

II. *A Mathematical Theory of Natural and Artificial Selection.*

PART I.

By J. B. S. HALDANE, M.A.

Trinity College, Cambridge.

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:

- (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), and
- (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 (1), Warren (2), 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 (3).

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 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-Mendelian 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 (1.0)—(1.2).

SPECIFICATION OF 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 $k = \cdot 01$, 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, when $k = -\infty$ no A 's reproduce. It will be convenient to refer to these two cases 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 $k = \cdot 01$ 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 (2) 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 $\frac{1}{m}$ 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.

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 (4) have shown that some embryos of any litter perish in utero. Their deaths are certainly sometimes selective. In litters from the mating yellow \times yellow one-quarter of the embryos die in the blastula stage, yet as Durham (5) 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 *Dinophylus* 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.

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(6) has shown that *Oenothera muricata* transmits certain characters by the pollen only, others by the ovules only. Schmidt(7) has found a character in *Lebistes* transmitted by males to males only, and Goldschmidt(8) 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 $(n + 1)$ th, will be $u_n A : (1 - k) B$.

$$\therefore u_{n+1} = \frac{u_n}{1 - k}, \dots\dots\dots(1.0)$$

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

$$y_n = \frac{1}{1 + (1 - k)^{-n}} \dots\dots\dots(1.1)$$

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

$$\left. \begin{aligned} y_n &= \frac{1}{1 + e^{kn}} \\ kn &= \log_e \left(\frac{1 - y_n}{y_n} \right) \end{aligned} \right\} \dots\dots\dots(1.2)$$

or

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 $k = .001$, i.e. 999 B 's survive for every 1000 A 's. 9,184 generations are needed for the proportion of A 's to increase from 1% to 99%. Equation (1.2) gives an error of only 4 in this number.

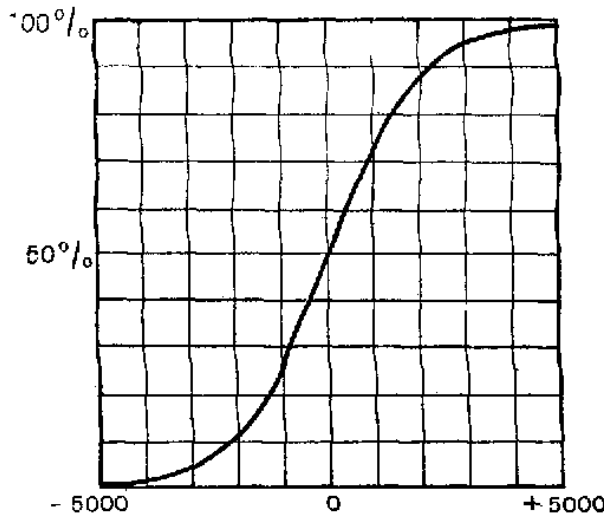


Fig. 1. Effect of selection on a non-amphimictic character. $k = .001$.
 Abscissa = generations.
 Ordinate = percentage of population with the favoured character.

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.

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 (9) and Hardy (10) 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 $(n - 1)$ 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 :

$$y_n = (1 + u_n)^{-2} \dots \dots \dots (2.0)$$

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 :

$$(Nu_n^2 + Nu_n) A, \text{ and } (Nu_n + N(1 - k)) a.$$

So the ratio

$$u_{n+1} = \frac{u_n(1 + u_n)}{1 + u_n - k} \dots \dots \dots (2.1)$$

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,	$2 \times u_n^2 \times 2u_n$,, $4u_n^3$,
$AA \times aa$,, ,,	$2 \times u_n^2 \times (1 - k)$,, $2(1 - k)u_n^2$,
$Aa \times Aa$	$2u_n \times 2u_n$,, $4u_n^2$,
$Aa \times aa$ and reciprocally,	$2 \times 2u_n \times (1 - k)$,, $4(1 - k)u_n$,
$aa \times aa$	$(1 - k) \times (1 - k)$,, $(1 - k)^2$.

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	$2u_n^3 + 2(1 - k)u_n^2 + 2u_n^2 + 2(1 - k)u_n$,, $2u_n(1 + u_n)(1 + u_n - k)$,
aa	$u_n^2 + 2(1 - k)u_n + (1 - k)^2$,, $(1 + u_n - k)^2$.

