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 back-cross 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

$$D_n = \left[1 - \frac{1}{(u_0+n+1)^2}\right]^k. \qquad (1.1^*)$$

But in F_2, $u_2 = 1$, whence $u_n = n-1$, and

$$D_n = (1 - n^{-2})^k. \qquad (1.2)$$

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

$$g_m{}^1 = \frac{2^{-k} \sum_{i=0}^{m} 2^i \, {}_mC_i \sum_{j=1}^{k-m} 2j \, {}_{k-m}C_i \, Z(i, j)}{1 - D_n}. \qquad (3.2^*)$$

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

$$\begin{aligned} D_n = {} & 2g_6 - g_6{}^2 + 6g_5(2g_1 + 10g_2 + 20g_3 + 20g_4 + 5g_5) \\ & + 30g_4(g_2 + 4g_3 + 3g_4) + 20g_3{}^2 \end{aligned}$$

$$\begin{aligned} g_0{}^1 = {} & \frac{1}{8(1-D_n)}[g_0(8g_0 + 48g_1 + 60g_2 + 40g_3 + 15g_4 + 3g_5) \\ & + 15g_1(4g_1 + 8g_2 + 4g_3 + g_4) + 15g_2(3g_2 + 2g_3)] \end{aligned}$$

$$\begin{aligned} g_1{}^1 = {} & \frac{1}{16(1-D_n)}[g_0(16g_1 + 40g_2 + 40g_3 + 20g_4 + 5g_5) \\ & + g_1(56g_1 + 200g_2 + 160g_3 + 65g_4 + 10g_5) + 10g_2(14g_2 + 17g_3 + 4g_4) + 30g_3{}^2] \end{aligned}$$

$$\begin{aligned} g_2{}^1 = {} & \frac{1}{4(1-D_n)}[g_0(2g_2 + 4g_3 + 3g_4 + g_5) + g_1(2g_1 + 20g_2 + 28g_3 + 17g_4 + 4g_5) \\ & + g_2(29g_2 + 62g_3 + 28g_4 + 4g_5) + 12g_3(2g_3 + g_4)] \end{aligned}$$

$$\begin{aligned} g_3{}^1 = {} & \frac{1}{16(1-D_n)}[g_0(4g_3 + 6g_4 + 3g_5) + 3g_1(4g_2 + 16g_5 + 17g_4 + 6g_5) \\ & + 6g_2(7g_2 + 31g_3 + 24g_4 + 6g_5) + 2g_3(71g_3 + 78g_4 + 12g_5) + 24g_4{}^2] \end{aligned}$$

$$\begin{aligned} g_4{}^1 = {} & \frac{1}{8(1-D_n)}[g_0(g_4 + g_5) + g_1(4g_3 + 13g_4 + 8g_5) + g_2(3g_2 + 34g_3 + 56g_4 + 24g_5) \\ & + 8g_3(6g_3 + 13g_4 + 4g_5) + 8g_4(5g_4 + 2g_5)] \end{aligned}$$

$$\begin{aligned} g_5{}^1 = {} & \frac{1}{16(1-D_n)}[g_0g_5 + 5g_1(g_4 + 2g_5) + 10g_2(g_3 + 4g_4 + 4g_5) \\ & + 10g_3(3g_3 + 12g_4 + 8g_5) + 80g_4(g_4 + g_5) + 16g_5{}^2]. \end{aligned}$$

Hence $D_2=\dfrac{729}{4096}=.1780$, 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 $D_3=\dfrac{1{,}416{,}512}{11{,}336{,}689}=.1249$, 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

$$\begin{aligned} n+C=S_n+\log\,(S_n-1)&+\frac{1}{2S_n}-\frac{1}{3S_n^2}-\frac{5}{36S_n^3}-\frac{13}{240S_n^4}\\ &-\frac{193}{1800S_n^5}-\frac{947}{7560S_n^6}-\cdots\cdots \end{aligned} \tag{3.4}$$

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 $\Delta x_n=\dfrac{-k(x_n-1)^2}{x_n}$ 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

$$\begin{aligned}(1 - R_n)g_m^1 &= 2^{-k}\sum_{i=0}^{m} 2^i\,{}_mC_i \sum_{j=0}^{k-m} 2^j\,{}_{k-m}C_j\, Z(i, j), \text{ when } m \neq 0, \text{ and}\\ (1 - R_n)g_0^1 &= 2^{-k}\sum_{j=0}^{k-1} 2^j\,{}_kC_j\, Z(0, j).\end{aligned} \qquad (4.1^*)$$

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

$$z_{n+1} = \frac{z_n - r_n^2}{1 - r_n^2}$$

$$r_{n+1} = \frac{\frac{1}{2}(r_n + z_n^2) - r_n^2}{1 - r_n^2} \qquad 4.2^*)$$

$$R_{n+1} = r_n^2.$$

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 $p_n AA : 2q_n Aa : r_n aa$, 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:—

$$p_{n+1} = \frac{p_n + \frac{2}{3}(q_n^2 - p_n r_n)}{1 - r_n^2}$$

$$q_{n+1} = \frac{q_n - \frac{2}{3}(q_n^2 - p_n r_n)}{1 - r_n^2}$$

$$r_{n+1} = \frac{r_n - r_n^2 + \frac{2}{3}(q_n^2 - p_n r_n)}{1 - r_n^2}.$$

Putting $z_n = q_n + r_n$, we have

$$z_{n+1} = \frac{z_n - r_n^2}{1 - r_n^2}$$

$$r_{n+1} = \frac{\frac{1}{3}(r_n + 2z_n^2) - r_n^2}{1 - r_n^2} \qquad (5.1^*)$$

$$R_n = r_n^2.$$

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 $(\frac{1}{2}-2^{1-nk})$. The additional multiple recessives removed from F_n would therefore be $(\frac{1}{2}-2^{-n})^k-(\frac{1}{2}-2^{1-n})^k$ taken from a population of $1-(\frac{1}{2}-2^{1-n})^k$ of the original total. Thus the proportion of recessives appearing in F_n is

$$\begin{aligned} R_n &= \frac{(\frac{1}{2}-2^{-n})^k-(\frac{1}{2}-2^{1-n})^k}{1-(\frac{1}{2}-2^{1-n})^k} \\ &= \frac{(2^{n-1}-1)^k-(2^{n-1}-2)^k}{2^{nk}-(2^{n-1}-2)^k}. \end{aligned} \tag{6.1}$$

