CONTINUOUSLY DISTRIBUTED FACTORS AFFECTING FITNESS[1]

JACK LESTER KING

Lawrence Radiation Laboratory, University of California, Berkeley, California 94720

Received October 1, 1966

HUBBY and LEWONTIN (1966; LEWONTIN and HUBBY 1966) have demonstrated the probability that a population of diploid organisms can support more than an estimated 1000 polymorphisms. Their calculations agree with those of KIMURA and CROW (1964) and VAN VALEN (1963) in indicating that such widespread polymorphism cannot be maintained through single locus heterosis without entailing an absurdly high segregational load. The present paper is intended to demonstrate one way in which a high level of polymorphism might be maintained through selection favoring heterozygosis, with a modest segregational load. The segregational load will be considered as a component in the total variance in viability, rather than as a proportion of the population lost through selection. A mathematical model will be presented and several numerical examples will be given.

For the sake of simplicity, one aspect of fitness will be treated as if it represented the sum of all aspects of fitness; this aspect of fitness is zygote to adult viability, which can be treated as a probability. The simplifying assumption is that the genes being considered do not influence the fecundity of fertile adults.

Very severe genetic effects, like severe environmental accidents, will result in death under most circumstances. Minor decrements in the probability of survival, such as those due to overdominance and to small fluctuations in the environment, are subject to many modifying factors, and are cumulative in nature. One of the principal factors affecting the expression of minor genetic and environmental effects is the environmental opportunity afforded as a result of the density of the population. In most natural populations, the reproductive potential far exceeds the environmental opportunity, and natural selection proceeds by culling to what the habitat can support.

In the present model it is assumed that the genetic and nongenetic factors affecting the probability of survival are cumulative in each individual, and that individuals can be ordered in a linear array according to the sum of the factors affecting their survival. Natural selection proceeds by culling a proportion of the zygotes, taking those with the worst combinations of genes, environment and luck. The remainder, with better combinations of genes, environment and luck, survive to reproduce. Thus this is a threshold model.

A few more simplifying assumptions must be made. It is assumed that the probability of homozygosis is the same at each of N polymorphic loci, and that the selective advantage of the heterozygote is the same over any homozygote at

[1] This work was done under the auspices of the United States Atomic Energy Commission.

that locus, and is the same at each polymorphic locus. A list of symbols is given below.

N: The number of polymorphic loci in the population.
h: The probability of homozygosis at each locus.
σ^2: The variance of a survival factor parameter.
M: The mean of the survival factor parameter.
r: The relative contribution, to the total survival factor variance, of the component due to homozygosity at polymorphic loci.
W_o: The proportion of zygotes surviving to reproduce among all zygotes conceived.
W_a: The proportion of zygotes surviving to reproduce among all zygotes that are homozygous at a given locus.
W_b: The proportion of zygotes surviving to reproduce among all zygotes that are heterozygotes at a given locus.
s: The selective advantage of heterozygosis over homozygosis at a given locus. Thus $s = 1 - W_a/W_b$.
t: The threshold value of the survival factor parameter.
z: The distance between the mean and the threshold in standard deviations.

The number of homozygous loci per individual is binomially distributed about a mean Nh, with a variance $Nh(1-h)$. It is considered that Nh is large enough that the distribution can be considered to be normal.

Each homozygous locus contributes one unit to a measure that shall be termed the "survival factor parameter." This parameter is such that zygotes with values greater than t fail to survive, while those with survival factor values less than t do survive. Sources other than homozygosity for overdominant loci contribute to the parameter, among them environmental factors, stochastic events, and other genetic factors with minor and cumulative effects on fitness.

