This dissertation has been
microfilmed exactly as received

69-20,665

ROUX, Carl Ziervogel, 1937-
THE MATHEMATICAL THEORY OF GENETIC LOADS.

Iowa State University, Ph.D., 1969
Biology-Genetics

University Microfilms, Inc., Ann Arbor, Michigan

THE MATHEMATICAL THEORY OF GENETIC LOADS

by

Carl Ziervogel Roux

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subjects: Statistics
Animal Breeding

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Heads of Major Departments

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Of Science and Technology
Ames, Iowa
1969

TABLE OF CONTENTS

I. INTRODUCTION

The mechanisms for the maintenance of genetic variability in sexually reproducing species present problems that continue to fascinate biologists, for it is true that whenever the facade of the relatively homogeneous phenotypic appearance of cross fertilizing populations was probed an amazing degree of genetic diversity was revealed. The explanations for the maintenance of genetic diversity range all the way from phenomena manifested primarily on an individual basis to the relationship between a population and its environment. Among the phenomena that manifest themselves on an individual basis can be counted the biochemical superiority of heterozygotes and meiotic drive or non-random segregation. Next comes inter-individual relationships like the mating of unlike parents and various forms of competion. The relationships between a population and its environment that can have some influence on the maintenance of genetic heterogeneity include factors such as the population being part of a geographic gradient, the preference of different genotypes for different ecological niches and the regular cyclic alternation of selection pressure.

In this study we will focus our attention on two major hypotheses to account for the maintenance of genetic heterogeneity in cross fertilizing diploid populations. These two hypotheses are gene frequency equilibrium theories based upon 1) mutation-selection balance and 2) heterozygote superiority. The technique of estimating genetic loads was proposed by Morton, Crow and Muller (1956) and Crow (1958) as a possibility for distinguishing between these two hypotheses.

In their development of the topic, these authors made the following assumptions:

(i) Hardy-Weinberg frequencies exist at all loci.

(ii) Gene action between loci is independent and multiplicative, with each locus having a small effect on fitness.

(iii) The fitness of the fittest genotype is known or estimable.

(iv) All selection taking place is of the constant selection coefficient type.

(v) No sex difference in fitness exists.

(vi) The description of the genotypic array by Wright's coefficient inbreeding is adequate for the purpose of the estimation of genetic loads.

The purpose of this study is to provide a framework to clarify the use of these assumptions in the derivation of load theory and which will allow the consideration of alternative assumptions that may be more realistic. Such a framework will also enable one to see what the effect of likely departures from the basic assumptions of the load theory may be.

In order to keep the mathematics at a manageable level we will assume a population with only two life phases designated as infancy and adulthood, non-overlapping generations, random mating, a uniform environment and infinite population size. The validity of this type of simplifying approach derives from a feeling that in the advance of science it is not simplification that leads to error, but rather the absence of a rigorous and clear analysis of the problem at hand. A simplified approach may also reveal whether a theory holds enough promise to warrant further investigation. On the other hand, stumbling blocks may be revealed that can preclude

further development.

The headings of the different chapters and sections in this study are believed to be such that a reasonable idea of the content can be formed. Nevertheless a few remarks about the method of approach followed in each chapter may be pertinent here.

In Chapter II a review of the literature is given. The treatment is deliberately brief and sketchy since the formulation at present employed in the treatment of fitness is considered to be inadequate. In Chapter III an attempt is therefore made to develop a theory of the genetic structure of populations that can serve as a basis for the theory of genetic loads. In much of Chapter III two-loci mathematics will be used, since the notation is simplified in this manner and because in most instances the extension to the case of n loci will be obvious.

Much of the material in Chapter III is of interest in its own right, so that no serious attempt was made to remain within the limits of the prerequisites of load theory. The main results of this chapter are given by Theorems 1 through 5. Theorems 1 and 2 appear in Section D and Theorems 3 and 4 in Section E. Theorem 4 has two corollaries. The work on border points in the multiplicative case and on mutation and selection in Section F has not been formalized in theorems. Theorem 5 is given in Section G.

The load theory in the case of reproductive fitness is given in Chapter IV. The approach here is to work from the single locus results to a generalized n-loci approach. This approach has the advantage that it allows the derivation of the n-loci results with a minimum of notational difficulties. It is also easy to follow for anybody acquainted with the existing theory.

The application of the load theory is considered in Chapter V. Thus far load theory has been applied mainly to the trait viability. The theory for loads as applied to viability is therefore derived in this chapter, and its application to experimental situations is considered.

Special attention is given to the problem of estimation of the load ratios in relation to the experimental error structure, since this aspect has received scant attention in the literature. This consideration of the error structure is part of the attempt throughout to be explicit with regard to the underlying assumptions of the load theory.

The equations are indexed according to the major subheadings. For example, Equation A.3 refers to the third numbered equation of that section of a particular chapter. In a reference to an equation in a different chapter, the chapter number is given with the equation. For example the reference Equation III.A.3 refers to the third equation in Section A of Chapter III.

II. REVIEW OF LITERATURE

The general idea of genetic loads dates back to a paper by J.B.S. Haldane (1937) entitled "The effect of variation on fitness." Haldane defined the fitness of a genotype in a hermaphrodite organism as half the mean number of progeny left by an individual of that genotype. Progeny due to self-fertilization are to be counted twice over, and individuals must be counted at the same stage of the life cycle, e.g. at birth or maturity. Also in determining average fitness arithmetic means are to be taken in space and geometric means in time. Without being more explicit about it Haldane then proceeded operationally in the same way as did Crow (1958) in the paper in which the load concept was formally defined.

Crow defined fitness as the expected number of progeny of a genotype where the offspring are counted at the same age as the parents. From the way in which Haldane and Crow, and also most other authors in the field of population genetics, employ a mathematical variable called fitness in their derivations, it is clear that they handle fitness in the same way as a metric trait, i.e., that they suppose it to be the property of a single individual. This is true since the idea behind simple quantitative genetic theory of a metric trait is that there is a genotypic value associated with each genotype, after averaging over the relevant population of environments. However, sexual reproduction, except in the case of selfing, is the product of the interaction between two genotypes. It is clear, therefore, that the concept of fitness needs to be reformulated, and that subsequent derivations of the properties of the trait "fitness" will have to be done in the light of such a reformulation. In particular, the idea of defining fitness as

half the number of offspring encounters great difficulties except under very special assumptions.

This study is essentially an attempt to place the study of the trait "fitness" on a sound footing and is based on an expansion of the work of Kempthorne and Pollak (1966). However, since in this chapter we are concerned with the genesis of the load concept, the load ideas will be represented in their original form, and the reformulation will be presented in later chapters after the relevant concepts have been introduced and developed.

Haldane (1937) showed that the effect of mutation on population fitness depends mainly on the mutation rate and not on the harmfulness of the individual mutant. This is true provided that the average selective disadvantage of the mutant gene in various genotypic combinations is large relative to the mutation rate and that the mutant is deleterious or neutral in the heterozygous state. If the heterozygote is fitter than either homozygote at a locus, the average fitness, relative to that of the heterozygote, will usually be decreased by a considerably larger amount than at a locus where variability is maintained by recurrent mutation. These results will be derived through the use of the load concept as formalized by Crow (1958).

The proportion by which the fitness of the average genotype in the population is reduced in comparison with the best genotype has been designated as the genetic load (Crow, 1958),

$$L = \frac{x_{max} - \bar{x}}{x_{max}} , \qquad (1)$$

where x_{max} is the fitness of the maximal genotype and $\bar{x}$ is the mean fitness of the population. In the case where genetic heterogeneity is

maintained by a balance between the forces of mutation and selection we have the mutation load, and in the case where genetic heterogeneity is maintained by a balance between selective forces we speak of a segregation load.

It follows that if there is some way by which it can be determined whether one has mainly a mutation or mainly a segregation load in a population, one would be able to determine roughly the cause of genetic heterogeneity in the population under consideration. In the last analysis one can perform only two operations on a population. First, one can choose the parents of the next generation, and second, one can determine the pattern in which they are mated. It will be shown that the two loads show a different response under inbreeding, and that this provides us with a possibility of discriminating between the mutation and segregation loads. The work on this possibility will now be briefly reviewed, together with the development, in outline, of the load argument.

A. The Mutation Load

Consider a single locus with two alleles (1), (2) with frequencies p(1) and p(2). Fitness is taken to be the average number of offspring per parent, parent and offspring being counted at the same age. The random mating load or the expressed load (L_R) is then calculated from the population described as follows:

Genotype	(11)	(12)	(22)
Frequency	$p^2(1)$	$2p(1)p(2)$	$p^2(2)$
Fitness	x(1)	x(2)	x(3)
or	c	c(1-hs)	c(1-s) .

We assume (1) to be dominant over (2), and c is a constant factor common to all genotypes, $0 \leq h \leq 1$, $0 < s \leq 1$. The average fitness of the population is $x = x(1)p^2(1) + 2x(2)p(1)p(2) + x(3)p^2(2) = c[(1-2hsp(1)p(2) - sp^2(2)]$. By definition the load is

$$\frac{x_{max} - x}{x_{max}} ,$$

so that the random mating load in this case is

$$L_R = \frac{x(1)-x}{x(1)} = \frac{c(1-1+2hsp(1)p(2)+sp^2(2))}{c} = 2hsp(1)p(2) + sp^2(2) .$$

For fitnesses as is assumed here, and with small mutation rates, the equilibrium frequencies given in the literature are $p(1) = 1 - \frac{u}{hs}$ and $p(2) = \frac{u}{hs}$, where u is the probability of mutation from (1) to (2). The derivation of these frequencies in a more satisfactory way than is possible with the methodology at hand will be given in Section F of Chapter III. On substituting these equilibrium frequencies in the foregoing equation we get

$$L_R = 2u .$$

In the case where $h = 0$, we have that the equilibrium frequencies are $p(1) = 1 - \sqrt{\frac{u}{s}}$, $p(2) = \sqrt{\frac{u}{s}}$, and we find

$$L_R = u .$$

The foregoing are the values given for loads in the literature, and is a proof of Haldane's (1937) conclusion that the effect of mutation on population fitness depends mainly on the mutation rate and not on the harmfulness of the individual mutant.

To calculate the inbred or total load (in contrast to the random mating or expressed load) we assume that complete homozygosity is achieved without any change in gene frequencies. We get

Genotype	(11)	(22)
Frequency	p(1)	p(2)
Fitness	x(1)	x(3)
or	c	c(1-s)

$$x = x(1)p(1) + x(3)p(2) = c(1-p(2)s) \quad .$$

The inbred load is then

$$L_I = \frac{cp(2)s}{c} = p(2)s \quad .$$

We have then that

$$\frac{L_I}{L_R} = \frac{p(2)s}{2hsp(1)p(2)+sp^2(2)} = \frac{1}{2hp(1)+p(2)}$$

$$= \frac{1}{2h}$$

if $h >> p(2)$, $0 < h \leq 1$, $p(2)$ is small,

$$= \frac{1}{p(2)} \quad \text{if} \quad h \doteq 0 \; .$$

B. The Segregation Load

If the heterozygote is superior, it is convenient to let $x(1) = c(1-s)$, $x(2) = c$, $x(3) = c(1-t)$. Then there will be an equilibrium with $p(1) = \frac{t}{s+t}$ and $p(2) = \frac{s}{s+t}$. The proportion by which the average fitness of the population is reduced, in comparison with a hypothetical population composed

of (12) heterozygotes, is given by $sp^2(1) + tp^2(2)$. If we substitute the equilibrium values of $p(1)$ and $p(2)$ into this, we get for the random mating segregation load

$$L_R = \frac{st}{s+t} \quad .$$

Usually, s and t will be much larger than the mutation rates, so that L_R in the case of the segregation load will be much larger than in the case of the mutation load. This provides the basis for Haldane's (1937) conclusion, which was quoted earlier in this chapter, that the loss of fitness through variation is greater in the case of heterozygote superiority than in the case where variability is maintained by recurrent mutation.

We get the inbred load to be equal to $L_I = sp(1) + tp(2) = \frac{2st}{s+t}$. It follows that $\frac{L_I}{L_R} = 2$, for gene frequencies at equilibrium. Crow (1958) extended this result to the multiple allele case and showed that $\frac{L_I}{L_R} \leq m$, where m is the number of alleles, and where equality holds when the heterozygotes are equal in fitness, and superior to any homozygote. This result should be compared to the ratio $\frac{L_I}{L_R} = \frac{1}{2h}$ in the case of the mutation load.

The only information available on the value of the parameter h is from results of Drosophila experiments. By utilizing some of the experimental techniques which make Drosophila such a popular experimental organism, Hiraizumi and Crow (1960) were able to compare the effect on preadult viability of lethal and semi-lethal second chromosomes in the heterozygous state against controls. These authors found no significant difference between the lethal and semi-lethal classes, but found the viability of the lethal and semi-lethal chromosomes relative to the normal chromosomes to

be 0.9693, or that the selective disadvantage of these chromosomes relative to the normal is 0.0307.

Hiraizumi and Crow found that the proportion of chromosomes in their population containing lethals or semi-lethals was 0.288. They then assume the mutants to be distributed in a Poisson fashion on the chromosomes. This implies that the mean number of mutants among chromosomes with at least one mutant is $\frac{m}{1-e^{-m}}$, where e^{-m} = the proportion of normal chromosomes. It follows that $1-e^{-m} = 0.288$, from which we can solve for m to obtain $\frac{m}{1-e^{-m}} = 1.181$. Assuming the loci to act multiplicatively, the authors get the mean selective disadvantage per locus to be $\frac{.0307}{1.181} = 0.026$. This they consider to be in sufficient agreement with the value of 0.02 - 0.03 calculated indirectly from the data of Muller and Campbell on which Morton, Crow and Muller (1956) based their conclusions. The procedure of Hiraizumi and Crow (1960) is certainly appealing heuristically, but a more rigorous approach to the problem of estimating h is certainly desirable.

From the values of h quoted above one would expect the load ratio to be between 20 and 25 in the case of the mutation load. In the case of the segregation load, the load ratio is expected to be small, since it is believed that the number of cases where many alleles are involved in some system of heterozygote superiority must be small. The argument in support of this contention is that the greater the number of alleles per locus, the greater will the possibility be that some will be lost through sampling if the population is of a finite size. In the case of neutral genes Kimura (1955) proved that the effect of random drift increases approximately proportional to the square of the number of alleles. Unfortunately no work under the assumptions of diploidy and selection is available on this topic.

The existing results on mutation and selection under the assumption of finite population size depend, in the last analysis, on the assumption of haploidy. On the basis of this work Kimura and Crow (1964) argue that a population structure with many genes at which a segregation load exists is unlikely.

If we now accept the theoretical evidence on the likely size of the load ratios under the two hypotheses as reasonably compelling, we may try to estimate $\frac{L_I}{L_R}$ for some population, and if it were sufficiently large, conclude that the load was mainly mutational in the population under consideration.

However, before the foregoing theory could be applied, further extensions were necessary. First, it was extended (Morton, Crow and Muller, 1956) without change to many loci acting together on the assumption that their effects are independent, or not "synergistic." Second, L_I , the load for complete homozygosis, corresponds to an inbreeding coefficient, $F = 1$, and normally cannot be observed. However, it could be estimated from the load in individuals resulting from the mating of close relatives, which happens by chance in a finite population with small effective population size, or which can be induced at will in laboratory populations.

We now have a procedure which can shed some light on the genetic architecture of a population. If the estimated load ratio is high in magnitude we conclude that the genetic heterogeneity is maintained by a balance beteen mutation and selection. On the other hand, a load ratio of small magnitude would suggest that the genetic heterogeneity in a population is mainly the result of a balance between selective forces.

Levene (1963) noted a number of difficulties involved in applying the genetic load method to questions regarding the genetic architecture of a population. First, the studies in which the load methodology have been applied were all concerned with components of fitness instead of fitness itself. Second, studies involving all components of fitness over an entire generation in a population in genetic equilibrium are extraodinarily difficult, but even if they could be carried out, another difficulty would arise. One would need an estimate of the fitness of the maximal genotype in a population. With many loci it will be difficult to find an individual homozygous or heterozygous at all or nearly all loci.

Haldane and Jayakar (1965) point out that with respect to fitness there may be an overdominant relationship between the genotypes, but that other relationships may hold with respect to the components of fitness. An analysis based on the components of fitness will then give a false picture as to the nature of the loads.

The problems caused by the foregoing and other possible difficulties will become more intelligible after the relevant theory has been developed. An evaluation of experimental applications of the load theory is, therefore, best deferred until a later chapter.

Kimura, Maruyama and Crow (1963) and Kimura and Crow (1964) also did some theoretical work on loads in small populations. This work must, however, be regarded as of questionable value, since the gene frequency distribution theory on which it is based can be assumed to be applicable in the case of haploids only. Moran (1962, p. 102) has some instructive remarks about the applicability of haploid models in the case of diploids which may be pertinent here.

Li (1967, and other references contained therein) discusses some misinterpretations associated with the load concept. Often the idea of a load is associated with the implication that selection acts only through differential mortality; the image is one of a population carrying a burden of so many genetic deaths. The point is, of course, that selection may also be due to differences in fertility. Let us consider a population in Hardy-Weinberg equilibrium consisting of genotypes AA, Aa and aa with relative fitnesses 2:3:1. If the allele a does not exist, the population will consist of AA genotypes only and there will be no genetic selection and consequently, no load. Now, if there are two alleles and three genotypes with relative fitnesses 2:3:1, the average fitness of a stationary population is higher than before (for a general proof see Li, 1967), because of the higher fitness of Aa, in spite of the lower fitness of aa . Thus, we see that the situation may be equally arbitrarily described as a gain for the population rather than a load.

It is perhaps necessary to emphasize that the foregoing remarks pertain solely to the use of the load concept in the discussion of the biological aspects of adaptedness of populations. If the load theory is valid, the estimates of the load ratios will give us the information about the genetic architecture of a population which they were designed to reveal.

It is also necessary to note that one should be careful when comparing the magnitudes of loads in different populations. We have defined the load in a population to be

$$L = \frac{x_{max} - x}{x_{max}} ,$$

where x_{max} is the fittest genotype and $\bar{x}$ the mean of the population. This gives us $\bar{x} = x_{max}(1-L)$. We see, therefore, that a comparison between loads will be equivalent to a comparison between mean fitnesses only when the maximal genotypes are the same in all populations.

III. EQUILIBRIA IN DIPLOID RANDOM MATING POPULATIONS

A. A General Description of the Model

A model should have the virtue of clarity in the statement of its basic assumptions and allow for further development to take place without any concealed features. As there are ambiguities involved with the overlapping generations model, we will assume non-overlapping generations. The problem is that Fisher's (1930), measure of "fitness", the Malthusian parameter, is ill-defined for the case of a genetically segregating population. Alternative formulations of the overlapping generations model are those of Haldane (1927a) and Norton (1928). However, this work is obscure with regard to assumptions, and is furthermore not sufficiently general for an examination of the problem at hand.

To achieve sufficient clarity in our model, we consider a population in which the members have two life phases, infancy and adulthood. The probability of survival of an individual from infancy to adulthood depends only on its genotype. Mating takes place at random among adults. The fecundity of a mating is described by a discrete probability distribution, giving the probabilities associated with $0,1, \ldots ,q$ offspring. To ensure the possibility of Hardy-Weinberg frequencies we assume that the expected number of offspring of a mating pair is the product of two means, one corresponding to each parent. These depend solely on the genotype of a parent and not on sex. The viability and fecundity probabilities associated with pairs of individuals depend only on their genotypes and are therefore independent of the genotypic composition of a population.

B. The Panmictic Structure of Gamete Frequencies

1. The panmictic structure of gamete frequencies in the two-loci case

The two-loci case exhibits most of the complexities of the n-loci case without causing the notation to become too cumbersome. It also allows easy specialization to the single locus case. We therefore give the two-loci derivations in detail, with the other cases following as straightforward extensions.

Assume that the members of the population are counted at two life phases, designated as infant and adult. Suppose we have two loci with alleles $(i^1_{x_1})$, $i^1_{x_1} = 1,2, \ldots, m^1$ at the first locus and alleles $(i^2_{x_2})$, $i^2_{x_2} = 1,2, \ldots, m^2$ at the second locus, where $x_1 = 0$ or 1 and $x_2 = 0$ or 1. Let us denote an arbitrary genotype as $(i^1_0 i^1_1 i^2_0 i^2_1)$, where it is assumed that $(i^1_0 i^2_0)$ is on one chromosome and $(i^1_1 i^2_1)$ is on the other chromosome. Let us also suppose that gametes are produced in the following way:

$(i^1_0 i^2_0)$ with frequency γ_{00},

$(i^1_0 i^2_1)$ with frequency γ_{01},

$(i^1_1 i^2_0)$ with frequency γ_{10},

and

$(i^1_1 i^2_1)$ with frequency γ_{11}.

We have that $\gamma_{00} + \gamma_{01} + \gamma_{10} + \gamma_{11} = 1$ and assume that $\gamma_{00} = \gamma_{11}$ and $\gamma_{01} = \gamma_{10}$.

Let us assume that the probability that a pair of genes $(i^1_{x_1} i^2_{x_2})$ mutates to the pair $(i^1_{y_1} i^2_{y_2})$ is $v(i^1_{x_1} i^1_{y_1} i^2_{x_2} i^2_{y_2})$. In general we will have that

$$v(i^1_{x_1} i^1_{y_1} i^2_{x_2} i^2_{y_2}) \neq v(i^1_{x_1} i^1_{y_1} i^2_{y_2} i^2_{x_2}) \neq v(i^1_{y_1} i^1_{x_1} i^2_{x_2} i^2_{y_2}) \neq v(i^1_{y_1} i^1_{x_1} i^2_{y_2} i^2_{x_2}) ,$$

for $x_1, x_2, y_1, y_2 = 0$ or 1. We define

$$v(i^1_{x_1} i^1_{x_1} i^2_{x_2} i^2_{x_2}) = 1 - \sum_{\substack{i^1_{y_1} \neq i^1_{x_1} \\ \text{or} \\ i^2_{y_2} \neq i^2_{x_2}}} v(i^1_{x_1} i^1_{y_1} i^2_{x_2} i^2_{y_2}) ,$$

so that it follows that

$$\sum_{i^1_{y_1} i^2_{y_2}} v(i^1_{x_1} i^1_{y_1} i^2_{x_2} i^2_{y_2}) = 1 .$$

If the alleles at different loci mutate independently from each other, we can write that $v(i^1_{x_1} i^1_{y_1} i^2_{x_2} i^2_{y_2}) = v(i^1_{x_1} i^1_{y_1}) v(i^2_{x_2} i^2_{y_2})$. In this case we define $v(i^a_{x_a} i^a_{x_a})$ for $a = 1,2$, to be equal to

$$1 - \sum_{i^a_{y_a} \neq i^a_{x_a}} v(i^a_{x_a} i^a_{y_a}) ,$$

from which it follows that $\sum_{i^a_{y_a}} v(i^a_{x_a} i^a_{y_a}) = 1$.

Sometimes it may be necessary to write $v^{12}(i^1_{x_1} i^1_{y_1} i^2_{x_2} i^2_{y_2})$ instead of $v(i^1_{x_1} i^1_{y_1} i^2_{x_2} i^2_{y_2})$ in order to indicate clearly that a mutation probability refers to both the first and the second locus. In general we will prefer to delete the superscripts in order to ease the notational problem. These

remarks are applicable to all symbols employed in this study.

Extending the work of Kempthorne and Pollak (1966), we take selection into account in the fashion listed below.

(a) Suppose that individuals have survival from infancy to adulthood based on genotype, and that the probability that an infant of genotype $(i_0^1 i_1^1 i_0^2 i_1^2)$ survives to adulthood is $\ell(i_0^1 i_1^1 i_0^2 i_1^2)$.

(b) Suppose mating takes place at random among the adults, and that the probability that a mating of genetic type $(i_0^1 i_1^1 i_0^2 i_1^2)$ with $(j_0^1 j_1^1 j_0^2 j_1^2)$ gives t infants is $b_t(i_0^1 i_1^1 i_0^2 i_1^2 j_0^1 j_1^1 j_0^2 j_1^2)$. We observe that fertility is solely determined by the parental genotypes without any interaction between genotype of offspring and genotype of parent. The fertility probabilities are defined in such a way that $\sum_{t=0}^{q} b_t(i_0^1 i_1^1 i_0^2 i_1^2 j_0^1 j_1^1 j_0^2 j_1^2) = 1$. It follows that the mean number of offspring at infancy due to the mating between the adults $(i_0^1 i_1^1 i_0^2 i_1^2)$ and $(j_0^1 j_1^1 j_0^2 j_1^2)$ is

$$E(t) = \sum_{t=0}^{q} t b_t(i_0^1 i_1^1 i_0^2 i_1^2 j_0^1 j_1^1 j_0^2 j_1^2) = b(i_0^1 i_1^1 i_0^2 i_1^2 j_0^1 j_1^1 j_0^2 j_1^2) \quad .$$

In order to simplify our further derivations it is necessary to assume

$$b(i_0^1 i_1^1 i_0^2 i_1^2 j_0^1 j_1^1 j_0^2 j_1^2) = b(i_0^1 i_1^1 i_0^2 i_1^2) b(j_0^1 j_1^1 j_0^2 j_1^2) \quad .$$

We will refer to this assumption by the term product fecundity.

(c) Assume that with the restriction on fecundity of a mating, and ignoring mutation, the probability distribution of the genotype of the offspring is exactly that from ordinary Mendelian segregation at each locus considered separately. The usual generalization of Mendelian segregation to two loci considered jointly is also assumed.

(d) We assume that generations do not overlap.

Suppose that in a generation of infants the genotypic array is

$$\sum_{j_0^1 j_1^1 j_0^2 j_1^2} p(j_0^1 j_1^1 j_0^2 j_1^2)(j_0^1 j_1^1 j_0^2 j_1^2) \quad .$$

Then the survival probabilities lead to the adult population

$$\frac{\sum_{j_0^1 j_1^1 j_0^2 j_1^2} p(j_0^1 j_1^1 j_0^2 j_1^2)\ell(j_0^1 j_1^1 j_0^2 j_1^2)(j_0^1 j_1^1 j_0^2 j_1^2)}{\sum_{j_0^1 j_1^1 j_0^2 j_1^2} p(j_0^1 j_1^1 j_0^2 j_1^2)\ell(j_0^1 j_1^1 j_0^2 j_1^2)} \quad .$$

This population now mates at random. The frequency of the mating

$$(r_0^1 r_1^1 r_0^2 r_1^2)x(s_0^1 s_1^1 s_0^2 s_1^2)$$

is proportional to

$$p(r_0^1 r_1^1 r_0^2 r_1^2)\ell(r_0^1 r_1^1 r_0^2 r_1^2)p(s_0^1 s_1^1 s_0^2 s_1^2)\ell(s_0^1 s_1^1 s_0^2 s_1^2)$$

for all possible values of r_0^1, r_1^1, r_0^2, r_1^2, s_0^1, s_1^1, s_0^2, and s_1^2. The conditional probability of t offspring is

$$b_t(r_0^1 r_1^1 r_0^2 r_1^2 s_0^1 s_1^1 s_0^2 s_1^2) \quad .$$

To get the conditional probability arrays of the different genetic types of offspring, it is necessary first to get the gametic types given by each of the genotypes when mutation is taken into account. The genotype $(r_0^1 r_1^1 r_0^2 r_1^2)$ will give gametes of the type $(i_0^1 i_0^2)$ with frequencies $v(r_0^1 i_0^1 r_0^2 i_0^2)\gamma_{00}$, $v(r_0^1 i_0^1 r_1^2 i_0^2)\gamma_{01}$, $v(r_1^1 i_0^1 r_0^2 i_0^2)\gamma_{10}$, and $v(r_1^1 i_0^1 r_1^2 i_0^2)\gamma_{11}$.

