24

XXI.—**On the Dominance Ratio.** By **R. A. Fisher**, M.A., Fellow of Gonville and Caius College. *Communicated* by Professor J. ARTHUR THOMSON.

(MS. received March 8, 1922. Read June 19, 1922.)

INTRODUCTION.

IN 1918, in a paper published in the *Transactions of the Royal Society of Edinburgh*, the author attempted an examination of the statistical effects in a mixed population of a large number of genetic factors, inheritance in which followed the Mendelian scheme. At that time, two misapprehensions were generally held with regard to this problem. In the first place, it was generally believed that the variety of the assumptions to be made about the individual factors—which allelomorph was dominant; to what extent did dominance occur; what were the relative magnitudes of the effects produced by the different factors; in what proportion did the allelomorphs occur in the general population; were the factors dimorphic or polymorphic; to what extent were they coupled,—besides the more general possibilities of preferential mating (homogamy), preferential survival (selection), and environmental effects, rendered it possible to reproduce any statistical resultant by a suitable specification of the population. It was, therefore, important to prove that when the factors are sufficiently numerous, the most general assumptions as to their individual peculiarities lead to the same statistical results. Although innumerable constants enter into the analysis, the constants necessary to specify the statistical aggregate are relatively few. The total variance of the population in any feature is made up of the elements of variance contributed by the individual factors, increased in a calculable proportion by the effects of homogamy in associating together allelomorphs of like effect. The degree of this association, together with a quantity which we termed the Dominance Ratio, enter into the calculation of the correlation coefficients between husband and wife, and between blood relations. Special causes, such as epistacy, may produce departures, which may in general be expected to be very small from the general simplicity of the results; the whole investigation may be compared to the analytical treatment of the Theory of Gases, in which it is possible to make the most varied assumptions as to the accidental circumstances, and even the essential nature of

 Proceedings of the Royal Society of Edinburgh, 42: 321-341, (1922).

the individual molecules, and yet to develop the general laws as to the behaviour of gases, leaving but a few fundamental constants to be determined by experiment.

In the second place, it was widely believed that the results of biometrical investigation ran counter to the general acceptance of the Mendelian scheme of inheritance. This belief was largely due to the narrowly restricted assumptions as to the Mendelian factors, made by Pearson in his paper of 1903 (6). It was there assumed that the factors were all equally important, that the allelomorphs of each occurred in equal numbers, and that all the dominant genes had a like effect. The effect of homogamy was also left out of consideration, and it is to this that must be ascribed the much lower correlations given by calculation, compared to those actually obtained. When the more general system came to be investigated, it was found to show a surprisingly complete agreement with the experimental values, and to indicate with an accuracy which could not otherwise be attained, how great a proportion of the variance of these human measurements is to be ascribed to heritable factors.

At the time when the paper of 1918 was written, it was necessary, therefore, to show that the assumption of multiple, or cumulative, factors afforded a working hypothesis for the inheritance of such apparently continuous variates as human stature. This view is now far more widely accepted: Mendelian research has with increasing frequency encountered characters which are evidently affected by many separate factors. In some fortunate circumstances, as in *Drosophila*, it has been possible to isolate and identify the more important of these factors by experimental breeding on the Mendelian method; more frequently, however, and especially in the case of the economically valuable characters of animals and plants, no such analysis has been achieved. In these cases we can confidently fall back upon statistical methods, and recognise that if a complete analysis is unattainable it is also unnecessary to practical progress.

This fact is meeting with increasing recognition in the United States, and a considerable number of mathematical investigations have been published dealing with the statistical effects of various systems of mating (Wentworth and Remick, 1916; Jennings, 1916, 1917; Robbins, 1917, 1918). A number of the simpler results of my 1918 paper have since been confirmed by independent American investigators (Wright, 1921). The present note is designed to discuss the distribution of the frequency ratio of the allelomorphs of dimorphic factors, and the conditions under which the variance of the population may be maintained. A number of points of general interest are shown to flow from purely statistical premises.

Recent work in genetics (East and Jones, 1920) leads unavoidably to the conclusion that inbreeding is not harmful in itself, but is liable to appear harmful only through the emergence of harmful recessive characters. This raises the question as to why recessive factors should tend to be harmful, or why harmful factors should tend to be recessive: unless this association exist we should expect to obtain great improvements by inbreeding ordinarily crossbred species, as often as great deterioration. The statistical reason for this association is clear from the distribution of the ratio of allelomorph frequency which occurs under genotypic selection, for, if we assume that adaptation is the result of selection, the majority of large mutations must be harmful, and these can only be incorporated in the common stock in the sheltered region where the rare recessives accumulate (fig. 4).* Similarly there are many well-attested cases of the crossbred (heterozygous) individual showing surprising vigour; but it is not obvious that there is any biological reason for the heterozygote to be more vigorous than the two homozygotes. From a consideration of the stability of the frequency ratios, however, it appears that there will only be stable equilibrium if the heterozygote is favoured by selection against both the homozygotes: naturally this will occur only in a minority of factors, but when it occurs such a factor will be conserved. In the opposite case it will certainly be eliminated.

Cases in which the heterozygote is favoured by selection in preference to both homozygous forms have an additional interest in that these cases, when the selection is intense, may form the basis upon which is built up a system of balanced lethal factors. Muller (1918) has shown that such systems will tend to be built up when selection strongly favours the heterozygote, and has explained how in the light of such systems the majority of the phenomena, including the "mutations," of *Œnothera*, find a genetic explanation.