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

$$u_{n+1} = \frac{u_n(1 + u_n)}{1 + u_n - k} \dots \dots \dots (2.1)$$

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

$$u_0 = y_0^{-\frac{1}{2}} - 1,$$

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 (2.0). Thus if we start with 25% of recessives, and $k = .5$, i.e. the recessives are only half as viable as the dominants, then $u_0 = 1$, and

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

$$y_1 = (1 + \frac{4}{3})^{-2} = \frac{9}{49} = .184, \text{ or } 18.4\% .$$

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 (11) and Hardy have solved the case where the recessives all die. Here $k = 1$, and

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

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

$$\therefore y_n = (n+1+u_0)^{-2}$$

$$= (n+y_0^{-\frac{1}{2}})^{-2}$$

$$= y_0(1+ny_0^{\frac{1}{2}})^{-2}. \dots\dots\dots(2.2)$$

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 $\frac{1}{(n+1)^2}$. 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:

$$u_{n+1} = \frac{u_n(1+u_n)}{1+u_n-k}; \dots\dots\dots(2.1)$$

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

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 $\Delta u_n = \frac{du_n}{dn}$.

$$\therefore \frac{du_n}{dn} = \frac{ku_n}{1+u_n} \text{ approximately;}$$

$$\therefore kn = \int_{u_0}^{u_n} \frac{1+u}{u} du$$

$$= u_n - u_0 + \log_e \left(\frac{u_n}{u_0} \right). \dots\dots\dots(2.3)$$

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

$$kn = u_n + \log_e u_n - 1. \dots\dots\dots(2.4)$$

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

$$kn = u_n + (1-k) \log_e u_n - 1 \dots\dots\dots(2.5)$$

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

$$kn = u_n + \left(1 - \frac{k}{2}\right) \log_e u_n - 1 \dots\dots\dots(2.6)$$

gives results with a very small error. But for every case so far observed equation (2.4) 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. Fig. 2 shows graphically the rate of increase of dominants and recessives respectively when $k = \pm .001$, 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 ten times as intense, the population will clearly change ten times as fast, and so on.

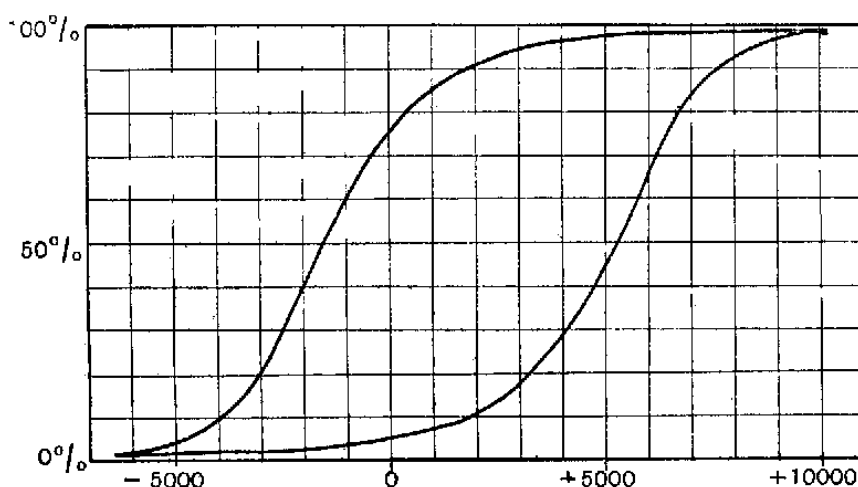


Fig. 2. Effect of selection on an autosomal Mendelian character. $k = .001$.
Upper curve, dominants favoured; lower curve, recessives favoured.
Abscissa = generations. Ordinate = percentage of population with the favoured character.

TABLE I.

Effect of slow selection on an autosomal Mendelian character.

kn (number of generations $\times k$)	-1000	-100	-50	-20	-15	-10			
% of recessives when dominants are favoured	99.9998	99.975			
" " " recessives	" "	" "	.0001	.0105	.0427	.2773	.4215	1.036			
-9	-8	-7	-6	-5	-4.5	-4	-3.5	-3	-2.5	-2	-1.5
99.933	99.82	99.50	98.68	96.50	94.38	91.14	86.36	79.71	71.24	61.53	50.68
1.254	1.545	1.940	2.497	3.308	..	4.537	...	6.528	...	9.718	12.11
-1	-0.5	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5
40.98	32.05	25.0	19.53	15.30	12.11	9.718	...	6.528	...	4.537	...
15.30	19.53	25.0	32.05	40.98	50.68	61.53	71.24	79.71	86.36	91.14	94.38
5	6	7	8	9	10	15	20	50	100	1000	
3.308	2.497	1.940	1.545	1.254	1.036	.4215	.2773	.0427	.0105	.0001	
96.50	98.68	99.50	99.82	99.933	99.975	99.9998	

In Table I the values of y_n calculated from equations (2.4) and (2.0) are given in terms of kn . In Table II 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 (12) 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 $k = .041$. 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 II, when $y = .9$, $kn = -3.863$, $\therefore n = -94.2$. So 94.2 generations of selection will bring the recessives down to 25%. The remaining 5.8 generations give $kn = .238$, and from Table I by interpolation we find $y = .224$, 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%? $y_n = .01$, hence, from Table II, $kn = 10.197$, $\therefore n = 248.7$. 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 (13) 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.