The assumption is made that all genetic and nongenetic factors with small effects on fitness are additive with respect to the parameter, and that total values of the parameter are normally distributed in the population about a mean M, with a variance of σ^2. The component of variance due to balanced heterosis, $Nh(1-h)$ in units of effect, is some small proportion r of the total variance. Thus,

$$\sigma^2 = \frac{1}{r} Nh(1-h).$$

The proportion of all zygotes surviving to reproduce, W_0, is determined by the reproductive potential of the organism and the population size that the environment can sustain; in this model it is the relative area of the normal curve to the left of the threshold t:

$$W_0 = (2\pi\sigma^2)^{-\frac{1}{2}} \int_{x=-\infty}^{x=t} e^{-\left[\frac{(M-x)^2}{2\sigma^2}\right]} dx.$$

This is readily transformed to:

$$W_0 = (2\pi)^{-\frac{1}{2}} \int_{x=-\infty}^{x=z} e^{-\left[\frac{x^2}{2}\right]} dx, \qquad \text{where } z = \frac{t-M}{\sigma}.$$

Among zygotes homozygous at a given locus, the probability of survival is slightly less than W_0:

$$W_a = (2\sigma\pi^2)^{-\frac{1}{2}} \int_{x=-\infty}^{x=t} e^{-\left[\frac{(M+1-h-x)^2}{2\sigma^2}\right]} dx;$$

$$W_a = (2\pi)^{-\frac{1}{2}} \int_{x=-\infty}^{x=z-\frac{1-h}{\sigma}} e^{-\frac{x^2}{2}} dx.$$

Similarly, among zygotes heterozygous at a given locus, the probability of survival is slightly greater than the average:

$$W_b = (2\pi\sigma^2)^{-\frac{1}{2}} \int_{x=-\infty}^{x=t} e^{-\left[\frac{(M-h-x)^2}{2\sigma^2}\right]} dx;$$

$$W_b = (2\pi)^{-\frac{1}{2}} \int_{x=-\infty}^{x=z+\frac{h}{\sigma}} e^{-\frac{x^2}{2}} dx.$$

The selective advantage of heterozygosity over homozygosity is

$$s = 1 - W_a/W_b .$$

From these formulae numerical examples can be calculated (Table 1). It is intneded that these examples should utilize fairly plausible values of N, h, W_0 and s, sufficient to maintain polymorphism. For example, ROBERTSON (1962) indicates that a selective advantage of .005 over either homozygote is sufficient to maintain polymorphism, in a population of effective size of 1,000, at an equilibrium frequency of one half that predicted for an infinite population; providing that the mutation rate of one allele to the other is 10^{-5}. Actually, 10^{-5} is probably much too high an estimate for a mutation rate to fully functional isoalleles; and many populations exist that are effectively smaller than 1,000, and yet have appreciable polymorphism.

The threshold concept and the concept of fitness: Because many biologists are hostile toward the concept of genetic threshold, it is useful at this point to relate it to the more familiar concept of fitness as a function of genotype. To some the abruptness of the threshold seems unduly artificial. However, subtleties of chance

TABLE 1

Numerical examples of the threshold model of multiple balanced polymorphism

Number of polymorphic loci	Frequency of heterozygosis	Selective advantage of heterozygosity	Proportion of zygotes surviving	Effect of one substitution	Proportion of variance due to overdominance	Survival with 25% inbreeding	Inbreeding depression with 25% inbreeding
N	$1-h$	s	W_0	$1/\sigma$	r	W_F	$1-W_F/W_0$
1000	0.3	.001	0.5	.00125	.0003	.4627	7.5%
1000	0.3	.0025	0.5	.00313	.0021	.4075	18.5%
1000	0.3	.005	0.5	.00627	.0083	.3202	36%
1000	0.3	.010	0.5	.01253	.0330	.1711	66%
1000	0.3	.005	0.2	.00361	.0021	.1525	25%
1000	0.1	.005	0.5	.00627	.0035	.4397	12%
500	0.3	.0025	0.5	.00313	.0010	.4533	9%
500	0.3	.005	0.5	.00627	.0041	.4121	18%
500	0.1	.010	0.5	.01253	.0071	.3133	37%

and gradation are accounted for in the nongenetic component of variance. The threshold is not intended to represent a physiological phenomenon, but rather a statistical projection.