Similarly the genotype $(s_0^1 s_1^1 s_0^2 s_1^2)$ will produce gametes of the type $(i_1^1 i_1^2)$ with frequencies $v(s_0^1 i_1^1 s_0^2 i_1^2)\gamma_{00}$, $v(s_0^1 i_1^1 s_1^2 i_1^2)\gamma_{01}$, $v(s_1^1 i_1^1 s_0^2 i_1^2)\gamma_{10}$, and $v(s_1^1 i_1^1 s_1^2 i_1^2)\gamma_{11}$. It is clear that we assume here that mutation occurs independently from crossing over between chromosomes. It is also apparent that we assume mutation to occur independently from the process of selection, and that it can occur at any moment between one fertilization and the next. It follows that the conditional probability array of the different genetic types of offspring from the mating of $(r_0^1 r_1^1 r_0^2 r_1^2)$ with $(s_0^1 s_1^1 s_0^2 s_1^2)$ is

$$\Big(\sum_{x_1,x_2=0,1} v(r_{x_1}^1 i_0^1 r_{x_2}^2 i_0^2)\gamma_{x_1x_2}(i_0^1 i_0^2)\Big) X \Big(\sum_{x_1,x_2=0,1} v(s_{x_1}^1 i_1^1 s_{x_2}^1 i_1^2)\gamma_{x_1x_2}(i_1^1 i_1^2)\Big).$$

Let us now denote $Q(r_0^1 r_1^1 r_0^2 r_1^2) = p(r_0^1 r_1^1 r_0^2 r_1^2)\ell(r_0^1 r_1^1 r_0^2 r_1^2)$, so that it follows that the resulting infant population has the probability array

$$\sum_{t=0}^{q} \frac{t}{D} \sum_{\substack{i_0^1 i_1^1 \\ i_0^2 i_1^2}} \Bigg[\sum_{\substack{r_0^1 r_1^1 s_0^1 s_1^1 \\ r_0^2 r_1^2 s_0^2 s_1^2}} \Big(\sum_{x_1,x_2=0,1} v(r_{x_1}^1 i_0 r_{x_2}^2 i_0^2)\gamma_{x_1x_2} \Big)$$

$$X \Big(\sum_{x_1,x_2=0,1} v(s_{x_1}^1 i_1^1 s_{x_2}^2 i_1^2)\gamma_{x_1x_2} \Big) Q(r_0^1 r_1^1 r_0^2 r_1^2) Q(s_0^1 s_1^1 s_0^2 s_1^2) \qquad \text{(B.1)}$$

$$X\, b_t(r_0^1 r_1^1 r_0^2 r_1^2 s_0^1 s_1^1 s_0^2 s_1^2) \Bigg] (i_0^1 i_0^1 i_0^2 i_1^2),$$

where

$$D = \sum_{\substack{i_0^1 i_1^1 \\ i_0^2 i_1^2}} \Bigg[\sum_{\substack{r_0^1 r_0^1 s_0^1 s_1^1 \\ r_0^2 r_1^2 s_0^2 s_1^2}} \Big(\sum_{x_1,x_2=0,1} v(r_{x_1}^1 i_0^1 r_{x_2}^2 i_0^2)\gamma_{x_1x_2} \Big) X$$

$$\left(\sum_{x_1,x_2 = 0,1} v(s^1_{x_1} i^1_1 s^2_{x_2} i^2_1)\gamma_{x_1x_2} \right) Q(r^1_0 r^1_1 r^2_0 r^2_1) Q(s^1_0 s^1_1 s^2_0 s^2_1) \bar{x}$$

$$\left. \sum_t t b_t(r^1_0 r^1_1 r^2_0 r^2_1 s^1_0 s^1_1 s^2_0 s^2_1) \right] .$$

Now recall that we assumed

$$\sum_t t b_t(r^1_0 r^1_1 r^2_0 r^2_1 s^1_0 s^1_1 s^2_0 s^2_1) = b(r^1_0 r^1_1 r^2_0 r^2_1) b(s^1_0 s^1_1 s^2_0 s^2_1) .$$

We also find it convenient to write

$$W(r^1_0 r^1_1 r^2_0 r^2_1) = Q(r^1_0 r^1_1 r^2_0 r^2_1) b(r^1_0 r^1_1 r^2_0 r^2_1) .$$

From the definition of the mutation probabilities we have that

$$\sum_{i^1_0 i^2_0} v(r^1_{x_1} i^1_0 r^2_{x_2} i^2_0) = 1 ,$$

and hence it is easy to see that the denominator of Equation B.1 is equal to $W^2\ldots$, where we follow the usual convention to indicate with a dot that a variable has been summed over. It follows that the probability array of the infants can be written as

$$\sum_{\substack{i^1_0 i^1_1 \\ i^2_0 i^2_1}} \left[\sum_{r^1_0 r^1_1 r^2_0 r^2_1} \frac{\left(\sum_{x_1,x_2 = 0,1} v(r^1_{x_1} i^1_0 r^2_{x_2} i^2_0)\gamma_{x_1x_2} \right) W(r^1_0 r^1_1 r^2_0 r^2_1)}{W\ldots} \right] \tag{B.2}$$

$$X \left[\sum_{s^1_0 s^1_1 s^2_0 s^2_1} \frac{\left(\sum_{x_1,x_2 = 0,1} v(s^1_{x_1} i^1_1 s^2_{x_2} i^2_1)\gamma_{x_1x_2} \right) W(s^1_0 s^1_1 s^2_0 s^2_1)}{W\ldots} \right] (i^1_0 i^1_1 i^2_0 i^2_1) ,$$

which is of the form

$$\sum_{i_0^1 i_0^2} \sum_{i_1^1 i_1^2} p(i_0^1 i_0^2)p(i_1^1 i_1^2)(i_0^1 i_1^1 i_0^2 i_1^2) \quad .$$

Hence, under the present model the infant population has what we will define as being the panmictic structure of gamete frequencies. We may, therefore, take the numbers $p(i_0^1 i_1^1 i_0^2 i_1^2)$ to be of the form $p(i_0^1 i_0^2)p(i_1^1 i_1^2)$ for every generation after the first.

Before we pass on to other matters it is of importance to note that the assumption of product fecundity, $b(i_0^1 i_1^1 i_0^2 i_1^2 j_0^1 j_1^1 j_0^2 j_1^2) = b(i_0^1 i_1^1 i_0^2 i_1^2)$ X $b(j_0^1 j_1^1 j_0^2 j_1^2)$, implies symmetry in the way in which the genotypes enter the b-function that refers to a mating pair. It is therefore clear that with sex differences there can not be product fecundity in the sense which we use the term in this study.

The assumption of product fecundity merely implies in mathematical terms that a conditional expectation can be written as a product of two functions. An obvious, but certainly not the only, way in which this can happen will now be given here. Let us denote the probability of an adult $(i_0^1 i_1^1 i_0^2 i_1^2)$ contributing r offspring to a mating by $b_r(i_0^1 i_1^1 i_0^2 i_1^2)$. We also assume that all these probabilities add to one, and we denote $\sum_r r b_r(i_0^1 i_1^1 i_0^2 i_1^2)$ by $b(i_0^1 i_1^1 i_0^2 i_1^2)$. We now assume the probability of t infant offspring from the adult mating pair $(i_0^1 i_1^1 i_0^2 i_1^2)$ X $(j_0^1 j_1^1 j_0^2 j_1^2)$ to be given by

$$b_t(i_0^1 i_1^1 i_0^2 i_1^2 j_0^1 j_1^1 j_0^2 j_1^2) = \sum_{\text{all } rs=t} b_r(i_0^1 i_1^1 i_0^2 i_1^2) b_s(j_0^1 j_1^1 j_0^2 j_1^2) \quad ,$$

so that we can write

$$b(i_0^1 i_1^1 i_0^2 i_1^2 j_0^1 j_1^1 j_0^2 j_1^2) = \sum_t t b_t(i_0^1 i_1^1 i_0^2 i_1^2 j_0^1 j_1^1 j_0^2 j_1^2)$$

$$= \sum_t t \sum_{\text{all } rs=t} b_r(i_0^1 i_1^1 i_0^2 i_1^2) b_s(j_0^1 j_1^1 j_0^2 j_1^2)$$

$$= \sum_{rs} rs\, b_s(i_0^1 i_1^1 i_0^2 i_1^2) b_s(j_0^1 j_1^1 j_0^2 j_1^2)$$

$$= b(i_0^1 i_1^1 i_0^2 i_1^2) b(j_0^1 j_1^1 j_0^2 j_1^2) \quad .$$

In this context $\sum_t$ means summation over all values that the integer t assumes. These values are not necessarily the consecutive integers as can easily be seen by displaying the case where $r = 1,2$ and $s = 1,2$. Another way of dealing with this problem in notation would, of course, simply be to define $b_t(i_0^1 i_1^1 i_0^2 i_1^2 j_0^1 j_1^1 j_0^2 j_1^2)$ to be equal to zero for $t \neq rs$, in which case the summation $\sum_t$ can go over consecutive integers.

2. The panmictic structure of gamete frequencies in the n-loci case

It is clear that the method of proof employed in the 2-loci case is completely general and can, except for notational difficulties, easily be extended to the case of n loci. By the choice of the two-loci case the use of too cumbersome a notation was avoided. A full discussion on the notation used to accommodate linkage can be found in van Aarde (1963) or Schnell (1961).

3. The change in gamete frequencies

We have from Equation B.2 that

$$W\ldots\ p'(i_0^1 i_0^2) = \sum_{\substack{r_0^1 r_1^1 \\ r_0^2 r_1^2}} \left[\sum_{x_1, x_2 = 0,1} v(r_{x_1}^1 i_0^1 r_{x_2}^2 i_0^2) \gamma_{x_1 x_2} \right] W(r_0^1 r_1^1 r_0^2 r_1^2). \quad \text{(B.3)}$$

The extension of the Equations B.3 to the case of n-loci becomes now immediately obvious. We get

$$W_{\ldots\ldots}p'(i_0^1 i_0^2 \ldots i_0^n) = \sum_{r_0^1 r_1^1 r_0^2 r_1^2 \ldots r_0^n r_1^n} \Big[\sum_{x_1, x_2, \ldots, x_n = 0,1} v(r_{x_1}^1 i_0^1 r_{x_2}^2 i_0^2 \ldots r_{x_n}^n i_0^n) \times \gamma_{x_1 x_2 \ldots x_n} \Big] W(r_0^1 r_1^1 r_0^2 r_1^2 \ldots r_0^n r_1^n) \qquad \text{(B.4)}$$

where the quantities involved are such obvious extensions from the two-loci case that they require no comment. The case of no mutation is accommodated by defining

$$v(r_{x_1}^1 i_0^1 r_{x_2}^2 i_0^2 \ldots r_{x_n}^n i_0^n) = 1 \quad \text{if} \quad r_{x_1}^1 = i_0^1 \quad \text{and} \quad r_{x_2}^2 = i_0^2 \quad \text{and} \ldots\ldots \text{and}$$

$$r_{x_n}^n = i_0^n$$

$$= 0 \quad \text{otherwise.}$$

Unfortunately, there does not seem to be a general way by which the Equations B.4 can be simplified when the restriction of no mutation is taken into account. To see what sort of simplification can be accomplished, let us write the two-loci case out in detail. We denote

$$\ell(r_0^1 r_1^1 r_0^2 r_1^2) b(r_0^1 r_1^1 r_0^2 r_1^2) = a(r_0^1 r_1^1 r_0^2 r_1^2) .$$

It follows that $W_{\ldots\ldots} = \sum_{\substack{r_0^1 r_0^2 \\ r_1^1 r_1^2}} p(r_0^1 r_0^2) a(r_0^1 r_1^1 r_0^2 r_1^2) p(r_1^1 r_1^2)$.

From the Equations B.3 we have that

$$W_{\ldots\ldots} p'(i_0^1 i_0^2) = \sum_{\substack{r_0^1 r_1^1 \\ r_0^2 r_1^2}} [v(r_0^1 i_0^1 r_0^2 i_0^2)\gamma_{00} + v(r_0^1 i_0^1 r_1^2 i_0^2)\gamma_{01} + v(r_1^1 i_0^1 r_0^2 i_0^2)\gamma_{10} + v(r_1^1 i_0^1 r_1^2 i_0^2)\gamma_{11}] W(r_0^1 r_1^1 r_0^2 r_1^2) ,$$

$$W....p'(i_0^1 i_0^2) = \gamma_{00} \sum_{r_0^1 r_0^2} v(r_0^1 i_0^1 r_0^2 i_0^2) p(r_0^1 r_0^2) \sum_{r_1^1 r_1^2} p(r_1^1 r_1^2) a(r_0^1 r_1^1 r_0^2 r_1^2)$$

$$+ \gamma_{01} \sum_{r_0^1 r_1^2} v(r_0^1 i_0^1 r_1^2 i_0^2) \sum_{r_1^1 r_0^2} p(r_0^1 r_0^2) a(r_0^1 r_1^1 r_0^2 r_1^2) p(r_1^1 r_1^2)$$

$$+ \gamma_{10} \sum_{r_1^1 r_0^2} v(r_1^1 i_0^1 r_0^2 i_0^2) \sum_{r_0^1 r_1^2} p(r_0^1 r_0^2) a(r_0^1 r_1^1 r_0^2 r_1^2) p(r_1^1 r_1^2)$$

$$+ \gamma_{11} \sum_{r_1^1 r_1^2} v(r_1^1 i_0^1 r_1^2 i_0^2) p(r_1^1 r_1^2) \sum_{r_0^1 r_0^2} p(r_0^1 r_0^2) a(r_0^1 r_1^1 r_0^2 r_1^2) . \qquad (B.5)$$

From the symmetry of the situation we have that $a(i_0^1 i_1^1 i_0^2 i_1^2)$ $= a(i_1^1 i_0^1 i_1^2 i_0^2)$. By our initial assumptions we also have that $\gamma_{00} = \gamma_{11}$ and $\gamma_{01} = \gamma_{10}$. Hence it follows from the Equations B.5 that

$$W....p'(i_0^1 i_0^2) = 2\gamma_{00} \sum_{r_0^1 r_0^2} v(r_0^1 i_0^1 r_0^2 i_0^2) p(r_0^1 r_0^2) \sum_{r_1^1 r_1^2} p(r_1^1 r_1^2) a(r_0^1 r_1^1 r_0^2 r_1^2)$$

$$+ 2\gamma_{01} \sum_{r_0^1 r_1^2} v(r_0^1 i_0^1 r_1^2 i_0^2) \sum_{r_1^1 r_0^2} p(r_0^1 r_0^2) a(r_0^1 r_1^1 r_0^2 r_1^2) p(r_1^1 r_1^2) . \qquad (B.6)$$

To ease the notational problem, and because the previous notation served its main purpose in the derivation of the foregoing equations, we write in the Equations B.6, $i_0^1 = i^1$, $i_0^2 = i^2$, $r_0^1 = j^1$, $r_0^2 = j^2$, $r_1^1 = k^1$, and $r_1^2 = k^2$ to get

$$W....p'(i^1 i^2) = 2\gamma_{00} \sum_{j^1 j^2} v(j^1 i^1 j^2 i^2) p(j^1 j^2) \sum_{k^1 k^2} p(k^1 k^2) a(j^1 k^1 j^2 k^2) \qquad (B.7)$$

$$+ 2\gamma_{01} \sum_{j^1 k^2} v(j^1 i^1 k^2 i^2) \sum_{k^1 j^2} p(j^1 j^2) a(j^1 k^1 j^2 k^2) p(k^1 k^2) .$$

We note that the order of the subscripts is still important, and that, thus far, we only assumed that $a(j^1k^1j^2k^2) = a(k^1j^1k^2j^2)$.

In the case of no mutation we have that $v(j^1i^1j^2i^2) = 1$ if $j^1 = i^1$ and $j^2 = i^2$, so that it follows from the Equations B.7 that

$$W....p'(i^1i^2) = 2\gamma_{00}p(i^1i^2)\sum_{k^1k^2}p(k^1k^2)a(i^1k^1i^2k^2) + 2\gamma_{01}\sum_{k^1k^2}p(i^1k^2)a(i^1k^1k^2i^2)p(k^1i^2) \quad . \tag{B.8}$$

We note that it was unnecessary in the foregoing derivations to assume the absence of any sort of position or cis-trans effect. If no position effect is assumed we write

$$a(i^1k^1i^2k^2) = a(i^1k^1k^2i^2) = a(k^1i^1i^2k^2) \quad .$$

The Equations B.8 correspond for the case of $i^1 = 1,2$ and $i^2 = 1,2$ to formulas first derived by Lewontin and Kojima (1960).

4. Hardy-Weinberg structure in the single locus case

The simplicity of the single locus case derives from the absence of cross-overs and from the coincidence of gene and gamete frequencies. The single locus case can be derived from the two-loci case by the assumption of no recombination between loci. We assume that the gamete or chromosome denoted by $(i_0^1i_0^2)$ becomes the member (i) of a single locus allelic series with $m^1m^2 = m$ allelic forms. To this purpose we construct a one-to-one correspondence between the pair of numbers (i_0^1, i_0^2) and the number denoted by (i), where $i_0^1 = 1, \ldots, m^1$, $i_0^2 = 1, \ldots, m^2$ and where $(i) = 1,2, \ldots, m$, where $m = m^1m^2$. Likewise we set up the following one-to-one correspondences (i_1^1, i_1^2) and (j) , (r_0^1, r_0^2) and (r) , (r_1^1, r_1^2)

and (s), (s_0^1, s_0^2) and (t) and (s_1^1, s_1^2) and (u). We write $v(i_0^1 r_0^1 i_0^2 r_0^2) = v(ir)$, where $\sum_r v(ir) = 1$ and where $v(ii) = 1 - \sum_{j \neq i} v(ij)$. Since we assume no crossing-over to take place we have $\gamma_{00} = \gamma_{11} = \frac{1}{2}$ and $\gamma_{01} = \gamma_{10} = 0$. Equation B.2 then becomes

$$\frac{\sum_{ij} \sum_{rstu} \left(\frac{v(ri)+v(si)}{2}\right) \left(\frac{v(tj)+v(uj)}{2}\right) W(rs)W(ut)(ij)}{\sum_{ij} \sum_{rstu} \left(\frac{v(ri)+v(si)}{2}\right)\left(\frac{v(tj)+v(uj)}{2}\right) W(rs)W(ut)}, \qquad (B.9)$$

where $W(rs) = p(rs)\ell(rs)b(rs)$, so that $W(rs) = W(sr)$ since $p(rs) = p(sr)$, and since the same property holds for ℓ and b. Again we use the notation that $\sum_s W(rs) = W(r.) = W(.r)$ and $\sum_r W(r.) = W..$. Hence, Equation B.9 becomes on simplification

$$\frac{\sum_{ij} \sum_r v(ri)W(r.) \sum_t v(tj)W(t.)(ij)}{W_{..}^2} \qquad (B.10)$$

which is of the form $\sum_{ij} p(i)p(j)(ij)$. Thus, under the present model the resulting infant population has Hardy-Weinberg structure.

We denote now $\ell(ij)b(ij) = a(ij)$, and let $p(i)$ be the frequency of (i) in a given generation and $p'(i)$ the frequency in the succeeding generation. Hence, it follows from Equation B.10 that we can write for the change in gene frequency

$$p'(i) = \frac{\sum_r v(ri)p(r) \sum_s a(rs)p(s)}{\sum\sum_{tu} p(t)a(tu)p(u)} . \qquad (B.11)$$

In the case of no mutation we have, as before, $v(ij) = 1$ if $i = j$ and $v(ij) = 0$ otherwise, so that the Equations B.11 become

$$p'(i) = \frac{p(i) \sum_j a(ij)p(j)}{\sum\sum_{tu} p(t)a(tu)p(u)} . \qquad (B.12)$$

5. Linkage and genetic fine structure

Since the late 1940's it has been clear that the old image of genes being like "beads on a rosary" on a chromosome no longer holds. In the place of the "classical" gene, Benzer (1962) proposed respectively mutons, recons and cistrons as units of mutation, recombination by crossing-over and of the phenotype as determined by cis-trans tests. A unit of biological function is still sometimes referred to as a locus, which then apparently will include one or more cistrons. Griffing (1960) proposed two ways by which the knowledge of genetic fine structure can be dealt with in population genetics. Let us consider the genetic situation at a complex locus, which has two genetic conditions (mutant and normal) at each of two mutational sites. In the first method, the locus can be subdivided into two subloci, one for each of the mutational sites. This approach yields two sets of alleles, each set being the genetic alternatives at each sublocus. In this case, the gene model for quantitative inheritance must accommodate a position effect which may occur between alleles at different subloci. From the Equations B.8 we can see that it is only when recombination between the two subloci takes place that the consideration of a possible position effect becomes important, for with $\gamma_{01} = 0$ (no recombination) the first part of the equations are equivalent to the Equations B.12, as follows from the discussion in Section 3 on the Equations B.4.

The alternative method is to consider the overall locus as the basic entity, and to regard all possible genetic structures at this locus as the set of multiple alleles. A resultant complication of this approach is that mutation, in this case, will have to include both point mutation in the

conventional sense and intralocus recombination. It is clear that the overall locus will have a Hardy-Weinberg structure of "allele" frequencies, but that with any significant degree of recombination, the change in "allele" frequency would be given incorrectly. It will be obvious from the later discussions on equilibria that a pooling of the effects of mutation and recombination may have far from innocuous results in many cases.

It is clear that the mutation probabilities can be defined to cover any number of mutation sites, and that a mutation probability can therefore be defined for any recon to mutate to any other recon. It follows then by the foregoing discussion that the recon (to use Benzer's terminology) must be the fundamental unit of inheritance for the discussion of the type of selection considered in this work. The conventional term locus as it is commonly used in this chapter must therefore be considered equivalent to a recon in Benzer's terminology. Modifications in the approach to this matter, which is of importance in the assumptions necessary for the theory of genetic loads, will be discussed in Chapter V, Section A.4.b.

C. Hardy-Weinberg Equilibria of Gene Frequencies in the Single Locus Case with Selection and no Mutation

We have from our previous work (the Equations B.12) that the formula for change in gene frequency is

$$Vp'(i) = p(i) \sum_j a(ij)p(j) \qquad (C.1)$$

where $p'(i)$ and $p(i)$ are the frequencies of the allele (i) in two successive generations, and where we write $V = \sum_t \sum_u p(t)a(tu)p(u)$ in order to make our notation consonant with that used by Kingman (1961a).

As was mentioned before, we see that by defining $\ell(ij)b(ij)$ as the fitness of the individual (ij) we derive formulas equivalent to those usually given for populations in which the genotype (ij) has "viability" or "fitness" $a(ij)$. It also follows from the way in which they are defined that the $a(ij)$'s are symmetric.

Scheuer and Mandel (1959), Mulholland and Smith (1959), Atkinson, Watterson and Moran (1960), and Kingman (1961b) have been able to establish the inequality

$$V' = \sum_{ij} p'(i)p'(j)a(ij) \geq \sum_{ij} p(i)p(j)a(ij) = V \ , \qquad (C.2)$$

i.e., the mean fitness increases from generation to generation. Using this result, Mulholland and Smith (1959) have proved that the gene frequencies $p(i)$ tend to a limit after a large number of generations, and that this limit is a local maximum of V as a function of the $p(i)$. Discussions of the stability of non-trivial equilibria have in general been based on this property of V as a function of the $p(i)$.

Moran (1964) has shown that the property of increase in the mean fitness does not hold in the multi-loci case. For the purposes of later generalization to the multi-loci case, we therefore follow Kingman (1961a) in a discussion of stability of gene frequencies based directly on the Equations C.1. Since we will need many of the results later on, we will give the derivation of Kingman's results in detail.

We call $p(i)(i=1, \ldots, m)$ an equilibrium if $p'(i) = p(i)$. From the Equations C.1 this means that, for each i, either $p(i) = 0$, or $V = \sum_i a(ij)p(j)$. We assume $p(i) \neq 0$, and write the equilibrium as $p(i) = P(i)$. We therefore inquire if there exists a nontrivial equilib-

rium satisfying

$$P(i) > 0 \ , \ \sum_i P(i) = 1$$

$$\sum_j a(ij)P(j) = V_E \ , \qquad (C.3)$$

where the subscript "E" denotes "equilibrium value." The symbol for the equilibrium value of the quadratic form $\sum_{ij} p(i)a(ij)p(j)$, V_E , should throughout this study be distinguished from the quadratic form $\sum_{ij} e(i)a(ij)e(j)$, which we will denote by $V(e)$ or $V(e(i))$, wherever it will be necessary to do so.

Let us assume the non-trivial equilibrium $P(i)$ exists and put

$$p(i) = P(i) + e(i) \ , \ p'(i) = P(i) + e'(i) \ .$$

Then using the Equations C.1 and ignoring second order terms in $e(i)$, we obtain

$$e'(i) = e(i) + \frac{P(i)}{V_E} \sum_j a(ij)e(j) \ . \qquad (C.4)$$

Let

$$y(i) = \frac{e(i)}{\sqrt{P(i)}} \quad \text{and} \quad b(ij) = a(ij)\sqrt{P(i)P(j)}\,/V_E \ .$$

Then the Equations C.4 become

$$y'(i) = y(i) + \sum_j b(ij)y(j) \qquad (C.5)$$

and the condition $\Sigma\, e(i) = 0$ becomes $\sum_i \sqrt{P(i)}y(i) = 0$. Let us denote the vector with elements $\sqrt{P(i)}$ as $\sqrt{P}$ and the matrix with elements $b(ij)$ as $[b(ij)]$. It follows that $\sqrt{P}$ is a characteristic vector of the matrix $[b(ij)]$ with characteristic value unity, since

$$\sum_j b(ij)\sqrt{P(j)} = \frac{1}{V_E}\sum_j a(ij)\sqrt{P(i)}\,P(j) = \sqrt{P(i)} \ .$$

Let λ,α be another characteristic value and corresponding characteristic vector. Then

$$\sum_j b(ij)\alpha(j) = \lambda\alpha(i) \ .$$

Take i to be that index for which $\left|\frac{\alpha(i)}{\sqrt{P(i)}}\right|$ is greatest. Then

$$\left|\lambda\alpha(i)\right| \le \sum_j \left|b(ij)\alpha(j)\right| = \sum_j b(ij)\sqrt{P(j)}\left|\frac{\alpha(j)}{\sqrt{P(j)}}\right| \le \frac{\left|\alpha(i)\right|}{\sqrt{P(i)}}\sum_j b(ij)\sqrt{P(j)} = \left|\alpha(i)\right| .$$

Hence, $|\lambda| \le 1$. Let $Y = \sum_i \alpha(i)y(i)$ be the length of the projection of the vector y on α , where α is taken to have unit length. Then the Equations C.5 imply

$$Y' = Y + \lambda Y = (1+\lambda)Y \ . \qquad \text{(C.6)}$$

Here y is any vector orthogonal to $\sqrt{P}$, and hence, in general, $Y \neq 0$. Hence, the condition for the stability of Y is $|1 + \lambda| \le 1$, or, since $|\lambda| \le 1$, $\lambda \le 0$ (λ is real, since $[b(ij)]$ is symmetric).

Hence, a necessary and sufficient condition for stability is that $[b(ij)]$ has exactly one positive characteristic value. For any $e(i)$

$$\sum_i \sum_j b(ij)e(i)e(j) = \sum_i \sum_j a(ij)x(i)x(j)$$

if $x(i) = e(i)\sqrt{\frac{P(i)}{V_E}}$. Hence, we can deduce that the matrices $[a(ij)]$ and $[b(ij)]$ must have the same rank and signature.

It is also of some interest to look at the case where $p(i)$ may be equal to zero at equilibrium. From the Equations C.1 it follows that for each i , either $p(i) = 0$ or $V = \sum_j a(ij)p(j)$. Hence, any equilibrium

may be written as $p(i) = P(i)$ with

$$P(i) > 0 \ (i \in I) \ , \ P(i) = 0 \ (i \notin I)$$
$$\sum_{i \in I} P(i) = 1 \ , \ \sum_{i \in I} a(ij)P(i) = V(I) \quad (j \in I) \ , \tag{C.7}$$

where I is some subset of $(1,2, \ldots ,m)$.

Kingman (1961a) shows that a necessary and sufficient condition for stability is that

$$\sum_{j} a(ij)P(j) \leq V(I) \quad (i \notin I)$$

and that the sub-matrix $[a(ij)]$ $(i,j \in I)$ has exactly one positive characteristic value. Since the proof for this statement is closely duplicated in the case of additive gene action between two linked loci, it was deemed unnecessary to reproduce it here.

D. Hardy-Weinberg Equilibria of Gene Frequencies in the Two-loci Case with Selection and no Mutation

1. General considerations

We have from our previous work (the Equations B.8) that

$$V^{12}p'(i^1i^2) = 2\gamma_{00}p(i^1i^2) \sum_{k^1k^2} p(k^1k^2)a(i^1k^1i^2k^2)$$
$$+ 2\gamma_{01} \sum_{k^1k^2} p(i^1k^2)a(i^1k^1k^2i^2)p(k^1i^2) \ , \tag{D.1}$$

where $V^{12} = \sum_{j^1j^1k^1k^2} p(j^1j^2)a(j^1k^1j^2k^2)p(k^1k^2)$, and where $p(i^1i^2)$ and $p'(i^1i^2)$ are gamete frequencies in two successive generations at infancy.