TABLE II.

Effect of slow selection on an autosomal Mendelian character.

% of favoured type0001	.001	.01	.05	.1	.2	.5
kn when dominants are favoured	-15.51	-13.21	-10.90	-9.294	-8.600	-7.905	-6.996		
kn ,, recessives			-1005	-320.0	-102.60	-45.50	-33.04	-23.42	-14.72

1	2	3	5	10	15	20	25	30	35	40
-6.286	-5.580	-5.161	-4.624	-3.863	-3.290	-2.979	-2.712	-2.439	-2.180	-1.964
-10.197	-6.875	-4.976	-3.717	-1.933	-1.041	-.448	0	+ .366	+ .681	+ .962

45	50	55	60	65	70	75	80	85	90	95
-1.708	-1.467	-1.220	-.962	-.681	-.366	0	+ .448	+ 1.041	+ 1.933	+ 3.717
+ 1.220	+ 1.467	+ 1.708	+ 1.964	+ 2.180	+ 2.439	+ 2.712	+ 2.979	+ 3.290	+ 3.863	+ 4.620

97	98	99	99.5	99.8	99.9	99.95	99.99	99.999	99.9999
+ 4.976	+ 6.875	+ 10.197	+ 14.72	+ 23.42	+ 33.04	+ 45.50	+ 102.60	+ 320.0	+ 1005
+ 5.161	+ 5.580	+ 6.286	+ 6.996	+ 7.905	+ 8.600	+ 9.294	+ 10.90	+ 13.21	+ 15.51

Assuming that there were not more than 1% of dominants in Manchester in 1848, nor less than 99% in 1898, we have, from Table II, $kn = 16.58$ as a minimum. But $n = 50$, since this moth usually has one brood per year. $\therefore k = .332$ 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 (2.1) shows that 48 generations are needed for the change if $k = .3$. 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 (14). Another possible explanation would be a large excess of dominants begotten in mixed families, as occurs in human night-blindness according to Bateson (15). 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.

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 $AA \times AA \dots$	p_n^2	0	0
„ „ $AA \times Aa \dots$	$2p_n q_n$	$2p_n q_n$	0
„ „ $AA \times aa \dots$	0	$2p_n r_n$	0
„ „ $Aa \times Aa \dots$	$\frac{4q_n^2}{4 - k}$	$\frac{8q_n^2}{4 - k}$	$\frac{(4 - 4k)q_n^2}{4 - k}$
„ „ $Aa \times aa \dots$	0	$\frac{2q_n r_n}{2 - k}$	$\frac{(2 - 2k)q_n r_n}{2 - k}$
„ „ $aa \times aa \dots$	0	0	r_n^2

$$\left. \begin{aligned} \therefore [AA] = p_{n+1} &= (p_n + q_n)^2 + \frac{kq_n^2}{4 - k} \\ \frac{1}{2}[Aa] = q_{n+1} &= (p_n + q_n)(q_n + r_n) + kq_n \left(\frac{q_n}{4 - k} + \frac{r_n}{2 - k} \right) \\ [aa] = r_{n+1} &= (q_n + r_n)^2 - kq_n \left(\frac{3q_n}{4 - k} + \frac{2r_n}{2 - k} \right) \end{aligned} \right\} \dots\dots\dots(3.0)$$

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 $25^\circ, 56.25^\circ, 66.02^\circ, 84.25^\circ$, etc.

In the more interesting case when k is small we can solve approximately, as follows. From equation (3.0) 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 $u_n = \frac{p_n + q_n}{q_n + r_n}$, then q_n only differs from $\frac{u_n}{(1 + u_n)^2}$ by a small quantity of the order of k .

Now

$$\begin{aligned}
 u_{n+1} &= \frac{p_{n+1} + q_{n+1}}{q_{n+1} + r_{n+1}} \\
 &= \frac{p_n + q_n + kq_n \left(\frac{2q_n}{4-k} + \frac{r_n}{2-k} \right)}{q_n + r_n - kq_n \left(\frac{2q_n}{4-k} + \frac{r_n}{2-k} \right)} \\
 &= \frac{p_n + q_n + \frac{1}{2}kq_n(q_n + r_n)}{q_n + r_n - \frac{1}{2}kq_n(q_n + r_n)} \text{ approximately} \\
 &= \frac{u_n + \frac{1}{2}kq_n}{1 - \frac{1}{2}kq_n} \\
 &= u_n + \frac{1}{2}kq_n(1 + u_n) \text{ approximately} \\
 &= u_n + \frac{k u_n}{2(1 + u_n)} \dots \dots \dots (3.1)
 \end{aligned}$$