The interesting speculation has recently been put forward that random survival is a more important factor in limiting the variability of species than preferential survival (Hagedoorn, 4). The ensuing investigation negatives this suggestion. The decay in the variance of a species breeding at random without selection, and without mutation, is almost inconceivably slow: a moderate supply of fresh mutations will be sufficient to maintain the variability. When selection is at work even to the most trifling extent, the new mutations must be much more numerous in order to

* On the Lamarckian theory of evolution, on the other hand, where most, or all, mutations are assumed to be beneficial, we should expect by inbreeding, which uncovers the accumulated mutations in this region, to make great and immediate progress.

maintain equilibrium. That such is the actual state of the case in mankind may be inferred from the fact that the frequency distribution of the numerical proportion of the allelomorphs, calculated on the assumption of selection maintained in equilibrium by occasional mutation, leads to the value of the Dominance Ratio which is actually observed. In all cases it is worth noting that the rate of mutation required varies as the variance of the species, but diminishes as the number of individuals is increased. Thus a numerous species, with the same frequency of mutation, will maintain a higher variability than will a less numerous species: in connection with this fact we cannot fail to remember the dictum of Charles Darwin, that "wide ranging, much diffused and common species vary most" (1, chap. ii).

1. Equilibrium under Selection.

Let the three phases of a dimorphic factor be born in any generation in the proportion

$$P : 2Q : R,$$

then the proportion of the two allelomorphic genes will be

$$P+Q : Q+R, \quad \text{or} \quad p : q;$$

if by selection those that become parents are in the proportion

$$aP : 2bQ : cR, \qquad \text{where} \qquad aP+2bQ+cR=1,$$

then the proportion born in the next generation will be

$$(aP+bQ)^2 : 2(aP+bQ)(bQ+cR) : (bQ+cR)^2;$$

equilibrium is thus only possible if $Q^2=PR$, *i.e.* $P=p^2$, $Q=pq$, $R=q^2$, and if $aP+bQ=p$, $bQ+cR=q$.

Hence it follows that, if

$$a=1+\alpha,\ b=1+\beta,\ c=1+\gamma,$$

$$\frac{\alpha}{p^2}=-\frac{\beta}{pq}=\frac{\gamma}{q^2}$$

specifies the condition of equilibrium.

If selection favours the homozygotes, no stable equilibrium will be possible, and selection will then tend to eliminate whichever gene is below its equilibrium proportion; such factors will therefore not commonly be found in nature: if, on the other hand, the selection favours the heterozygote, there is a condition of stable equilibrium, and the factor will continue in the stock. Such factors should therefore be commonly found, and may explain instances of heterozygote vigour, and to some extent the deleterious effects sometimes brought about by inbreeding.

* p^2 and q^2 should be interchanged.

If the selective action is sufficiently powerful, it may lead in these cases to the establishment of a balanced lethal system.

2. The Survival of Individual Genes.

If we consider the survival of an individual gene in such an organism as an annual plant, we may suppose that the chance of it appearing in the next generation in 0, 1, 2, 3 individuals to be

$$p_0, p_1, p_2, \ldots$$

where

$$p_0+p_1+p_2+\ldots = 1.$$

If

$$f(x)=p_0+p_1x+p_2x^2+\ldots$$

then evidently if there were two such genes in the first generation, the chance of occurrence in r individuals, or more strictly, in r homologous loci, in the second generation, will be the coefficient of x^r in

$$(f(x))^2.$$

It follows that the chance of a single gene occurring in r homologous loci, in the third generation, will be coefficient of x^r in

$$f(f(x)).$$

The form of $f(x)$ will vary from species to species, and in the same species according to the stage of development on which we fix our attention. For simplicity we shall suppose that the successive generations are enumerated at the same stage of development. For the purpose of an evolutionary argument it is indifferent at what stage of development the enumeration is made: in general it will be most convenient to fix our attention on that stage at which the species is least numerous.

In certain important cases the form of $f(x)$ may be calculated. In a field of cross-fertilised grain each mature and ripened plant is the mother of a considerable number of grains, and the father, possibly, of an almost unlimited number. If the number of the species is nearly constant, the average number of its progeny which are destined to become mature is very nearly 2. Or since each gene of a homologous pair occurs in half the gametes, the average number of mature plants in the second generation in which it occurs is 1. Each ovule, therefore, or each pollen grain has individually a very small chance of surviving, and the proportions p_0, p_1, p_2, will be closely given by the Poisson series

$$e^{-1}\left(1, 1, \frac{1}{2!}, \frac{1}{3!}, \ldots\right)$$

In the more general case in which the number of the species is not stationary but increases in each generation in the ratio $m : 1$, m being near to unity, the series will be

$$e^{-m}\left(1, m, \frac{m^2}{2!}, \frac{m^3}{3!}, \ldots\right)$$

and $f(x) = e^{m(x-1)}$. The chance of extinction of a single gene in one generation is e^{-m}, where m is near to unity. In other species in which an individual may survive for many breeding seasons, or in which the generation is of indeterminate length, the form of the function $f(x)$ will be modified: it is sufficiently clear, however, that if we consider that stage in an animal's or plant's life-history at which reproduction is about to commence, the form of the function will not be very different, and the chance of extinction of a particular gene, thus far established in the species, will be

$$e^{-l},$$

where l is a small number not greatly different from unity.* The arbitrary element thus introduced into the question of the survival of a mutant gene is due to the fact that in the first place its survival depends on that of the individual in which it occurs, and this chance is variable from species to species; once, however, it has reached the point of existing in an adult individual capable of leaving many offspring, the conditions of its survival are closely similar in all cases. While it is rare, its survival will be at the mercy of chance, even if it is well fitted to survive. Using the above expression,

$$f(x) = e^{x-1},$$

it may be seen that only about 2 per cent. will survive 100 generations, while those that do will on the average be represented in some 50 individuals. Only when the number of individuals affected becomes large will the effect of selection predominate over that of random survival, though even then only a very small minority of the population may be affected.

3. Factors not acted on by Selection.

If p be the proportion of any gene, and q of its allelomorph in a dimorphic factor, then in n individuals of any generation we have $2np$ genes scattered at random. Let

$$\cos\theta = 1 - 2p$$

where θ lies between 0 and π.

* An upper limit can be set to l by the mere fact of segregation, for in the case of the most uniform possible reproduction, when each individual bears 2 offspring the chance of extinction of any gene is $\frac{1}{4}$, so that l cannot exceed 1·4.