Again $p(i^1i^2)$ $(i^1 = 1,2, \ldots, m^1,\ i^2 = 1,2, \ldots, m^2)$ is called an equilibrium if $p'(i^1i^2) = p(i^1i^2)$. As noted previously, we write $a(i^1j^1i^2j^2)$ for $a^{12}(i^1j^1i^2j^2)$ in the cases where no great possibility for confusion exists. We write any equilibrium as $p(i^1i^2) = P(i^1i^2)$. At equilibrium the Equations D.1 will be a system of cubic equations for which, in general, no explicit solutions can be found. However, in the simple case of two alleles per locus and symmetrical selection patterns explicit solutions can be found for the gamete frequencies at equilibrium, as well as conditions for stable equilibria. An excellent review of this work together with a couple of simple numerical examples to illustrate the general features of the situation is given by Li (1967).

We have only been able to prove that gamete frequencies have panmictic structure, and consideration of the examples in Li (1967) shows that even at equilibrium under selection the gamete frequencies will, in general, not be equal to the product of the gene frequencies, i.e., $p(i^1i^2) \neq p(i^1)p(i^2)$. The mean fitness of the population is therefore a function of gamete frequencies as indicated by the formula for V^{12} in the Equations D.1.

Consideration of the results given in the Equations D.1 will also show that the population equilibria (or stationary states) do not necessarily correspond to stationary values of the mean fitness. The stationary states in mean fitness will be given by $\dfrac{\partial F}{\partial p(i^1i^2)} = 0$, where

$$F = V^{12} - 2\lambda\Big(\sum_{j^1j^2} p(j^1j^2) - 1\Big)$$

and where λ is a Lagrange multiplier. We get that

$$\sum_{k^1 k^2} p(k^1 k^2) a(i^1 k^1 i^2 k^2) = \lambda$$

and (D.2)

$$\lambda = V^{12} \quad .$$

If we compare Equations D.1 and D.2 we see that they will lead to equivalent equilibrium states if $\gamma_{01} = 0$, or, in other words, if no cross-overs occur.

It is also easy to show by means of counter-examples that in the two-loci case the theorem that mean fitness increases from one generation to the next, no longer holds in general. This point is very well discussed by Moran (1964) and by Li (1967).

It is obvious from the foregoing discussion that in the multi-loci situation the effects of selection must in general be discussed in terms of gamete frequencies. The description of a population by gamete frequencies, however, involves many more parameters than a description by gene frequencies. Let us assume we have ℓ loci and m alleles per locus. Then we will need $\ell^m - 1$ parameters to describe the population in terms of gamete frequencies, whereas if the description can be reduced to gene frequencies we will need only $\ell(m-1)$ parameters. It is therefore of interest to inquire under which circumstances a description of a population in terms of gene frequencies will be valid in the two-loci situation. The validity of a gene frequency description in the multi-loci case is of special importance for the extension of genetic load theory from the single locus case to the multi-loci case.

A clue for the conditions under which, at equilibrium, a gene frequency description of a population will be valid is given by the work of Bodmer and Felsenstein (1967). In our notation these authors define gene action to be additive between loci if $a(i^1j^1i^2j^2) = a(i^1j^1) + a(i^2j^2)$ and to be multiplicative between loci if $a(i^1j^1i^2j^2) = a(i^1j^1)a(i^2j^2)$. In the case of two loci each with two alleles, they found that the gamete frequencies are equal to the product of the gene frequencies under the following conditions:

(i) if fitness is additive between loci, and if the equilibrium at each locus considered alone is due to heterozygote superiority, i.e., if

$$a^1(12) > a^1(11), a^1(22), \text{ and if } a^2(12) > a^2(11), a^2(22),$$

and

(ii) if fitness is multiplicative between loci under the same conditions as given under (i) and the recombination value exceeds a simple function of the selection coefficients. These results of Bodmer and Felsenstein will be derived as special cases of more general results.

Let us now consider the possibility of an equilibrium $P(i^1i^2) = P(i^1)P(i^2)$, where $P(i^1)$ is the equilibrium frequency of the allele (i^1), $P(i^2)$ is the equilibrium frequency of the allele (i^2) and $P(i^1i^2)$ is the equilibrium frequency of the gamete (i^1i^2). Equations D.1 become then

$$V_E^{12}P(i^1)P(i^2) = 2\gamma_{00}P(i^1)P(i^2) \sum_{k^1k^2} P(k^1)P(k^2)a(i^1k^1i^2k^2)$$

$$+ 2\gamma_{01}P(i^1)P(i^2) \sum_{k^1k^2} P(k^1)P(k^2)a(i^1k^1k^2i^2),$$

which if there is no position effect become

$$V_E^{12}P(i^1)P(i^2) = P(i^1)P(i^2) \sum_{k^1k^2} P(k^1)P(k^2)a(i^1k^1i^2k^2) \quad . \qquad (D.3)$$

Hence, if $P(i^1)P(i^2) > 0$, the Equations D.3 imply

$$V_E^{12} = \sum_{k^1k^2} P(k^1)P(k^2)a(i^1k^1i^2k^2) \quad , \qquad (D.4)$$

where

$$V_E^{12} = \sum_{j^1k^1} \sum_{j^2k^2} P(j^1)P(j^2)a(j^1k^1j^2k^2)P(k^1)P(k^2) \quad ,$$

by definition, and where the subscript "E" to V^{12} denotes, as before, "equilibrium value." V_E^{12} should not be confused with $V^{12}(e(j^1j^2))$ or $V^{12}(e)$ which will be used to denote

$$\sum_{j^1k^1} \sum_{j^2k^2} e(j^1j^2)a(j^1k^1j^2k^2)e(k^1k^2) \quad ,$$

if the need to do so arises.

We see that V_E^{12} is common to all m^1m^2 equations in the Equations D.4. If we can regard V_E^{12} as a constant then the Equations D.4 would be a set of linear equations for which solutions are easily obtained if they exist. We will prove that V_E^{12} is a constant determined by the matrix with elements $a(i^1j^1i^2j^2)$ for all solutions to which the restriction, $\sum_{i^1i^2} P(i^1)P(i^2) = 1$, applies. The solutions to which this restriction apply, are, of course, the only set of solutions of interest in the present circumstances, since the Equations D.4 refer to gene frequencies.

For the purposes of the argument employed here, let us denote the $m^1m^2 \times m^1m^2$ matrix with elements $a(i^1j^1i^2j^2)$ by A, and the vector with

m^1m^2 elements $P(i^1)P(i^2)$ by P. It follows that we can write the Equations D.4 in matrix notation as $AP = P'AP\iota$ or as $AP = V_E^{12}\iota$, where ι denotes a vector of unit elements of dimension m^1m^2. We follow Penrose (1955) in defining the Moore inverse of the matrix A as the unique matrix A^+ for which it is true that

a) $AA^+A = A$,

b) $A^+AA^+ = A^+$,

c) $(AA^+)' = AA^+$,

d) $(A^+A)' = A^+A$.

Let us assume c to be a constant and let the other terms in the equation $AP = c\iota$ have the same meaning as before. The general solution of the consistent equation $AP = c\iota$ is of the form

$$\hat{P} = cA^+\iota + (A^+A-I)Z \quad ,$$

where Z is an arbitrary vector, and where I is the identity matrix. We notice that if A is of full rank, then $A^+ = A^{-1}$, and $\hat{P} = cA^{-1}$.

We now prove that $\hat{P}'A\hat{P} = c^2\iota'A^+\iota$. We have that

$$\hat{P}'A\hat{P} = (cA^+\iota + (A^+A-I)Z)'A(cA^+\iota + (A^+A-I)Z)$$

$$= c^2\iota'A^{+\prime}AA^+\iota + Z'(A^+A-I)'A(A^+A-I)Z + 2c\iota'A^{+\prime}A\,(A^+A-I)Z \quad .$$

From the property (a) of the Moore inverse it follows that the last two terms in the foregoing equation are equal to zero. We can also see that A is symmetric if and only if A^+ is symmetric, as follows from (a) and (b) and the uniqueness of A^+. A is symmetric since we assumed $a(i^1j^1i^2j^2) = a(j^1i^1j^2i^2)$ as was discussed in Section B of this chapter. Hence,

$$\hat{P}'A\hat{P} = c^2\iota'A^+\iota \quad .$$

Note now that if we impose the restriction $P'\iota = 1$, it follows from the equation $AP = c\iota$ that $P'AP = c$. It follows immediately in all cases where $P'\iota = 1$, that $\hat{P}'A\hat{P} = (\iota'A^{+}\iota)^{-1}$.

We note that in the full rank case the Equations D.4 will have a unique solution as follows from the form of the general solution to the equation $AP = P'AP\iota$. It is also necessary to note that not all classes of solutions to the foregoing equation are acceptable to us, since in the class of solutions which satifies the restriction $\hat{P}'\iota = 1$ not all elements of the vector P will be greater than zero, as we require.

2. Additive and multiplicative gene action between loci

The definitions of additive and multiplicative gene action between loci can be handled formally by writing $a(i^1j^1i^2j^2) = a(i^1j^1) + a(i^2j^2)$ in the case of additive gene action and $a(i^1j^1i^2j^2) = a(i^1j^1)a(i^2j^2)$ in the case of multiplicative gene action between loci. These definitions merely imply that a function of the variables i^1, j^1, i^2 and j^2 must be expresible in a certain form.

It is, however, of interest to inquire how these properties of additivity and multiplicativity may arise, since these modes of gene action between loci determine some important conditions for genetic equilibria.

The obvious way in which multiplicative gene action between loci can arise is closely related to the manner in which product fecundity can arise by independent fecundity distributions. The discussion on product fecundity followed on the derivation of Equation B.2. It is now necessary to recall that we denoted the probability of an adult of genotype $(i^1j^1i^2j^2)$ contributing u infant offspring to a mating by $b_u(i^1j^1i^2j^2)$. We proceed then by defining $b_u(i^1j^1i^2j^2) = \sum_{\text{all } rs=u} b_r(i^1j^1)b_s(i^2j^2)$.

Hence, by our previous notation for the mean number of infants from an adult with genotype $(i^1j^1i^2j^2)$ we write

$$b(i^1j^1i^2j^2) = \sum_u u b_u(i^1j^1i^2j^2) = \sum_u u \sum_{\text{all } rs=u} b_r(i^1j^1)b_s(i^2j^2)$$

$$= \sum_{rs} rsb_r(i^1j^1)b_s(i^2j^2)$$

$$= [\sum_r rb_r(i^1j^1)][\sum_s sb_s(i^2j^2)]$$

$$= b(i^1j^1)b(i^2j^2) ,$$

where $b(i^1j^1)$ and $b(i^2j^2)$ are conditional means which can be ascribed to the different adult genotypes at the relevant loci.

We note that in this context $\sum_u$ means summation over all values that the integer u assumes. These values are not necessarily the consecutive integers. The reason for this can easily be seen in the case where $r = 1,2$ and $s = 1,2$ which will cause $b(i^1j^1i^2j^2)$ to be equal to $[b_1(i^1j^1) + 2b_2(i^1j^1)][b_1(i^2j^2) + 2b_2(i^2j^2)]$, as follows from the foregoing equation. Another way of dealing with this problem in notation would, of course, simply be to define $b_u(i^1j^1i^2j^2)$ to be equal to zero for $u \neq rs$, in which case the summation $\sum_u$ can go over consecutive integers.

We also assume the probability of survival of an infant $(i^1j^1i^2j^2)$, which we denote by $\ell(i^1j^1i^2j^2)$, to be equal to $\ell(i^1j^1)\ell(i^2j^2)$. It follows that we can write

$$a(i^1j^1i^2j^2) = \ell(i^1j^1i^2j^2)b(i^1j^1i^2j^2) = \ell(i^1j^1)b(i^1j^1)\ell(i^2j^2)b(i^2j^2)$$

$$= a(i^1j^1)a(i^2j^2) .$$

In essence, then, we assumed here that multiplicativity of gene action between loci results from independent probability distributions for offspring being ascribed to genes at different loci, and from independence at the different loci of events leading to survival from infancy to adulthood.

In the case of additive gene action between loci let us define

$$b_u(i^1j^1i^2j^2) = \sum_{t=0}^{u} b_t(i^1j^1)b_{u-t}(i^2j^2) \quad ,$$

or in other words we define the sequence $b_k(i^1j^1i^2j^2)$ to be the convolution of the two sequences $b_k(i^1j^1)$ and $b_k(i^2j^2)$. It follows from the properties of the probability generating functions of convolutions that

$$b(i^1j^1i^2j^2) = \sum_u u b_u(i^1j^1i^2j^2) = \sum_r r b_r(i^1j^1) + \sum_s s b_s(i^2j^2) = b(i^1i^1) + b(i^2j^2).$$

In essence what we assumed here is that the event of $u = x$ offspring can occur in any of the following mutually exclusive ways $(r = 0,\ s = x)$, $(r = 1,\ s = x-1)$ ..... $(r = x,\ s = 0)$.

The problem in the case of additive gene action is that there does not seem to be a way in which the events associated with the probability functions,

$$a(i^1j^1i^2j^2) = \ell(i^1j^1i^2j^2)b(i^1j^1i^2j^2) = \ell(i^1j^1i^2j^2)(b(i^1j^1) + b(i^2j^2)) \quad ,$$

can be interpreted in order to give meaning to the relationship denoted by

$$a(i^1j^1i^2j^2) = a(i^1j^1) + a(i^2j^2) \quad ,$$

except by assuming $\ell(i^1j^1i^2j^2) = k$, identically, where k is a constant number such that $0 < k \leq 1$. Even the relationship

$$\ell(i^1j^1i^2j^2) = \ell(i^1j^1) + \ell(i^2j^2) \quad ,$$

simply does not make sense, as it implies that survival can be due to the effect of one locus or the other. This point is also brought out clearly if we try to use the same argument on viability as the one that we used on fecundity. The reason for this is that we will have to consider the composite event of survival due to gene action at both loci, and that this event will imply the survival of two adults from one infant. We are, therefore, forced to conclude that if we assume additive gene action between loci it implies that natural selection can only operate through differences in fecundity. Another reason for regarding the assumption of additive gene action between loci in the present set-up as biologically not very realistic is that the only infertile case would be due to the event with probability $b_0(i^1j^1)b_0(i^2j^2)$, or, in other words, a lethal effect can only be produced by "lethal" gene action at both loci.

The foregoing explanations of the mechanisms of multiplicative and additive gene action between loci are certainly not the only ones that are theoretically possible. The reason for this is that although the postulated fecundity distributions imply $b(i^1j^1i^2j^2) = b(i^1j^1)b(i^2j^2)$ and $b(i^1j^1i^2j^2) = b(i^1j^1) + b(i^2j^2)$ respectively for the two cases under consideration, these forms of the fecundity means do not imply the postulated fecundity distributions.

From our definitions of additive and multiplicative gene action between loci the Equations D.4 become in the case of additive gene action between loci

$$V_E^1 + V_E^2 = \sum_{k^1} a(i^1k^1)P(k^1) + \sum_{k^2} a(i^2k^2)P(k^2)\,, \qquad (D.5)$$

and in the case of multiplicative gene action between loci

$$V_E^1 V_E^2 = \sum_{k^1} a(i^1 k^1) P(k^1) \sum_{k^2} a(i^2 k^2) P(k^2) \quad , \qquad (D.6)$$

where

$$V_E^1 = \sum_{j^1} \sum_{k^1} P(j^1) a(j^1 k^1) P(k^1)$$

and

$$V_E^2 = \sum_{j^2} \sum_{k^2} P(j^2) a(j^2 k^2) P(k^2) \quad .$$

Since they are special cases of Equations D.4, which we have shown to have a unique solution in the full-rank case, it follows that the Equations D.5 and D.6 have unique solutions when the matrix $[a(i^1 j^1 i^2 j^2)]$ is of full rank. The solutions to the single locus equations

$$\sum_{k^1} a(i^1 k^1) P(k^1) = V_E^1$$

and

$$\sum_{k^2} a(i^2 k^2) P(k^2) = V_E^2$$

will therefore give unique solutions to the Equations D.5 and D.6 in the case where the matrix $[a(i^1 j^1 i^2 j^2)]$ is of full rank.

If the matrix $[a(i^1 j^1 i^2 j^2)]$ is written out, it is easy to see that $[a(i^1 j^1) a(i^2 j^2)]$ is the Kronecker product of the matrices $[a(i^1 j^1)]$ and $[a(i^2 j^2)]$. A simple example of the Kronecker product of two matrices is given in Section D.2b. Bellman (1960) shows that the characteristic roots of $[a(i^1 j^1) a(i^2 j^2)]$ are $\lambda_i \mu_j$, where λ_i are the characteristic roots of $[a(i^1 j^1)]$ and μ_j are the characteristic roots of $[a(i^2 j^2)]$. It follows that if $[a(i^1 j^1 i^2 j^2)]$ is of full rank, so are $[a(i^1 j^1)]$ and $[a(i^2 j^2)]$ and vice versa.

a. Stable equilibria in the multiplicative case The problem is now to determine if conditions exist under which the equilibrium $P(i^1i^2) = P(i^1)P(i^2)$ will be stable. Analogously to the single locus case, we write $p(i^1i^2) = P(i^1)P(i^2) + e(i^1i^2)$ and $p'(i^1i^2) = P(i^1)P(i^2) + e'(i^1i^2)$. We assume no position effects, and by substitution in the Equations D.1, ignoring quadratic terms in the $e(i^1i^2)$'s and remembering that

$$\sum_{i^1i^2} e(i^1i^2) = 0 \ ,$$

we obtain

$$V_E^{12}e'(i^1i^2) = 2\gamma_{00}e(i^1i^2)V_E^{12} + 2\gamma_{00}P(i^1)P(i^2) \sum_{k^1k^2} a(i^1k^1i^2k^2)e(k^1k^2)$$

$$+ 2\gamma_{01} P(i^2) \sum_{k^1k^2} e(i^1k^2)a(i^1k^1i^2k^2)P(k^1)$$

$$+ 2\gamma_{01} P(i^1) \sum_{k^1k^2} e(k^1i^2)a(i^1k^1i^2k^2)P(k^2) \ . \qquad \text{(D.7)}$$

It does not seem to be possible to derive equilibrium conditions from the Equations D.7 for arbitrary fitnesses. We therefore proceed to the multiplicative case where $a(i^1j^1i^2j^2) = a(i^1j^1)a(i^2j^2)$. We use the single locus solutions to the Equations D.6 to write

$$V_E^1V_E^2e'(i^1i^2) = 2\gamma_{00}\, e(i^1i^2)V_E^1V_E^2 + 2\gamma_{00}\, P(i^1)P(i^2) \sum_{k^1k^2} a(i^1k^1)a(i^2k^2)e(k^1k^2)$$

$$+ 2\gamma_{01}\, P(i^2) \sum_{k^2} e(i^1k^2)a(i^2k^2)V_E^1$$

$$+ 2\gamma_{01}\, P(i^1) \sum_{k^1} e(k^1i^2)a(i^1k^1)V_E^2 \ . \qquad \text{(D.8)}$$

We then make the following transformations:

$$y(i^1i^2) = \frac{e(i^1i^2)}{\sqrt{P(i^1)P(i^2)}} \quad ,$$

$$b(i^1j^1) = \frac{a(i^1j^1)}{V_E^1}\sqrt{P(i^1)P(j^1)} \quad ,$$

and

$$b(i^2j^2) = \frac{a(i^2j^2)}{V_E^2}\sqrt{P(i^2)P(j^2)} \quad . \qquad \text{(D.9)}$$

Substituting Equations D.9 in the Equations D.8 we get after simplification

$$y'(i^1i^2) = 2\gamma_{00}\, y(i^1i^2) + 2\gamma_{00} \sum_{k^1k^2} y(k^1k^2)b(i^1k^1)b(i^2k^2)$$

$$+ 2\gamma_{01} \sum_{k^2} y(i^1k^2)b(i^2k^2) + 2\gamma_{01} \sum_{k^1} y(k^1i^2)b(i^1k^1) \quad . \qquad \text{(D.10)}$$

The condition

$$\sum_{i^1i^2} e(i^1i^2) = 0 \quad ,$$

becomes now

$$\sum_{i^1i^2} \sqrt{P(i^1)P(i^2)}\; y(i^1i^2) = 0 \quad .$$

Let us denote the vector with elements $\sqrt{P(i^1)P(i^2)}$ as $\sqrt{P^1P^2}$ and the matrix with elements $b(i^1j^1)b(i^2j^2)$ as $[b(i^1j^1)b(i^2j^2)]$. It follows that $\sqrt{P^1P^2}$ is a characteristic vector of $[b(i^1j^1)(b(i^2j^2)]$ with characteristic value unity, since

$$\sum_{j^1j^2} b(i^1j^1)b(i^2j^2)\sqrt{P(j^1)P(j^2)}$$

$$= \sum_{j^1j^2} \frac{a(i^1j^1)}{V_E^1}\sqrt{P(i^1)}\,P(j^1)\,\frac{a(i^2j^2)}{V_E^2}\sqrt{P(i^2)}\,P(j^2) = \sqrt{P(i^1)P(i^2)}.$$

In the same way we have that $\sqrt{P^1}$ is a characteristic vector of $[b(i^1j^1)]$ and $\sqrt{P^2}$ is a characteristic vector of $[b(i^2j^2)]$, both with characteristic value unity. In the same way as in the single locus case (Section C of this chapter), it can now easily be proved that for any other characteristic value, say λ, of any matrices of the form $[b(i^1j^1)]$, $[b(i^2j^2)]$ or $[b(i^1j^1)b(i^2j^2)]$, it must be true that $|\gamma| \leq 1$.

Upon inspection it will be seen that the matrix designated as $[b(i^1j^1)b(i^2j^2)]$ can be written as the Kronecker product of the matrices $[b(i^1j^1)]$ and $[b(i^2j^2)]$. (This statement will become obvious when the two-allele case is considered in detail). It follows therefore that the characteristic roots of $[b(i^1j^1)b(i^2j^2)]$ are $\mu_r^1\mu_s^2$, where μ_r^1 are the characteristic roots of $[b(i^1j^1)]$ and μ_s^2 the characteristic roots of $[b(i^2j^2)]$. The form of the characteristic vectors of Kronecker matrices also deserves some attention. Bellman (1960) shows that the components of a characteristic vector of the matrix $[b(i^1j^1)b(i^2j^2)]$ can always be written as a product of two components, one from a characteristic vector of $[b(i^1j^1)]$ and the other from $[b(i^2j^2)]$. It follows that we can write for the components of the characteristic vector y^{12} that $y(i^1i^2) = y(i^1)y(i^2)$, where $y(i^1)$ is a component of the characteristic vector y^1 of $[b(i^1j^1)]$ and $y(i^2)$ is a component of the vector y^2 of $[b(i^2j^2)]$. Since the characteristic vectors associated with a certain characteristic value form a subspace of the vector space associated with a matrix, it follows from the single locus case, where the y-vectors have the property that $\sum_i \sqrt{P(i)}\, y(i) = 0$, that $\sum_{i^1} \sqrt{P(i^1)}\, y(i^1) = 0$ and that $\sum_{i^2} \sqrt{P(i^2)}\, y(i^2) = 0$. It follows therefore that the maximal dimension of the

vector space spanned by characteristic vectors of the type y^{12} is $(m^1-1)(m^2-1)$. There will be $(m^1-1)(m^2-1)$ roots of the type $\lambda^1\lambda^2$ given by the equations

$$\sum_{k^1k^2} y(k^1k^2)b(i^1k^1)b(i^2k^2) = \sum_{k^1} y(k^1)b(i^1k^1) \sum_{k^2} y(k^2)b(i^2k^2)$$

$$= \lambda^1 y(i^1)\lambda^2 y(i^2) = \lambda^1\lambda^2 y(i^1i^2) \quad ,$$

and hence a $(m^1-1)(m^2-1)$ dimensional subspace of the matrix $[b(i^1j^1)b(i^2j^2)]$ composed of vectors of the type y^{12}.

The vector space of $[b(i^1j^1)b(i^2j^2)]$ must also contain a subspace of dimension (m^1-1) of characteristic vectors of the type $y^1\sqrt{P^2}$ and a subspace of dimension (m^2-1) of characteristic vectors of the type $y^2\sqrt{P^1}$, as can readily be discerned from the equations

$$\sum_{k^1k^2} y(k^1)\sqrt{P(k^2)}\, b(i^1k^1)b(i^2k^2) = \sum_{k^1} y(k^1)b(i^1k^1) \sum_{k^2}\sqrt{P(k^2)}\, b(i^2k^2)$$

$$= \lambda^1 y(i^1)\, 1\sqrt{P(i^2)} = \lambda^1 y(i^1)\sqrt{P(i^2)} \quad .$$

and

$$\sum_{k^1k^2}\sqrt{P(k^1)}y(k^2)b(i^1k^1)b(i^2k^2) = \sum_{k^1}\sqrt{P(k^1)}\, b(i^1k^1) \sum_{k^2} y(k^2)b(i^2k^2)$$

$$= 1\sqrt{P(i^1)}\ \ \lambda^2 y(i^2) = \lambda^2\sqrt{P(i^1)}\ y(i^2) \quad .$$

Since $\sqrt{P^1P^2}$ is also a characteristic vector of $[b(i^1j^1)b(i^2j^2)]$ we now have accounted for m^1m^2 linearly independent characteristic vectors to span the vector space associated with the matrix $[b(i^1j^1)b(i^2j^2)]$. It now remains to formalize our definitions of the vectors $y^1\sqrt{P^2}$, $y^2\sqrt{P^1}$ and $\sqrt{P^1P^2}$ by associating with them vectors consisting respectively of

the components $y(i^1)\sqrt{P(i^2)}$, $y(i^2)\sqrt{P(i^1)}$ and $\sqrt{P(i^1)P(i^2)}$.

It follows from the foregoing discussion that for characteristic vectors of the matrix $[b(i^1j^1)b(i^2j^2)]$ of the type y^{12} , we can write the Equations D.10 as follows

$$y'(i^1i^2) = 2\gamma_{00}y(i^1i^2) + 2\gamma_{00}\lambda^1\lambda^2y(i^1i^2) + 2\gamma_{01}\lambda^2y(i^1i^2) + 2\gamma_{01}\lambda^1y(i^1i^2) . \tag{D.11}$$

Let $Y = \sum_{i^1i^2} \alpha(i^1i^2)y(i^1i^2)$ be the length of the projection of the vector y^{12} on α , where α is taken to have unit length. Then the Equations D.11 imply

$$Y' = Y[2\lambda_{00}(1 + \lambda^1\lambda^2) + 2\gamma_{01}(\lambda^1 + \lambda^2)] , \tag{D.12}$$

or remembering that $\gamma_{00} = \frac{1}{2}(1-\rho)$ and that $\gamma_{01} = \frac{1}{2}\rho$, where ρ is the coefficient of recombination

$$Y' = Y[1 + \lambda^1\lambda^2 + \rho(\lambda^1 + \lambda^2 - 1 - \lambda^1\lambda^2)] . \tag{D.13}$$

Since they have to be associated with distinct characteristic roots, y^{12} is any vector orthogonal to $\sqrt{P^1P^2}$, and hence, in general, $Y \neq 0$. Hence, a necessary and sufficient condition for the stability of Y is that

$$\left|1 + \lambda^1\lambda^2 + \rho(\lambda^1 + \lambda^2 - 1 - \lambda^1\lambda^2)\right| \leq 1 . \tag{D.14}$$

We also have that it is true for $\lambda^1\lambda^2$, λ^1, λ^2 that $|\lambda| \leq 1$. If we assume that the equilibria at the two loci considered separately are stable, we have $\lambda^1 \leq 0$, $\lambda^2 \leq 0$, as was shown in the previous section. It follows that $\lambda^1\lambda^2 \geq 0$, $|\lambda^1\lambda^2| \leq |\lambda^1|$ and $|\lambda^1\lambda^2| \leq |\lambda^2|$.