It is a more common procedure to ascribe a specific fitness value to a specific genotype, and to treat this value as an independent probability. This has led to absurdities with regard to the problem of segregational load in that fitness-probability values have been illegitimately assigned to components of genotypes, which are not independent. Still, it is quite possible to interpret the threshold model in terms of the fitness values of specific genotypes.

Figure 1 illustrates the threshold model, a density distribution with additive genetic and nongenetic survival factors on the abscissa and a threshold at t. In Figure 2 the scale is the same, but the fitness distributions of several specific genotypes are given; each has a specific mean, and a distribution about its mean. The variance of this distribution is the nongenetic component of total variance. Each of the distributions intersects the threshold. The probability of survival is the proportion of the area to the left of the threshold. The probability of survival of the genotype on the threshold is one half.

Each genotype, then, has a specific fitness value in this model. The fitness values are plotted against the genotypes in Figure 3. The maximum value of $1/W_0$ is approached asymptotically on the left, and the minimum of zero fitness is approached asymptotically on the right. The curve is in fact a familiar one: it is the cumulative normal distribution.

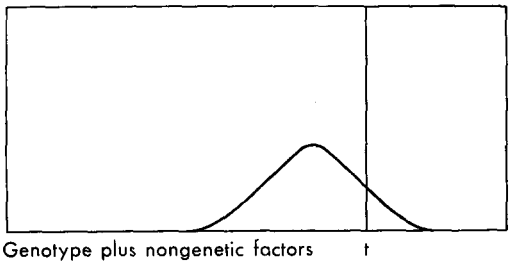

FIGURE 1.—The threshold model: normally distributed genetic and nongenetic factors affecting survival.

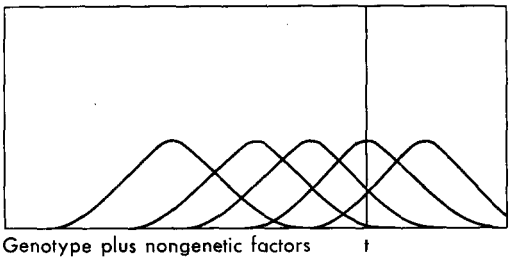

FIGURE 2.—Schematic representation of the distribution of nongenetic survival factors about each genotypic mean.

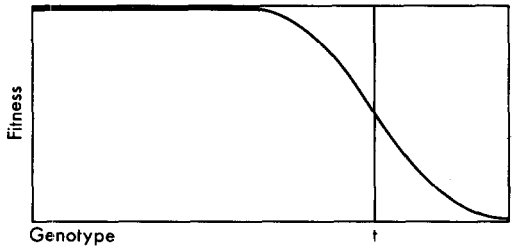

FIGURE 3.—The probability of survival as a function of the genoytpe. The curve is identical with the cumulative normal distribuiton.

Other aspects of fitness: To this point the model has been concerned with only one aspect of fitness, namely preadult survival. This is a relatively easy aspect to deal with mathematically. The probability of survival, like all probabilities, has a clear upper limit (one). It is an all-or-none phenomenon, and thus subject to threshold analysis. Part of the aim of this study has been to demonstrate that it is possible for a population to maintain many balanced heterotic polymorphisms, and it is *possible* that these are all maintained through effects on preadult viability. However, it is also desirable to extend the model to total fitness.

One can postulate, with SVED, REED and BODMER (1967) and WALLACE (1958) that for each population there is an upper limit to total fitness. Now, either there is an upper limit or there is not; if there is not, there is no problem in the maintenance of many heterotic polymorphisms, and it is not nonsensical to postulate a fly that lays two billion fertile eggs (LEWONTIN and HUBBY 1966). If there is an upper limit, it can be represented by a sharp cut-off and a level plateau at a certain level, as in SVED's model, or it can be approached asymptotically as in WALLACE's. The asymptotic approach to maximal fitness that is illustrated in Figure 3 is derived from a threshold model of survival, but it is congruent with WALLACE's hypothetical curve for total fitness. It is not an unreasonable assumption that the relationship between total fitness and genetic factors affecting fitness is as shown in Figure 3: an extensive range in which genetic variation does not measurably affect fitness, and a range in which additional increments of deleterious genetic factors overreach the limits of homoestasis and do affect fitness.

