/49 give 50 percent, and 8/49 give 25 percent. So $D_3=8/49$, and 41/49 of F_3 is available for mating. Only class 3 matings give further double dominants. A mating of class 3 gives 25 percent double dominants, and matings which can be symbolized by

$$\frac{3}{4}\cdot\frac{1}{9}(Aabb+aaBb+aabb)^2.$$

If there is a proportion p_n of such matings among the parents of F_n, then $D_n=\frac{1}{4}p_n$,
and

$$p_{n+1}=\frac{\frac{1}{6}p_n}{1-\frac{1}{4}p_n}.$$

If we put

$$\frac{x_{n+1}}{x_n}=1-\frac{1}{4}p_n$$

we find

$$6x_{n+2}-7x_{n+1}+x_n=0,$$

so that

$$x_n=a+6^{-n}$$

Since 6/41 of the matings of the parents of F_4 are derived from families of the type (*Aabb*+*aaBb*+*aabb*),

$$p_4=\frac{2}{9}\cdot\frac{6}{41}=\frac{4}{123}, \text{ and } \frac{x_5}{x_4}=\frac{122}{123}.$$

Hence $a=609\times 6^{-5}$ and

$$\frac{x_{n+1}}{x_n}=\frac{1+609\times 6^{n-4}}{6+609\times 6^{n-4}}.$$

So

$$D_n=1-\frac{x_{n+1}}{x_n}=\frac{5}{6+609\times 6^{n-5}} \tag{10.1}$$

from F_4 onwards. Hence the successive values of D_n, the proportion of double dominants in successive generations from F_2 onwards are $\frac{9}{16}$, $\frac{8}{49}$, $\frac{1}{123}$, $\frac{1}{732}$, $\frac{1}{3186}$ and so on. The corresponding values for a random mating population, derived from equation (3.3), are $\frac{9}{16}$, $\frac{8}{49}$, $\frac{200}{1681}$, $\frac{192{,}200}{2{,}193{,}361}$, .0658, et cetera which diminish far more slowly.

The final population can readily be calculated. Such a mating as *Aabb* ×*aabb* ultimately gives a population of 1/4 *AAbb*, 3/4 *aabb* and so on. The mating *Aabb*×*aaBb* gives 1/6 similar matings and 7/12 matings ultimately leading to

$$\frac{1}{7}AAbb,\ \frac{1}{7}aaBB,\ \frac{5}{7}aabb$$

So the final population is

$$\frac{1}{12}\left(1+\frac{1}{6}+\frac{1}{6^2}+\cdots\right)(AAbb+aaBB+5aabb)$$

or 1/10 (*AAbb*+*aaBB*+5*aabb*). The ultimate population from all the F_2 mating types is therefore in the ratios 54 *AAbb*:54 *aaBB*:95 *aabb*.

In general the proportion of class 4 matings, which never give any multiple dominants, among the survivors of F_2 can easily be shown to be $\frac{16^k-15^k}{(4^k-3^k)^2}$ a quantity which rises from 63.3 percent when $k=2$ to 86.6 percent when $k=10$. In all cases a majority of the matings of F_2 never give multiple dominants, and thus if a few lines are started, most of them will give the desired phenotype only, though they will give different genotypes. There is thus no practical value in working out expressions for D_n for different values of k.

ELIMINATION OF MULTIPLE RECESSIVES; BROTHER-SISTER MATING

The principles to be employed are illustrated when $k=2$. The mating population consists of 8 genotypes other than *aabb*. Let the matings among the survivors of F_n occur in the following proportions:

c	*AABB* × *AABB*	g	*AABb* × *Aabb**
	AABB × *AABb**		*AaBb* × *AAbb**
	AABB × *AAbb**	h	*AABb* × *aaBb**
	AABb × *AABb*	i	*AAbb* × *aaBB*
	AABb × *AAbb**	j	*AAbb* × *Aabb**
	AAbb × *AAbb**	k	*AAbb* × *aaBb**
d	*AABB* × *AaBb*	l	*AaBb* × *AaBb*
	AABb × *AaBB*	m	*AaBb* × *Aabb**
e	*AABB* × *Aabb**	p	*Aabb* × *Aabb**
	AABb × *aaBB**	q	*Aabb* × *aaBb*
f	*AABb* × *AaBb**		