Solving as for equation (2.1) we find

$$\frac{1}{2}kn = u_n + \log_e u_n - 1. \dots \dots \dots (3.2)$$

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

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

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

Then families will be begotten as follows :

	<i>AA</i>	<i>Aa</i>	<i>aa</i>
From <i>AA</i> females ...	$\frac{p_n u_n}{1 + u_n}$	$\frac{p_n}{1 + u_n}$	0
.. <i>Aa</i> ..	$\frac{q_n u_n}{1 + u_n}$	q_n	$\frac{q_n}{1 + u_n}$
.. <i>aa</i> ..	0	$\frac{r_n u_n}{1 + u_n}$	$\frac{r_n}{1 + u_n}$

After selection and replacement the proportions will be :

	<i>AA</i>	<i>Aa</i>	<i>aa</i>
From <i>AA</i> females ...	$\frac{p_n u_n}{1 + u_n}$	$\frac{p_n}{1 + u_n}$	0
.. <i>Aa</i> ..	$\frac{q_n u_n}{1 + u_n - \frac{1}{2}k}$	$\frac{q_n(1 + u_n)}{1 + u_n - \frac{1}{2}k}$	$\frac{q_n(1 - k)}{1 + u_n - \frac{1}{2}k}$
.. <i>aa</i> ..	0	$\frac{r_n u_n}{1 + u_n - k}$	$\frac{r_n(1 - k)}{1 + u_n - k}$

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 $p_{n+1} = \frac{p_n(1-p_n)}{1+p_n}$. Starting from the standard population, successive values of r_n are 25% , 75% , 87.5% , 99.7% , etc. Where k is small we obtain approximate equations analogous to (3.0) whose solution is

$$\frac{3}{4}kn = u_n + \log_e u_n - 1. \dots\dots\dots(3.3)$$

Thus selection proceeds at $\frac{3}{4}$ of the rate given by equation (2.4).

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 $(n-1)$ th generation form spermatozoa in the ratio $u_nA : 1a$, eggs in the ratio $v_nA : 1a$. Then the n th generation consists of zygotes in the ratios

$$u_n v_n AA : (u_n + v_n) Aa : 1aa. \\ \therefore y_n = (1 + u_n)^{-1} (1 + v_n)^{-1}. \dots\dots\dots(4.0)$$

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

$$\left. \begin{aligned} u_{n+1} &= \frac{2u_n v_n + u_n + v_n}{u_n + v_n + 2} \\ v_{n+1} &= \frac{2u_n v_n + u_n + v_n}{u_n + v_n + 2 - 2k} \end{aligned} \right\} \dots\dots\dots(4.1)$$

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

$$\therefore u_n = \left[2^{n-1} \left(1 + \frac{1}{u_0} \right) - 1 \right]^{-1}, \\ \text{and} \quad y_n = 1 + 2^{1-n} (y_0^{\frac{1}{2}} - 1). \dots\dots\dots(4.2)$$

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;

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

and $\Delta u_n = \frac{k u_n}{2(1 + u_n)}$ approximately.

$$\therefore \frac{1}{2}kn = u_n + \log_e u_n - 1, \dots\dots\dots(4.3)$$

and selection proceeds at half the rate given by equation (2.4), a result stated by Punnett (3).

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 (16) 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

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

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

$$\text{♂ } u_n v_n AA : (u_n + v_n) Aa : (1 - k) aa.$$

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

$$\left. \begin{aligned} \therefore u_{n+1} &= \frac{2u_n v_n + u_n + v_n}{u_n + v_n + 2 - 2k} \\ v_{n+1} &= \frac{2}{1 - k} \frac{u_n v_n + u_n + v_n}{u_n + v_n + 2} \end{aligned} \right\} \dots \dots \dots (5.0)$$

Whilst
$$y_n \text{ (for males)} = (1 + u_n)^{-1} (1 + v_n)^{-1} \dots \dots \dots (5.1)$$

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.

$$\left. \begin{aligned} \therefore y_n &= 0 \\ z_n &= 2^{1-n} z_1 \end{aligned} \right\} \dots \dots \dots (5.2)$$

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.

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

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

$$\therefore kn = 2 \log_e u_n \dots \dots \dots (5.3)$$

if $u_0 = 1$, so selection occurs on the whole more rapidly than by equation (2.4). (See Table V.) y_n is given by equation (2.0).