Further, if a second generation of n individuals be now formed at random, the standard departure of p from its previous value will be

$$\sigma_p = \sqrt{\frac{pq}{2n}},$$

hence,

$$\sigma_\theta = \sqrt{\frac{pq}{2n}}\frac{d\theta}{dp} = \frac{1}{\sqrt{2n}}.$$

The fact that this is independent of θ makes it easy to calculate the changes in the distribution of θ, in the absence of selection, for let $y(\theta)\,d\theta$ represent the distribution of θ in any one generation, the distribution in the next will be given by

$$y + \Delta y = \int_0^\pi \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{\delta\theta^2}{2\sigma^2}}\left(y + y'\delta\theta + \frac{\delta\theta^2}{2!}y'' + \ldots\right)$$
$$= y + \frac{\sigma^2}{2}y'' + \ldots$$

Now σ^2 is very small, being $\frac{1}{2n}$, so that measuring time in generations, we have

$$\frac{\partial y}{\partial T} = \frac{1}{4n}\frac{\partial^2 y}{\partial \theta^2}.$$ *

Since we have drawn no distinction between the gene and its allelomorph, we are only concerned with symmetrical solutions: the stationary case is

$$y = \frac{A}{\pi},$$

where A is the number of factors present.

Besides this, we have when y is increasing

$$y = A_0 e^{kT}\frac{p}{2\sinh \frac{1}{2}p\pi}\cdot\cosh p\left(\theta - \frac{\pi}{2}\right),$$

and when y is decreasing

$$y = A_0 e^{-kT}\frac{p}{2\sin \frac{1}{2}p\pi}\cdot\cos p\left(\theta - \frac{\pi}{2}\right),$$

for which

$$k = \frac{p^2}{4n}.$$

4. Terminal Conditions.

If we represent by e^{-l} the chance that a particular gene borne by a single individual will not be represented in the next generation, the chance of extinction for a factor of which b genes are in existence will be

$$e^{-bl}.$$

* See Paper 86 for correction.

When θ is near to 0, p which is always equal to $\sin^2\frac{\theta}{2}$, will be very nearly equal to $\frac{1}{4}\theta^2$. Let

$$t = \sin\tfrac{1}{2}\theta,$$

then the number of genes in existence is $2nt^2$, and the chance of their extinction in one generation is e^{-2nlt^2}.

This chance is negligible save when t is very small, and may be equated to $\frac{1}{2}\theta$; hence the number of genes exterminated in any one generation

$$2\int_0 ye^{-2nlt^2}d\theta$$

$$= 4\int_0 ye^{-2nlt^2}dt.$$

In the stationary case $y = \frac{A}{\pi}$, and the number of genes exterminated will be

$$\frac{A}{\pi}\cdot\frac{2\sqrt{2\pi}}{\sqrt{4ln}} = A\sqrt{\frac{2}{\pi ln}},$$

if new mutations occur at a rate $n\mu$, then this equilibrium will be possible if

$$A = \sqrt{\frac{\pi l}{2}}n^{\frac{3}{2}}\mu.$$

For species in this stationary state the variance will vary (1) as the rate of mutation, (2) as the number of the population raised to the power of $\frac{3}{2}$, (3) as $\sqrt{l}$, a quantity which will seldom differ much from unity. Using the variate $z = \log_e\frac{p}{q}$, the distribution for this case is shown in fig. 1.

5. The Hagedoorn Effect.

In the absence of mutation, extinction will still go on, and the number of factors must diminish, hence we may put for this case

$$y = A_0e^{-kT}\cdot\frac{p}{2\sin\frac{1}{2}p\pi}\cdot\cos p\left(\theta - \frac{\pi}{2}\right).$$

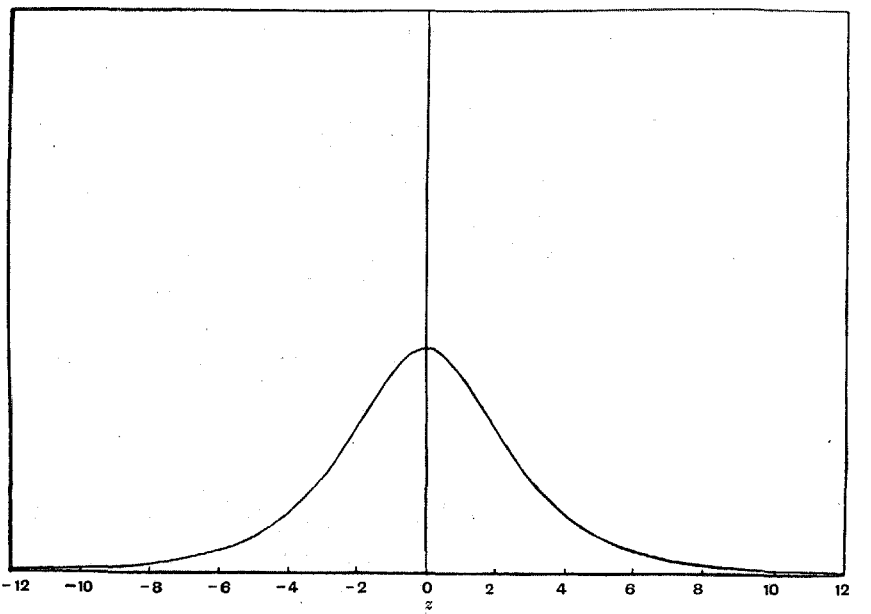
If θ is small,

$$\cos p\left(\theta - \frac{\pi}{2}\right) = \cos\tfrac{1}{2}p\pi + p\theta\sin\tfrac{1}{2}p\pi - \tfrac{1}{2}p^2\theta^2\cos\tfrac{1}{2}p\pi \ \ldots$$

$$= \cos\tfrac{1}{2}p\pi + 2p\sin\tfrac{1}{2}p\pi\,.\,t - 2p^2\cos\tfrac{1}{2}p\pi\,.\,t^2 \ \ldots,$$

so that the rate of extinction is

$$A_0e^{-kT}\frac{p}{2\sin\frac{1}{2}p\pi}\cdot\sqrt{\frac{2\pi}{ln}}\left\{\cos\tfrac{1}{2}p\pi + 2p\sin\tfrac{1}{2}p\pi\cdot\sqrt{\frac{2}{4\pi ln}}\right\}$$



Distribution of the logarithmic frequency ratio $\left(z=\log\frac{p}{q}\right)$ of the allelomorphs of a dimorphic factor.