Here reciprocals are included together. The symbol c denotes c_n, c' denotes c_{n+1}, and so on. An asterisk denotes that there is another similar mating obtained by substituting *A* for *B*. The various matings whose total frequency is c never produce double recessives. The number of these,

$$R' = \frac{1}{16}l_n + \frac{1}{8}m_n + \frac{1}{4}p_n + \frac{1}{4}q_n. \tag{10.1*}$$

The remaining mating types (d to k) give no double recessives in the first generation, but do so later. Also $l_1=1$. Hence

$$\begin{aligned}
(1-R')c' &= c_n + \frac{7}{16}d_n + \frac{1}{4}e_n + \frac{19}{64}f_n + \frac{1}{4}g_n + \frac{1}{16}h_n + \frac{1}{4}j_n + \frac{31}{240}l_n + \frac{1}{14}m_n + \frac{1}{12}p_n\\
(1-R')d' &= \frac{1}{4}d_n + \frac{1}{8}f_n + \frac{1}{15}l_n\\
(1-R')e' &= \frac{1}{16}f_n + \frac{1}{15}l_n\\
(1-R')f' &= \frac{1}{4}d_n + \frac{1}{2}2_n + \frac{3}{16}f_n + \frac{1}{8}g_n + \frac{1}{4}h_n + \frac{2}{15}l_n + \frac{1}{14}m_n\\
(1-R')g' &= \frac{1}{8}f_n + \frac{1}{4}g_n + \frac{2}{15}l_n + \frac{1}{7}m_n\\
(1-R')h' &= \frac{1}{32}f_n + \frac{1}{8}h_n + \frac{1}{15}l_n + \frac{1}{28}m_n\\
(1-R')i' &= \frac{1}{120}l_n
\end{aligned}$$

$$\begin{aligned}
(1-R')j' &= \tfrac{1}{8}f_n+\tfrac{1}{8}g_n+\tfrac{1}{2}j_n+\tfrac{1}{30}l_n+\tfrac{1}{14}m_n+\tfrac{1}{3}p_n\\
(1-R')k' &= \tfrac{1}{30}l_n+\tfrac{1}{28}m_n\\
(1-R')l' &= \tfrac{1}{12}d_n+\tfrac{1}{4}e_n+\tfrac{1}{16}f_n+\tfrac{1}{16}g_n+\tfrac{1}{4}h_n+i_n+\tfrac{1}{4}k_n+\tfrac{1}{15}l_n+\tfrac{1}{14}m_n+\tfrac{1}{12}q_n\\
(1-R')m' &= \tfrac{1}{16}f_n+\tfrac{1}{8}g_n+\tfrac{1}{4}h_n+\tfrac{1}{2}k_n+\tfrac{2}{15}l_n+\tfrac{3}{14}m_n+\tfrac{1}{3}q_n\\
(1-R')p' &= \tfrac{1}{64}f_n+\tfrac{1}{16}g_n+\tfrac{1}{16}h_n+\tfrac{1}{4}j_n+\tfrac{1}{4}k_n+\tfrac{1}{30}l_n+\tfrac{5}{56}m_n+\tfrac{1}{3}p_n+\tfrac{1}{6}q_n\\
(1-R')q' &= \tfrac{1}{30}l_n+\tfrac{1}{14}m_n+\tfrac{1}{6}q_n
\end{aligned}\tag{10.2*}$$

Hence

$$R_2=\frac{1}{16},\quad R_3=\frac{1}{25},\quad R_4=\frac{4649}{103680}=4.48 \text{ percent}$$

whereas with random mating (from equation 4.2)

$$R_2=\frac{1}{16},\quad R_3=\frac{1}{25},\quad R_4=\frac{361}{11664}=3.10 \text{ percent}.$$

Thus at first somewhat more double recessives appear as the result of inbreeding. However it is clear that in the case of inbreeding R_n ultimately approximates to a geometrical series, and consequently diminishes far more rapidly than in the case of random mating. For practical purposes the value of c_n is even more important than that of $R_n \cdot c_2=\frac{31}{225}=13.8$ percent, while $c_3=\frac{14319}{45360}=31.6$ percent. That is to say 31.6 percent of all F_3 mated pairs will give no double recessives. Hence even two generations of brother-sister mating will have eliminated the possibility of producing double recessives from many lines.