If this assumption can be accepted—and I see no alternative to it—then the relationships between genes and fitness are the same in preadult viability and in other aspects of fitness, and the model achieves generality. The cumulative normal distribution is described by the same mean, variance, and standard deviation; the threshold is dispensed with, and with it, hopefully, many intellectual reservations.

The optimal genotype, maximal fitness, and gene interaction: If it is legitimate to ascribe fitness values to specific genotypes, then these values can be ordered, and either one genotype has the highest value or a number of genotypes share this value. In the present model maximal fitness is approached asymptotically. With a thousand polymorphic loci, there are trillions of hypothetically possible genotypes with essentially the same maximal fitness, and hence there is no meaningful single "optimal genotype."

The effect on fitness of a substitution of homozygosity for heterozygosity depends largely on the genetic background. Within the broad range of genotypes with near maximal fitness, such a substitution has essentially no effect on fitness. Against the average genetic background the same substitution reduces fitness sufficiently to maintain polymorphism.

The total reduction in fitness due to homozygous overdominant loci is much greater than the sum of the effects the individual substitutions would have in the maximal fitness range; it is much less than the sum of the effects the individual substitutions would have, on the average, in the range of probable genotypes. Thus gene interaction with respect to fitness can be said with equal justification to be positively or negatively synergistic, depending on which value is chosen. With the average decrement in fitness associated with each substitution as a standard, there is considerable negative synergism between loci in the total effect.

Within the range of probable genotypes there is scarcely any observable synergism. Most genotypes occur in a very short range of the fitness distribution (Figure 3); over short distances, the cumulative normal distribution is nearly linear. The expected reduction in fitness associated with a genotype with exactly one more homozygous locus than the mean number is essentially half the expected reduction in fitness associated with a genotype with exactly two more homozygous loci than the mean number.

Continuosly distributed mutational decrements: The individual effects on fitness associated with homozygosity of balanced heterotic loci are nearly always small, continuously distributed and (according to the present hypothesis) cumulative. Deleterious alleles kept in the population because of mutation pressure alone can be minor or drastic in their effects on fitness. Drastic effects, principally lethality, are neither cumulative nor continuously distributed, and are outside of the range of the present model. Of course, it is believed at present that many or most drastic recessives have minor deleterious effects on the fitness of heterozygotes, and behave in populations mostly in accord with their minor dominant effects. The present model is concerned with all cumulative and continuously distributed factors affecting fitness, of whatever origin.

Let Q be the mean number of minor deleterious effects due to mutation (deleterious dominants or pairs of homozygous recessives) per zygote. The distribution approximates the Poisson, so that the variance is also Q and the standard deviation is $Q^{1/2}$, in number of effects per zygote. For algebraic simplicity, assume that the mutational effects are of equal magnitude; the magnitude of one such effect is equivalent to x standard deviations of the total distribution of genetic and nongenetic factors affecting fitness. One standard deviation of the component of variance due to mutational effects is equivalent then to $x = Q^{1/2}$ standard deviations of the total distribution, and the component of variance due to mutational effects, relative to the total variance, is $r = x^2Q$ (the parameter r is comparable to h^2, heritability, in selection studies).

One numerical example should suffice. Suppose that $Q = 25$ and the selection

coefficient against each effect is $s = .01$ (e.g., with $N = 12{,}500$ loci and a mutation rate of $u = 10^{-5}$ per locus to dominant deleterious alleles of such magnitude, the equilibrium frequency per locus is $q = u/s = 10^{-3}$, and $2Nq = 25$).

When $W_0 = 0.5$, the selection coefficient of $s = .01$ is equivalent to $x = .01253$ standard deviations of the total distribution of fitness factors. The component of variance due to deleterious minor mutant effects, relative to total variance, is $r = x^2Q = .004$. This is approximately of the same order of magnitude as most of the numerical examples given in Table 1 of the relative variance component due to overdominance.