It is now advantageous to write $\beta = 1 - 2\rho$. The Equation D.14 becomes then

$$\left| \frac{1}{2} (1 + \lambda^1\lambda^2 + \lambda^1 + \lambda^2) + \frac{\beta}{2}(1 + \lambda^1\lambda^2 - \lambda^1 - \lambda^2) \right| \leq 1 . \qquad (D.15)$$

Under the assumption of single locus stability, i.e., $-1 \leq \lambda^1 \leq 0$, $-1 \leq \lambda^2 \leq 0$, the quantity inside the absolute sign in Equation D.15 is always positive. Consequently we can write

$$\beta \leq \frac{1 - \lambda^1\lambda^2 - \lambda^1 - \lambda^2}{1 + \lambda^1\lambda^2 - \lambda^1 - \lambda^2} \qquad (D.16)$$

or

$$\rho \geq \frac{\lambda^1\lambda^2}{1 + \lambda^1\lambda^2 - \lambda^1 - \lambda^2} \quad . \qquad (D.17)$$

It is now essential to make a distinction between the observed or induced perturbation vectors and those which are characteristic vectors of the matrix $[b(i^1j^1)b(i^2j^2)]$, although in some cases the two may coincide. Let us denote

$$\frac{e(i^1i^2)}{\sqrt{P(i^1)P(i^2)}}$$

by $z(i^1i^2)$ if it is an observed perturbation, and by $y(i^1i^2)$ if it is a component of a characteristic vector. Since $\sum_{i^1i^2} e(i^1i^2) = 0$, it is true that

$$\sum_{i^1i^2} \sqrt{P(i^1)P(i^2)}\, y(i^1i^2) = 0 = \sum_{i^1i^2} \sqrt{P(i^1)P(i^2)}\, z(i^1i^2) \quad . \qquad (D.18)$$

We showed that $y(i^1i^2) = y(i^1)y(i^2)$ and that consequently the space of y^{12} vectors has a dimension of $(m^1-1)(m^2-1)$. This will, however, not be

true in the case of the z-vectors, as the space of the z-vectors has in general a dimension of m^1m^2-1 , as follows from Equation D.18. Hence, vectors of all the types y^{12} , $y^1\sqrt{P^2}$, $y^2\sqrt{P^1}$ are needed to form a basis for the space of z-vectors.

We now note that the components of the z-vectors change from generation to generation according to the Equations D.10, with $z(i^1i^2)$ substituted for $y(i^1i^2)$. We can, therefore, study the change in the components of the z-vectors by observing the change in components of the basis vectors on the assumption that they change according to the Equations D.10. Hence, we substitute $\sqrt{P(i^1)}\, y(i^2)$ for $y(i^1i^2)$ in the Equations D.10 to get on simplification that

$$\sqrt{P(i^1)}\, y'(i^2) = 2\gamma_{00}\sqrt{P(i^1)}\, y(i^2) + 2\gamma_{00}\sqrt{P(i^1)}\, \lambda^2 y(i^2) + 2\gamma_{01}\sqrt{P(i^1)}\, y(i^2) + 2\gamma_{01}\sqrt{P(i^1)}\lambda^2 y(i^2) ,$$

which, on remembering that $2\gamma_{00} + 2\gamma_{01} = 1$, reduces to

$$y'(i^2) = y(i^2)(1 + \lambda^2) .$$

In a precisely analogous manner we get on substituting $y(i^1)\sqrt{P(i^2)}$ for $y(i^1i^2)$ in the Equations D.10 that

$$y'(i^1) = y(i^1)(1 + \lambda^1) .$$

These results correspond to that given as Equation C.6 in the single locus case.

We summarize the foregoing results in Theorem 1.

Theorem 1 In the case of multiplicative gene action between loci the Hardy-Weinberg equilibrium $p(i^1i^2) = P(i^1)P(i^2) > 0$ is stable if and only if the single locus fitnesses are such that both the single locus cases are stable, and the linkage between the two loci is not too tight.

In quantitative terms this means that the characteristic values of $[b(i^1j^1)]$ and $[b(i^2j^2)]$, which we denoted by λ^1 and λ^2, must be such that $\lambda^1 \leq 0$ and $\lambda^2 \leq 0$ and that the amount of recombination must be large enough so that the inequalities given as Equations D.16 or D.17 hold.

It must be emphasized that this is a local result and does not say anything about the possibility for the existence of other types of equilibria other than that of $P(i^1i^2) = P(i^1)P(i^2)$. This matter will be discussed further in Section E.5.

b. Multiplicative gene action in the two-allele case It is instructive to consider the case of two alleles at each of the two loci in detail. We follow the pattern used by Bellman (1960) for the characteristic vectors of Kronecker product matrices and write the vector y^{12} as

$$\begin{bmatrix} y(11) \\ y(12) \\ y(21) \\ y(22) \end{bmatrix}.$$

We can write

$$\sum_{k^1k^2} y(k^1k^2)b^1(i^1k^1)b^2(i^2k^2)$$

for $i^1 = 1,2$, $i^2 = 1,2$ in matrix notation as follows:

$$\begin{bmatrix} b^1(11)b^2(11) & b^1(11)b^2(12) & b^1(12)b^2(11) & b^1(12)b^2(12) \\ b^1(11)b^2(21) & b^1(11)b^2(22) & b^1(12)b^2(21) & b^1(12)b^2(22) \\ b^1(21)b^2(11) & b^1(21)b^2(12) & b^1(22)b^2(11) & b^1(22)b^2(12) \\ b^1(21)b^2(21) & b^1(21)b^2(22) & b^1(22)b^2(21) & b^1(22)b^2(22) \end{bmatrix} \begin{bmatrix} y(11) \\ y(12) \\ y(21) \\ y(22) \end{bmatrix}$$

(D.19)

From the form of the matrix $[b^1(i^1k^1)b^2(i^2k^2)]$ in the Equations D.19 we see clearly that it is the Kronecker product of the matrices $[b^1(i^1k^1)]$ and $[b^2(i^2k^2)]$. It follows that the characteristic values of $[b^1(k^1k^1)b^2(i^2k^2)]$ can be found by finding the characteristic values of $[b^1(i^1k^1)]$ and $[b^2(i^2k^2)]$.

Let us write the fitness matrices as:

$$\begin{bmatrix} a^1(11) & a^1(12) \\ a^1(21) & a^1(22) \end{bmatrix} = \begin{bmatrix} 1 - s^1 & 1 \\ 1 & 1 - t^1 \end{bmatrix}$$

and

$$\begin{bmatrix} a^2(11) & a^2(12) \\ a^2(21) & a^2(22) \end{bmatrix} = \begin{bmatrix} 1 - s^2 & 1 \\ 1 & 1 - t^2 \end{bmatrix} .$$

At equilibrium $P^1(1) = \dfrac{t^1}{s^1 + t^1}$, $P^1(2) = \dfrac{s^1}{s^1 + t^1}$ and $V_E^1 = \dfrac{t^1 + s^1 - s^1t^1}{s^1 + t^1}$.

It follows that we can write

$$\begin{bmatrix} b^1(11) & b^1(12) \\ b^1(21) & b^1(22) \end{bmatrix} = \begin{bmatrix} \dfrac{a^1(11)}{V_E^1}\sqrt{P^1(1)P^1(1)} & \dfrac{a^1(12)}{V_E^1}\sqrt{P^1(1)P^1(2)} \\ \dfrac{a^1(12)}{V_E^1}\sqrt{P^1(1)P^1(2)} & \dfrac{a^1(22)}{V_E^1}\sqrt{P^1(2)P^1(2)} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{(1-s^1)t^1}{t^1+s^1-s^1t^1} & \frac{\sqrt{s^1t^1}}{t^1+s^1-s^1t^1} \\ \frac{\sqrt{s^1t^1}}{t^1+s^1-s^1t^1} & \frac{(1-t^1)s^1}{t^1+s^1-s^1t^1} \end{bmatrix} .$$

The characteristic roots of the matrix $[b^1(i^1j^1)]$ can be derived to be

$$\mu^1 = 1 \quad \text{or} \quad \mu^1 = \frac{-s^1t^1}{t^1+s^1-s^1t^1} ,$$

and in the same way the characteristic roots of $[b^2(i^2j^2)]$ can be seen to be

$$\mu^2 = 1 \quad \text{or} \quad \mu^2 = \frac{-s^2t^2}{t^2+s^2-s^2t^2} .$$

The characteristic roots of $[b^1(i^1j^1)b^2(i^2j^2)]$ are therefore 1 ,

$$\lambda^1 = \frac{-s^1t^1}{t^1+s^1-s^1t^1} , \quad \lambda^2 = \frac{-s^2t^2}{t^2+s^2-s^2t^2} ,$$

and

$$\lambda^1\lambda^2 = \frac{s^1t^1s^2t^2}{(t^1+s^1-s^1t^1)(t^2+s^2-s^2t^2)} .$$

From the definition of a genetic load (see Chapter II, Equation 1) we see that the load at the first locus at equilibrium is

$$L^1 = \frac{s^1t^1}{t^1+s^1} ,$$

and at the second locus is

$$L^2 = \frac{s^2 t^2}{t^2 + s^2} .$$

It follows that

$$\lambda^1 = \frac{L^1}{L^1 - 1} , \quad \lambda^2 = \frac{L^2}{L^2 - 1} .$$

The Equation D.17 becomes after simplification

$$\rho \geq L^1 L^2 . \tag{D.20}$$

This is the special case of Theorem 1 derived by Bodmer and Felsenstein (1967). These authors also show that with $\rho < L^1 L^2$ and $\lambda^1 \leq 0$, $\lambda^2 \leq 0$, there exists one or more stable equilibria with $P(i^1 i^2) > 0$ and $P(i^1 i^2) \neq P(i^1)P(i^2)$. In some cases $P(i^1 i^2)$ may be appreciably different from $P(i^1)P(i^2)$.

c. Stable equilibria in the additive case It is now easy to derive the conditions for stability for the case of additive gene action between loci. We assume $a(i^1 j^1 i^2 j^2) = a(i^1 j^1) + a(i^2 j^2)$ and use the results given in the Equations D.5 to obtain, by virtue of the Equations D.7, that

$$V_E^{12} e'(i^1 i^2) = 2\gamma_{00} V_E^{12} e(i^1 i^2) + 2\gamma_{00} P(i^1)P(i^2) \sum_{k^1 k^2} [a(i^1 k^1) + a(i^2 k^2)] e(k^1 k^2)$$

$$+ 2\gamma_{01} P(i^2) [\sum_{k^2} e(i^1 k^2) V_E^1 + \sum_{k^2} e(i^1 k^2) a(i^2 k^2)]$$

$$+ 2\gamma_{01} P(i^1) [\sum_{k^1} e(k^1 i^2) a(i^1 k^1) + \sum_{k^1} e(k^1 i^2) V_E^2] .$$

Hence, by the substitution of the Equations D.9 in the foregoing, we obtain

$$V_E^{12}\sqrt{P(i^1)P(i^2)}\; y'(i^1i^2) = 2\gamma_{00}V_E^{12}\sqrt{P(i^1)P(i^2)}\; y(i^1i^2)$$

$$+ 2\gamma_{00}P(i^1)P(i^2) \sum_{k^1k^2} \left[\frac{V_E^1 b(i^1k^1)}{\sqrt{P(i^1)P(k^1)}} + \frac{V_E^2 b(i^2k^2)}{\sqrt{P(i^2)P(k^2)}}\right]\sqrt{P(k^1)P(k^2)}\; y(k^1k^2)$$

$$+ 2\gamma_{01}P(i^2)\left[\sum_{k^2}\sqrt{P(i^1)P(k^2)}\; y(i^1k^2)V_E^1 + \sum_{k^2}\sqrt{P(i^1)P(k^2)}\; y(i^1k^2)\frac{b(i^2k^2)V_E^2}{\sqrt{P(i^2)P(k^2)}}\right]$$

$$+ 2\gamma_{01}P(i^1)\left[\sum_{k^1}\sqrt{P(k^1)P(i^2)}\; y(k^1i^2)\frac{b(i^1k^1)V_E^1}{\sqrt{P(i^1)P(k^1)}} + \sum_{k^1}\sqrt{P(k^1)P(i^2)}\; y(k^1i^2)V_E^2\right].$$

It is now necessary to remember that in the case of additive gene action between loci, $V_E^{12} = V_E^1 + V_E^2$, so that on rearranging terms we obtain

$$(V_E^1 + V_E^2)\sqrt{P(i^1)P(i^2)}\; y'(i^1i^2) = 2\gamma_{00}(V_E^1 + V_E^2)\sqrt{P(i^1)P(i^2)}\; y(i^1i^2)$$

$$+ 2\gamma_{00}(V_E^1+V_E^2)\sqrt{P(i^1)P(i^2)}\sum_{k^1k^2} y(k^1k^2)\; X$$

$$\left[\frac{V_E^1 b(i^1k^1)\sqrt{P(i^2)P(k^2)} + V_E^2 b(i^2k^2)\sqrt{P(i^1)P(k^1)}}{V_E^1 + V_E^2}\right]$$

$$+ 2\gamma_{01}(V_E^1 + V_E^2)\sqrt{P(i^1)P(i^2)}\sum_{k^2} y(i^1k^2)\left[\frac{V_E^1\sqrt{P(i^2)P(k^2)} + V_E^2 b(i^2k^2)}{V_E^1 + V_E^2}\right],$$

$$+ 2\gamma_{01}(V_E^1 + V_E^2)\sqrt{P(i^1)P(i^2)}\sum_{k^1} y(k^1i^2)\left[\frac{V_E^1 b(i^1k^1)+V_E^2\sqrt{P(i^1)P(k^1)}}{V_E^1 + V_E^2}\right],$$

which on the cancellation of common terms become

$$y'(i^1i^2) = 2\gamma_{00}y(i^1i^2)$$

$$+ 2\gamma_{00}\sum_{k^1k^2} y(k^1k^2)\left[\frac{V_E^1 b(i^1k^1)\sqrt{P(i^2)P(k^2)} + V_E^2 b(i^2k^2)\sqrt{P(i^1)P(k^1)}}{V_E^1 + V_E^2}\right]$$

$$+ 2\gamma_{01}\sum_{k^2} y(i^1k^2)\left[\frac{V_E^1\sqrt{P(i^2)P(k^2)} + V_E^2 b(i^2k^2)}{V_E^1 + V_E^2}\right] \qquad \text{(D.21)}$$

$$+ 2\gamma_{01}\sum_{k^1} y(k^1i^2)\left[\frac{V_E^1 b(i^1k^1) + V_E^2\sqrt{P(i^1)P(k^1)}}{V_E^1 + V_E^2}\right] .$$

In the consideration of the matrices and vectors associated with the Equations D.21, we will follow the same notational conventions as in the case of multiplicative gene action between loci. Let us first study the matrices $[\sqrt{P(i^1)P(k^1)}]$ and $[\sqrt{P(i^2)P(k^2)}]$. $\sqrt{P^1}$ is a characteristic vector of $[\sqrt{P(i^1)P(k^1)}]$ with characteristic value unity as can be seen from the equations $\sum_{k^1}\sqrt{P(i^1)P(k^1)} = \sqrt{P(i^1)}$. In the same way $\sqrt{P^2}$ is a characteristic vector of $[\sqrt{P(i^2)P(k^2)}]$ with characteristic value unity. We also note that $[\sqrt{P(i^1)P(k^1)}]$ and $[\sqrt{P(i^2)P(k^2)}]$ both have the rank of unity. It follows that both these matrices have one characteristic value of unity and the rest equal to zero. Also, inspection of the matrix $[b(i^1k^1)\sqrt{P(i^2)P(k^2)}]$ shows it to be the Kronecker product of the matrices $[b(i^1k^1)]$ and $[\sqrt{P(i^2)P(k^2)}]$.

In the same way as in the case of multiplicative gene action between loci we now make a distinction between the observed or induced perturbation vectors and the vectors which are characteristic vectors of the matrix $[b(i^1j^1)b(i^2j^2)]$. In this case we note that the components of the z-

vectors change from generation to generation according to the Equations D.21, with, of course, $z(i^1i^2)$ substituted for $y(i^1i^2)$. As before we study the change in the components of the z-vectors by observing the change in the components of the basis vectors y^{12}, $y^1\sqrt{P^2}$ and $y^2\sqrt{P^1}$, which will also take place according to the Equations D.21.

We consider again the vectors y^{12} to be characteristic vectors of the Kronecker product matrices $[b(i^1j^1)b(i^2j^2)]$, so that we can write for their components $y(i^1i^2) = y(i^1)y(i^2)$, where $y(i^1)$ is a component of a characteristic vector y^1 of $[b(i^1j^1)]$ and likewise $y(i^2)$ is a component of a characteristic vector y^2 of $[b(i^2j^2)]$. Since the vector $\sqrt{P^2}$ belongs to the range space of the matrix $[\sqrt{P(i^2)P(j^2)}]$, y^2 has to belong to the null space of $[\sqrt{P(i^2)P(j^2)}]$ as this matrix has only one non-zero characteristic root. In the same way vectors of the type y^1 has to belong to the null space of $[\sqrt{P(i^1)P(j^1)}]$.

Since by the foregoing argument $\sum_{k^2} y(k^2)\sqrt{P(i^2)P(k^2)} = 0$ $= \sum_{k^1} y(k^1)\sqrt{P(i^1)P(k^1)}$, it follows that for all elements belonging to vectors of the type y^{12}, the Equations D.21 can be written as

$$y'(i^1i^2) = 2\gamma_{00}y(i^1i^2) + 2\gamma_{01}y(i^1i^2)\frac{V_E^2\lambda^2}{V_E^1 + V_E^2} + 2\gamma_{01}y(i^1i^2)\frac{V_E^1\lambda^1}{V_E^1 + V_E^2}$$

or since $2\gamma_{00} = (1-\rho)$ and $2\gamma_{01} = \rho$

$$y'(i^1i^2) = y(i^1i^2)[(1-\rho) + \rho(\frac{\lambda^1V_E^1 + \lambda^2V_E^2}{V_E^1 + V_E^2})] \quad . \qquad (D.22)$$

Define, in the same manner as before, that $Y = \sum_{i^1i^2} \alpha(i^1i^2)y(i^1i^2)$.

We get

$$Y' = Y[(1-\rho) + \rho(\frac{\lambda^1 V_E^1 + \lambda^2 V_E^2}{V_E^1 + V_E^2})] \quad . \tag{D.23}$$

If we assume single locus stability, we have $-1 \leq \lambda^1 \leq 0$, $-1 \leq \lambda^2 \leq 0$. We see from Equation D.23 that this is enough to ensure the stability of Y , since in the regular case $\rho \leq \frac{1}{2}$.

It is now necessary to remember from our work on the single locus case and from the case of multiplicative gene action between loci that $\Sigma_{k^1} y(k^1)b(i^1k^1) = \lambda^1 y(i^1)$ and that $\Sigma_{k^2}\sqrt{P(k^2)}b(i^2k^2) = \sqrt{P(i^2)}$. We shall use these facts to study the change in the absolute magnitude of components of vectors of the type $y^1\sqrt{P^2}$. On substituting $y(i^1)\sqrt{P(i^2)}$ for $y(i^1i^2)$ in the Equations D.21, we obtain on simplification that

$$y'(i^1)\sqrt{P(i^2)} = 2\gamma_{00}y(i^1)\sqrt{P(i^2)} + 2\gamma_{00}y(i^1)\sqrt{P(i^2)}\,\frac{V_E^1\lambda^1}{V_E^1 + V_E^2}$$

$$+ 2\gamma_{01}y(i^1)\sqrt{P(i^2)} + 2\gamma_{01}y(i^1)\sqrt{P(i^2)}\,\frac{V_E^1\lambda^1}{V_E^1 + V_E^2} \quad .$$

On recalling that $2\gamma_{00} + 2\gamma_{01} = 1$ we obtain easily that

$$y'(i^1)\sqrt{P(i^2)} = y(i^1)\sqrt{P(i^2)}\,(1 + \frac{\lambda^1 V_E^1}{V_E^1 + V_E^2}) \quad . \tag{D.24}$$

In precisely a similar fashion we deduce that

$$y'(i^2)\sqrt{P(i^1)} = y(i^2)\sqrt{P(i^1)}\,(1 + \frac{\lambda^2 V_E^2}{V_E^1 + V_E^2}) \quad . \tag{D.25}$$

Since $\frac{V_E^2}{V_E^1 + V_E^2}$ and $\frac{V_E^2}{V_E^1 + V_E^2}$ are always positive it is now obvious that in order to secure stability from the Equations D.24 and D.25, λ^1 and λ^2 must always non-positive. We also recall that $|\lambda^1| \leq 1$ and that $|\lambda^2| \leq 1$. It follows that the systems defined by the Equations D.24 and D.25 are stable if and only if $1 \leq \lambda^1 \leq 0$ and $-1 \leq \lambda^2 \leq 0$. Hence, we conclude from the Equations D.23, D.24, and D.25 that the equilibrium $p(i^1 i^2) = P(i^1)P(i^2) > 0$ is stable if and only if the single locus fitnesses are such that both the single locus cases are stable, i.e., that $\lambda^1 \leq 0$ and $\lambda^2 \leq 0$, where λ^1 and λ^2 are the non-unity characteristic roots of the matrices $[b(i^1 j^1)]$ and $[b(i^2 j^2)]$ respectively.

We now succeeded in proving Theorem 2.

Theorem 2 In the two-loci case with additive gene action between loci, stable Hardy-Weinberg equilibria exist if and only if both the constituent single locus cases are stable.

We note again that this is a local result. The conditions under which Theorem 2 becomes a global result will be discussed in Sections E.1 and E.2.

E. The Relationship between Selection Equilibria and Stationary Points in Mean Fitness in the Two-loci Case

1. The increase in the population mean fitness in the case of additive gene action between loci

It has been mentioned in Section C that several authors were able, in the single locus case, to establish the inequality

$$V' = \sum_i \sum_j a(ij)p'(i)p'(j) \geq V = \sum_i \sum_j a(ij)p(i)p(j) \quad ,$$

i.e., the mean fitness increases from generation to generation. Using this result Mulholland and Smith (1959) have proved that the gene frequencies $p(i)$ tend to a limit after a large number of generations, and that this limit is a local maximum of V as a function of the $p(i)$. Kingman (1961b) showed that $V' = V$ if and only if for each i either $p(i) = 0$, or $\sum_j a(ij)p(j) = V$. This last condition is equivalent to the condition for a stationary point that we get when we put $p'(i) = p(i) = P(i) > 0$ in the formula for gene frequency change (Equation B.12). Thus, the mean fitness must increase if the initial point is an internal point which is not stationary, and a stable stationary point must be local maximum of V.

We will now prove that in the case of additive gene action between loci the mean fitness increases from generation to generation, i.e., $V'^{12} \geq V^{12}$. Let us substitute from the Equations D.1 in

$$\sum_{i^1i^2} \sum_{j^1j^2} p'(i^1i^2)[a(i^1j^1 + a(i^2j^2)]p'(j^1j^2)$$

to get

$$V'^{12} = \sum_{i^1i^2} \sum_{j^1j^2} (V^{12})^{-2}\{2\gamma_{00}p(i^1i^2) \sum_{k^1k^2} p(k^1k^2)[a(i^1k^1) + a(i^2k^2)]$$

$$+ 2\gamma_{01} \sum_{k^1k^2} p(i^1k^2)[a(i^1k^1) + a(i^2k^2)]p(k^1i^2)\}[a(i^1j^1) + a(i^2j^2)]$$

$$\{2\gamma_{00}p(j^1j^2) \sum_{k^1k^2} p(k^1k^2)[a(j^1k^1) + a(j^2k^2)]$$

$$+ 2\gamma_{01} \sum_{k^1k^2} p(j^1k^2)[a(j^1k^1) + a(j^2k^2)]p(k^1j^2)\} \quad . \qquad (E.1)$$

Since we can, for example, argue from the symmetry of the variables in the different functions and from the property $2\gamma_{00} + 2\gamma_{01} = 1$, that

$$\sum_{i^1 i^2} a(i^1 j^1)\{2\gamma_{00} p(i^1 i^2) \sum_{k^1 k^2} p(k^1 k^2)[a(i^1 k^1) + a(i^2 k^2)]$$

$$+ 2\gamma_{01} \sum_{k^1 k^2} p(i^1 k^2)[a(i^1 k^1) + a(i^2 k^2)]p(k^1 i^2)\}$$

$$= \sum_{i^1} a(i^1 j^1)\{2\gamma_{00} \sum_{i^2} p(i^1 i^2) \sum_{k^1 k^2} [a(i^1 k^1) + a(i^2 k^2)]p(k^1 k^2)$$

$$+ 2\gamma_{01} \sum_{k^2} p(i^1 k^2) \sum_{k^1 i^2} [a(i^1 k^1) + a(i^2 k^2)]p(k^1 i^2)\}$$

$$= \sum_{i^1 i^2} a(i^1 j^1)p(i^1 i^2) \sum_{k^1 k^2} p(k^1 k^2)[a(i^1 k^1) + a(i^2 k^2)] \ ,$$

it follows that we can write Equation E.1 as

$$V'^{12} = \sum_{i^1 i^2} \sum_{j^1 j^2} (V^{12})^{-2}\{p(i^1 i^2) \sum_{k^1 k^2} p(k^1 k^2)(a(i^1 k^1) + a(i^2 k^2))\}$$

$$\{a(i^1 j^1) + a(i^2 j^2)\}\{p(j^1 j^2) \sum_{k^1 k^2} p(k^1 k^2)(a(j^1 k^1) + a(j^2 k^2))\} \ . \qquad \text{(E.2)}$$

Let us relabel in a one-to-one fashion i^1, i^2 with i, and k^1, k^2 with k to get $a(i^1 k^1) + a(i^2 k^2) = a(i^1 k^1 i^2 k^2) = a(ik)$. Equation E.2 then has the same form as the single locus case where the different partial averages give rise to the result $V' \geq V$. It follows that with additive gene action between loci $V'^{12} \geq V^{12}$. It also follows from Kingman's (1961b) result in the case of a single locus, that $V'^{12} = V^{12}$ if and only

if $p(i^1i^2) = 0$ or

$$\sum_{j^1j^2} p(j^1j^2)[a(i^1j^1) + a(i^2j^2)] = V^{12} \qquad (E.3)$$

for all i^1, i^2.

The interesting thing to take note of in the argument presented here is that it is the assumption of additive gene action between loci which allows the Equation E.1 to take the form of the Equation E.2. The change in gamete frequencies in the case of no recombination between loci is, as apparent from the Equations D.1, given by

$$V^{12}p'(i^1i^2) = p(i^1i^2) \sum_{k^1k^2} p(k^1k^2)a(i^1k^1i^2k^2) \quad .$$

Thus, it is only in the case of no recombination between loci that the Equations E.1 and E.2 will be equivalent for arbitrary gene action between loci. This conclusion is in agreement with the result of Kojima and Kelleher (1961) that in the case of two alleles per locus, $V'^{12} \geq V^{12}$ holds in general only if there is no epistasis.

From the discussion on Equations D.4 and D.5, we have that $p(i^1i^2) = P(i^1)P(i^2)$ is always a solution to the Equations E.3, where $P(i^1)$ $(i^1 = 1, \ldots, m^1)$ is the solution to the single locus equations $\sum_{j^1} P(j^1)a(i^1j^1) = \sum_{i^1} \sum_{j^1} P(i^1)a(i^1j^1)P(j^1)$ and likewise $P(i^2)$ $(i^2 = i, \ldots, m^2)$ is the solution to the single locus equations at the second locus. We also have from the previous discussion (Section D of this chapter) that the Equations E.3 have an unique solution of the form $p(i^1i^2) = P(i^1)P(i^2)$ when $[a(i^1j^1) + a(i^2j^2)]$ is of full rank, so that in this case the Equations E.3 are equivalent to the Equations D.5. Hence, in the case of

additive gene action between loci, it follows that in the full rank case the population mean fitness increases until Hardy-Weinberg frequencies are reached at both loci. Hence, there are no equilibria of the form $P(i^1i^2) \neq P(i^1)(P(i^2)$ in the case of additive gene action between loci when the matrix $[a(i^1j^1) + a(i^2j^2)]$ is of full rank. We have not been able to obtain a proof of this result for the non-full rank case. However, we have that fitness increases until $\sum_{j^1j^2} a(i^1j^1i^2j^2)p(j^1j^2) = V^{12}$, which implies that at equilibrium the Equations D.1 will have to be of the same form. From a comparison of Equations D.1 with Equations D.4 and D.5, it seems that in most cases the only way in which the Equations D.1 will be equivalent to $\sum_{j^1j^2} a(i^1j^1i^2j^2)P(j^1j^2) = V^{12}$ (the Equations E.3), would be for $P(i^1i^2) = P(i^1)P(i^2)$ at equilibrium.

We summarize the results of this section in the theorem given below.

Theorem 3 In the case of additive gene action between loci and the matrix $[a(i^1j^1) + a(i^2j^2)]$ being of full rank, the population mean fitness increases until Hardy-Weinberg frequencies are reached at both loci.

2. The relationship between equilibria and stationary points in mean fitness in the case of additive gene action between loci

The fact that $V'^{12} \geq V^{12}$ does not necessarily imply that fitness will increase until the maximum of the function $V^{12}(p)$ is reached. From a comparison of the Equations D.2 and E.3 it is clear that the point at which $V'^{12} = V^{12}$ may equally be a saddle point. It is therefore necessary for us to prove that $p(i^1i^2) = P(i^1)P(i^2) > 0$ is a local maximum of $V^{12}(p)$ if $p(i^1i^2) = P(i^1)P(i^2) > 0$ is a stable equilibrium for the additive case.