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
$k=2, 100R_n$	6.25	$8.\dot{3}$	$5.\dot{9}\dot{0}$	3.502	1.909	0.9968	0.5095	0.2435	0.1295
$k=3, 100R_n$	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 $\frac{k}{2-2^{1-k}}$, 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:—

$$x_{n+1} = \frac{x_n + \frac{1}{4}y_n + \frac{1}{12}z_n}{d_n}$$

$$y_{n+1} = \frac{\frac{1}{2}y_n + \frac{1}{3}z_n}{d_n} \tag{8.1*}$$

$$z_{n+1} = \frac{\frac{1}{4}y_n + \frac{1}{3}z_n}{d_n}.$$

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),

$$r_{n+3}=\frac{\left(2-\frac{7}{\sqrt{13}}\right)(5-\sqrt{13})^{-n}+\left(2+\frac{7}{\sqrt{13}}\right)(5+\sqrt{13})^{-n}}{2+\left(1+\frac{1}{\sqrt{13}}\right)(5-\sqrt{13})^{-2}+\left(1-\frac{1}{\sqrt{13}}\right)(5+\sqrt{13})^{-n}} \cdot \qquad (8.5)$$

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 RECESSIVES	RANDOM MATING HETEROZYGOTES	BROTHER-SISTER MATING RECESSIVES (r_n)	BROTHER-SISTER MATING 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

$$D_n = (1 - r_n)^k \qquad (9.1^*)$$

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 $a=609\times6^{-5}$ and

$$\frac{x_{n+1}}{x_n}=\frac{1+609\times6^{n-4}}{6+609\times6^{n-4}}.$$

So

$$D_n=1-\frac{x_{n+1}}{x_n}=\frac{5}{6+609\times6^{n-5}} \tag{10.1}$$

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,

$$R' = \frac{1}{16}l_n + \frac{1}{8}m_n + \frac{1}{4}p_n + \frac{1}{4}q_n. \qquad (10.1^*)$$

The remaining mating types (d to k) give no double recessives in the first generation, but do so later. Also $l_1 = 1$. Hence

$$\begin{aligned}
(1-R')c' &= c_n + \frac{7}{16}d_n + \frac{1}{4}e_n + \frac{19}{64}f_n + \frac{1}{4}g_n + \frac{1}{16}h_n + \frac{1}{4}j_n + \frac{31}{240}l_n + \frac{1}{14}m_n + \frac{1}{12}p_n \\
(1-R')d' &= \frac{1}{4}d_n + \frac{1}{8}f_n + \frac{1}{15}l_n \\
(1-R')e' &= \frac{1}{16}f_n + \frac{1}{15}l_n \\
(1-R')f' &= \frac{1}{4}d_n + \frac{1}{2}2_n + \frac{3}{16}f_n + \frac{1}{8}g_n + \frac{1}{4}h_n + \frac{2}{15}l_n + \frac{1}{14}m_n \\
(1-R')g' &= \frac{1}{8}f_n + \frac{1}{4}g_n + \frac{2}{15}l_n + \frac{1}{7}m_n \\
(1-R')h' &= \frac{1}{32}f_n + \frac{1}{8}h_n + \frac{1}{15}l_n + \frac{1}{28}m_n \\
(1-R')i' &= \frac{1}{120}l_n
\end{aligned}$$

$$\begin{aligned}
(1-R')j' &= \frac{1}{8}f_n+\frac{1}{8}g_n+\frac{1}{2}j_n+\frac{1}{30}l_n+\frac{1}{14}m_n+\frac{1}{3}p_n\\
(1-R')k' &= \frac{1}{30}l_n+\frac{1}{28}m_n\\
(1-R')l' &= \frac{1}{12}d_n+\frac{1}{4}e_n+\frac{1}{16}f_n+\frac{1}{16}g_n+\frac{1}{4}h_n+i_n+\frac{1}{4}k_n+\frac{1}{15}l_n+\frac{1}{14}m_n+\frac{1}{12}q_n\\
(1-R')m' &= \frac{1}{16}f_n+\frac{1}{8}g_n+\frac{1}{4}h_n+\frac{1}{2}k_n+\frac{2}{15}l_n+\frac{3}{14}m_n+\frac{1}{3}q_n\\
(1-R')p' &= \frac{1}{64}f_n+\frac{1}{16}g_n+\frac{1}{16}h_n+\frac{1}{4}j_n+\frac{1}{4}k_n+\frac{1}{30}l_n+\frac{5}{56}m_n+\frac{1}{3}p_n+\frac{1}{6}q_n\\
(1-R')q' &= \frac{1}{30}l_n+\frac{1}{14}m_n+\frac{1}{6}q_n
\end{aligned}\tag{10.2*}$$

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. $c_2=\frac{31}{225}=13.8$ 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 11.1, 13.7, 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.