Estimates of the number of mutable loci for any organism still range widely. It is possible that at some time sound estimates of gene number, sound estimates of mutation rates, and sound estimates of the probable fitness of mutant heterozygotes will be combined to give a calculated mutational load manifestly too great for the survival of the species in question, just as past estimates of heterotic effects have been calculated to give unbearable segregational loads. I hope that the present model can obviate such a question before it has properly been raised.

The phenodeviant hypothesis: The threshold model is particularly suited to examination of some aspects of the phenodeviant hypothesis (LERNER 1954). One can make the extremely simplifying assumption that the nongenetic effects are independent of the balanced heterotic effects, and that in the latter case selection is solely against individuals homozygous for more than a certain threshold number of polymorphic alleles. The model is the same, except that $r = 1$ and W_0 becomes the proportion of zygotes that are not multiply homozygous phenodeviants. With 1,000 polymorphic loci, h assumed to be 0.5 and W_0 assumed to be .95, the selective advantage of heterozygosis at each locus is .006. Or, with 50 loci and a phenodeviant frequency of .01, the selective advantage of heterozygosis is .008; there might be a number of independent developmental systems, each producing phenodeviants in low frequency.

LERNER did not rule out the influence of nongenetic factors in the production of phenodeviants, however. The essential feature of the present model—the ability of the organism to tolerate homozygosis at many, but not too many, loci—is derived directly from his *Genetic Homeostasis* (1954).

Inbreeding depression: SVED, REED and BODMER (1967) raise a very important point which has previously been overlooked, namely, the rather large inbreeding depression that is the consequence of the assumption of heterozygous advantages at many loci. LEWONTIN (private communication) states, "What (SVED *et al.*) here really accomplished in their paper is to turn our attention away from the spurious problem of the optimum genotype that has so long plagued us, and instead turned our attention to the *real* problem, which is 'Why isn't inbreeding depression more severe if there is all that heterosis?' "

In the threshold model, inbreeding of a degree F changes the probability of homozygosis per polymorphic locus from h to $h + F(1-h)$ and the mean of the distribution from Nh to $N(h+F(1-h))$. The total variance in the distribution of survival factors is nearly unchanged. The inbreeding depression due to addi-

tional homozygosis for overdominant alleles can be defined as $1 - W_F/W_0$, where

$$W_F = (2\pi\sigma^2)^{-\frac{1}{2}} \int_{x=-\infty}^{x=t} e^{-\left[\frac{(M+NF-NFh-x)^2}{2\sigma^2}\right]} dx;$$

$$W_F = (2\pi)^{-\frac{1}{2}} \int_{x=-\infty}^{x=z-\frac{NF-NFh}{\sigma}} e^{-\frac{x^2}{2}} dx.$$

Values of predicted inbreeding depression with $F = 0.25$ are given for several numerical examples in Table 1. The threshold model and the model of SVED *et al.* do not differ greatly with respect to the predicted effect of inbreeding.

For a recent empirical comparison, SITTMANN *et al.* (1966) found the probability of egg to fertile adult survival of domesticated Japanese quail to be reduced from $W_0 = 0.296$ to $W_F = 0.100$ when $F = .25$, for a calculated inbreeding depression of 0.662. LERNER (1954) cites similar data for chickens, with an average inbreeding depression of about 0.50. Other organisms, however, have shown relatively small amounts of inbreeding depression.

Population measurements under artificial conditions: Removal of a natural population to an artificial environment—e.g., the laboratory—drastically alters the fitness and natural selection parameters. In a stable population, for instance, the proportion of zygotes achieving maturity is probably never greater than 0.5, and must often be much smaller. In an artificial environment most populations capable of reproducing at all increase explosively.