For values of k exceeding 2 the equations become quite excessively complex. However c_2, the probability that a given breeding pair in F_2 (after eliminating the multiple recessives) will never, in any later generation, yield multiple recessives, is

$$c_2=\frac{16^k-15^k}{(4^k-1)^2}\,. \tag{10.3}$$

The values of c_2 as percentages from $k=1$ to 5 inclusive are $11.\dot{1}$, $13.\dot{7}$, 18.17, 22.93, 27.63. Thus if $k=5$ rather over 1/4 of all the matings of surviving F_2 would give no recessives. c_2 increases rather slowly, only reaching 47.55 percent when $k=10$, and 72.49 percent when $k=20$.

Just as in the case where multiple dominants are eliminated, the final population is not genetically homogeneous in the case of brother-sister mating. If mating were re-started between different lines, multiple recessives might appear in the second, though not in the first, generation.

The rather delicate problem of brother-sister mating in an autopolyploid is reserved for a future publication.

SUMMARY

Expressions are found for the effects of selection on populations, both random mating, selfed and inbred, where the character selected depends upon several genes, and (as in the case of crosses between pure lines) each gene pair is present in the same ratio.

LITERATURE CITED

BENNETT, A. A., 1924 Problems and solutions No. 3005. Amer. Math. Monthly **31:** 255–257.

CLOUDMAN, A. M., 1932 A comparative study of transplantability of eight mammary gland tumors arising in inbred mice. Amer. J. Cancer **16:** 568–630.

HALDANE, J. B. S., 1926 A mathematical theory of natural and artificial selection, Part III. Proc. Cambridge Phil. Soc. **23:** 363–372.

1930 Theoretical genetics of autopolyploids. J. Genet. **22:** 359–372.

1932 A mathematical theory of natural and artificial selection, Part IX. Proc. Cambridge Phil. Soc. **28:** 244–248.

ZIRKLE, C., 1926 Some numerical results of selection. Genetics **11:** 531–583.

A mathematical theory of natural and artificial selection. Part IX. Rapid selection. By Mr J. B. S. HALDANE, Trinity College.

[*Received* 2 February, *read* 7 March, 1932.]

In Part I of this series (1) it was proved that in a random mating population in which the ratio between the numbers of two genes in the nth generation was u_n, and the proportion of recessives therefore $(u_n+1)^{-2}$, the coefficient of selection being k, then

$$u_{n+1}=\frac{u_n(u_n+1)}{u_n+1-k}. \tag{1}$$

Hence, if $x_n=1+1/u_n$, we have

$$x_{n+1}=x_n-\frac{k(x_n-1)^2}{x_n}.$$

It has since been shown (2) that, if y is a given function of x, $y_r=\left(\frac{d}{dx}\right)^r y$, and $x_{n+1}=x_n+ky$, then

$$n=\int_{x_0}^{x_n}\Big[k^{-1}y^{-1}+\tfrac{1}{2}y^{-1}y_1-\frac{k}{12}(y^{-1}y_1^2+y_2)+\frac{k^2}{24}(y^{-1}y_1^3+2y_1y_2)$$
$$-\frac{k^3}{720}(19y^{-1}y_1^4+59y_1^2y_2+yy_2^2-2yy_1y_3-y^2y_4)+\dots\Big]\,dx,$$

provided that y is regular and does not vanish in the interval considered, and that the series converges uniformly. It was also pointed out that the coefficient of $y^{-1}y_1^r$ is the $(r+1)$th term in the expansion of $\{\log(1+k)\}^{-1}$.

In our case $y=-(x-1)^2/x$, hence

$$n=\int_{x_0}^{x_n}\Big[\frac{-x}{k(x-1)^2}+\frac{x+1}{2x(x-1)}+\frac{k(x^2+2x+3)}{12x^3}$$
$$+\frac{k^2(x^2-1)(x^2+2x+5)}{24x^5}$$
$$+\frac{k^3(x-1)^2(19x^4+76x^3+220x^2+360x+105)}{720x^7}+\dots\Big]\,dx$$
$$=C+\frac{1}{k}\left[\frac{1}{x_n-1}-\log(x_n-1)\right]+\tfrac{1}{2}\Big[2\log(x_n-1)-\log x_n\Big]$$
$$+\frac{k}{12}(\log x_n-2x_n^{-1}-\tfrac{3}{2}x_n^{-2})$$
$$+\frac{k^2}{24}(\log x_n-2x_n^{-1}-2x_n^{-2}+\tfrac{2}{3}x_n^{-3}+\tfrac{5}{4}x_n^{-4})$$
$$+\frac{k^3}{720}(19\log x_n-38x_n^{-1}-\tfrac{87}{2}x_n^{-2}+\tfrac{4}{3}x_n^{-3}+\tfrac{395}{4}x_n^{-4}$$
$$-30x_n^{-5}-\tfrac{35}{2}x_n^{-6})+\dots, \tag{2}$$