FIG. 1.—$df=\frac{1}{2\pi}\operatorname{sech}\frac{1}{2}z\,dz$;

represents the distribution when, in the absence of selection, fortuitous extinction is counterbalanced by mutation. Dominance Ratio = ·2308.

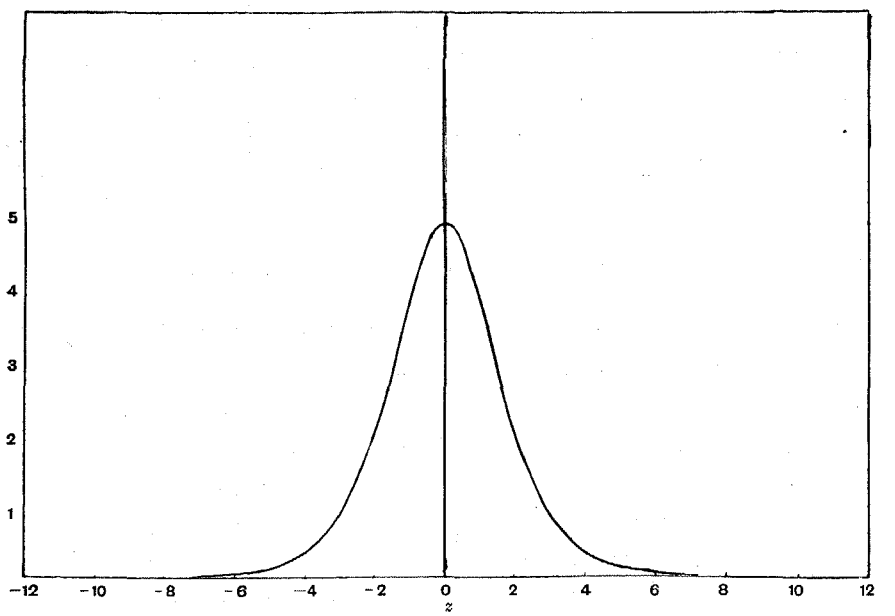


FIG. 2.—$df=\frac{1}{4}\operatorname{sech}^2\frac{1}{2}z\,dz$;

represents the distribution when, in the absence of selection and mutation, the variance is steadily decaying owing to fortuitous extinction of genes. Dominance Ratio = ·2500. This is the condition emphasised by Hagedoorn.

the third term being evidently negligible compared to the first. For equilibrium, therefore,

$$k = p\sqrt{\frac{2\pi}{ln}}\left\{\tfrac{1}{2}\cot\tfrac{1}{2}p\pi + \frac{p}{2\pi ln}\right\}.$$

Remembering that $k = \frac{p^2}{4n}$, we have

$$\frac{p^2}{n}\left(\tfrac{1}{4} - \frac{1}{l}\right) = \sqrt{\frac{2\pi}{ln}}\frac{p}{2}\cot\tfrac{1}{2}p\pi.$$

Hence $\cot\frac{1}{2}p\pi$ is of the order $\frac{1}{\sqrt{n}}$ and is very small, so that p is near to 1. Then

$$k = \frac{1}{4n}.$$

This is a very slow rate of diminution, a population of n individuals breeding at random would require $4n$ generations to reduce its variance in the ratio 1 to e, or 2·8 n generations to halve it. As few specific groups contain less than 10,000 individuals between whom interbreeding takes place, the period required for the action of the Hagedoorn effect, in the entire absence of mutation, is immense. It will be noticed that since l is always less than 1·4 in species stationary in number, the solution above makes p slightly greater than l, which strictly would indicate negative frequencies at the extremes: the value of k is, however, connected with the curvature in the central portion of the curve, and the small distortion at the extremes, where the assumptions, upon which our differential equation is based, break down, will not affect its value. (Fig. 2 shows the distribution of $z = \log\frac{p}{q}$.)

The number by which the number of factors current is reduced in each generation is $\frac{A}{4n}$, and since this number depends on the general form of the distribution curve, it will not be diminished by a number of mutations of the same order. The effect of such very rare mutations would merely be to adjust the terminal of the curve until the rate of extinction is increased sufficiently to counterbalance the additional mutations. It is probable, however, that μ is always far greater than is necessary to make this state of affairs impossible, save in the case of a small colony recently isolated from a very variable species. In this case, with n small and A large, μ might for a time be of the order An^{-2}, rather than of the order $An^{-\frac{3}{2}}$, or An^{-1}.

In the case of a population with A factors, with a supply of fresh

* For $2\pi ln$, read $\sqrt{2\pi ln}$.

† For l, read 1.

mutations sufficient only to be in equilibrium with a smaller number B factors, we may put

$$B\sqrt{\frac{2}{\pi ln}}=\frac{Ap}{2\sin\frac{1}{2}p\pi}\cdot\cos\tfrac{1}{2}p\pi\cdot\sqrt{\frac{2\pi}{ln}},$$

or,

$$\frac{B}{A}=\tfrac{1}{2}p\pi\cot\tfrac{1}{2}p\pi,$$

so that if

$$\frac{a}{\tan a}=\frac{B}{A},\quad 0<a<\frac{\pi}{2},$$

$$p=\frac{2a}{\pi},$$

and

$$k=\frac{a^2}{\pi n}.$$

Similarly, if B>A, the rate of increase in variance may be calculated from the equations

$$\frac{a}{\tanh a}=\frac{B}{A},$$

$$k=\frac{a^2}{\pi n}.$$

The rate of decrease, therefore, cannot, in the absence of selection, exceed the value indicated by $k=\frac{1}{4n}$; no such limit can be assigned to the rate of increase.