We define a local maximum here such that if $P(i^1i^2)$ is a local maximum of $V^{12}(p)$, every point in a neighborhood of $P(i^1i^2)$ will give the function $V^{12}(p)$ a lesser or equal value than it has at $P(i^1i^2)$. We note that since $\sum_{i^1i^2} P(i^1i^2) = 1$, the function $V^{12}(p)$ will be defined on an open region for the $p(i^1i^2)$ which are such that $P(i^1i^2)$ are > 0. The necessity for an open region arises from our definition of a maximum. The inclusion of border points will cause no great difficulty, but a discussion on them is best deferred to a consideration of border points in conjunction with the survival of new genes in Section E.6.

Let us write

$$p(i^1i^2) = P(i^1)P(i^2) + e(i^1i^2) \quad ,$$

where $P(i^1)$ is the equilibrium frequency of the gene (i^1) and $P(i^2)$ is the equilibrium frequency of the gene (i^2). Then, on the assumption of the existence of a stable Hardy-Weinberg equilibrium under additive gene action between loci, it follows from the Equations D.4, and from $\sum_{i^1i^2} e(i^1i^2) = 0$, that

$$\begin{aligned} V^{12}(p) = & \sum_{i^1i^2} \sum_{j^1j^2} P(i^1)P(i^2)a(i^1j^1i^2j^2)P(j^1)P(j^2) \\ & + \sum_{i^1i^2} \sum_{j^1j^2} e(i^1i^2)a(i^1j^1i^2j^2)e(j^1j^2) \quad . \end{aligned} \tag{E.4}$$

Hence, $P(i^1)P(i^2)$ is a local maximum of $V^{12}(p)$ if and only if

$$\sum_{i^1i^2} \sum_{j^1j^2} e(i^1i^2)a(i^1j^1i^2j^2)e(j^1j^2) \leq 0 \quad , \tag{E.5}$$

for all $e(i^1i^2)$ satisfying $\sum_{i^1i^2} e(i^1i^2) = 0$.

We can write

$$\sum_{i^1 i^2} \sum_{j^1 j^2} e(i^1 i^2) a(i^1 j^1 i^2 j^2) e(j^1 j^2) = \sum_{i^1 j^1} e(i^1 .) a(i^1 j^1) e(j^1 .) + \sum_{i^2 j^2} e(.i^2) a(i^2 j^2) e(.j^2) , \tag{E.6}$$

where, e.g., $e(i^1 .) = \Sigma_{i^2} e(i^1 i^2)$, and consequently $\Sigma_{i^1} e(i^1 .) = 0$. We have from the assumptions necessary to write Equation E.4 in the given form that $\Sigma_{i^1} e(i^1 .) \Sigma_{j^1} P(j^1) a(i^1 j^1) + \Sigma_{i^2} e(.i^2) \Sigma_{j^2} P(j^2) a(i^2 j^2) = 0$. Since we assume $p(i^1 i^2) = P(i^1) P(i^2)$ to be stable, it follows from Theorem 2, Section D.2.c, that we can assume both single locus situations to give rise to stable equilibria, so that we can write

$$\sum_{i^1} e(i^1 .) \sum_{j^1} P(j^1) a(i^1 j^1) = 0 = \sum_{i^2} e(.i^2) \sum_{j^2} P(j^2) a(i^2 j^2) , \tag{E.7}$$

by reason of

$$\sum_{j^1} P(j^1) a(i^1 j^1) = V_E^1 \quad \text{and} \quad \sum_{j^2} P(j^2) a(i^2 j^2) = V_E^2 .$$

By using Equation E.7 and $\Sigma_{i^1} e(i^1 .) = 0$, it follows from the single locus case covered by Kingman (1961a) that $\Sigma_{i^1 j^1} e(i^1 .) a(i^1 j^1) e(j^1 .) \leq 0$ if and only if the matrix $[a(i^1 j^1)]$ has exactly one positive characteristic value. A similar result holds for $\Sigma_{i^2 j^2} e(.i^2) a(i^2 j^2) e(.j^2)$. The requirement of exactly one positive characteristic value in the fitness matrix is a necessary and sufficient condition for single locus stability. We proved, however, that in the case of additive gene action between loci the

equilibrium $p(i^1i^2) = P(i^1)P(i^2) > 0$ is stable if and only if the constituent single locus cases are stable. Hence if $p(i^1i^2) = P(i^1)P(i^2) > 0$ is a stable equilibrium it follows from Equations E.4, E.5, and E.6 that $P(i^1)P(i^2) > 0$ is a local maximum of

$$V^{12}(p) = \sum_{i^1i^2} \sum_{j^1j^2} p(i^1i^2)[a(i^1j^1) + a(i^2j^2)]p(j^1j^2) \quad .$$

We shall now prove that if $V^{12}(p)$ has a local maximum on the points with $P(i^1i^2) > 0$, then this maximum implies a stable equilibrium on the points where $P(i^1i^2) = P(i^1)P(i^2)$.

Let us denote the local maximum of $V^{12}(p)$ by V_E^{12}. We now assume for all $i^1 = 1,2, \ldots ,m^1$ and all $i^2 = 1,2, \ldots ,m^2$ that $P(i^1i^2) > 0$ and impose the restriction that $\sum_{i^1i^2} p(i^1i^2) = 1$. It then follows from the derivation of the Equations D.2 that it will be true at V_E^{12} for $\lambda = V_E^{12}$, that

$$\frac{\partial V^{12}(p)}{\partial p(i^1i^2)} = \sum_{j^1j^2} a(i^1j^1i^2j^2)p(j^1j^2) = \lambda \quad , \tag{E.8}$$

where $i^1 = 1,2, \ldots ,m^1$ and $i^2 = 1,2, \ldots ,m^2$.

If we compare the Equations E.8 with the Equations D.4 and D.5 it is easy to see that, under the assumption of additivity, $P(i^1i^2) = P(i^1)P(i^2)$ will always be a solution to the Equations E.8. Consequently the assumption of a local maximum implies that there exist points on the maximum for which it will be true that $P(i^1i^2) = P(i^1)P(i^2)$, where $P(i^1)$ and $P(i^2)$ are solutions to

$$\sum_{j^1} a(i^1j^1)P(j^1) = V_E^1 \quad \text{and} \quad \sum_{j^2} a(i^2j^2)P(j^2) = V_E^2 \quad , \tag{E.9}$$

respectively.

Since we assume $p(i^1i^2) = P(i^1i^2)$ to be a local maximum it follows analogously to the Equations E.4 and E.5 and by virtue of the Equations E.8 that

$$\sum_{i^1j^2} \sum_{i^2j^2} e(i^1i^2)a(i^1j^1)e(j^1j^2) + \sum_{i^1j^1} \sum_{i^2j^2} e(i^1i^2)a(i^2j^2)e(j^1j^2) \leq 0 \ . \quad \text{(E.10)}$$

It also follows from the Equations E.9 that

$$\sum_{i^1} e(i^1.) \sum_{j^1} a(i^1j^1)P(j^1) = 0 = \sum_{i^2} e(.i^2) \sum_{j^2} a(i^2j^2)P(j^2) \ , \quad \text{(E.11)}$$

where $\sum_{i^2} e(i^1i^2) = e(i^1.)$ and where $\sum_{i^1} e(i^1i^2) = e(.i^2)$.

Equation E.10 must hold for all $e(i^1i^2)$ such that $\sum_{i^1i^2} e(i^1i^2) = 0$. Let us assume that for some i^1, $e(i^1i^2) \neq 0$ for $i^2 = k^2$, and $e(i^1i^2) = 0$ for $i^2 \neq k^2$, where k^2 is an arbitrary value of i^2. From the assumption of $\sum_{i^1i^2} e(i^1i^2) = 0$, it follows immediately that $\sum_{i^1} e(i^1k^2) = 0$. If we now substitute the $e(i^1i^2)$ as specified above in the Equation E.10, we obtain the result that the relationship

$$\sum_{i^1j^1} e(i^1.)a(i^1j^2)e(j^1.) \leq 0 \ , \quad \text{(E.12)}$$

must hold if the Equation E.10 is to be satisfied for all $e(i^1i^2)$ such that $\sum_{i^1i^2} e(i^1i^2) = \sum_{i^1} e(i^1.) = 0$.

In the same way we can show that the Equation E.10 implies that

$$\sum_{i^2j^2} e(.i^2)a(i^2j^2)e(.j^2) \leq 0 \quad \text{(E.13)}$$

for all $e(i^1i^2)$ such that $\sum_{i^2} e(.i^2) = 0$.

The Equations E.9, E.11, E.12 and E.13 describe two single locus situations to which the theory developed by Kingman (1961a) can be applied. It follows from Kingman's results that the inequalities given as Equations E.12 and E.13 will hold if and only if the two single locus systems are :

stable. It follows from Theorem 2 that the stability of the single locus systems implies stability of the Hardy-Weinberg system associated with both loci.

The results which we obtained, thus far, in this section can now be summarized by Theorem 4.

Theorem 4 In the case of additive gene action between two loci the Hardy-Weinberg equilibrium is stable if and only if the mean fitness function has a local maximum.

It is also of interest to inquire under which conditions the function $V^{12}(p)$ will have a strict local maximum, that is, we inquire under which conditions every point in the neighborhood of the equilibrium point $P(i^1)P(i^2) > 0$ will give the function $V^{12}(p)$ a strictly lesser value than it has at $P(i^1)P(i^2)$. It was pointed out in our discussion following on the Equations D.4 that a general solution to equations of the type given in the Equations E.8 is given by

$$P = \lambda A^{+} \iota + (A^{+}A - I)Z \quad , \qquad (E.14)$$

where P is a vector with elements $P(i^1 i^2)$, A^{+} is the unique Moore inverse of the matrix A with elements $a(i^1 j^1 i^2 j^2)$, ι is the column vector with unit elements, I is the identity matrix and where Z is an arbitrary vector. It is clear that the restriction $\sum_{i^1 i^2} P(i^1 i^2) = 1$ forces $V_E^{12} = P'AP$ to be equal to the constant λ for all vectors P satisfying the Equations E.8. Hence, the requirement of a strict local maximum for $V^{12}(p)$ implies that the Equation E.14 must have a unique solution. This requires the matrix A (with elements $a(i^1 j^1 i^2 j^2)$) to be of full-rank, in which case the Moore inverse becomes the conventional inverse, A^{-1}.

It is clear that the assumption of full rank for A will also imply a strict local maximum of $V^{12}(p)$.

From the relationship between the Equations E.9 and the Equations D.4 and D.5 it is clear that if $[a(i^1j^1) + a(i^2j^2)]$ is of full rank, then $[a(i^1j^1)]$ and $[a(i^2j^2)]$ are of full rank. The reason for this is that the Equations E.9 always represent solutions to the Equations D.5, so that if the Equations D.5 have a unique solution, then the Equations E.9 must have unique solutions.

It is easy to see that in the case of the fitness matrix being of full rank Kingman's (1961a) single locus results on local maxima hold for strict inequalities so that a relationship between stable equilibria and strict local maxima can be established.

The same type of argument as that which was employed in the derivation of Theorem 4 can therefore be employed to prove Corollary 4.1.

Corollary 4.1 In the case of the matrix $[a(i^1j^1) + a(i^2j^2)]$ being of full rank the two-loci Hardy-Weinberg equilibrium is stable if and only if the mean fitness function has a strict local maximum, with all the internal points being positive.

Corollary 4.1 is pertinent to our understanding of the relationship between Theorems 3 and 4. In the case of a singular matrix of fitness-values the position on the possibility of non-Hardy-Weinberg equilibria remains ambiguous under the present approach, as follows from the remarks in the derivation of Theorem 3. It should also be noted that the method of approach which we followed in the derivation of Theorem 2 does not allow the consideration of non-Hardy-Weinberg equilibria.

So we have proved results giving the relationship between stable equilibria and maxima of the mean fitness function, substantially equivalent to those of Mulholland and Smith (1959) in the single locus case. In this respect, then, the introduction of another locus does not change the picture substantially as long as gene action remains additive between loci.

It is of some interest to note that all local maxima will be global maxima on the points where the gamete frequencies are greater than zero. This result follows immediately from the discussion on Equation E.14, where it was pointed out that at equilibrium the mean fitness is equal to a constant on all internal points.

3. The relationship between equilibria and stationary points in mean fitness in the case of multiplicative gene action between loci

In the relationship between equilibria and stationary points in mean fitness, the results of the multiplicative case are diametrically opposed to that of the additive case, for it is easy to show that the stable equilibrium, $p(i^1i^2) = P(k^1)P(i^2) > 0$, where these terms have the same meaning as before, does not constitute a local maximum of the mean fitness function, $V^{12}(p)$.

In the same way as in Equation E.4 we can write

$$V^{12}(p) = \sum_{i^1 i^1}^{\Sigma_2} \sum_{j^1 j^1}^{\Sigma_2} P(i^1)P(i^2)a(i^1j^1)a(i^2j^2)P(j^1)P(j^2)$$

$$+ \sum_{i^1 i^1}^{\Sigma_2} \sum_{j^1 j^1}^{\Sigma_2} e(i^1i^2)a(i^1j^1)a(i^2j^2)e(j^1j^2) \quad . \qquad \text{(E.15)}$$

As before we write the equilibrium fitness as V_E^{12}. From Equation E.15 it follows that V_E^{12} is a local maximum if and only if the quadratic form in the perturbations is less than or equal to zero. From Kingman's (1961a) work it is easy to see that the quadratic form in the perturbations will be less than or equal to zero if and only if the matrix $[a(i^1j^1)a(i^2j^2)]$ has only one positive characteristic root. We showed before (Section D.2.b) that the matix $[a(i^1j^1)a(i^2j^2)]$ is a Kronecker product of the matrices $[a(i^1j^1)]$ and $[a(i^2j^2)]$, and therefore that the characteristic values of $[a(i^1j^1)a(i^2j^2)]$ are equal to the product of the characteristic values of $[a(i^1j^1)]$ and $[a(i^2j^2)]$. We also have that stability of the form $P(i^1i^2) = P(i^1)P(i^2)$ implies the conditions for stability in the constituent single locus cases, and these necessary and sufficient conditions are that $[a(i^1j^1)]$ and $[a(i^2j^2)]$ have only one positive characteristic root each. It follows from the properties of Kronecker product matrices that the matrix $[a(i^1j^1)a(i^2j^2)]$ has $(m^1-1)(m^2-1) + 1$ positive characteristic values. Hence, the point $p(i^1i^2) = P(i^1)P(i^2) > 0$ cannot be a local maximum of $V^{12}(p)$ in the case of multiplicative gene action between loci.

4. Non-Hardy-Weinberg equilibria and maximum mean fitness

Let us consider the case of a two-loci model of two alleles per locus under the restriction that $a(i^1j^111) = a(i^1j^122)$ for $i^1 = 1,2$, $j^1 = 1,2$ and $a(11\ i^2j^2) = (22\ i^2j^2)$ for $i^2 = 1,2$ and $j^2 = 1,2$. We follow Lewontin (1964) in denoting $a(11\ 11) = a$, $a(11\ 12) = c$, $a(12\ 11) = b$ and $a(12\ 12) = d$.

Under these simplifying conditions Lewontin could prove that

$$V_E^{12} = V_{H\text{-}W}^{12} + 4D^2(a + d - b - c) \quad ,$$

where $D = [p(11)p(22) - p(12)p(21)]$, and where V_E^{12} denotes the equilibrium mean fitness and $V_{H\text{-}W}^{12}$ the mean fitness on the assumption of Hardy-Weinberg frequencies, or in other words, on the assumption that the gamete frequencies are equal to the product of the gene frequencies. We now have to note from the work of Lewontin and Kojima (1960) that non-Hardy-Weinberg stability of the foregoing system requires

$$(a + d - b - c) > 0 \quad .$$

It follows, for the system under discussion, that Lewontin's (1964) result implies that at any stable equilibrium of the form $p(i^1i^2) = P(i^1i^2) \neq P(i^1)P(i^2)$ the mean fitness at equilibrium is always greater than the mean fitness at the point $p(i^1i^2) = P(i^1)P(i^2)$.

This is quite a remarkable result, which will be of great importance to our understanding of the genetic structure of populations if it can be shown to be generally true. Attempts to generalize this theorem of Lewontin's have been unsuccessful this far.

In the case of additive gene action between loci it is easy to see that

$$a + d - b - c = 0 \quad .$$

This would suggest that in the additive case the mean fitness at equilibrium is always equal to that at the Hardy-Weinberg equilibrium point $P(i^1i^2) = P(i^1)P(i^2)$. This result can be seen to be true, since the equations giving maximal fitness (Equations E.3) and extreme values (Equations D.2 and E.8) have the same form and since we proved in the discussion on the Equations D.4 that $V^{12}(p)$ has the same value for all solutions to the equations mentioned. We also have that $p(i^1i^2)$

$= P(i^1)P(i^2)$ is always a solution to the Equations E.3. Hence, $V^{12}(p)$ has in all cases at equilibrium the value it would have at the point $P(i^1)P(i^2)$.

5. Local and global equilibria

In the case of additive gene action between loci it follows from the proofs of Theorems 3 and 4 that in the full-rank case $p(i^1i^2) = P(i^1)P(i^2) > 0$ is the only stable equilibrium, or in other words that $p(i^1i^2) = P(i^1)P(i^2) > 0$ is a global equilibrium.

In the case of multiplicative gene action between loci no result like $V'^{12} \geq V^{12}$ holds, so that the only way to determine whether $p(i^1i^2) = P(i^1)P(i^2)$ is a global equilibrium would be a direct examination of the Equations D.1. Unfortunately this is a system of cubic equations, for which literal solutions cannot be in general obtained. In the simple case of two alleles at each of two loci, and with the restriction that $a(11\ i^2j^2) = a(22\ i^2j^2)$ and $a(i^1j^1\ 11) = a(i^1j^1\ 22)$ for i^1 , j^1 , i^2 , $j^2 = 1,2$, Lewontin and Kojima (1960) found literal solutions together with their stability conditions. This result of Lewontin and Kojima shows that $p(i^1i^2) = P(i^1)P(i^2)$ is a global equilibrium if it is stable. A similar result holds in the cases considered in the numerical work of Nei (1964). Nei considered a case of multiplicative gene action between loci with the following selection patterns at the first and second loci respectively

(11)	(12)	(22)
$1 - s^1$	1	$1 - t^1$
(11)	(12)	(22)
$1 - s^2$	1	$1 - t^2$

where $t^1 = t^2 = 1$ and $s^1 = s^2 = \frac{1}{2}$. The recombination values taken into consideration are 0.01, 0.05, 0.10, 0.20, and 0.50. The product of the loads at each locus is equal to 0.111. Stable equilibria of the type $p(i^1i^2) = P(i^1)P(i^2)$ are found only in the cases of the recombination value being equal to 0.20 and 0.50, as we would expect from the inequality given as Equation D.20. These two stable equilibria are also global equilibria. It is only in the cases where $p(i^1i^2) = P(i^1)P(i^2)$ is unstable that two alternative stable equilibria of the type $p(i^1i^2) = P(i^1i^2) \neq P(i^1)P(i^2)$ exist as is also true in the model considered by Lewontin and Kojima (1960). These results give hope that the general stable local equilibria of the form $p(i^1i^2) = P(i^1)P(i^2)$ may also in the multiplicative case be global stable equilibria, but no method of proof seems to be evident.

6. The stability of border points and the survival of new genes

a. Border points in the general case The stability of border points is another instance where interesting contrasts between the single locus and two-loci cases exist. Let us, therefore, in a fashion analogous to the single locus case, consider the situation when $P(i^1i^2)$ may be equal to zero. We again confine ourselves to the case in which there is no position effect. Let I^1 be some subset of $(1,2, \ldots ,m^1)$ and I^2 be some subset of $(1,2, \ldots ,m^2)$, and let I be the cartesian product of the sets I^1 and I^2. Let us consider the general equilibrium $P(i^1i^2) > 0$ for $i^1 \in I^1$ and $i^2 \in I^2$, and $P(i^1i^2) = 0$ for $i^1 \notin I^1$ or $i^2 \notin I^2$. As before, let $p(i^1i^2) = P(i^1i^2) + e(i^1i^2)$. Then it follows that $e(i^1i^2) \geq 0$ for $i^1 \notin I^1$ or $i^2 \notin I^2$, and that $\sum_{i^1i^2} e(i^1i^2) = 0$, where the summation goes over all possible values of the ordered pair (i^1i^2).

In general, we will indicate in our notation when the summation is restricted to all members or to all non-members of the sets I^1, I^2, or I, whatever the case may be. Without any such indication the subscripts attached to the summation sign will indicate summation over all members of the set $(1,2, \ldots ,m^1)$ or $(1,2, \ldots ,m^2)$ or their cartesian product, as the case may be. From the Equations D.1 it follows that for $i^1 \epsilon I^1$ and $i^2 \epsilon I^2$ that

$$V^{12}(I)P(i^1i^2) = 2\gamma_{00}P(i^1i^2) \sum_{k^1k^2} P(k^1k^2)a(i^1k^1i^2k^2) + 2\gamma_{01} \sum_{k^1k^2} P(i^1k^2)a(i^1k^1i^2k^2)P(k^1i^2) \quad . \tag{E.16}$$

We also have that

$$\begin{aligned} V^{12}(p) &= \sum_{j^1j^2} \sum_{k^1k^2} [P(j^1j^2) + e(j^1j^2)]a(j^1k^1j^2k^2)[P(k^1k^2) + e(k^1k^2)] \\ &= V^{12}(I) + 2 \sum_{j^1j^2} \sum_{k^1k^2} e(j^1j^2)a(j^1k^1j^2k^2)P(k^1k^2) \\ &\quad + \sum_{j^1j^2} \sum_{k^1k^2} e(j^1j^2)a(j^1k^1j^2k^2)e(k^1k^2) \quad . \end{aligned} \tag{E.17}$$

Let us define

$$\sum_{k^1k^2} P(k^1k^2)a(j^1k^1j^2k^2) = V^{12}(I) + \alpha(j^1j^2) \quad , \tag{E.18}$$

and ignore products of e's. Then we can write Equations E.17 as

$$V^{12}(p) = V^{12}(I) + 2 \sum_{j^1j^2} e(j^1j^2)\alpha(j^1j^2) \quad . \tag{E.19}$$

Let us first consider the case where $i^1 \notin I^1$ and $i^2 \notin I^2$. We get on substituting $p(j^1j^2) = P(j^1j^2) + e(j^1j^2)$ with $P(j^1j^2) = 0$ for $j^1 \notin I^1$ or $j^2 \notin I^2$ in the Equations D.1 that

$$V^{12}(I)e'(i^1i^2) = 2\gamma_{00}e(i^1i^2) \sum_{k^1k^2} P(k^1k^2)a(i^1k^1i^2k^2) \quad ,$$

which by Equations E.18 and E.19 yield for $i^1 \notin I^1$ and $i^2 \notin I^2$

$$e'(i^1i^2) = \gamma_{00}e(i^1i^2)\left(1 + \frac{\alpha(i^1i^2)}{V^{12}(I)}\right) \quad . \tag{E.20}$$

It follows that for stability

$$2\gamma_{00}\left(1 + \frac{\alpha(i^1i^2)}{V^{12}(i)}\right) \leq 1 \quad ,$$

or

$$\alpha(i^1i^2) \leq \frac{\rho}{1-\rho} V^{12}(I) \quad . \tag{E.21}$$

Hence, it follows from the Equations E.18 that for the stability of the Equations E.20, we have to have

$$\sum_{k^1k^2} P(k^1k^2)a(i^1k^1i^2k^2) \leq \frac{1}{1-\rho} V^{12}(I) \quad , \tag{E.22}$$

for $i^1 \notin I^1$ and $i^2 \notin I^2$.

Next, consider the case $i^1 \in I^1$, $i^2 \notin I^2$. In the same way as for the Equations E.20, we get

$$\begin{aligned} V^{12}(I)e'(i^1i^2) = {} & 2\gamma_{00}e(i^1i^2) \sum_{k^1k^2} P(k^1k^2)a(i^1k^1i^2k^2) \\ & + 2\gamma_{01} \sum_{k^1k^2} P(i^1k^2)a(i^1k^1i^2k^2)e(k^1i^2) \quad . \end{aligned} \tag{E.23}$$

For the case $i^1 \notin I^1$ and $i^2 \in I^2$, we get

$$V^{12}(I)e'(i^1i^2) = 2\gamma_{00}e(i^1i^2) \sum_{k^1k^2} P(k^1k^2)a(i^1k^1i^2k^2)$$
$$+ 2\gamma_{01} \sum_{k^1k^2} e(i^1k^2)a(i^1k^1i^2k^2)P(k^1i^2) \quad . \tag{E.24}$$

There does not seem to be much to be said about Equations E.23 and E.24 in the general case. In the case where $i^1 \in I^1$ and $i^2 \in I^2$, we get on substituting $p(i^1i^2) = P(i^1i^2) + e(i^1i^2)$ in the Equations D.1 and utilizing Equations E.16, E.17, and E.18 that

$$V^{12}(I)e'(i^1i^2) = 2\gamma_{00}e(i^1i^2) \sum_{k^1k^2} P(k^1k^2)a(i^1k^1i^2k^2)$$
$$+ 2\gamma_{00}P(i^1i^2) \sum_{k^1k^2} e(k^1k^2)a(i^1k^1i^2k^2)$$
$$+ 2\gamma_{01} \sum_{k^1k^2} e(i^1k^2)a(i^1k^1i^2k^2)P(k^1i^2) \tag{E.25}$$
$$+ 2\gamma_{01} \sum_{k^1k^2} P(i^1k^2)a(i^1k^1i^2k^2)e(k^1i^2)$$
$$- 2P(i^1i^2) \sum_{j^1j^2} \alpha(j^1j^2)e(j^1j^2) \quad .$$

The Equations E.25, E.24, and E.23 certainly show why it is so difficult to make general statements about stability when linkage is involved. The only result that can be deduced in general is that the inequality of the Equations E.22 must hold for the stability of border points.

b. Border points in a system with Hardy-Weinberg structure We now return to the more amenable cases of multiplicative and additive gene action between loci, where the existence of Hardy-Weinberg equilibria is possible.

Let us therefore assume that for $P(i^1i^2) > 0$, $P(i^1i^2) = P(i^1)P(i^2)$. Hence, the Equations E.18 become in the case of multiplicative gene action

$$\sum_{k^1}\sum_{k^2} a(j^1k^1)a(j^2k^2)P(k^1)P(k^2) = V^1(I^1)V^2(I^2) + \alpha(j^1j^2) = [V^1(I^1) + \alpha(j^1)][V^2(I^2) + \alpha(j^2)] . \tag{E.26}$$

where $V^1(I^1)$ and $V^2(I^2)$ have the same definition as V_E^1 and V_E^2 in the Equations D.6 and where $\alpha(j^1j^2) = 0$ if $j^1 \epsilon I^1$ and $j^2 \epsilon I^2$ and where $\alpha(j^1) = 0$ if $j^1 \epsilon I^1$ and $\alpha(j^2) = 0$ if $j^2 \epsilon I^2$, by Equations D.4 and D.6. In the case of additive gene action between loci the Equations E.18 become

$$\sum_{k^1}\sum_{k^2} [a(j^1k^1) + a(j^2k^2)]P(k^1)P(k^2) = V^1(I^1) + V^2(I^2) + \alpha(j^1j^2) = [V^1(I^1) + \alpha(j^1)] + [V^2(I^2) + \alpha(j^2)] . \tag{E.27}$$

where $\alpha(j^1j^2)$, $\alpha(j^1)$, and $\alpha(j^2)$ are as defined in the Equations E.26 and where their properties follow from Equations D.4 and D.5.