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 (17) 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

$$\text{♀'s } u^2 AA : 2u Aa : 1aa; \quad \text{♂'s } uA : 1a$$

is stable during random mating, and other populations approach it asymptotically. In any population let the eggs of the $(n - 1)$ th generation be $u_n A : 1a$, the ♀-producing spermatozoa $v_n A : 1a$. Then the n th generation will be: ♀'s $u_n v_n AA : (u_n + v_n) Aa : 1aa$: ♂'s $u_n A : 1a$.

$$\left. \begin{aligned} \therefore u_{n+1} &= \frac{2u_n v_n + u_n + v_n}{2 + u_n + v_n} \\ v_{n+1} &= u_n \end{aligned} \right\}, \dots\dots\dots(6.0)$$

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

$$\left. \begin{aligned} y_n &= (1 + u_n)^{-1} (1 + v_n)^{-1} \\ z_n &= (1 + u_n)^{-1} \end{aligned} \right\}, \dots\dots\dots(6.1)$$

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

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

Solving as usual for recurring series, we have

$$\left. \begin{aligned} 3z_n &= z_0 + 2z_1 + (-\frac{1}{2})^n (2z_0 - 2z_1) \\ y_n &= z_n z_{n-1} \end{aligned} \right\}, \dots\dots\dots(6.2)$$

$$\left. \begin{aligned} \therefore z_\infty &= \frac{1}{3} (z_0 + 2z_1) = \frac{1}{3} (z_n + 2z_{n+1}) \\ y_\infty &= z_\infty^2 \end{aligned} \right\}, \dots\dots\dots(6.3)$$

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.

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

$$\left. \begin{aligned} u_{n+1} &= \frac{2u_n v_n + u_n + v_n}{u_n + v_n + 2 - 2k} \\ v_{n+1} &= \frac{u_n}{1 - k} \end{aligned} \right\}, \dots\dots\dots(7.0)$$

and
$$\left. \begin{aligned} y_n &= (1 + u_n)^{-1} (1 + v_n)^{-1} \\ z_n &= (1 + u_n)^{-1} \end{aligned} \right\}, \dots\dots\dots(6.1)$$

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,

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

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

and
$$\left. \begin{aligned} z_n &= 2^{-n} z_0 \\ y_n &= 0 \end{aligned} \right\}, \dots\dots\dots(7.1)$$

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 equations (4.1), 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.

$$kn = \log_e u_n + 2 \log \left(\frac{3 + u_n}{4} \right), \dots\dots\dots(7.2)$$

$$\left. \begin{aligned} y_n &= (1 + u_n)^{-2} \\ z_n &= (1 + u_n)^{-1} \end{aligned} \right\} \dots\dots\dots(7.3)$$

Table III 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

TABLE III.

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

° of recessives of homozygous sex	99.998	99.98	99.80	99.60	99.00	98.01
" " " heterozygous sex	99.999	99.99	99.9	99.8	99.5	99
kn (number of generations $\times k$) ...	-12.09	-9.786	-7.481	-6.787	-5.866	-5.164

96.04	90.25	81	64	49	36	25	16	10	6.25	4
9.8	95	90	80	70	60	50	40	31.62	25	20
-4.454	-3.485	-2.700	-1.802	-1.156	-.580	0	.619	1.282	1.910	2.506

2	1	.5	.25	.1	.01	.001	.0001	.0.1	.0.1
14.14	10	7.071	5	3.162	1	.3162	.1	.01	.001
3.441	4.394	5.366	6.353	7.679	11.07	14.50	17.95	24.86	31.76

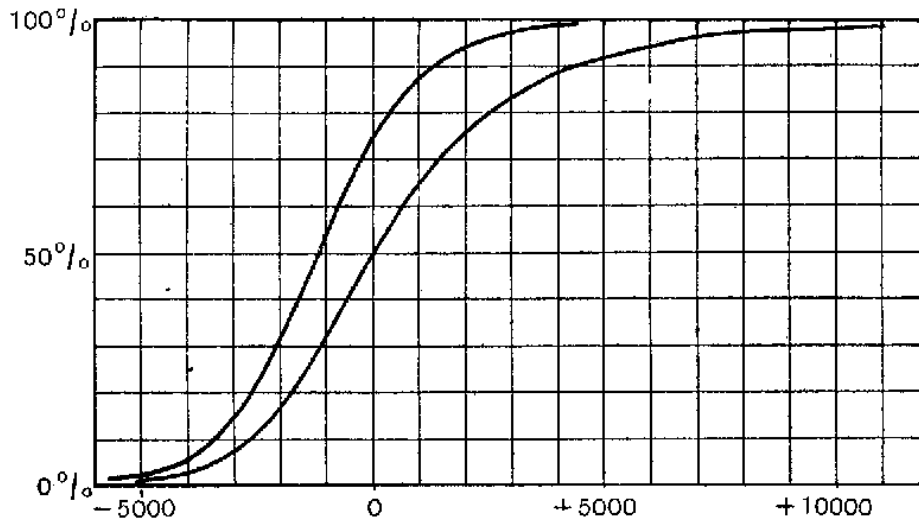


Fig. 3. Effect of selection on a sex-linked character. $k = .001$. Dominants favoured. Upper curve, homozygous sex; lower curve, heterozygous sex. Abscissa = generations. Ordinate = percentage of sex with the favoured character.

autosomal character. In the heterozygous sex selection occurs at about the same rate in the two cases. However, as appears from Table V, 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 III 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.