6. Uniform Genetic Selection.

In section 1 we have seen that the effects of selection on any Mendelian factor may be expressed by the triple ratio $a:b:c$ representing the relative fitness of the three phases. Only when b exceeds both a and c is there a condition of stable equilibrium; when b is less than both a and c there is a condition of unstable equilibrium; and such factors will tend rapidly to disappear from the stock. Generally, however, we may expect that either b will be intermediate, or equal to a, the value for the dominant homozygote. Two hypothetical cases may, therefore, be considered: (1), in which b is the geometric mean of a and c, and the selection merely affects the proportion of the allelomorphic genes; we may call this uniform genetic selection; and (2), in which b is equal to a, which we may call uniform genotypic selection.

In uniform genetic selection the genetic ratio will be altered in a constant ratio r in each generation, so that after n generations of selection we have

$$\frac{p}{q}=r^n\frac{p_0}{q_0},$$

evidently $r=\frac{a}{b}=\frac{b}{c}$ of section 1.

We may suppose that usually r is near to unity, and $\log r$, which may be positive or negative, may be considered to be of the order of 1 per cent. Let $\log r = a$, then for different factors a will have different values, indifferently positive and negative, since we have no reason to suppose that the selection favours either dominant or recessive characters. The mean square value of a for different factors we shall write σ_a^2.

For any factor

$$\frac{d}{d\mathrm{T}} \log \frac{p}{q} = a;$$

therefore

$$\frac{dp}{d\mathrm{T}} = pqa,$$

$$\frac{d\theta}{d\mathrm{T}} = a\sqrt{pq}.$$

The factors which in one generation are at θ, will in the next be scattered owing to two causes: (1) random survival causing variance, $\frac{1}{2n}$; (2) selection causing variance, $pq\,\sigma_a^2 (= \frac{1}{4}\sin^2\theta \,.\, \sigma_a^2)$. The total variance at any point will be

$$\frac{1}{2n} + \tfrac{1}{4}\sigma_a^2 \sin^2\theta;$$

and so long as σ_a^2 is small as we have supposed, the equilibrium distribution will be

$$y \propto \frac{1}{\sqrt{\sin^2\theta + \dfrac{2}{n\sigma_a^2}}},$$

or nearly

$$y = \frac{\mathrm{A}}{2\log(\sigma_a\sqrt{8n})} \cdot \frac{1}{\sqrt{\sin^2\theta + \dfrac{2}{n\sigma_a^2}}}.$$

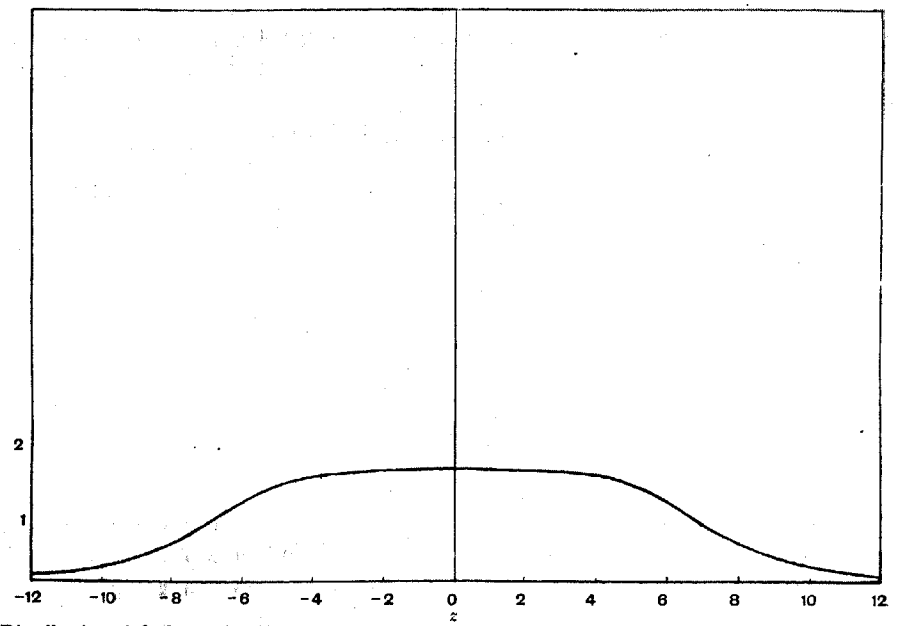
n being large compared with $\frac{1}{\sigma_a^2}$, the effects of selection are, for the more important factors, much more influential than those of random survival. At the extremes, however, for very unequally divided factors the latter is the more important cause of variation. (The distribution of $z = \log \frac{p}{q}$ is shown in fig. 3.)

The amount of mutation needed to maintain the variability with this amount of selection may be calculated from the terminal ordinate

$$\frac{\mathrm{A}\sigma_a\sqrt{\dfrac{n}{2}}}{2\log(\sigma_a\sqrt{8n})},$$

whence

$$n\mu = \sqrt{\frac{2\pi}{ln}} \cdot \frac{\mathrm{A}\sigma_a\sqrt{\dfrac{n}{2}}}{2\log(\sigma_a\sqrt{8n})} = \frac{\mathrm{A}\sigma_a\sqrt{\dfrac{\pi}{l}}}{2\log(\sigma_a\sqrt{8n})}.$$



Distribution of the logarithmic frequency ratio.

FIG. 3.—$df \propto \frac{dz}{\sqrt{1+k^2\cosh^2 \frac{1}{2}z}}$; $k = \cdot 1$;

genetic selection counterbalanced by mutation. Dominance Ratio, ·2000.