1) Border points under multiplicative gene action We now proceed to discuss Equations E.23 and E.24 under the assumption of multiplicative gene action between loci and under the assumption that the Equations E.22 hold and so $e(i^1i^2) \to 0$ for $i^1 \notin I^1$ and $i^2 \notin I^2$. The case of $i^1 \epsilon i^1$ and $i^2 \notin I^2$ (the Equations E.23) becomes then

$$V^{12}(I)e'(i^1i^2) = (1-\rho)e(i^1i^2) \sum_{k^1} a(i^1k^1)P(k^1) \sum_{k^2} a(i^2k^2)P(k^2)$$

$$+ \rho P(i^1) \sum_{k^2} a(i^2k^2)P(k^2) \sum_{k^1 \epsilon I^1} a(i^1k^1)e(k^1i^2) \quad .$$

Hence, by the Equations E.26 it follows that

$$V^1(I^1)V^2(I^2)e'(i^1i^2) = (1-\rho)e(i^1i^2)V^1(I^1)[V^2(I^2) + \alpha(i^2)]$$

$$+ \rho P^1(i^1)[V^2(I^2) + \alpha(i^2)] \sum_{k^1 \epsilon I^1} a(i^1k^1)e(k^1i^2) \quad .$$

Let us now make the transformations

$$z(i^1i^2) = \frac{e(i^1i^2)}{\sqrt{P(i^1)}} \quad \text{and} \quad b(i^1j^1) = \frac{1}{V^1(I^1)} a(i^1j^1)\sqrt{P(i^1)P(j^1)} \quad ,$$

to get

$$V^1(I^1)V^2(I^2)z'(i^1i^2)\sqrt{P(i^1)} = (1-\rho)z(i^1i^2)\sqrt{P(i^1)}\, V^1(I^1)[V^2(I^2)+\alpha(i^2)].$$

$$+ \rho\sqrt{P(i^1)}\, V^1(I^1)[V^2(I^2)+\alpha(i^2)] \sum_{k^1 \epsilon I^1} z(k^1i^2)b(i^1k^1).$$

We note that $z(i^1i^2) \geq 0$ when $i^1 \epsilon I^1$ and $i^2 \notin I^2$, as follows from the definition, $p(i^1i^2) = P(i^1)P(i^2) + e(i^1i^2)$, where $P(i^1)P(i^2) = 0$ for $i^1 \notin I^1$ or $i^2 \notin I^2$. It follows that, for a specific value of $i^2 \notin I^2$, all the vectors z cannot be characteristic vectors of the matrix $[b(i^1j^1)]$, since some of these vectors are known to have negative elements. However, for a fixed i^2 we can study the behaviour of the z-vectors in terms of a linear combination of vectors in the set of characteristic vectors of the matrix $[b(i^1j^1)]$. We will denote the characteristic vectors of

$[b(i^1j^1)]$ by the vectors y with elements $y(i^1i^2)$.

The characteristic vectors of $[b(i^1j^1)]$ will change in the same fashion from generation to generation as the vectors z , so that we get

$$y'(i^1i^2) = y(i^1i^2)(1 + \frac{\alpha(i^2)}{V^2(I^2)})(1-\rho(1-\lambda^1)) \ , \qquad (E.28)$$

where $\sum_{k^1} y(k^1i^2)b(i^1k^1) = \lambda^1 y(i^1i^2)$, where λ^1 is a characteristic value of the matrix $[b(i^1j^1)]$ and where $|\lambda^1| \leq 1$ as was shown in Section C.

Since the coefficient of $y(i^1i^2)$ is positive, the system given by Equations E.28 will be stable if and only if

$$[1 + \frac{\alpha(i^2)}{V^2(I^2)}][1-\rho(1-\lambda^1)] \leq 1 \ ,$$

which is equivalent to

$$\alpha(i^2) \leq \frac{\rho(1-\lambda^1)}{1-\rho(1-\lambda^1)} V^2(I^2) \qquad (E.29)$$

and which in turn implies, from Equations E.26 that

$$\sum_{k^2} a(i^2k^2)P(k^2) \leq \frac{V^2(I^2)}{1-\rho(1-\lambda^1)} \ , \qquad (E.30)$$

or

$$\sum_{k^1k^2} a(i^1k^1)a(i^2k^2)P(k^1)P(k^2) \leq \frac{V^1(I^1)V^2(I^2)}{1-\rho(1-\lambda^1)} \ , \qquad (E.31)$$

for $i^1 \epsilon I^1$ and all i^2 .

For $i^1 \notin I^1$ and $i^2 \epsilon I^2$ we can in the same way deduce that if the Equations E.22 hold the Equations E.24 will give rise to a stable system if and only if

$$\alpha(i^1) \leq \frac{\rho(1-\lambda^2)}{1-\rho(1-\lambda^2)} V^1(I^1) \quad , \tag{E.32}$$

or equivalently from the Equations E.26, if and only if

$$\sum_{k^1} a(i^1k^1)P(k^1) \leq \frac{V^1(I^1)}{1-\rho(1-\lambda^2)} \, , \tag{E.33}$$

or

$$\sum_{k^1k^2} a(i^1k^1)a(i^2k^2)P(k^1)P(k^2) \leq \frac{V^1(I^1)V^2(I^2)}{1-\rho(1-\lambda^2)} \quad , \tag{E.34}$$

for all i^1 and for $i^2 \in I^2$.

Still under the assumption of multiplicative gene action between loci and with the further assumption that the inequalities of Equations E.22, E.29, and E.32 hold so that $e(j^1j^2) \to 0$ for $j^1 \notin I^1$ or $j^2 \notin I^2$, the Equations E.25 become for $i^1 \in I^1$ and $i^2 \in I^2$

$$\begin{aligned} V^1(I^1)V^2(I^2)e'(i^1i^2) &= 2\gamma_{00}e(i^1i^2)V^1(I^1)V^2(I^2) + 2\gamma_{00}P(i^1)P(i^2) \sum_{\substack{k^1 \in I^1 \\ k^2 \in I^2}} a(i^1k^1)a(i^2k^2)e(k^1k^2) \\ &\quad + 2\gamma_{01}P(i^2) \sum_{k^2 \in I^2} e(i^1k^2)a(i^2k^2)V^1(I^1) \\ &\quad + 2\gamma_{01}P(i^1) \sum_{k^1 \in I^1} e(k^1i^2)a(i^1k^1)V^2(I^2) \, , \end{aligned} \tag{E.35}$$

by virtue of the Equations E.26. The Equations E.35 are of precisely the same form as Equations D.8. Hence, a necessary and sufficient condition for Hardy-Weinberg equilibrium of the genetic system defined on the points

$(1,2, \ldots, m^1)$ and $(1,2, \ldots, m^2)$ is that the inequalities of Equations E.22, E.29, and E.32 or their equivalents hold and that for the system defined on the points belonging to the set I, $\lambda^1 \leq 0$, $\lambda^2 \leq 0$ and that the Equation D.16 hold, where λ^1 and λ^2 are non-unity characteristic roots of the matrices $[b(i^1j^1)]$ and $[b(i^2j^2)]$ for which $i^1, j^1 \epsilon I^1$ and $i^2, j^2 \epsilon I^2$, and where $[b(i^1j^1)]$ and $[b(i^2j^2)]$ are as defined previously in Section D.

2) The increase of a new gene under multiplicative gene action

From the foregoing the necessary and sufficient condition for the initial increase of a new gene which is introduced in small numbers in a population can be deduced. Consider the genes (α^1) as being introduced at the first locus. All other genes are supposed to belong to the set I.

We note here that, as $\sum_{i^2 \epsilon I^2} e(i^1i^2) \neq 0$ for $i^1 \notin I^1$, the vector with elements $z(i^1i^2)$, which we defined at the beginning of the previous section, will have to have the characteristic vector with elements $c\sqrt{P(i^2)}$ in its basis of characteristic vectors of $[b(i^2j^2)]$, where c is a multiplicative constant. We showed in Sections C and D that the characteristic root associated with the vector with elements $c\sqrt{P(i^2)}$ has the value of unity. An analogous result holds for the case of $i^1 \epsilon I^1$ and $i^2 \notin I^2$.

It follows that a necessary and sufficient condition for the genes (α^1), $\alpha^1 > 0$, to start to die out immediately is, from the Equations E.33 and E.34 with $\lambda^2 = 1$, that $\sum_{k^1} a(\alpha^1k^1)P(k^1) < V^1(I^1)$ and $\sum_{k^1k^2} a(\alpha^1k^1)a(j^2k^2) \times P(k^1)P(k^2) < V^1(I^1)V^2(I^2)$. The new genes (α^1) will increase if and only if

$$\sum_{k^1} a(\alpha^1k^1)P(k^1) > V^1(I^1) \qquad \text{(E.36)}$$

and

$$\sum_{k^1} \sum_{k^2} a(\alpha^1k^1)a(j^2k^2)P(k^1)P(k^2) > V^1(I^1)V^2(I^2) \quad . \qquad \text{(E.37)}$$

Equations E.36 and E.37 pertain only to an initial increase of a new gene in a population with Hardy-Weinberg equilibrium. One might suspect that if a new Hardy-Weinberg equilibrium were possible, the population would eventually attain it, but at present there is no way of proving it.

3) Border points under additive gene action The case of additive gene action between loci is somewhat more tractable than the other cases because of the results following from the relationship $V'^{12} \geq V^{12}$. We proceed to discuss Equations E.23 and E.24 under the assumptions of $p(i^1i^2) = P(i^1)P(i^2)$ at equilibrium and of $a(i^1j^1i^2j^2) = a(i^1j^1) + a(i^2j^2)$, and under the assumption that the Equations E.22 hold so that $e(i^1i^2) \to 0$ for $i^1 \notin I^1$ and $i^2 \notin I^2$. The case of $i^1 \in I^1$ and $i^2 \notin I^2$ (Equations E.23) becomes then by the application of the Equations E.27

$$[V^1(I^1) + V^2(I^2)]e'(i^1i^2) = (1-\rho)e(i^1i^2)[V^1(I^1) + V^2(I^2) + \alpha(i^2)] + \rho P(i^1)[\sum_{k^1 \in I^1} a(i^1k^1)e(k^1i^2) + (V^2(I^2) + \alpha(i^2))(\sum_{k^1 \in I^1} e(k^1i^2))] .$$

Then we get in the same fashion as was used for the derivation of the Equations E.28, that we can write

$$[V^1(I^1) + V^2(I^2)]y'(i^1i^2) = (1-\rho)y(i^1i^2)[V^1(I^1) + V^2(I^2) + \alpha(i^2)] + \rho[V^1(I^1) \sum_{k^1 \in I^1} y(k^1i^2)b(i^1k^1) + \sqrt{P(i^1)}\,(V^2(I^2) + \alpha(i^2))(\sum_{k^1 \in I^1} y(k^1i^2)\sqrt{P(k^1)})].$$

(E.38)

For a specific value of i^2, let y^{12} be a characteristic vector of the matrix $[b(i^1j^1)]$. It follows that $\sum_{k^1 \epsilon I^1} y(k^1i^2)b(i^1k^1) = \lambda^1 y(i^1i^2)$, where λ^1 is, as before, a characteristic root of $[b(i^1j^1)]$. From the properties of characteristic vectors of $[b(i^1j^1)]$ in Section C it follows that $\sum_{k^1 \epsilon I^1} y(k^1i^2)\sqrt{P(k^1)} = 0$, for vectors associated with the non-unity characteristic roots. Hence we write for elements of these vectors, that

$$y'(i^1i^2) = y(i^1i^2)\left[\frac{(1-\rho)[V^1(I^1) + V^2(I^2) + \alpha(i^2)] + \rho V^1(I^1)\lambda^1}{V^1(I^1) + V^2(I^2)}\right] .$$

Since $|\lambda^1| \leq 1$ the coefficient of $y(i^1i^2)$ in the foregoing equations is always positive so that we have for elements of the vectors under consideration that a necessary and sufficient condition for stability of the genetic system on the points $i^1 \epsilon I^1$ and $i^2 \notin I^2$ is

$$\left[\frac{(1-\rho)[V^1(I^1) + V^2(I^2) + \alpha(i^2)] + \rho V^1(I^1)\lambda^1}{V^1(I^1) + V^2(I^2)}\right] \leq 1 ,$$

or

$$\alpha(i^2) \leq \frac{\rho[V^1(I^1)(1-\lambda^1) + V^2(I^2)]}{(1-\rho)} ,$$

or by the Equations E.27

$$\sum_{j^2} a(i^2j^2)P(j^2) \leq \frac{1}{1-\rho} V^2(I^2) + \frac{\rho}{1-\rho} V^1(I^1)(1-\lambda^1) , \qquad (E.39)$$

or

$$\sum_{j^1}\sum_{j^2} [a(i^1j^1) + a(i^2j^2)]P(j^1)P(j^2) \leq \frac{1}{1-\rho}[V^1(I^1)(1-\lambda^1\rho) + V^2(I^2)] , \qquad (E.40)$$

for all i^2 and for $i^1 \epsilon I^1$.

In the same way we can show that a necessary condition for stability of the genetic system on the points $i^1 \notin I^1$ and $i^2 \epsilon I^2$ is that

$$\alpha(i^1) \leq \frac{\rho}{1-\rho} [V^1(I^1) + V^2(I^2)(1-\lambda^2)]$$

or

$$\sum_{j^1} a(i^1 j^1) P(j^1) \leq \frac{1}{1-\rho} V^1(I^1) + \frac{\rho}{1-\rho} V^2(I^2)(1-\lambda^2) \qquad (E.41)$$

or

$$\sum_{j^1 j^2} (a(i^1 j^1) + a(i^2 j^2)) P(j^1) P(j^2) \leq \frac{1}{1-\rho} [V^1(I^1) + V^2(I^2)(1-\lambda^2\rho)] \ , \qquad (E.42)$$

for all i^1 and for $i^2 \epsilon I^2$.

For the same reason as we noted in the derivation of the Equations E.36, we also have to consider in the Equations E.38 the case where the elements of the vector y^{12} are equal to $c\sqrt{P(i^1)}$, where c is a constant common to all elements of the vector. In this case it is easy to see that we obtain from the Equations E.38 that

$$[V^1(I^1) + V^2(I^2)]c' \sqrt{P(i^1)} = c\sqrt{P(i^1)} \ [V^1 I^1) + V^2(I^2) + \alpha(i^2)] \qquad (E.43)$$

We see that the Equations E.43 result in stability under the assumption of $\alpha(i^2) \leq 0$, which give rise to the same result as that embodied in the Equations E.39 under the assumption of $\rho = 0$.

For the case of $i^1 \notin I^1$ and $i^2 \epsilon I^2$ it is possible to derive in the same way that for stability we have to have that

$$\alpha(i^1) \leq 0 \ . \qquad (E.44)$$

The Equations E.44 give rise to a result equivalent to that of Equations E.41 and E.42 with $\rho = 0$.

We note further that if the inequalities given by Equations E.43 and E.44 hold, the inequalities of Equations E.39 and E.41 or their equivalents must also hold, since $|\lambda^a| \leq 1$ for $a = 1,2$ and $0 \leq \rho \leq \frac{1}{2}$.

Hence, under the assumption of additive gene action between loci and with the further assumption that the inequalities of Equations E.22, E.43 and E.44 hold so that $e(j^1j^2) \to 0$ for $j^1 \notin I^1$ or $j^2 \notin I^2$, the Equations E.25 become for $i^1 \epsilon I^1$ and $i^2 \epsilon I^2$ of precisely the same general form as the Equations D.7, from which the Equations D.21 were derived. This argument is entirely analogous to that employed in the multiplicative case. It follows that a necessary and sufficient condition for the Hardy-Weinberg stability of the genetic system defined on the points $(1,2,\ldots,m^1)$ and $(1,2,\ldots,m^2)$ is that the inequalities of Equations E.22, E.43, and E.44 or their equivalents hold and that for the system defined on the points belonging to the set I, $\lambda^1 \leq 0$ and $\lambda^2 \leq 0$, where λ^1 and λ^2 are non-unity characteristic roots of the matrices $[b(i^1j^1)]$ and $[b(i^2j^2)]$ for which $i^1, j^1 \epsilon I^1$ and $i^2, j^2 \epsilon I^2$, and where $[b(i^1j^1)]$ and $[b(i^2j^2)]$ are as defined previously in Section D. This last condition is precisely that the two constituent single locus cases on the set I must both be stable.

We are now in a position to extend Theorem 4 of Section E.2 to the case of border points under the assumption of additive gene action between loci. We shall prove, first, that if $p(i^1i^2) = P(i^1)P(i^2)$ is a stable equilibrium on the set of points $(1, \ldots, m^1)$ and $(1, \ldots, m^2)$ with $P(i^1)P(i^2) > 0$ only if $i^1 \epsilon I^1$ and $i^2 \epsilon I^2$, where I^1 and I^2 are

subsets of $(1, \ldots, m^1)$ and $(1, \ldots, m^2)$ respectively, then $p(i^1i^2)$ $= P(i^1)P(i^2)$ is a local maximum of $V^{12}(p)$ in which only the $p(i^1i^2)$'s with $P(i^1)P(i^2) > 0$ are allowed to vary.

The proof follows immediately from the fact that with stability on the border points, $e(i^1i^2) \to 0$ for $i^1 \notin I^1$ or $i^2 \notin I^2$, where $e(j^1j^2)$ are defined as before.

On the converse side let us assume that we have a local maximum on the points $i^1 \in I^1$ and $i^2 \in I^2$, while the $p(i^1i^2)$ are not allowed to vary away from zero on the points $i^1 \notin I^1$ or $i^2 \notin I^2$. This implies that $e(i^1i^2)$ $= 0$ always for $i^1 \notin I^1$ or $i^2 \notin I^2$. This is enough to imply the stability of the border points, whereas the local maximum on the internal points will imply stability there.

We state the foregoing results as Corollary 4.2.

Corollary 4.2 The equilibrium $p(i^1i^2) = P(i^1)P(i^2)$ with $P(i^1)P(i^2) > 0$ for $i^1 \in I^1$ and $i^2 \in I^2$ and $P(i^1)P(i^2) = 0$ for $i^1 \notin I^1$ or $i^2 \notin I^2$ is stable if and only if it is a local maximum of $V^{12}(p)$ in which only the $p(i^1i^2)$'s with $i^1 \in I^1$ and $i^2 \in I^2$ are allowed to vary away from zero values.

In the full rank case, i.e., where the matrix $[a(i^1j^1) + a(i^2j^2)]$ is of full rank for $i^1, j^1 \in I^1$ and $i^2, j^2 \in I^2$, it follows from Corollary 4.1 that Corollary 4.2 can be stated in terms of a strict local maximum of $V^{12}(p)$.

The foregoing extension to Theorem 4 and the results, following on the Equations E.44, that preceded it, effectively summarizes the results that can be obtained with the inclusion of border points under additivity between loci. The conclusions following the Equations E.35 effectively summarized the results that can be obtained with the inclusion of border

points under the assumption of multiplicative gene action between loci.

4) The survival of new genes under additivity From the foregoing the conditions for the increase of new genes and their persistence can be deduced. Suppose that a system $P(i^1)P(i^2) > 0$, ($i^1 = 1, \ldots, m^1-1$ and $i^2 = 1, \ldots, m^2-1$), is in a stable non-trivial equilibrium. Then, if V^{12} is the equilibrium mean fitness we know that

$$\sum_{j^1} \sum_{j^2} P(j^1)P(j^2)[a(i^1 j^1) + a(i^2 j^2)] = V^{12} \quad , \qquad \text{(E.45)}$$

for all i^1 and all i^2.

Now introduce small proportions of the genes (m^1) and (m^2) into the population. Clearly $(P(1)P(1), \ldots, P(m^1-1)P(m^2-1), \underbrace{0\ 0\ 0\ 0\ 0}_{m^1+m^2-1})$ is an equilibrium for the system and is stable if by the Equations E.22

$$\sum_{j^1 j^2} P(j^1)P(j^2)[a(j^1 m^1) + a(j^2 m^2)] < \frac{1}{1-\rho} V^{12}$$

and by the Equations E.44

$$\sum_{j^1} P(j^1)a(m^1 j^1) < V^1$$

and by the Equations E.43

$$\sum_{j^2} P(j^2)a(m^2 j^2) < V^2 \quad .$$

If this is the case, the system will resume its original equilibrium, so that (m^1) and (m^2) will die out. If any of the foregoing inequalities are reversed the equilibrium is not stable, and hence will not be resumed after perturbation.

We now assume that the following inequalities hold:

$$\Sigma_{j^1} P(j^1)a(m^1 j^1) > V^1 \qquad \text{(E.46)}$$

and

$$\Sigma_{j^2} P(j^2)a(m^2 j^2) > V^2 \quad . \qquad \text{(E.47)}$$

Then if the system on the points $(1,2, \ldots ,m^1)$ and $(1,2, \ldots ,m^2)$ is stable, the genes m^1 and m^2 will increase until the equilibrium frequencies of this system are reached. In this case, then, the new genes will persist without displacing any of the other genes.

If the system on $(1,2, \ldots ,m^1)$ and $(1,2, \ldots ,m^2)$ is not stable, there is, at present, no telling in advance what may happen. The new genes may increase at first and later perhaps die out. However, we proved that a two-loci system under additivity is stable if and only if its constituent single locus systems would have been stable on their own, so that we can use some of Kingman's (1961a) results here. We know that the matrix $[a(i^1 j^1)]$ in the single locus case can give rise to a stable equilibrium if and only if it has only one positive characteristic value. Now if $[a(ij)]$ has p positive characteristic values, it was shown by Kingman that a least $(p-1)$ genes must die out before equilibrium can be reached.

If only one of the Equations E.46 and E.47 hold, then only one of the genes (m^1) and (m^2) will have a chance to become established. Hence, we can conclude that a necessary and sufficient condition for a new gene, say (m^1), to increase is

$$\Sigma_{j^1} P(j^1)a(m^1 j^1) > V^1 \quad ,$$

and that a sufficient condition for it to persist is that the system in which it is included must be stable. The foregoing condition corresponds to that of Kingman (1961a) for the single locus case.

F. Hardy-Weinberg and other Equilibria of Gene Frequencies with Selection and Mutation

1. The single locus case

An examination of the equations giving the change in gene frequencies (the Equations B.11) will show then to be a system of cubic equations, for which, except in the simplest cases, no literal solutions are available. For the purpose of taking mutation into account, we will, therefore, have to make do with approximations of various sorts. However, the two-allele case will be examined in detail and an example of Haldane's (1927b) will be worked.

a. Stability conditions We again make the transformation $p(i) = P(i) + e(i)$, and writing $V_E = \sum_{ij} P(i)a(ij)P(j)$, we get by substitution in the Equations B.11 that

$$V_E e'(i) = \sum_s \sum_t v(si)e(s)a(st)P(t) + \sum_s \sum_t v(si)P(s)a(st)e(t) - 2P(i) \sum_s \sum_t e(s)a(st)P(t) , \quad (F.1)$$

where the perturbations $e(i)$ are considered to be small enough so that squares and cross products can be neglected. As before we define $P(i)$ to be the equilibrium value when, in the Equations B.11, $p'(i) = p(i) = P(i)$. We assume that, when $i \neq j$, the $v(ij)$ are so small that their cross products with the perturbations are negligible. The Equations F.1 then become

$$V_E e'(i) = e(i) \sum_t a(it)P(t) + P(i) \sum_s a(is)e(s) - 2P(i) \sum_s \sum_t e(s)a(st)P(t) \ . \tag{F.2}$$

In the case of two alleles we denote $v(11) = 1-u$, $v(12) = u$ and $v(21) = v$ and remembering that $e(1) = -e(2)$ and $P(1) = 1 - P(2)$ we can write the Equations F.1 as

$$\begin{aligned} V_E e'(1) = e(1)[&2P(1)P(2)(a(11) - 2a(12) + a(22)) \\ &+ 2P(1)(u(a(12) - a(11)) + v(a(22) - a(12))) \\ &+ a(12)(1-u+v) - 2va(22)] \ . \end{aligned} \tag{F.3}$$

When the mutation rates are small Equation F.3 can be written as

$$V_E e'(1) = e(1)[2P(1)P(2)(a(11) - 2a(12) + a(22)) + a(12)] \ . \tag{F.4}$$

On intuitive grounds one would expect that, in a multi-allelic situation, if one of the homozygotes, say (11), is superior to all other gene combinations, then there would be an equilibrium of the sort where $P(1)$ would be near to unity and where the other gene frequencies would be small. Since this situation is of importance in the consideration of the mutation load, it will be considered here. With $P(i)$ small for $i \neq 1$, so that $P(i)e(j)$ can be neglected for $i \neq 1$, we get from the Equations F.2 for $i \neq 1$ that

$$a(11)e'(i) = e(i)a(il) \ .$$

Hence, since $a(11) > a(1i)$ for $i \neq 1$, the genetic system on the points $i \neq 1$, will be stable. Since $\sum_i e(i) = 0$ and the $e(i)$ for $i \neq 1$ will approach zero after sufficient time elapsed, the same would be true for $e(11)$. Therefore the genetic system described by the assumed gene

frequencies will be stable.

b. The approximation of a selection system with mutation by a system without mutation Since the inclusion of mutation complicates matters so much it is of interest to consider under which circumstances a system without mutation would be a good approximation to a selection system in which mutation is taken into account.

Let us define an equilibrium

$$p(i) = P(i) = Q(i) + Z(i) \quad , \tag{F.5}$$

where $Q(i) > 0$ if $i \in I$ and $Q(i) = 0$ if $i \notin I$, where I is, as before, some subset of $(1,2, \ldots ,m)$. The $Q(i) > 0$ are further defined as a solution to the Equations B.12 which are the equations giving the change in gene frequencies in the absence of mutation. As before, we adopt the convention that a subscript to a summation sign indicates that summation takes place over all members of the set $(1,2, \ldots ,m)$. If summation takes place over a subset of $(1,2, \ldots ,m)$ only, the subset under consideration will be indicated.

It now follows that we can write

$$\begin{aligned} V(P) &= \sum_{ij} (Q(i) + Z(i))a(ij)(Q(j) + Z(j)) \\ &= \sum_{ij} Q(i)a(ij)Q(j) + 2 \sum_{ij} Z(i)a(ij)Q(j) + \sum_{ij} Z(i)a(ij)Z(j) \\ &= V(I) + 2 \sum_{ij} Z(i)a(ij)Q(j) + V(Z) \quad . \end{aligned} \tag{F.6}$$

We define

$$\sum_{j} a(ij)Q(j) = V(I) + \alpha(i) \quad . \tag{F.7}$$

We then substitute the Equations F.5 and F.6 in the Equations F.2 to get

$$(V(I) + 2\sum_{ij} Z(i)a(ij)Q(j) + V(Z))e'(i) = e(i)\sum_{t} a(it)(Q(t) + Z(t))$$

$$+ (Q(i) + Z(i))\sum_{s} a(is)e(s) - 2(Q(i) + Z(i))\sum_{s}\sum_{t} e(s)a(st)(Q(t) + Z(t)). \qquad (F.8)$$

Consider now the case where $i \notin I$ and where the $Z(i)$ are so small that $V(Z)$ and products of the type $Z(i)e(j)$ can be neglected. The Equations F.8 then become by virtue of the Equations F.7

$$V(I)e'(i) = e(i)(V(I) + \alpha(i)) \quad . \qquad (F.9)$$

In the case where $i \in I$ and where the assumption of stability on the system defined by F.9 causes $e(i) \to 0$ for $i \notin I$, the Equations F.8 become by reason of the Equations F.7

$$V(I)e'(i) = e(i)V(I) + Q(i)\sum_{s \in I} a(is)e(s) \quad . \qquad (F.10)$$

As can be seen from Section C the Equations F.9 and F.10 agree precisely with Kingman's (1961a) equations for the stability of the selection system defined on the points $(1,2, \ldots ,m)$. Thus, when the $Z(i)$ are small the systems with or without mutation have precisely the same stability conditions.

What now remains to be done is to try and determine conditions under which it would be likely for the $Z(i)$ of the Equations F.5 to be small.

Let us substitue $p(i) = Q(i) + z(i)$, where the $Q(i)$ are the selection without mutation equilibrium and the $z(i)$ are deviations from that equilibrium, in $V(p)p'(i) = \sum_{st} v(si)p(s)a(st)p(t)$ (the Equations B.11)

to get by reason of the Equations F.7 that

$$V(p)p'(i) = \sum_{st} v(si)(Q(s) + z(s))(V(I) + \alpha(s))$$

$$+ \sum_{s} v(si)(Q(s) + z(s)) \sum_{t} a(st)z(t) \quad . \qquad \text{(F.11)}$$

If we neglect terms containing squares and products of the $z(i)$ with the $v(ij)$ $(i \neq j)$ the Equations F.11 become

$$V(p)(Q(i) + z'(i)) = z(i)(V(I) + \alpha(i)) + Q(i) \sum_{t} a(it)z(t)$$

$$+ \sum_{s} v(si)Q(s)V(I) \quad , \qquad \text{(F.12)}$$

since $v(ii) = 1 - \sum_{j \neq i} v(ij)$ by definition, and since $Q(i) = 0$ for $i \notin I$.