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

$$\text{♀ } p_n AA : 2q_n Aa : r_n aa; \quad \text{♂ } s_n A : t_n a.$$

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

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

Hence, writing $k' = \frac{k}{2 - k}$,

$$\left. \begin{aligned} p_{n+1} &= (p_n + q_n) s_n \\ 2q_{n+1} &= (p_n + q_n) t_n + (q_n + r_n) s_n + k' q_n t_n \\ r_{n+1} &= (q_n + r_n) t_n - k' q_n t_n \\ s_{n+1} &= p_n + q_n + k' q_n \\ t_{n+1} &= q_n + r_n - k' q_n \end{aligned} \right\} \dots\dots\dots(8.0)$$

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

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

where
$$\phi(n) = \frac{2^{-n}}{\sqrt{5}} [(1 + \sqrt{5})^n - (1 - \sqrt{5})^n], \dots\dots\dots(8.1)$$

i.e. $\phi(n)$ is the n th term of Fibonacci'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 succeeding 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 equations (3.0), and find

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

$$\left. \begin{aligned} \therefore kn &= 3 \log_e \left[\frac{u_n(2+u_n)}{3} \right] \\ r_n &= (1+u_n)^{-2} \\ t_n &= (1+u_n)^{-1} \end{aligned} \right\}, \dots\dots\dots(8.2)$$

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 V.

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 (18).] 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

$$\left. \begin{aligned} u_{n+1} &= \frac{2u_nv_n + u_n + v_n}{u_n + v_n + 2 - 2k} \\ v_{n+1} &= u_n \end{aligned} \right\}, \dots\dots\dots(9.0)$$

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 (7.1)). When k is small we solve as in (7.0), and find

$$\begin{aligned} 3\Delta u_n &= \frac{2ku_n}{1+u_n} \text{ approximately.} \\ \therefore \frac{2}{3}kn &= u_n + \log_e u_n - 1, \dots\dots\dots(9.1) \\ \left. \begin{aligned} y_n &= (1+u_n)^{-2} \\ z_n &= (1+u_n)^{-1} \end{aligned} \right\}, \dots\dots\dots(7.3) \end{aligned}$$

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

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 (19) in poultry can only show in the female sex. If selection is limited to the heterozygous sex,

$$\left. \begin{aligned} u_{n+1} &= \frac{2u_nv_n + u_n + v_n}{u_n + v_n + 2} \\ v_{n+1} &= \frac{u_n}{1-k} \end{aligned} \right\}, \dots\dots\dots(10.0)$$

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

$$\begin{aligned}
 u_{n+1} &= \frac{u_n}{2 + u_n} \text{ (after the second generation),} \\
 v_n &= 0. \\
 \therefore u_n &= \left[2^{n-1} \left(1 + \frac{1}{u_1} \right) - 1 \right]^{-1}, \text{ but } u_1 = u_0; \\
 \therefore y_n = z_n &= 1 - 2^{1-n} z_0. \dots\dots\dots(10\cdot1)
 \end{aligned}$$

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

$$\begin{aligned}
 u_{n+1} &= 1 + 2u_n \text{ (after the second generation),} \\
 v_{n+1} &= \infty. \\
 \therefore u_n &= 2^{n-1} (1 + u_1) - 1; \\
 \therefore y_n = 0 \\
 z_n &= 2^{1-n} z_0 \left. \vphantom{\begin{matrix} \therefore u_n \\ \therefore y_n \\ z_n \end{matrix}} \right\} \dots\dots\dots(10\cdot2)
 \end{aligned}$$

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

$$\begin{aligned}
 \Delta u_n &= \frac{1}{3} k u_n \text{ approximately;} \\
 \therefore kn &= 3 \log u_n \dots\dots\dots(10\cdot3)
 \end{aligned}$$

if $u_0 = 1$; and y_n, z_n are given by equations (7\cdot3). 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.

CERTATION, OR GAMETIC SELECTION OF AN AUTOSOMAL CHARACTER.