where C does not vary with n. Now since $x_n > 1$, the error due to neglecting all terms except the logarithmic terms and $\frac{1}{k(x_n-1)}$ is very small when x_n and x_0 are large, and is always less than

$$\frac{7k}{24}+\frac{25k^2}{288}-\frac{347k^3}{8640}+\dots.$$

Although I have been unable to prove that this series converges, it will be seen that its sum can be neglected without serious error, even when $|k|>1$. The remaining terms, when $1>k\geqslant -1$, give

$$n=C+\frac{1}{k(x_n-1)}+\frac{\log x_n}{\log(1-k)}+\frac{1-k}{k}\log\frac{x_n}{x_n-1}, \text{ approximately.}$$

This summation is only justified in theory if $1>k\geqslant -1$, but it will be seen later that it holds even when $k<-1$. Since $x_n=1+1/u_n$, we can write the above equation as

$$n=\frac{u_n-u_0}{k}+\log\left(\frac{1+1/u_n}{1+1/u_0}\right)\Big/\{\log(1-k)\}+\frac{1-k}{k}\log\left(\frac{1+u_n}{1+u_0}\right). \quad (3)$$

Though series (2) is doubtless more accurate, equation (3) is quite satisfactory for ordinary purposes. k cannot exceed 1, and the equation is exact when $k=1$, or is infinitesimal, and approximate for intermediate values. When k is negative it is also approximate. It can also be shown to attain any desired degree of accuracy if both u_n and u_0 are sufficiently large or small. Its accuracy can readily be tested by substituting values of u_0 and u_1 calculated from equation (1) for any values of u_0 and k, and then calculating n, which should equal unity were equation (3) exact. If $k=\frac{1}{2}$,

$u_0=1,\quad u_1=\frac{4}{3},\quad n=1{\cdot}0135,$

$u_0=100,\quad u_1=100{\cdot}4975,\quad n=1{\cdot}0011,$

$u_0={\cdot}01,\quad u_1={\cdot}019804,\quad n=1{\cdot}0032.$

Hence in this case the error probably does not exceed $2\,^{\circ}\!/_{\circ}$. If $k=0{\cdot}9$,

$u_0=1,\quad u_1=1{\cdot}\dot{8}\dot{1},\quad n=0{\cdot}9670,$

$u_0=100,\quad u_1=100{\cdot}8991,\quad n=1{\cdot}0068,$

$u_0={\cdot}01,\quad u_1={\cdot}09\dot{1}\dot{8},\quad n=1{\cdot}0277.$

The error is thus under $4\,^{\circ}\!/_{\circ}$. Finally, if $k=-4$,

$u_0=1,\quad u_1={\cdot}\dot{3},\quad n=1{\cdot}1042,$

$u_0=100,\quad u_1=96{\cdot}1905,\quad n=0{\cdot}9606,$

$u_0={\cdot}01,\quad u_1={\cdot}002016,\quad n=1{\cdot}0361.$

The error here exceeds 10 °/₀ over a certain range of values of u. But it is to be noted that these values of k refer to extremely intense selection. Thus, when $k = -4$, five recessives survive for every dominant, and when $k = 0{\cdot}9$, ten dominants for every recessive. Such intense selection can hardly ever, in the course of evolution, have been the direct cause of large changes in a population.

Elton (3) has expressed the opinion that occasional intense selection, for example during periodic famines and plagues, may be more efficient than less intense selection acting in every generation. This view can be examined quantitatively. Consider two populations, in the first of which selection of intensity k occurs in every generation, whilst in the second selection occurs with intensity mk in every mth generation, where $|mk| < 1$. Then from equation (2) it follows that in the first population the number of generations needed to change u_0 to u_n is approximately

$$n = \frac{1}{k}(u_n - u_0 + \log u_n - \log u_0) - \tfrac{1}{2}[\log(u_n^2 + u_n) - \log(u_0^2 + u_0)],$$

and in the second population, approximately,

$$n' = \frac{1}{k}(u_n - u_0 + \log u_n - \log u_0) - \frac{m}{2}[\log(u_n^2 + u_n) - \log(u_0^2 + u_0)],$$

assuming that squares of km can be neglected. Since $m > 1$, the second time is shorter than the first if $u_n(u_n + 1) > u_0(u_0 + 1)$, regardless of the sign of k, i.e. if $u_n > u_0$. Thus when selection is favouring dominants, it is more efficient if concentrated in a series of cataclysms, but when it is favouring recessives the opposite is the case. But unless $|mk|$ is fairly large the difference is unimportant.