Now, if $i \notin I$, then the Equations F.12 become

$$V(I)z'(i) = z(i)(V(I) + \alpha(i)) + \sum_{s} v(si)Q(s)V(I)$$

or

$$z'(i) = z(i)(1 + \frac{\alpha(i)}{V(I)}) + \sum_{s} v(si)Q(s) \quad . \qquad \text{(F.13)}$$

The pertinent solutions to the recurrence equations given in Equation F.13 are

$$z_k(i) = z_0(i)(1 + \frac{\alpha(i)}{V(I)})^k - \frac{V(I)}{\alpha(i)} (\sum_{s} v(si)Q(s))[1 - (1 + \frac{\alpha(i)}{V(Q)})^k] \ ,$$

where k denotes the k-th generation and 0 the 0-th generation. If k becomes large we obtain, if $1 + \frac{\alpha(i)}{V(Q)} < 1$, that

$$z(i) \to - \frac{V(I)}{\alpha(i)} (\sum_{s \in I} v(si)Q(s)) \quad . \qquad \text{(F.14)}$$

We note that if we want to use the results of Section C, we need the restriction that $\sum_{i \in I} z(i) = 0$. This, of course, implies that we have to have the $z(i)$ for $i \notin I$ to be approximately zero. It follows that the mutation probabilities in the Equations F.14 must be very small.

In the case where $i \in I$, and under the assumption that $z(i) \to 0$ if $i \notin I$, we can write the Equations F.12 by virtue of the Equations F.6 and F.7 as

$$z'(i) = z(i) + \frac{Q(i)}{V(I)} \sum_{t \in I} a(it)z(t) + \sum_{s} v(si)Q(s) - Q(i) . \qquad (F.15)$$

Let us now transform $y(i) = \frac{z(i)}{\sqrt{Q(i)}}$ and $b(ij) = \frac{1}{V(I)} a(ij)\sqrt{Q(i)Q(j)}$ for $i,j \in I$, so that the foregoing equations become

$$y'(i) = y(i) + \sum_{j \in I} b(ij)y(j) + \frac{1}{\sqrt{Q(i)}}\left(\sum_{s \in I} v(si)Q(s) - Q(i)\right) . \qquad (F.16)$$

The correct solution of the Equations F.16 is more difficult than the homogeneous case represented by the Equations C.6. In the case of the solution of the Equations C.6 we used a somewhat heuristic approach which had the advantage of laying bare the nature of the characteristic roots and vectors. This procedure was especially helpfull in the more complex cases covered in Section D. However, for the solution of the Equations F.16 we will need a more rigorous matrix algebra approach to the solution of the recurrence equations.

In matrix notation we can write the Equations F.16 as

$$y_t = (I + B)y_{t-1} + c$$

or

$$y_t = Hy_{t-1} + c \quad ,$$

where the subscript t denotes the generation and y_t is the $p \times 1$ column vector with elements $y(i)$, B is the $p \times p$ matrix with elements $b(ij)$, I is the $p \times p$ identity matrix and c is a vector with elements

$$\frac{1}{\sqrt{Q_i}} \left(\sum_{s \in I} v(si)Q(s) - Q(i) \right) ,$$

and where $i \in I$, $i = 1,2, \ldots ,p$.

We recall from Section C that B has a characteristic root equal to unity and that the vector with elements $\sqrt{Q(i)}$ is associated with it. From the restriction that $\sum_{i \in I} z(i) = \sum_{i \in I} \sqrt{Q(i)}\, y(i) = 0$, we have that the perturbation vectors belong to the subspace generated by the characteristic vectors of B associated with the roots unequal to unity. We assume that the rows and columns of B are arranged in such a fashion that the root $\lambda_p = 1$ occupies the last position in the diagonal matrix composed of the roots of B.

It is clear that it follows that we can express any perturbation vector as

$$y_t = w_t(1)\, X_1 + w_t(2)\, X_2 + \ldots + w_t(p-1)\, X_{p-1} ,$$

where $X_1, \ldots ,X_{p-1}$ are orthogonal $p \times 1$ vectors associated with the non-unity characteristic roots of B and where the $w_t(i)$ $i = 1, \ldots ,p-1$ are scalars. It follows that we can write

$$w_t = X'HXw_{t-1} + X'c ,$$

where H is the $p \times p$ matrix $(B + I)$, X is a $p \times (p-1)$ matrix with columns consisting of characteristic vectors of B and where w_t is a $(p-1) \times 1$ vector.

It is clear that the characteristic vectors of B are also the characteristic vectors of H, with the characteristic roots of H being equal to $\lambda_s + 1$, where the λ_s are the non-unity characteristic roots of B.

It follows that we can write $w_t = D_1 w_{t-1} + X'c$, where D_1 is a $(p-1) \times (p-1)$ diagonal matrix consisting of the characteristic roots of H. The solution of the foregoing difference equation is known to be $w_t = D_1^t w_o + (I-D_1^t)(I-D_1)^{-1}X'c$. We now assume the genetic system on the points belonging to the set I to be stable, that is, we assume $D_1^t = 0$ after a sufficient number of generations elapsed, where 0 is the matrix with elements equal to zero. The necessary and sufficient condition for this to be the case is that $|1 + \lambda_s| < 1$, which is in agreement with the condition for stability derived in Section C.

It follows that with t becoming very large we can write

$$w_t = - D^{-1}X'c \quad ,$$

or (F.17)

$$y_t = - XD^{-1}X'c \quad ,$$

where D is a diagonal matrix consisting of the non-unity characteristic roots of B, which we denoted by λ_s, $s = 1,2, \ldots ,p-1$. The vector c we defined before as the vector with elements

$$\frac{1}{\sqrt{Q(i)}} \left(\sum_{s \in I} v(si)Q(s) - Q(i) \right) \quad ,$$

for $i = 1,2, \ldots ,p$.

It should be emphasized that the foregoing results apply only to the description of a selection system with mutation in a neighborhood of a selection system without mutation that is small enough so that the squares of deviations from the selection without mutation equilibrium are negligible. From the form of the Equations F.14 it is clear that the mutation rates from the internal points to the boundary points must be very small. In the case of Equation F.17 it is more difficult to judge how large the mutation rates in the vector c must be so that it will be true that $z(i)z(j) = \sqrt{Q(i)}\, y(i) \sqrt{Q(j)}\, y(j)$ will be negligible as we assumed in the derivation of the Equations F.10 and F.16. One would guess that the mutation probabilities will have to be rather small.

To extend the foregoing local result to a global result will be extremely difficult. The reason for this is that with mutation and selection, the result that the mean fitness of a population never decreases from one generation to the next, no longer holds. It was pointed out in Sections C and E that all global properties derive from this non-decreasing property of the mean fitness.

However, in the next section we will derive some two-allele results that are pertinent here. In the discussion following on the Equation F.22 we will show that in the case of non-zero mutation rates and non-zero fitnesses there can be only one stable equilibrium if $x = a(11) - 2a(12) + a(22)$ is negative. It is known that the condition $a(12) > a(11)$ and $a(12) > a(22)$ is necessary and sufficient for stability in a two-allele selection system. It follows that if the mutation rates are small enough, the selection without mutation equilibrium will be an adequate approximation to the global equilibrium of the selection system with mutation.

For the purposes of the mutation load we are also interested in how a border point selection system would be modified by mutation. We write

$$p(i) = Q(i) + z(i) \quad ,$$

where

$$Q(i) = 1 \quad \text{if} \quad i = 1$$

and where

$$Q(i) = 0 \quad \text{if} \quad i \neq 1 \quad .$$

By substitution in the Equations F.12 we obtain

$$\overline{V(p)}(Q(i) + z'(i)) = v(1i)a(11) + z(i)a(il) + Q(i) \sum_t a(it)z(t) \quad .$$

Hence, if $i \neq 1$ we have by assuming small $z(i)$'s that

$$a(11)z'(i) = v(1i)a(11) + z(i)a(il) \quad ,$$

which implies for $a(il) < a(11)$ that $\lim(z(i)) = \frac{a(11)v(1i)}{a(11)-a(1i)}$ for all $i \neq 1$. Since $\sum_i z(i) = 0$, it follows therefore that

$$\lim(z(1)) = - \sum_{i \neq 1} \frac{a(11)v(1i)}{a(11) - a(1i)} \quad . \tag{F.18}$$

It is of some interest to note that for the parametrization used in the two allele system (Section c) we have that

$$\lim(z(1)) = - \frac{u}{hs} \quad ,$$

which gives $p(2) = \frac{u}{hs}$ in agreement with Equation F.19.

c. The two-allele system

1) Approximate solutions in a special case In load theory an important case is that of no overdominance, where the fitnesses are denoted as $a(11) = 1$, $a(12) = 1 - hs$ and $a(22) = 1 - s$ for $0 \leq h \leq 1$, $0 < s \leq 1$. We denote the gene frequencies as $p(1)$ and $p(2)$, and write $v(11) = 1 - u$, $v(12) = u$, $v(21) = v$, and $v(22) = 1 - v$. Then we can deduce from the Equations B.11 that the change in $p(2)$ denoted as $\Delta p(2)$ must be

$$\Delta p(2) = \frac{1}{D} [(1-v)p(2)(1-sp(2)-hsp(1)) + up(1)(1-hsp(2))$$

$$- p(2)(1-2hsp(1)p(2)-sp^2(2)] \ ,$$

where

$$D = 1-2hsp(1)p(2) - sp^2(2) \ .$$

To get the equilibrium frequency we put $\Delta p(2) = 0$. The resulting cubic equation has a general solution in terms of cube roots which is not very informative. We therefore resort to approximations. For the range of values assumed for h and s we expect on the basis of the theory of selection without mutation that, for the usual values of u and v , $p(1)$ will be very near to unity after sufficient time elapsed. $p(2)$ will then be so small that $p^2(2)$ can be considered negligible compared to $p(2)$ and $p^3(2)$ negligible compared to $p^2(2)$. Also the magnitude of $up(2)$ and $vp(2)$ will be considered small enough to be negligible. In the case of $h \gg p(2)$ the equation $\Delta p(2) = 0$ becomes under these conditons approximately

$$-hsp(2) + u = 0$$

or

$$p(2) = \frac{u}{hs} \ . \qquad \text{(F.19)}$$

In the case $h = 0$ we get approximately

$$p^2(2) = \frac{u}{s}$$

or

$$p(2) = \sqrt{\frac{u}{s}} \quad . \tag{F.20}$$

If we neglect the cubic terms in p_2 in $\Delta p(2) = 0$, together with terms containing $hp^2(2)$, we can examine the resulting quadratic in order to try and see under which conditions the Equation F.20 will be a satisfactory approximation. This procedure, then, makes the assumption of h being of the same order as $p(2)$. The result of this analysis shows that we have to consider terms containing $u^{3/2}$, $v^{3/2}$ and $u^{1/2}hs^{3/2}$ to be of negligible magnitude. We conclude, therefore, that h must be of the same order as u, v and $p(2)$ in order for the Equation F.20 to be a satisfactory solution to the equation $\Delta p(2) = 0$.

By substituting the solutions given by Equations F.19 and F.20 in Equation F.4 we can test for the stability of the given genetic systems. In the case of $p(2) = P(2) = \frac{u}{hs}$, Equation F.4 becomes

$$e'(1) = e(1)[1 - 2u - 2\frac{u}{h} - hs][1 + 2u]$$

or

$$e'(1) = e(1)[1 + 2u - 2\frac{u}{h} - hs + 2u - 2hu] ,$$

from which it is apparent that the system will be stable for $hs > 2u$, since $0 < h \leq 1$ by assumption. In the case of $p_2 = P_2 = \sqrt{\frac{u}{s}}$ Equation F.4 becomes

$$e'(1) = e(1)[1 + 2u - 2\sqrt{us}][1 + u] \doteq e(1)[1 + 3u - 2\sqrt{us}] \quad ,$$

so that the system will be stable if and only if $2\sqrt{us} > 3u$.

2) The location and number of roots in the equilibrium equation

Let us again consider two alleles (1) and (2) with frequencies $p(1)$ and $p(2)$. As before, we denote the fitnesses of the different genotypes by $a(11)$, $a(12)$ and $a(22)$, and we deduce the change in $p(2)$ from the Equations B.11 and denote it as $\Delta p(2)$. Putting $\Delta p(2) = 0$ and writing $v(12) = u$ and $v(21) = v$, we get

$$[a(11) - 2a(12) + a(22)]p^3(2) + [a(12)(3-v+u)-(1-v)a(22)-(2+u)a(11)]p^2(2)$$
$$+ [a(11)(1+2u)-(1-v+u)a(21)]p(2) - ua(11) = 0 \quad . \qquad (F.21)$$

Let us define

$$x = a(11) - 2a(12) + a(22)$$
$$y = a(11) - a(21)$$
$$z = a(22) - a(21) \quad .$$

It follows that $x = y + z$, so that $\Delta p(2) = 0$ becomes

$$xp^3(2) - [y(2+u) + z(1-v)]p^2(2) + [y(1+u) + ua(11) + va(21)]p_2 - ua(11) = 0 \quad .$$
$$(F.22)$$

From the facts that if $p(2) = 0$, then $\Delta p(2) = -ua(11)$, and if $p(2) = 1$, then $\Delta p_2 = va(22)$, and from the assumption of $ua(11) > 0$ and of $va(22) \geq 0$, it follows that $\Delta p(2)$ has at least one real root in the interval $[0,1]$.

We now have to consider the three cases of $x > 0$, $x = 0$, $x < 0$. First, consider the case of $x > 0$. It follows by the definition of the quantities that one or both of y and z must be positive. Assume y

to be positive. Then if z is non-negative or if z is negative and $y(2+u) > -z(1-v)$, it follows by Descartes' rule of signs that there can be at most three positive real roots. Also, in this case there can be no negative real roots.

We would now like to know under which circumstances $p(2) = 1$ is an upper bound for the real roots of $\Delta p(2) = f(p(2))$. We require that $f(1) > 0$, $f'(1) > 0$, $f''(1) > 0$ and that $f'''(1) > 0$. Now, $f'(p(2)) = 3xp^2(2) - 2[y(2+u) + z(1-v)]p(2) + y(1+u) + ua(11) + va(21)$, $f''(p(2)) = 6xp(2) - 2[y(2+u) + z(1-v)]$ and $f'''(p(2)) = 6x$.

We have that $f(1) = va(22)$ and that $f'(1) = z(1+2v) + a(12)(u+v)$, which will be greater than zero if

$$z(1+2v) > -a(12)(u+v) ,$$

or

$$-z(1+2v) < a(12)(u+v) ,$$

or if

$$a(12) - a(22) < \frac{a(12)(u+v)}{(1+2v)} .$$

Also,

$$f''(1) = 2y(1-u) + 2z(2+v) ,$$

which will be greater than zero if

$$y(1-u) > -z(2+v) ,$$

or if

$$x > -z(1+v) + yu ,$$

and $f'''(1) = 6x$, which will be positive by assumption.

Now we thus have that for the case x positive the necessary conditions for three real roots in $[0,1]$ are that:

(i) $y(2+u) > -z(1-v)$ if z is negative, or $z \geq 0$,

(ii) $a(12) - a(22) < \frac{a(12)(u+v)}{(1+2v)}$,

(iii) $x > -z(1+v) + yu$.

If we add that the discriminant of the cubic equation $\Delta p(2) = 0$ must be greater or equal to zero it will ensure the existence of three real roots, which if the foregoing three conditions are fulfilled will all fall in the acceptable interval $[0,1]$.

In the case where x and y are positive and z negative with $-z(1-v) > y(2+u)$, there is by the rule of signs only one positive real root, which must be between zero and one.

The case of $x > 0$ has now been covered in enough generality, because if y is negative and z positive, one can obviate the difficulty simply by relabeling the genotypes.

Second, if $x = 0$, there will have to be two real roots for Equation F.22. If $ua(11) \neq 0$ and $va(22) \neq 0$, it follows from the values of $\Delta p(2)$ at $p(2) = 0$ and $p(2) = 1$ that there can be only one real root in the interval from zero to one. With either $ua(11)$ or $va(22)$ or both equal to zero, there may be two real roots in $[0,1]$.

Third, when $x < 0$, we again distinguish two cases:

(i) both y and z are negative,

(ii) only one of y and z is negative.

(i) If both y and z are negative we get from Descartes' rule that there can only be one negative real root with two positive real roots.

If $ua(11) \neq 0$ and $va(22) \neq 0$, we get only one root in $[0,1]$.

(ii) Without loss of generality we can assume y to be positive. Again we obtain the same result as in (i).

We thus come to the conclusion that unless $ua(11)$ and/or $va(22)$ are equal to zero, we can have more than one root in the interval $[0,1]$ only when $x > 0$ and when

(i) $y(2+u) > -z(1-v)$ if z is negative, or $z \geq 0$, and

(ii) $a(12) - a(22) < \frac{a(12)(u+v)}{(1+2v)}$, and

(iii) $x > -z(1+v) + yu$, and where we need the discriminant to be positive to ensure the existence of three real roots.

We now consider the influence of the different modes of gene action on the number of acceptable roots for the equation $\Delta p(2) = 0$. Let us first consider the case where y is positive and $z \leq 0$, so that we can use the parametrization $a(11) = 1$, $a(12) = 1 - hs$, $a(22) = 1 - s$, $0 < s < 1$ and $0 < h \leq 1$. Then, $x = 2hs - s$, $y = hs$ and $z = hs - s$. Hence, x is positive only when $h > \frac{1}{2}$, and therefore it is only when $h > \frac{1}{2}$ that there exists a possibility for three acceptable real roots to $\Delta p(2) = 0$, under the present parametrization. The other case where x can be positive is when both homozygotes are superior to the heterozygote. In the case of "overdominance" x is negative so that in most cases only one real root can exist for the equation $\Delta p(2) = 0$.

Haldane (1927b) considered the case where $a(11) = 1 + k$, $a(12) = 1$, $a(22) = 1$ and $k = .008$ with mutation rates $u = .000400$ and $v = .000001$. He found three real roots, which is in accordance with the given three necessary conditions. In this case the three roots of $\Delta p(2)$ are

(i) $p(2) = 0.053,926$

(ii) $p(2) = 0.949,824$

(iii) $p(2) = 0.997,368$.

Substituting these equilibria in Equation F.3 we find respectively that (i) $e'(2) = .993298\, e(2)$, that (ii) $e'(2) = 1.000341\, e(2)$ and that (iii) $e'(2) = .999640\, e(2)$. We therefore conclude, as did Haldane, that equilibria (i) and (iii) are the only stable ones. The possible evolutionary implications of this result are discussed by Haldane (op. cit.)

3) Some comments on the three and higher number of allele cases

An attempt was made to handle the three-allele case along the same lines as the two allele case. For this purpose let us denote the three alleles by (1), (2) and (3). We find then that we have only two independent equations for the change in gene frequency, since $p(1) + p(2) + p(3) = 1$, and, hence, $\Delta p_1 + \Delta p_2 + \Delta p_3 = 0$.

The problem can therefore be handled as follows:

(i) Substitute $p(3) = 1 - p(1) - p(2)$ in the two equations for $\Delta p(1)$ and $\Delta p(2)$, and

(ii) Write both $\Delta p(1)$ and $\Delta p(2)$ as polynomials in $p(1)$ in which $p(2)$ is regarded as a constant.

We write

$$\Delta p_1 = ap^3(1) + bp^2(1) + cp(1) + d = 0$$

$$\Delta p_2 = jp^2(1) + kp(1) + \ell = 0 .$$

In order for these two equations to have a common root in the variable $p(1)$ Sylvester's determinant must be equal to zero, i.e.

$$a^2D = \begin{vmatrix} a\ell - cj & b\ell - ck - dj & - dk \\ ak - bj & a\ell - cj & - dj \\ j & k & \ell \end{vmatrix},$$

must be equal to zero, where D denotes Sylvester's determinant.

By putting $D = 0$ we get a polynomial in $p(2)$, the roots of which can then be found. The roots of $D = 0$ with $0 \leq p(2) \leq 1$ can then be substituted into the equations $\Delta p(1) = 0$ and $\Delta p(2) = 0$ and the roots common to both equations in the interval $0 \leq p(1) \leq 1$ will then be regarded as acceptable solutions.

However, as is readily apparent from the outline of the argument presented here, the three-allele situation is too complex to yield any intelligible answers by the application of the methods employed in the two allele case. Numerical answers for specific cases can, however, be readily obtained by the method outlined here. In the case of more than three alleles substitution with the help of Sylvester's determinant will probably be too clumsy so that alternative electronic computer methods will have to be employed.

To test for stability in the three-allele case we substitute $e(3) = - e(1) - e(2)$ in the Equations F.2 to get the following two independent equations:

$$\begin{aligned} V_E e'(1) = e(1)\{\sum_{r=1}^{3}[(v(11)a(1r) - v(31)a(3r) + (v(r1)a(r1) - v(r1)a(r3)) \\ - 2P(1)(a(1r) - a(3r))]P(r)\} + e(2)\{\sum_{r=1}^{3}[v(21)a(2r) - v(31)a(3r) \\ + (v(r1)a(r2) - v(r1)a(r3)) - 2P(1)(a(2r) - a(3r))]P(r)\}, \end{aligned}$$

$$V_E e'(2) = e(1)\{\sum_{r=1}^{3}[(v(12)a(1r) - v(32)a(3r)) + (v(r2)a(r1) - v(r2)a(3r)$$
$$- 2P(2)(a(1r) - a(3r))]P(r)\}$$
$$+ e(2)\{\sum_{r=1}^{3}[v(22)a(2r) - v(32)a(3r) + v(r2)a(r2) - v(r2)a(r3)$$
$$- 2P(2)(a(2r) - a(3r))]P(r)\} .$$

In matrix notation we can write that

$$\begin{bmatrix} e'(1) \\ e'(2) \end{bmatrix} = \begin{bmatrix} c(11) & c(12) \\ c(21) & c(22) \end{bmatrix} \begin{bmatrix} e(1) \\ e(2) \end{bmatrix} .$$

If the characteristic values of the matrix $[c(ij)]$ are less than unity, the small deviations from equilibria will decrease, so that stability can be assumed.

2. The two-loci case

a. Introduction In the case of two loci with mutation and selection we again have a system of cubic equations (the Equations B.7), the solution of which will give the equilibrium state of gamete frequencies in a population. In the same way as in the case of selection without mutation we will inquire under which circumstances a description of the population in terms of gene grequencies instead of gamete frequencies will be legitimate.

b. Stability conditions Let us again make the transformation $p(i^1i^2) = P(i^1i^2) + e(i^1i^2)$, where as before the $P(i^1i^2)$ are a equilibrium point (i.e., a solution of the Equations B.7 with $p'(i^1i^2) = p(i^1i^2)$) and where the $e(i^1i^2)$ are small perturbations. Again we write $V_E^{12} = \sum_{i^1i^2}\sum_{j^1j^2} P(i^1i^2)a(i^1j^1i^2j^2)P(j^1j^2)$. From the Equations B.7 it

follows then that by neglecting squares and products in the perturbations, we can write

$$V_E^{12}e'(j^1j^2) = 2\gamma_{00}\sum_{i^1i^2}v(i^1j^1i^2j^2)P(i^1i^2)\sum_{k^1k^2}e(k^1k^2)a(i^1k^1i^2k^2)$$

$$+ 2\gamma_{00}\sum_{i^1i^2}v(i^1j^1i^2j^2)e(i^1i^2)\sum_{k^1k^2}P(k^1k^2)a(i^1k^1i^2k^2)$$

$$+ 2\gamma_{01}\sum_{i^1k^2}v(i^1j^1k^2j^2)\sum_{k^1i^2}e(i^1i^2)a(i^1k^1i^2k^2)P(k^1k^2) \qquad \text{(F.23)}$$

$$+ 2\gamma_{01}\sum_{i^1k^2}v(i^1j^1k^2j^2)\sum_{k^1i^2}P(i^1i^2)a(i^1k^1i^2k^2)e(k^1k^2)$$

$$- 2P(j^1j^2)\sum_{i^1i^2k^1k^2}e(i^1i^2)a(i^1k^1i^2k^2)P(k^1k^2) \quad .$$

Hence, if the mutation rates are small enough so that the cross products of the $v(i^1j^1i^2j^2)$ with $i^1 \neq j^1$ or $i^2 \neq j^2$ with the perturbations $e(i^1i^2)$ are negligible, the Equations F.23 become

$$V_E e'(j^1j^2) = 2\gamma_{00}P(j^1j^2)\sum_{k^1k^2}e(k^1k^2)a(j^1k^1j^2k^2)$$

$$+ 2\gamma_{00}e(j^1j^2)\sum_{k^1k^2}P(k^1k^2)a(j^1k^1j^2k^2)$$

$$+ 2\gamma_{01}\sum_{k^1k^2}e(j^1k^2)a(j^1k^1k^2j^2)P(k^1j^2) \qquad \text{(F.24)}$$

$$+ 2\gamma_{01}\sum_{k^1k^2}e(k^1j^2)a(j^1k^1k^2j^2)P(j^1k^2)$$

$$- 2P(j^1j^2)\sum_{i^1i^2k^1k^2}e(i^1i^2)a(i^1k^1i^2k^2)P(k^1k^2) \quad .$$

For the purpose of the mutation load we postulate an equilibrium described as $p(j^1j^2) = P(j^1j^2) = P(j^1)P(j^2)$, with both $P(j^1)$ and $P(j^2)$ having frequencies near to unity for $j^1 = 1$ and $j^2 = 1$ and with all other gene frequencies near to zero. We want to know whether such an equilibrium will be stable.

Let us first consider the case where $j^1 \neq 1$ and $j^2 \neq 1$. By the assumption that products of the type $e(i^1i^2)P(k^1k^2)$ are negligible if $k^1 \neq 1$ or $k^2 \neq 1$ the Equations F.24 become

$$a(1111)e'(j^1j^2) = 2\gamma_{00}e(j^1j^2)a(j^1 1 j^2 1) \quad , \tag{F.25}$$

for $j^1 \neq 1$, $j^2 \neq 1$. We have that $2\gamma_{00} \leq 1$, so that if $a(1111) > a(j^1 1 j^2 1)$ for all $j^1 \neq 1$, $j^2 \neq 1$, it will be true that $e(j^1j^2) \rightarrow 0$ after enough generations of selection elapsed. The second case that should be considered is where $j^1 = 1$ and $j^2 \neq 1$. Under these circumstances we derived on neglecting terms like $e(i^1i^2)P(k^1k^2)$, if $k^1 \neq 1$ or $k^2 \neq 1$, and assuming that $e(i^1i^2)P(11)$ is approximately equal to $e(i^1i^2)$, that

$$a(1111)e'(j^1j^2) = 2\gamma_{00}e(j^1j^2)a(j^1 1 j^2 1) + 2\gamma_{01} \sum_{k^1} e(k^1j^2)a(j^1k^1 1 j^2) \ .$$

Hence, if we assume no position effect (see Section B.3) and that $e'(j^1j^2) \rightarrow 0$ for all $j^1 \neq 1$ and $j^2 \neq 1$ we obtain

$$a^{12}(1111)e'(j^1j^2) = e(j^1j^2)a(j^1 1 j^2 1) \quad , \tag{F.26}$$

for $j^1 = 1$ and $j^2 \neq 1$. It can easily be seen that the Equations F.26 also hold for $j^1 \neq 1$ and $j^2 = 1$. By the argument given for the Equations F.25, the genetic system on the points $j^1 = 1$, $j^2 \neq 1$ and on the points

$j^1 \neq 1$, $j^2 = 1$ will be stable if $a(1111) > a(1j^1 1j^2)$ for the specified j^1's and j^2's. Since $\sum_{i^1 i^2} e(i^1 i^2) = 0$, it follows that the system will be stable for $j^1 = 1$ and $j^2 = 1$ if all the other points are stable. We conclude that the postulated genetic system will be stable if $a(1111) > a(1j^1 1j^2)$ for $j^1 \neq 1$ or $j^2 \neq 1$.

c. The approximation of a selection system with mutation by a system without mutation In the case of two loci the same motivation as in the single locus case exists to try to approximate a mutation-selection system by a selection system with Hardy-Weinberg structure of gene frequencies.

Let us define an equilibrium

$$p(i^1 i^2) = P(i^1 i^2) = P(i^1)P(i^2) + Z(i^1 i^2) , \qquad (F.27)$$

where the $P(i^1 i^2)$ occur at the point where $p'(i^1 i^2) = p(i^1 i^2)$ in the Equations B.7, and where the $P(i^1)P(i^2)$ are a solution to the Equations D.3 with $P(i^1)P(i^2) > 0$ only if $i^1 \epsilon I^1$ and $i^2 \epsilon I^2$. As in Section E I^1 is some subset of $(1,2, \ldots ,m^1)$ and I^2 is some subset of $(1,2, \ldots ,m^2)$. As before, I is the cartesian product of I^1 and I^2. Again it is understood that $P(i^1)$ is the frequency of the gene (i^1) and $P(i^2)$ is the frequency of the gene (i^2) .