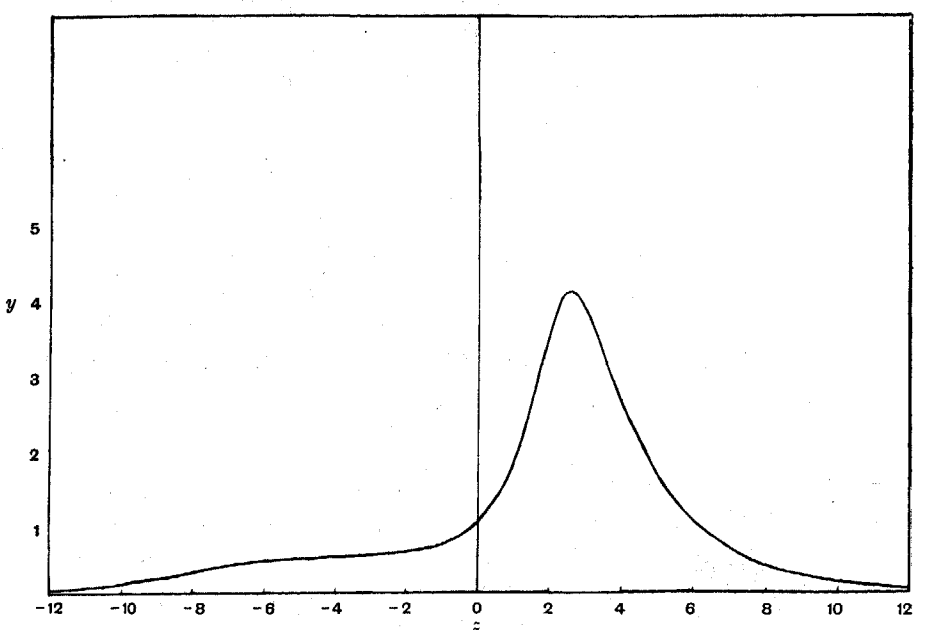


FIG. 4.—$df \propto \frac{dz}{\sqrt{e^{-z}\operatorname{sech}^2 \frac{1}{2}z + k^2\cosh^2 \frac{1}{2}z}}$; $k = \cdot 1$;

genotypic selection, with complete dominance, counterbalanced by mutation. Dominance Ratio, ·3333. This is the probable condition of natural species, including man. Note the accumulation of rare recessives.

Since the logarithm does not increase very rapidly, we may say approximately that A is proportional to $\frac{n\mu}{\sigma_a}$.

It will be seen that to maintain the same amount of variability, as in the case of equilibrium in the absence of selection (section 4), the rate of mutation must be increased by a factor of the order $\sigma_a \sqrt{n}$. Even in the low estimate we have made of the intensity of selection on the majority of factors, this quantity will usually be considerable. The existence of even the slightest selection is in large populations of more influence in keeping variability in check than random survival.

A further effect of selection is to remove preferentially those factors for which a is high, and to leave a predominating number in which a is low. In any factor a may be low for one of two reasons: (1) the effect of the factor on development may be very slight, or (2) the factor may effect changes of little adaptive importance. It is therefore to be expected that the large and easily recognised factors in natural organisms will be of little adaptive importance, and that the factors affecting important adaptations will be individually of very slight effect. We should thus expect that variation in organs of adaptive importance should be due to numerous factors, which individually are difficult to detect.

Owing to this preferential removal of important factors the above solution only truly represents an equilibrium of the variability of the species under absolutely uniform conditions of selection when the new mutations which arise have the same frequency distribution of relative importance as those removed by selection. It must be remembered, however, that the change of variability even by selection is a very slow process, and that gradual changes in the physical and biological environment of a species will alter the values of a for each factor, so tending to neutralise the tendency of selection to lower the value of σ_a. Nevertheless, a will be on the whole numerically smaller for factors in the current stock than it is for fresh mutations.

7. Uniform Genotypic Selection.

If the heterozygote is selected to the same extent as the dominant, or $b=a$, it is easy to see by writing down the first generation, that a genetic ratio $p:q$, becomes in one generation by selection $\frac{p}{q}\,\frac{a}{ap+cq}$; or, writing $1+\beta$ for $\frac{a}{c}$,

$$\frac{p}{q}\,\frac{1+\beta}{1+p\beta};$$

or, when β is small,

$$\frac{p}{q}(1+q\beta).$$

Such selection is therefore equivalent to a genetic selection

$$\alpha = q\beta.$$

Now

$$\frac{d\theta}{dT} = \alpha\sqrt{pq} = \beta q\sqrt{pq},$$

and for the variance caused by selection, instead of $pq\sigma_\alpha^2$, as in Section 6, we now write $pq^3\sigma_\beta^2$: we have then for the total variance produced in one generation in the value of θ,

$$\frac{1}{2n} + \frac{1}{16}\sin^2\theta(1+\cos\theta)^2\sigma_\beta^2$$

$$= \frac{1}{2n} + \sin^2\tfrac{1}{2}\theta\cos^6\tfrac{1}{2}\theta\,.\,\sigma_\beta^2,$$

and the equilibrium distribution will be

$$y \propto \frac{1}{\sqrt{\sin^2\tfrac{1}{2}\theta\cos^6\tfrac{1}{2}\theta + \dfrac{1}{2n\sigma_\beta^2}}}.$$

It is important to notice that this distribution, unlike those hitherto considered, is unsymmetrical, factors of which the dominant phase is in excess are in the majority. This has an important influence on the value of the dominance ratio.

If $2n\sigma_\beta^2$ is large, we can write with sufficient accuracy *

$$y = \frac{A}{1{\cdot}4022(2n\sigma_\beta^2)^{\frac{1}{4}} + \frac{2}{3}\log(8n\sigma_\beta^2) - \frac{2}{3}} \cdot \frac{1}{\sqrt{\sin^2\tfrac{1}{2}\theta\cos^6\tfrac{1}{2}\theta + \dfrac{1}{2n\sigma_\beta^2}}}.$$

The terminal ordinate therefore varies nearly as $(2n\sigma_\beta^2)^{\frac{1}{4}}$, and for large populations in equilibrium, μ varies as $n^{-\frac{1}{4}}$ and as $\sigma_\beta^{\frac{1}{2}}$.

Genotypic selection resembles genetic selection in diminishing the amount of variability which a given frequency of mutation can maintain, or *per contra*, increasing the amount of mutation needed to maintain a given amount of variability; it differs, however, in being comparatively inactive in respect of factors in which the dominant allelomorph is in excess, and consequently in allowing a far greater number of factors to exist in this region (see fig. 4).