The work of Renner (6) and Heribert-Nilsson (20) 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 $u_n A : 1a$, this proportion being reduced by selection in one gender to $u_n A : (1 - k)a$, the selection being general and not familial. Then the n th generation will be $u_n^2 AA : (2 - k)u_n Aa : (1 - k)aa$.

$$\therefore u_{n+1} = \frac{u_n(2u_n + 2 - k)}{(2 - k)u_n + 2 - 2k} \dots\dots\dots(11\cdot0)$$

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

$$\begin{aligned}
 u_{n+1} &= \frac{u_n}{2 + u_n}; \\
 \therefore y_n &= \frac{1}{1 + u_n} = 1 - 2^{1-n} (1 - y_1). \dots\dots\dots(11\cdot1)
 \end{aligned}$$

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

$$\begin{aligned}
 \Delta u_n &= \frac{1}{2} k u_n \text{ approximately;} \\
 \therefore kn &= 2 \log_e u_n, \dots\dots\dots(11\cdot2) \\
 y_n &= (1 + u_n)^{-2}. \dots\dots\dots(2\cdot0)
 \end{aligned}$$

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) $u_n A : 1a$, and selected gametes (say microgametes) $v_n A : 1a$, so its proportions are $u_n v_n AA : (u_n + v_n) Aa : 1aa$.

$$\left. \begin{aligned} \therefore u_{n+1} &= \frac{2u_n v_n + u_n + v_n}{u_n + v_n + 2} \\ v_{n+1} &= \frac{2u_n v_n + u_n + v_n + k'(u_n + v_n)}{u_n + v_n + 2 - k'(u_n + v_n)} \end{aligned} \right\} \dots\dots\dots(11.3)$$

where $k' = \frac{k}{2-k}$, as in equation (8.0). 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 :

$$25, 37.5, 54.69, 71.48, 84.16, 91.83, \text{ etc.}$$

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

$$25, 12.5, 4.56, 1.14, 0.17, .014, \text{ etc.}$$

When selection is slow, $\Delta = \frac{1}{2} k u_n$ approximately.

$$\therefore kn = 4 \log_e u_n \dots\dots\dots(11.4)$$

if $u_0 = 1$, so selection proceeds at half the rate given by equation (11.2), y_n being given by (2.0).

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 $u_n A : 1a$ before, or $u_n A : (1-k)a$ after selection, and ♀-producing spermatozoa in the ratio $v_n A : 1a$. Then the n th generation is

$$\text{♀ } u_n v_n AA : (u_n + v_n - k v_n) Aa : (1-k) aa : \text{♂ } u_n A : (1-k) a.$$

$$\left. \begin{aligned} \therefore u_{n+1} &= \frac{2u_n v_n + u_n + v_n - k v_n}{u_n + v_n + 2 - k v_n - 2k} \\ v_{n+1} &= \frac{u_n}{1-k} \end{aligned} \right\} \dots\dots\dots(12.0)$$

$$\therefore \Delta u_n = \frac{2}{3} k u_n \text{ approximately.}$$

$$\therefore \frac{2}{3} kn = \log_e u_n, \dots\dots\dots(12.1)$$

whilst y_n and z_n are given by equation (7.3), so selection proceeds twice as fast as in equation (10.3). In the more important case of familial selection (replacement in heterozygous individuals),

if $k' = \frac{k}{2-k}$, then

$$\left. \begin{aligned} u_{n+1} &= \frac{2u_n v_n + u_n + v_n + k'(u_n + v_n)}{u_n + v_n + 2 - k'(u_n + v_n)} \\ v_{n+1} &= u_n \end{aligned} \right\} \dots\dots\dots(12.2)$$

u_n being here the gametic ratio after selection.

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

$$\therefore \frac{1}{3} kn = \log_e u_n, \dots\dots\dots(12.3)$$

so selection proceeds as in equation (10.3).

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 (10·3).

COMPARATIVE RESULTS OF COMPLETE (ARTIFICIAL) SELECTION.

The results of complete selection in the more important cases are summarized in Table IV. 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

TABLE IV.
Effects of complete selection.

Character eliminated	Type of selection	Sex	° after 5 generations from 50°	° after 10 generations from 50°	Equation
Non-amphimictic ...	Any ...	Both ..	0	0	—
Autosomal dominant	Bisexual ...	„ ...	0	0	—
„ recessive	„ ...	„ ...	2·44	0·768	2·2
„ dominant	Unisexual...	„ ...	1·83	0·0572	4·2
„ recessive	„ ...	„ ...	8·88	3·27	—
Sex-linked dominant	Bisexual ...	{ Homozygous	0	0	—
		{ Heterozygous	0	0	—
„ recessive	„ ...	{ Homozygous	0	0	7·1
		{ Heterozygous	1·56	0·0484	7·1
„ dominant	In homozygous sex	{ Homozygous	0	0	—
		{ Heterozygous	0	0	—
„ recessive	„ „	{ Homozygous	5·34	1·74	—
		{ Heterozygous	18·5	13·28	—
„ dominant	In heterozygous sex	{ Homozygous	1·83	0·0572	10·1
		{ Heterozygous	3·125	0·0977	10·1
„ recessive	„ „	{ Homozygous	0	0	—
		{ Heterozygous	3·125	0·0977	10·2
Autosomal dominant	Gametic unisexual	Both ...	1·83	0·0572	11·1
„ recessive	„ „	„ ...	0	0	—

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

APPLICATIONS TO SLOW SELECTION.