We can also compare the time taken, with selection of the same intensity, to change u from a to b, when dominants are favoured (k positive), with the time taken for the change from b to a when recessives are favoured (k negative). From equation (3), putting $u_0 = a$, $u_n = b$, the time needed for the first change is

$$n = \frac{b-a}{k} + \log\left(\frac{1 + 1/b}{1 + 1/a}\right)\Big/\{\log(1-k)\} + \frac{1-k}{k}\log\left(\frac{1+b}{1+a}\right).$$

Changing the sign of k, and putting $u_0 = b$, $u_n = a$, we find that the time needed for the second change is

$$n' = \frac{b-a}{k} - \log\left(\frac{1 + 1/b}{1 + 1/a}\right)\Big/\{\log(1+k)\} + \frac{1+k}{k}\log\left(\frac{1+b}{1+a}\right).$$

Hence

$$n' - n = \left[\frac{-\log(1-k) - \log(1+k)}{\log(1-k)\log(1+k)}\right]\log\left(\frac{1 + 1/b}{1 + 1/a}\right) + 2\log\left(\frac{1+b}{1+a}\right).$$

This is positive, since k is positive and $b > a$. Hence on the above convention selection appears to be more effective when dominants are favoured. But it is illegitimate to regard a selection measured by $-k$ as the inverse of one measured by k unless both are very small. Thus when $k = \frac{1}{2}$, two dominants survive for every recessive, but when $k = -\frac{1}{2}$, only one-and-a-half recessives survive for every dominant. To obtain a fairer comparison we put $1 - k = e^{-\kappa}$, and change the sign of κ when selection is reversed. Equation (3) now becomes

$$n = \frac{u_n - u_0}{1 - e^{-\kappa}} - \frac{1}{\kappa} \log \left(\frac{1 + 1/u_n}{1 + 1/u_0}\right) + \frac{1}{e^{\kappa} - 1} \log \left(\frac{1 + u_n}{1 + u_0}\right). \quad (4)$$

Hence the time taken to change u from a to b when κ is positive is

$$n = \frac{b - a}{1 - e^{-\kappa}} - \frac{1}{\kappa} \log \left(\frac{1 + 1/b}{1 + 1/a}\right) + \frac{1}{e^{\kappa} - 1} \log \left(\frac{1 + b}{1 + a}\right),$$

and the time for the converse change is

$$n' = \frac{b - a}{e^{\kappa} - 1} - \frac{1}{\kappa} \log \left(\frac{1 + 1/b}{1 + 1/a}\right) - \frac{1}{1 - e^{-\kappa}} \log \left(\frac{1 + b}{1 + a}\right),$$

so that
$$n - n' = b - a + \log \left(\frac{1 + b}{1 + a}\right),$$

which is positive. Thus selection is more rapid if recessives are favoured. The difference is however only significant if b is large, that is to say recessives very rare in one of the populations considered. This result was to be expected, since if all dominants are killed off, i.e. $k = \kappa = -\infty$, selection is complete in one generation, whilst if all recessives are killed off, i.e. $k = 1$, $\kappa = \infty$, selection is a relatively slow process. The number κ occurring in equation (4) is the difference of the Malthusian parameters, as defined by Fisher (4), of the dominants and recessives. It approximates to k when both are small.

The problem which is here solved is the simplest, though perhaps the most important, of a large number. When selection acts at different rates in the two sexes, or when it acts on a sex-linked character, or one determined by several genes, or by one gene in a polyploid, we have to solve two or more simultaneous non-linear finite difference equations. When generations are not separate, we have, in general, to solve a set of at least four simultaneous non-linear integral equations. These equations have been stated in other parts of this series, and have been approximately solved when selection is not intense. But their complete solution is desirable for a discussion of problems raised by eugenics and artificial selection.

Summary.

Equations (3) and (4) describe the changes undergone by a Mendelian population mating at random, and under intense selection.

REFERENCES.

HALDANE (1). *Trans. Camb. Phil. Soc.* 23 (1924).

—— (2). *Proc. Camb. Phil. Soc.*, 28 (1932).

ELTON (3). *Ecology and Evolution.*

FISHER (4). *The Genetical Theory of Natural Selection.*