Through a change in the mean of nongenetic factors affecting fitness, the proportion of survivors is greatly increased and the threshold is moved well toward one tail of the distribution of fitness factors. Minor selection differentials, operative in a stable population, are greatly reduced in magnitude in an expanding population. Mutant alleles with drastic effects will continue to take their toll, but the continuously distributed genoypes will have been moved into the plateau asymptotically approaching optimal fitness. In this respect it is important to note that the inbreeding depressions predicted in Table 1 are based on the expected survival of inbred individuals in an environment in which only enough of the outbred individuals survive to replace the parental generation. The inbreeding depression should be much smaller in a rapidly expanding population. Similarly, experimentally determined selection differentials for either overdominant or mutant loci can be expected to be much smaller than the differentials operative under the conditions of stable population size. Note that the human population is currently in an expanding phase in an artificial environment of its own making.

DOBZHANSKY, SPASSKY and TIDWELL (1963) report the following egg to adult survival frequencies for laboratory populations of *Drosophila pseudoobscura*: outbred, 0.869; inbred, 0.738 with 25% inbreeding. The inbreeding depression for this component of fitness is thus 0.15 under optimal conditions. In the wild, however, nothing like 87% of Drosophila eggs reach adulthood. If the average number of fertile eggs per adult female is of the order of, say, 100, then the egg to adult survival in a stable population is of the order of $W_0 = 0.02$.

Let it be assumed (though improbable) that the laboratory and field are similar with respect to the *variance* of the distribution of nongenetic fitness factors, although the *means* differ; and further, that the inbreeding depression observed in the laboratory is due principally to continuously distributed minor factors rather than to recessive lethals. Then the survival frequency in the laboratory of $W_0 = 0.868$ is equivalent to a threshold distance of 1.12 standard deviations; with 25% inbreeding, the distance between the threshold and the mean is reduced to 0.64 standard deviations; thus the effect of 25% inbreeding is equivalent to 0.48 standard deviations of the total distribution of survival factors.

The same displacement of 0.48 SD, when $W_0 = .02$, results in a predicted inbreeding survival frequency of $W_F = .0055$. The inbreeding depression in egg to adult survival under these circumstances would be $1 - W_F/W_0 = 0.725$, more closely in accord with the data on Japanese quail and with the predictions of Table 1.

I am grateful for the advice and criticism of JOHN A. SVED, LEIGH VAN VALEN, R. C. LEWONTIN and TH. DOBZHANSKY.

SUMMARY

It is proposed that natural selection operates by culling a proportion of the population with the worst combination of genetic, nongenetic and stochastic factors affecting survival. A model is presented in which the effects of most of these factors are assumed to be additive, and the values of the sums of the effects are assumed to be normally distributed in the population. One thousand or more polymorphisms can be maintained through an average heterozygote superiority of about 1% per locus, with a very small total effect on the variance of fitness in the population.

LITERATURE CITED

DOBZHANSKY, TH., B. SPASSKY, and T. TIDWELL, 1963 Genetics of natural populations. XXXII. Inbreeding and the mutational and balanced genetic loads in natural populations of *Drosophila pseudoobscura*. Proc. Natl. Acad. Sci. U.S. **53**: 482–486.

HUBBY, J. L., and R. C. LEWONTIN, 1966 A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. Genetics **54**: 577–594.

KIMURA, M., and J. F. CROW, 1964 The number of alleles that can be maintained in a finite population. Genetics **49**: 725–738.

LERNER, I. M., 1954 *Genetic Homeostasis*. Oliver and Boyd, London.

LEWONTIN, R. C., and J. L. HUBBY, 1966 A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. Genetics **54**: 595–609.

ROBERTSON, A., 1962 Selection for heterozygotes in small populations. Genetics **47**: 1291–1300.

SITTMANN, K., H. ABPLANALP, and R. A. FRASER, 1966 Inbreeding depression in Japanese quail. Genetics **54**: 371–379.

SVED, J. A., T. E. REED, and W. F. BODMER, 1967 The number of balanced polymorphisms that can be maintained in a natural population. Genetics **55**: 469–481.

VAN VALEN, L., 1963 Haldane's dilemma, evolutionary rates, and heterosis. Am. Naturalist **97**: 185–190.

WALLACE, B., 1958 The role of heterozygosity in Drosophila populations. Proc. 10th Intern. Congr. Genet. **1**: 408–419.