With the exception of (1·1) 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 V 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

TABLE V.

Generations required for a given change with various types of slow selection. $k = .001$.

Dominant factor favoured	Type of selection	Sex	.001-1°	1-50°	50-99°	99-99.999°	Equations
Non-amphimictic	Bisexual racial	Both	6.921	4.592	4.592	6.921	1.1
Autosomal	" "	"	6.920	4.819	11,664	309,780	2.0. 2.4
"	{ Unisexual racial } Bisexual familial	"	13,841	9,638	23,328	619,560	2 0.3.2,4.3
"	" familial*	"	9.227	6.425	15,522	413,040	2.0. 3.3
" +	" racial	♂	13.831	8.819	6.157	7,112	2.0. 5.3
Sex-linked	" "	{ Homozygous Heterozygous	6.916 6.928	4.668 5.164	5,593 11,070	10.106 20.693	{ 7.2. 7.3
"	" familial	{ Homozygous Heterozygous	20.753 20.768	13.785 14.987	13.785 24.332	20.753 41.450	{ 8.2
"	{ Racial of homo- zygous sex	Homozygous Heterozygous	10.380 10,392	7.228 8.378	17.496 153,893	464.670 149,860.377	{ 7.3. 9.1
"	{ Racial of hetero- zygous sex	Homozygous Heterozygous	20.746 20.753	13.228 13.785	9.236 13.785	10.668 20.753	{ 7.3. 10.3
Autosomal	Unisexual gametic	Both	13.831	8.819	6.157	7,112	2.0. 11.2
"	" " ‡	"	27.661	17.638	12.314	14.224	2.0. 11.4
Sex-linked	{ Gametic of homo- zygous sex	Homozygous Heterozygous	10.373 10,377	6.619 6.892	4,618 6,892	5.334 10.377	{ 7.3. 12.1
"	{ Gametic of homo- zygous sex ‡	Homozygous Heterozygous	20.746 20.753	13.228 13.785	9.236 13.785	10.668 20.753	{ 7.3. 12.3

* 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 .001° to 1°, 11,664 generations to increase from 1° to 50°, and so on.

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:

A. 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.

B. Recessives may be helped to spread by assortative fertilization. This may take place in the following ways:

1. Psychological isolation. Thus Pearson and Lee (21) 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 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.

2. Anatomical isolation. Pearl (22) and Crozier and Snyder (23) have shown that in *Paramecium* 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.

3. 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.

4. 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.

5. Selective fertilization. If the results of Jones (24) 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.

C. 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 $\cdot 001$, their rate of increase will be speeded up elevenfold if the heterozygotes have an advantage of $\cdot 00001$ over the pure dominants.

D. 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 chromosome 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.

SUMMARY.

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.

REFERENCES.

1. PEARSON. *Proc. Roy. Soc.* 54—72.
2. WARREN. *Genetics*, 2, p. 305, 1917.
3. PUNNETT. *Mimicry in Butterflies*, p. 154.
4. IBSEN and STEIGLEDER. *Am. Nat.* 51, p. 740, 1917.
5. DURHAM. *Journ. Genetics*, 1, p. 107, 1911.
6. RENNER. *Zeit. Ind. Abst. u. Ver.* 18, p. 121, 1917.
7. SCHMIDT. *Comptes rendus trav. Lab. Carlsberg*, 14, 1920.
8. GOLDSCHMIDT. *Zeit. Ind. Abst. u. Ver.* 23, p. 1, 1920.
9. PEARSON. *Phil. Trans. Roy. Soc. A.* 203, p. 53, 1904.
10. HARDY. *Science*, 28, p. 49, 1908.
11. PUNNETT. *Journ. Hered.* 8, p. 464, 1917.
12. DETLEFSEN. *Genetics*, 3, p. 573, 1918.
13. BARRETT. *Lepidoptera of the British Islands*, 7, p. 130.
14. BATE. *Ent. Rec.* 1895, p. 27.
15. BATESON. *Mendel's Principles of Heredity*, p. 221.
16. WOOD. *Journ. Agric. Science*, 1, p. 364.
17. JENNINGS. *Genetics*, 1, p. 53, 1915.
18. MORGAN and BRIDGES. *Carn. Inst. Wash. Pub.* 237, 1916.
19. PEARL. *Am. Nat.* 46, p. 130.
20. HERIBERT-NILSSON. *Hereditas*, 1, p. 41, 1920.
21. PEARSON and LEE. *Biometrika*, 2, p. 371, 1903.
22. PEARL. *Biometrika*, 5, p. 213, 1907.
23. CROZIER and SNYDER. *Proc. Soc. Exp. Biol. & Med.* 19, p. 327, 1922.
24. JONES. *Biol. Bull.* 38, p. 251, 1920.