* I am indebted to Mr E. Gallop, Gonville and Caius College, Cambridge, for the value of the definite integral. Mr Gallop has shown that the three terms given are the heads of three series in descending powers of $n\sigma_\beta^2$, in which the integral may be expanded.

Now when dominance is complete, the dominance ratio from a group of factors having the same ratio $\frac{p}{q}$ is

$$\frac{1}{1+2\frac{q}{p}},$$

for in the notation of our previous paper

$$\delta^2 = 4p^2q^2a^2,$$

and

$$\alpha^2 = 4p^2q^2a^2\left(1+2\frac{q}{p}\right),$$

where a is half the difference between the two homozygous forms (3, p. 404).

The dominance ratio is therefore raised by an excess of factors in which the dominant gene is the more numerous, such as occurs under genotypic selection.

8. The Dominance Ratio.

The distribution found for the ratio $\frac{p}{q}$ or for the value of θ, which indicates the same quantity, in sections 3 to 7, enable us to calculate the value attained by the dominance ratio under each of the suppositions there considered.

1. In the Hagedoorn condition, where the variance is steadily decaying by random survival, in the absence of mutations or selection,

$$df = \tfrac{1}{2}\mathrm{A}\sin\theta d\theta,$$

writing $\phi = \frac{1}{2}\theta$, then $p = \sin^2\phi$, $q = \cos^2\phi$,
whence

$$\epsilon^2 = \mathrm{S}(\delta^2) = 8\mathrm{A}\overline{a^2}\int_0^{\frac{1}{2}\pi}\sin^5\phi\cos^5\phi d\phi,$$

$$\sigma^2 = \mathrm{S}(a^2) = 8\mathrm{A}\overline{a^2}\int_0^{\frac{1}{2}\pi}(\sin^5\phi\cos^5\phi + 2\sin^3\cos^7\phi)d\phi,$$

and

$$\frac{\epsilon^2}{\sigma^2} = \frac{1}{1+2\cdot\frac{8}{7}} = \cdot 2500.$$

2. When in the absence of selection, sufficient mutations take place to counteract the effect of random survival

$$df = \frac{2\mathrm{A}}{\pi}d\phi,$$

and we have to consider the ratio of the integrals

$$\int_0^{\frac{1}{2}\pi}\sin^4\phi\cos^4\phi d\phi, \quad \int_0^{\frac{1}{2}\pi}\sin^2\phi\cos^6\phi d\phi,$$

which are in the ratio 3 : 5.

The dominance ratio is therefore

$$\frac{3}{3+2(5)} = \cdot 2308\,;$$

the greater variation in the ratio $\frac{p}{q}$ showing itself in a lower dominance ratio.

3. In the third symmetrical case, when genetic selection is at work, the variation of $\frac{p}{q}$ is even greater (fig. 3); since both δ^2 and α^2 contain the factor p^2q^2, the factors in which p or q is very small, make no appreciable contribution to these quantities, consequently we only consider the central portion of the distribution, where

$$df \propto \frac{d\phi}{\sin\phi\cos\phi},$$

the intensity of selection appearing only as a constant factor, and therefore influencing the range of variation of the species, but not its dominance ratio. Here we have the integrals

$$\int_0^{\frac{1}{2}\pi} \sin^3\phi\cos^3\phi d\phi \quad \text{and} \quad \int_0^{\frac{1}{2}\pi} \sin\phi\cos^5\phi d\phi,$$

leading to a dominance ratio

$$\frac{1}{1+4} = \cdot 2000.$$

4. In the case of genotypic selection, which case most nearly reproduces natural conditions, the distribution in the centre of the range is

$$df \propto \frac{d\phi}{\sin\phi\cos^3\phi},$$

consequently the two integrals with which we are concerned

$$\int_0^{\frac{1}{2}\pi} \sin^3\phi\cos\phi d\phi, \quad \int_0^{\frac{1}{2}\pi} \sin\phi\cos^3\phi d\phi$$

are now equal, and the dominance ratio is raised to $\frac{1}{3}$.

In considering the interpretation of the dominance ratio, in our previous inquiry, we found that for symmetrical distributions the value $\frac{1}{3}$ occurred as a limiting value when the standard deviation of $z\left(=\log\frac{p}{q}\right)$ was made zero. Since the dominance ratio calculated from observed human correlations averaged ·32, with a standard error about ·03, we were led to consider that either the allelomorphs concerned occurred usually in nearly equal numbers, a supposition for which we saw no

rational explanation, or that the value of the dominance ratio had been raised by the prevalence of epistacy (non-linear interaction of factors), a suggestion for which no direct evidence could be adduced.

In the light of the above discussion in which we have deduced the distribution of allelomorphic ratios from the conditions of equilibrium with selective influences, from which condition it is probable that natural species do not widely depart, we find that the value $\frac{1}{3}$ for the dominance ratio is produced by the asymmetry of the distribution, and in such a manner as to be independent of the activity of the selective agencies, provided that this exceeds a certain very low level. When differential survival to the extent of only about 1 per cent. in a generation affects the different Mendelian factors, in a population of only a million, and far more for more powerful selection, or a larger population, the dominance ratio will be very close to its characteristic value of $\frac{1}{3}$.

The importance of the fact that this ratio is independent of the intensity of selection, lies not only in the fact that the intensity of selection is usually incapable of numerical estimation, but in the fact that factors having effects of different magnitudes on the soma, which are therefore exposed to selection of varying intensity, and contribute very different quota to the variance, are all affected in the same manner; those factors which by their insignificance might be exposed to selective influences which are not large compared to the effects of random survival will be precisely those which have little weight in computing the dominance ratio.

9. Assortative Mating.

With assortative mating it has been shown (3, p. 414) that the deviations from the mean of the three phases of any factor have, owing to association with similar factors, mean genotypic values given by the formula

$$\mathrm{I} = i + \frac{\mathrm{A}}{1-\mathrm{A}} \cdot \frac{i\mathrm{P} - k\mathrm{R}}{p},$$

$$\mathrm{J} = j - \frac{\mathrm{A}}{1-\mathrm{A}} \cdot \frac{p-q}{2pq}(i\mathrm{P} - k\mathrm{R}),$$

$$\mathrm{K} = k - \frac{\mathrm{A}}{1-\mathrm{A}} \cdot \frac{i\mathrm{P} - k\mathrm{R}}{q},$$

when i, j, k are the deviations in the absence of association, A measures the degree of association produced by assortative mating: p, q are the gene frequencies, and P, R the corresponding phase frequencies for the homozygous phases.

Writing $j=i$ to represent complete dominance, and $P=p^2$, $R=q^2$, since

$$(p^2+2pq)i+q^2k=0,$$

$$\frac{i}{q^2}=-\frac{k}{p(p+2q)}=\frac{i-k}{1}=\frac{p^2i-q^2k}{2pq^2};$$

and since $i-k=2a$, we have

$$I=i+\frac{A}{1-A}\cdot 4aq^2,$$

$$J=i-\frac{A}{1-A}\cdot 2aq(p-q),$$

$$K=k-\frac{A}{1-A}\cdot 4apq;$$

or

$$I-J=2a\cdot\frac{A}{1-A}\cdot q,$$

$$J-K=2a\left(1+\frac{A}{1-A}q\right).$$

If now the survival factors of the three phases are a, b, c, the effect of one generation's selection is given by

$$\frac{p_1}{q_1}=\frac{p_0}{q_0}\frac{ap+bq}{bp+cq}=\frac{p_0}{q_0}(1+p\overline{a-b}+q\overline{b-c}),$$

since a, b, and c are near to 1;
hence

$$\alpha=p(a-b)+q(b-c).$$

Now as $I-J$, $J-K$, the mean differences in any trait due to a single factor, are small compared with the whole variation within the population, we must take $a-b$, $b-c$ proportional to $I-J$ and $J-K$. In other words,

$$a-b=(I-J)\gamma,$$
$$b-c=(J-K)\gamma,$$

where γ measures the intensity of selection per unit change in the trait.

Hence

$$\alpha=\gamma(p\overline{I-J}+q\overline{J-K})$$
$$=\gamma\cdot\frac{2a}{1-A}\cdot q.$$

The general case of uniform genotypic selection when the mean values of the phases are modified by homogamy, therefore, reduces to the case already considered in which homogamy is absent. The total effect of homogamy is to increase the effect of selection by the factor $\frac{1}{1-A}$. The distribution of frequency ratios is unaltered, for although by introducing a difference between I and J the selective effect is made more intense when

p is large, which would tend to make the distribution more symmetrical, this effect is exactly balanced by the increased effect of selection when p is small. The dominance ratio is therefore unaltered by the direct effect of assortative mating.

SUMMARY.

The frequency ratio of the allelomorphs of a Mendelian factor is only stable if selection favours the heterozygote: such factors, though occurring rarely, will accumulate in the stock, while those of opposite tendency will be eliminated.

The survival of a mutant gene although established in a mature and potent individual is to a very large extent a matter of chance; only when a large number of individuals have become affected does selection, dependent on its contribution to the fitness of the organism, become of importance. This is so even for dominant mutants; for recessive mutants * selection remains very small so long as the mutant form is an inconsiderable fraction of the interbreeding group.

The distribution of the frequency ratio for different factors may be calculated from the condition that this distribution is stable, as is that of velocities in the Theory of Gases: in the absence of selection the distribution of $\log \frac{p}{q}$ is given in fig. 1. Fig. 2 represents the case of steady decay in variance by the action of random survival (the Hagedoorn effect).

Fig. 3 shows the distribution in the somewhat artificial case of uniform genetic selection: this would be the distribution to be expected in the absence of dominance. Fig. 4 shows the asymmetrical distribution due to uniform genotypic selection with or without homogamy.

Under genotypic selection the dominance ratio for complete dominance comes to be exactly $\frac{1}{3}$, in close agreement with the value obtained from human measurements.

The rate of mutation necessary to maintain the variance of the species may be calculated from these distributions. Very infrequent mutation will serve to counterbalance the effect of random survival; for equilibrium with selective action a much higher level is needed, though still mutation may be individually rare, especially in large populations.

It would seem that the supposition of genotypic selection balanced by occasional mutations fitted the facts deduced from the correlations of relatives in mankind.

* Before selection, insert the effect of.

REFERENCES.

(1) Darwin, Charles, *The Origin of Species.*

(2) East, E. A., and Jones, D. F., *Inbreeding and Outbreeding* (1920).

(3) Fisher, R. A., "The Correlation between Relatives on the Supposition of Mendelian Inheritance," *Trans. Royal Soc. Edin.*, 1918, lii, pp. 399–433.

(4) Hagedoorn, A. L. and A. C., *The Relative Value of the Processes causing Evolution* (The Hague: Martinus Nijhoff), p. 294 (1921).

(5) Jennings, H. S., "Numerical Results of Diverse Systems of Breeding," *Genetics*, i, 53–89 (1916).

Jennings, H. S., "The Numerical Results of Diverse Systems of Breeding with respect to two pairs of characters, linked or independent; with special relation to the effects of linkage," *Genetics*, ii, 97–154 (1917).

(6) Pearson, K., "On a Generalised Theory of Alternative Inheritance, with special reference to Mendel's Laws," *Phil. Trans.*, cciii, A, pp. 53–87 (1903).

(7) Muller, H. J., "Genetic Variability: twin hybrids and constant hybrids in a case of balanced lethal factors," *Genetics*, iii, 422–499 (1918).

(8) Robbins, R. B., "Some Applications of Mathematics to Breeding Problems," *Genetics*, ii, 489–504 (1917).

Robbins, R. B., "Applications of Mathematics to Breeding Problems, II," *Genetics*, iii, 73–92 (1918).

Robbins, R. B., "Some Applications of Mathematics to Breeding Problems, III," *Genetics*, iii, 375–306 (1918).

(9) Wentworth, E. N., and Remick, B. L., "Some Breeding Properties of the Generalised Mendelian Population," *Genetics*, i, 608–616 (1916).

(10) Wright, S., "Systems of Mating," *Genetics*, vi, 111–178 (1921).