In the same way as in Equations E.17 and E.18, we write

$$V^{12}(P^{12}) = V^{12}(I) + 2 \sum_{\substack{j^1 \epsilon I^1 \\ \text{or} \\ j^2 \epsilon I^2}} Z(j^1 j^2)\alpha(j^1 j^2) + V^{12}(Z) \quad . \qquad (F.28)$$

We also define $\alpha(i^1)$ and $\alpha(i^2)$ in the same way as they were defined before in Equations E.26 and E.27.

We now substitute the Equations F.27 in the Equations F.24 to get by virtue of the Equations E.18 that

$$\begin{aligned} V^{12}(P^{12})e'(j^1j^2) &= 2\gamma_{00}[P(j^1)P(j^2) + Z(j^1j^2)] \sum_{k^1k^2} e(k^1k^2)a(j^1k^1j^2k^2) \\ &+ 2\gamma_{00}e(j^1j^2) \sum_{k^1k^2} [P(k^1)P(k^2) + Z(k^1k^2)]a(j^1k^1j^2k^2) \\ &+ 2\gamma_{01} \sum_{k^1k^2} e(j^1k^2)a(j^1k^1j^2k^2)[P(k^1)P(j^2) + Z(k^1j^2)] \\ &+ 2\gamma_{01} \sum_{k^1k^2} e(k^1j^2)a(j^1k^1j^2k^2)[P(j^1)P(k^2) + Z(j^1k^2)] \\ &- 2[P(j^1)P(j^2) + Z(j^1j^2)][\sum_{\substack{i^1\epsilon I^1 \\ i^2\epsilon I^2}} e(i^1i^2)\alpha(i^1i^2) \text{ or} \\ &+ \sum_{i^1i^2k^1k^2} e(i^1i^2)a(i^1k^1i^2k^2)Z(k^1k^2)] \ . \end{aligned} \tag{F.29}$$

Since $j^1 \notin I^1$ and $j^2 \notin I^2$ imply $P(j^1)P(j^2) = 0$ and since we assume that the $Z(i^1i^2)$ are small enough so that their products with each other and their cross products with the $e(i^1i^2)$ can be ignored, it follows from the Equations F.28 that the Equations F.29 become

$$e'(j^1j^2) = 2\gamma_{00}e(j^1j^2)[1 + \frac{\alpha(j^1j^2)}{V^{12}(I)}] \ . \tag{F.30}$$

Now if $j^1 \epsilon I^1$ and $j^2 \notin I^2$ and if $e(i^1i^2) \to 0$ for all $i^1 \notin I^1$ and $i^2 \notin I^2$ the Equations F.29 become under the same conditions on the $Z(i^1i^2)$ as before

$$V^{12}(I)e'(j^1j^2) = 2\gamma_{00}e(j^1j^2) \sum_{\substack{k^1\in I^1\\ k^2\in I^2}} P(k^1)P(k^2)a(j^1k^1j^2k^2)$$

$$+ 2\gamma_{01} \sum_{\substack{k^1\in I^1\\ k^2\in I^2}} e(k^1j^2)a(j^1k^1j^2k^2)P(j^1)P(k^2) \quad . \qquad \text{(F.31)}$$

In the same way under the same conditions as in the Equations F.31 we get for $j^1 \notin I^1$ and $j^2 \in I^2$ that

$$V^{12}(I)e'(j^1j^2) = 2\gamma_{00}e(j^1j^2) \sum_{\substack{k^1\in I^1\\ k^2\in I^2}} P(k^1)P(k^2)a(j^1k^1j^2k^2)$$

$$+ 2\gamma_{01} \sum_{\substack{k^1\in I^1\\ k^2\in I^2}} e(j^1k^2)a(j^1k^1j^2k^2)P(k^1)P(k^2) \quad . \qquad \text{(F.32)}$$

Still assuming that the $Z(i^1i^2)$ are small and that $e(i^1i^2) \to 0$ for $i^1 \notin I^1$ or $i^2 \notin I^2$, we get from the Equations F.28 and F.29 that for $j^1 \in I^1$ and $j^2 \in I^2$

$$V^{12}(I)e'(j^1j^2) = 2\gamma_{00}P(j^1)P(j^2) \sum_{\substack{k^1\in I^1\\ k^2\in I^2}} e(k^1k^2)a(j^1k^1j^2k^2)$$

$$+ 2\gamma_{00}e(j^1j^2) \sum_{\substack{k^1\in I^1\\ k^2\in I^2}} a(j^1k^1j^2k^2)P(k^1)P(k^2)$$

$$+ 2\gamma_{01} \sum_{\substack{k^1\in I^1\\ k^2\in I^2}} e(j^1k^2)a(j^1k^1j^2k^2)P(k^1)P(j^2) \qquad \text{(F.33)}$$

$$+ 2\gamma_{01} \sum_{\substack{k^1\in I^1\\ k^2\in I^2}} e(k^1j^2)a(j^1k^1j^2k^2)P(j^1)P(k^2) \quad .$$

Hence, under the assumption of small $Z(i^1i^2)$ the Equations F.30, F.31, F.32 and F.33 define a genetic system on the points $(1,2, \ldots ,m^1)$ and $(1,2, \ldots ,m^2)$ equivalent to the system defined in Section E (Equations E.20, E.23, E.24 and E.25 and their subsequent development in Section E.5.b). We therefore conclude that under the assumption of small $Z(i^1i^2)$, the selection system without mutation is a good approximation to the system in which mutation is taken into consideration.

Analogously to the single locus case, we now would like to determine conditions under which the $Z(i^1i^2)$, as defined in the Equations F.27, will be small. We will give the development of the argument in outline only, since it closely parallels the single locus case which has been given in detail, and since the representation of the matrices involved is cumbersome compared to the single locus case. It is also believed that the definite conclusions which can be reached are discernible enough in the argument that will be given. Let us write

$$p(i^1i^2) = P(i^1)P(i^2) + z(i^1i^2) \quad , \qquad (F.34)$$

where the $P(i^1)P^2(i^2)$ are a selection without mutation solution to the Equations D.3 and where the $z(i^1i^2)$ are small deviations from Hardy-Weinberg equilibrium. As before we assume $P(i^1)P(i^2) > 0$ if $i^1 \epsilon I^1$ and $e^2 \epsilon I^2$ and $P(i^1)P(i^2) = 0$ otherwise, where I^1 is a subset of $(1,2, \ldots ,m^1)$ and I^2 a subset of $(1,2, \ldots ,m^2)$. The Equations F.34 are then substituted in the Equations B.7 to get, by virtue of the Equations E.26, and on the assumption that the $z(i^1i^2)$'s are small, so that their squares and cross-products with each other and with the $v(i^1j^1i^2j^2)$ with $i^1 \neq j^1$ or $i^2 \neq j^2$ can be ignored, that we can write

$$[V^{12}(I) + 2 \sum_{\substack{i^1 \notin I^1 \\ \text{or } i^2 \notin I^2}} z(i^1 i^2)\alpha(i^1 i^2)][P(i^1)P(i^2) + z(i^1 i^2)]$$

$$= \sum_{j^1 j^2} v(j^1 i^1 j^2 i^2)P(j^1)P(j^2) \sum_{k^1 k^2} P(k^1)P(k^2)a(j^1 k^1 j^2 k^2)$$

$$+ 2\gamma_{00}P(i^1)P(i^2) \sum_{k^1 k^2} z(k^1 k^2)a(i^1 k^1 i^2 k^2)$$

$$+ 2\gamma_{00}z(i^1 i^2) \sum_{k^1 k^2} P(k^1)P(k^2)a(i^1 k^1 i^2 k^2)$$

$$+ 2\gamma_{01} \sum_{k^1 j^2} z(i^1 j^2)a(i^1 k^1 i^2 j^2)P(k^1)P(i^2)$$

$$+ 2\gamma_{01} \sum_{k^1 j^2} P(i^1)P(j^2)a(i^1 k^1 i^2 j^2)z(k^1 i^2) \quad . \qquad \text{(F.35)}$$

In the case where $i^1 \notin I^1$ and $i^2 \notin I^2$ the Equations F.35 become

$$z(i^1 i^2) = \sum_{j^1 j^2} v(j^1 i^1 j^2 i^2)P(j^1)P(j^2) + 2\gamma_{00}z(i^1 i^2)[1 + \frac{\alpha(i^1 i^2)}{V^{12}(I)}] \quad . \qquad \text{(F.36)}$$

If we have a system with stability on the border points $i^1 \notin I^1$ and $i^2 \notin I^2$ we have that $2\gamma_{00}(1 + \frac{\alpha(i^1 i^2)}{V^{12}(I)}) < 1$ (see Equations E.20 and E.21 and their discussion), so that under this stability condition

$$z(i^1 i^2) \rightarrow -\left(\sum_{j^1 j^2} v(j^1 i^1 j^2 i^2)P(j^1)P(j^2) \frac{V^{12}(I)}{\alpha(i^1 i^2)}\right) = \lim(z(i^1 i^2)), \qquad \text{(F.37)}$$

say.

If we assume $z(i^1 i^2) \doteq \lim(z(i^1 i^2))$ for $i^1 \notin I^1$ and $i^2 \notin I^2$ the Equations F.35 become in the case of $i^1 \in I^1$ and $i^2 \notin I^2$

$$V^{12}(I)z'(i^1i^2) = \sum_{\substack{j^1\epsilon I^1\\ j^2\epsilon I^2}} v(j^1i^1j^2i^2)P(j^1)P(j^2)V^{12}(I)$$

$$+ 2\gamma_{00}z(i^1i^2)\sum_{\substack{k^1\epsilon I^1\\ k^2\epsilon I^2}} P(k^1)P(k^2)a(i^1k^1i^2k^2)$$

$$+ 2\gamma_{01}\sum_{\substack{k^1\epsilon I^1\\ j^2\epsilon I^2}} P(i^1)P(j^2)a(i^1k^1i^2j^2)z(k^1i^2) \qquad \text{(F.38)}$$

$$+ 2\gamma_{01}\sum_{\substack{j^2\epsilon I^2\\ k^1\notin I^1}} P(i^1)P(j^2)a(i^1k^1i^2j^2)\lim\,(z(k^1i^2)) \quad .$$

Under the same assumptions as for the Equations F.38 we get for $i^1 \notin I^1$ and $i^2 \epsilon I^2$

$$V^{12}(I)z'(i^1i^2) = \sum_{\substack{j^1\epsilon I^1\\ j^2\epsilon I^2}} v(j^1i^1j^2i^2)P(j^1)P(j^2)V^{12}(I)$$

$$+ 2\gamma_{00}z(i^1i^2)\sum_{\substack{k^1\epsilon I^1\\ k^2\epsilon I^2}} P(k^1)P(k^2)a(i^1k^1i^2k^2) \qquad \text{(F.39)}$$

$$+ 2\gamma_{01}\sum_{\substack{k^1\epsilon I^1\\ j^2\epsilon I^2}} z(i^1j^2)a(i^1k^1i^2j^2)P(k^1)P(i^2)$$

$$+ 2\gamma_{01}\sum_{\substack{k^1\epsilon I^1\\ j^2\notin I^2}} \lim(z(i^1j^2))a(i^1k^1i^2j^2)P(k^1)P(i^2) \quad .$$

Let there be k^1 elements in $(1,2, \ldots ,m^1)$ which belong to I^1 and similarly k^2 elements in $(1,2, \ldots ,m^2)$ which belong to I^2. Then for the same reason as was given in the derivation of Equation F.17 we have to

work in a $k^1k^2 - 1$ dimensional vector space. In order to secure this, we have to assume that $\lim(z(i^1i^2))$ for $i^1 \notin I^1$ or $i^2 \notin I^2$ is so near to zero that we can assume

$$\sum_{\substack{i^1\epsilon I^1 \\ i^2\epsilon I^2}} z(i^1i^2) = 0 \quad .$$

We see from the Equations F.36, F.38 and F.39 that in order to achieve this result we have to assume $\sum_{j^1j^2} v(j^1i^1j^2i^2)P(j^1)P(j^2)$ to be small enough to be negligible for all cases where $i^1 \notin I^1$ or $i^2 \notin I^2$. These conditions will give the Equations F.38 and F.39 the same form as the Hardy-Weinberg cases of the Equations E.23 and E.24. If one would like to see precisely what linear combinations of $\sum_{j^1j^2} v(j^1i^1j^2i^2)P(j^1)P(j^2)$ for $i^1 \notin I^1$ or $i^2 \notin I^2$ must be assumed negligible, the non-homogeneous analogs of the equations of Section E.6.b can be solved according to methods analogous to that employed in the matrix solution of the Equations F.16.

For the case of $i^1 \epsilon I^1$ and $i^2 \epsilon I^2$, then, and under the assumption of $z(i^1i^2) \doteq 0$ for $i^1 \notin I^1$ or $i^2 \notin I^2$, we deduce from the Equations F.35 that

$$\begin{aligned} V^{12}(I)z'(i^1i^2) = {} & \sum_{\substack{j^1\epsilon I^1 \\ j^2\epsilon I^2}} v(j^1i^1j^2i^2)P(j^1)P(j^2)V^{12}(I) - P(i^1)P(i^2)V^{12}(I) \\ & + 2\gamma_{00}z(i^1i^2)V^{12}(I) + 2\gamma_{00}P(i^1)P(i^2) \sum_{\substack{k^1\epsilon I^1 \\ k^2\epsilon I^2}} z(k^1k^2)a(i^1k^1i^2k^2) \\ & + 2\gamma_{01} \sum_{\substack{k^1\epsilon I^1 \\ j^2\epsilon I^2}} z(i^1j^2)a(i^1k^1i^2j^2)P(k^1)P(i^2) \\ & + 2\gamma_{01} \sum_{\substack{k^1\epsilon I^1_2 \\ j^2\epsilon I^2}} P(i^1)P(j^2)a(i^1k^1i^2j^2)z(k^1i^2) \quad . \end{aligned} \tag{F.40}$$

It is clear that the Equations F.40 are the non-homogeneous counterparts of the Equations D.7. With specialization to the cases of additive and multiplicative gene action between loci, it is possible to solve the Equations F.40 in a manner analogous to the solution of the recurrence relations given by the Equations F.16. The solution to the Equations F.40 will be in a similar form as the single locus solution given as Equation F.17, with the vector c in this case consisting of elements of the form

$$\frac{1}{\sqrt{P(i^1)P(i^2)}}\Big[\sum_{\substack{j^1\epsilon I^1\\ j^2\epsilon I^2}} v(j^1 i^1 j^2 i^2)P(j^1)P(j^2) - P(i^1)P(i^2)\Big] \ ,$$

for $i^1 \epsilon I^1$ and $i^2 \epsilon I^2$.

It does not seem possible to make more general statements about the implications of this type of answer than the ones that were given in the discussion on Equation F.17. The conclusions in the two-loci case are, therefore, very closely similar to that given in the case of a single locus.

For classical genes the mutation rate is usually assumed to be of the order of one to ten per million. If one could equate our units of inheritance that combine in an additive or a multiplicative way with classical genes, we would hazard the guess that in most cases the effect of mutation would be negligible.

For the purposes of the mutation load we are interested in the possibility of an equilibrium with

$$P(i^1)P(i^2) = 1 \quad \text{if} \quad i^1 = 1 \quad \text{and} \quad i^2 = 1$$

and

$$P(i^1)P(i^2) = 0 \quad \text{otherwise,}$$

or, in our previous notation, with $I^1 = (1)$ and $I^2 = (1)$.

For the case of $i^1 \neq 1$ and $i^2 \neq 1$ the Equations F.35 become

$$a(1111)z(i^1i^2) = v(1i^11i^2)a(1111) + 2\gamma_{00}a(i^11i^21)z(i^1i^2)$$

which implies for $2\gamma_{00}a(i^11i^21) < a(1111)$ that

$$z(i^1i^2) \to \frac{v(1i^11i^2)}{1 - 2\gamma_{00}\frac{a(i^11i^21)}{a(1111)}} \qquad (F.41)$$

$$= \lim(z(i^1i^2)) \text{ , say.}$$

If $i^I = 1$ and $i^2 \neq 1$ the Equations F.35 become

$$a(1111)z'(1i^2) = v(111i^2)a(1111) + 2\gamma_{00}z(1i^2)a(11i^21)$$

$$+ 2\gamma_{01}\sum_{k^1} a(1k^1i^21)z(k^1i^2) \text{ ,}$$

which, after a sufficient number of generations, we can write to be

$$a(1111)z'(1i^2) = a(1111)v(111i^2) + z(1i^2)a(11i^21)$$

$$+ \rho \sum_{k^1 \neq 1} a(1k^1i^21)\lim(z(k^1i^2)) \text{ ,}$$

by virtue of the Equations F.41 and the fact that $2\gamma_{00} = 1-\rho$ and $2\gamma_{01} = \rho$. It follows that

$$z(1i^2) \to \frac{a(1111)v(111i^2) + \rho \sum_{k^1 \neq 1} a(1k^1i^21)\lim(z(k^1i^2))}{a(1111) - a(11i^21)} \text{ ,} \qquad (F.42)$$

for $a(1111) > a(11i^21)$ with $i^2 \neq 1$.

We write

$$z(1i^2) \rightarrow \lim(z(1i^2)) \text{ , say.}$$

In the same way we have for $i^1 \neq 1$ and $i^2 = 1$ that

$$z(i^1 1) \rightarrow \frac{a(1111)v(1i^1 11) + \rho \sum_{j^2 \neq 1} \lim(z(i^1 j^2))a(i^1 11j^2)}{a(1111) - a(i^1 111)} \qquad \text{(F.43)}$$

$$\rightarrow \lim(z(i^1 1)) \text{ , say.}$$

The Equations F.43 hold only for $a(1111) > a(i^1 111)$, with $i^1 \neq 1$.

Since $\sum_{i^1 i^2} z(i^1 i^2) = 0$, it follows that we can solve for $z(11)$ from the Equations F.41, F.42 and F.43. However, we note in the Equations F.41 that $v(1i^1 1i^2)$ is the probability for mutations at both loci during the same generation. If we assume this probability to be so small as to be negligible we get from Equations F.42 and F.43 that

$$z(11) \rightarrow - \sum_{i^1 \neq 1} \left(\frac{a(1111)v(1i^1 11)}{a(1111) - a(i^1 111)}\right) - \sum_{i^2 \neq 1} \left(\frac{a(1111)v(111i^2)}{a(1111) - a(11i^2 1)}\right) . \qquad \text{(F.44)}$$

The Equation F.44 is clearly a straightforward generalization of Equation F.18 to the case of two loci.

G. Hardy-Weinberg Equilibria of Gene Frequencies in the n-Loci Case with Selection and no Mutation

The problem in the extension of the work in Section D to the case of n loci is mainly notational. However, if we take care to present the relevant equations in a form that will allow straightforward generalization to the n-loci case, the problem can in some instances be avoided by following the nature of the necessary manipulations in the 2-loci case.

We shall find it advantageous for the present purposes to return to the notation that was used at the beginning of Section B.

Recall now that if we assume no mutation to be taking place, $v(i^1_{x_1} i^2_{y_1} i^2_{x_2} i^2_{y_2})$ is equal to one only if $i^1_{x_1} = i^1_{y_1}$ and $i^2_{x_2} = i^2_{y_2}$, and is equal to zero otherwise. Then it is easy to see that we can rewrite the Equations B.5 under the assumptions of no mutation and no position effect as

$$W \cdots p'(i^1_0 i^2_0) = \sum_{i^1_1 i^2_1} \sum_{x_1,x_2=0,1} \gamma_{x_1 x_2} p(i^1_{x_1} i^2_{x_2}) a(i^1_0 i^1_1 i^2_0 i^2_1) p(i^1_{1-x_1} i^2_{1-x_2}) . \tag{G.1}$$

The Equations G.1 are analogous to the Equations B.8, which are the basic equations from which the formulas of Section D were developed.

We now proceed in a fashion precisely analogous to the derivation of the Equations D.7 by making the transformation $p(i^1_{x_1} i^2_{x_2}) = P(i^1_{x_1}) P(i^2_{x_2}) + e(i^1_{x_1} i^2_{x_2})$ in the Equations G.1 to get

$$V_E^{12} e'(i^1_0 i^2_0) = \sum_{i^1_1 i^2_1} \Big[\sum_{x_1,x_2=0,1} \gamma_{x_1 x_2} P(i^1_{x_1}) P(i^2_{x_2}) a(i^1_0 i^1_1 i^2_0 i^2_1) e(i^1_{1-x_1} i^2_{1-x_2})$$

$$+ \sum_{x_1,x_2=0,1} \gamma_{x_1 x_2} e(i^1_{x_1} i^2_{x_2}) a(i^1_0 i^1_1 i^2_0 i^2_1) P(i^1_{1-x_1}) P(i^2_{1-x_2}) \Big] .$$

By reason of symmetry it follows that we can write

$$V_E^{12}\, e'(i_0^1 i_0^2) = 2 \sum_{i_1^1 i_1^2} \Big[\sum_{x_1,x_2=0,1} \gamma_{x_1 x_2}\, e(i_{x_1}^1 i_{x_2}^2)\, a(i_0^1 i_1^1 i_0^2 i_1^2)\, P(i_{1-x_1}^1)\, P(i_{1-x_2}^2) \Big] \tag{G.2}$$

Like the Equations D.7 the Equations G.2 allow specialization to both the cases of additive and multiplicative gene action between loci. For reasons that will be stated later on, we will concentrate on the case of multiplicative gene action between loci.

We now utilize a set of transformations analogous to those given as the Equations D.9 in the Equations G.2 so that, on recalling $i_1^1 = 1,2, \ldots, m^1$ and $i_1^2 = 1,2, \ldots, m^2$, we obtain

$$y'(i_0^1 i_0^2) = \sum_{i_1^1 i_1^2} \sum_{x_1,x_2=0,1} 2\gamma_{x_1 x_2}\, y(i_{x_1}^1 i_{x_2}^2)\, b^{x_1}(i_0^1 i_1^1)\, b^{x_2}(i_0^2 i_1^2)\, (m^1)^{x_1-1}\, (m^2)^{x_2-1} \tag{G.3}$$

It is an easy job to convince oneself that the Equations G.3 are equivalent to the Equations D.10, if one recalls that we assumed $\gamma_{x_1 x_2} = \gamma_{1-x_1\, 1-x_2}$.

In Section D it was pointed out that elements of the characteristic vectors of the matrix $[b(i_0^1 i_1^1) b(i_0^2 i_1^2)]$ can be written as products of corresponding elements from the characteristic vectors of $[b(i_0^1 i_1^1)]$ and $[b(i_0^2 i_1^2)]$. Let us now assume $y(i_{x_1}^1 i_{x_2}^2)$ to be an element from a characteristic vector of $[b(i_0^1 i_1^1)\, b(i_0^2 i_1^2)]$, so that we can write $y(i_{x_1}^1 i_{x_2}^2) = w(i_{x_1}^1) w(i_{x_2}^2)$, where $w(i_{x_1}^1)$ and $w(i_{x_2}^2)$ are elements from the characteristic vectors of $[b(i_0^1 i_1^1)]$ and $[b(i_0^2 i_1^2)]$ respectively.

It follows therefore that for characteristic vectors of $[b(i_0^1 i_1^2)b(i_0^2 i_1^2)]$ we obtain from the Equations G.3

$$w'(i_0^1)w'(i_0^2) = w(i_0^1)w(i_0^2) \sum_{x_1 x_2 = 0,1} 2\gamma_{x_1 x_2} (\lambda^1)^{x_1} (\lambda^2)^{x_2} , \qquad (G.4)$$

where λ^1 and λ^2 are characteristic roots of $[b(i_0^1 i_1^1)]$ and $[b(i_0^2 i_1^2)]$ respectively. The Equations G.4 are precisely equivalent to the Equations D.11.

The form of the different characteristic roots of the matrix $[b(i_0^1 i_1^1)b(i_0^2 i_1^2)]$ and their associated vectors was given in the discussion following on the Equation D.18. It is clear from this discussion that the only characteristic vector of $[b(i_0^1 i_1^1)b(i_0^2 i_1^2)]$ that should be excluded from consideration in the Equations G.4 is the one associated with the root $\lambda^1\lambda^2 = 1$.

It follows that the necessary and sufficient condition for the system of perturbation vectors associated with the Equations G.4 to be stable is that

$$\left| \sum_{x_1, x_2 = 0,1} 2\gamma_{x_1 x_2} (\lambda^1)^{x_1} (\lambda^2)^{x_2} \right| \leq 1 , \qquad (G.5)$$

for all characteristic roots of $[b(i_0^1 i_1^1)]$ and $[b(i_0^2 i_1^2)]$ with the exception of the case where both λ^1 and λ^2 are equal to unity.

The extension of the result given by the inequality of the Equation G.5 to the case of n loci is now immediately obvious, and will be stated in the theorem given below.

Theorem 5 In the case of n loci with multiplicative gene action between loci the necessary and sufficient condition for the

existence of a stable Hardy-Weinberg equilibrium is that the condition,

$$\left| \sum_{x_1,x_2,\ldots,x_n=0,1} 2\gamma_{x_1x_2\ldots x_n} (\lambda^1)^{x_1} (\lambda^2)^{x_2} \ldots (\lambda^n)^{x_n} \right| \leq 1 \qquad (G.6)$$

hold, where not all of the characteristic roots λ^a, $a = 1,2, \ldots ,n$ may be taken to be equal to unity at the same time.

The λ^a are defined to be the characteristic roots of the matrices $[b(i_0^a i_1^a)]$, which were defined in the Equations D.9 for the specific cases of $a = 1$ and $a = 2$. The properties of the λ^a follow from Section C and from the derivations following on the Equations D.10. The $\gamma_{x_1x_2\ldots x_n}$ are defined by the natural extension of the $\gamma_{x_1x_2}$ which were defined at the beginning of Section B.

The proof of Theorem 5 follows immediately from the natural extension of the Equations G.1 through G.5 to the case of n loci.

Theorem 5, like Theorem 1, is of course applicable to internal points only, i.e. the points $P(i_0^1)P(i_0^2) \ldots P(i_0^n) = 0$ were excluded in its formulation. The border points can be handled by the techniques of Section E. Let us denote the set of internal points by the symbol I. Then we speculate on the basis of our work in Section E that, under the assumption of Hardy-Weinberg equilibria, a border point associated with the k-th locus will be stable if and only if $\Sigma_{i_1^k} P(i_1^k)a(i_0^k i_1^k) < V_E^k$, for $i_0^1 \notin I$, where $a(i_0^1 i_1^1)$ is the fitness at the k-th locus and where V_E^1 is the mean fitness at equilibrium at the k-th locus.

It is clear that, for example, γ_{11} and $\sum_{x_3,x_4,\ldots,x_n=0,1} \gamma_{11x_3x_4\ldots x_n}$ represent probabilities associated with the same events. The work of

Schnell (1961) and van Aarde (1963) show furthermore that probabilities like the ones mentioned above can be represented uniquely in terms of the same functions of the coefficients of recombination. It follows that in the consideration of terms containing $\gamma_{x_1 x_2 \ldots x_n}$ it will be natural to delete the variables that have been summed over.

Let us now consider the case where $\lambda^1 \neq 1$ and $\lambda^2 = \lambda^3 = \ldots = \lambda^n = 1$ in the Equation G.6. It is easy to convince oneself that $2\gamma_0 = 2\gamma_1 = 1$. Hence, Equation G.6 becomes $\left|1 + \lambda^1\right| \leq 1$. The same argument holds for all the other cases where only one characteristic root unequal to one is considered. We also have from Section C that $\left|1 + \lambda\right| \leq 1$ is the condition for stability in the case of a single locus. Hence, it follows that a necessary condition for Hardy-Weinberg stability for n loci is that all the constituent single locus cases must be stable.

Next, let us consider the case where $\lambda^1 \neq 1$ and $\lambda^2 \neq 1$, with $\lambda^3 = \lambda^4 = \ldots = \lambda^n = 1$. Equation G.6 becomes then

$$\left| 2\gamma_{00} + 2\gamma_{01}\lambda^2 + 2\gamma_{10}\lambda^1 + 2\gamma_{11}\lambda^1\lambda^2 \right| \leq 1 \quad ,$$

which is clearly equivalent to Equation D.14. The natural extension of this argument gives that all two-loci systems must be stable in order to allow stability for the n-loci system.

The foregoing argument extends obviously until the case is reached where all n roots must be unequal to unity. Hence, we can rephrase Theorem 5 as given below.

In the case of multiplicative gene action between loci the Hardy-Weinberg system associated with the set of loci (1,2, ... ,n) will be stable if and only if the Hardy-Weinberg systems associated with all subsets of

(1,2, ... ,n) are stable.

The problem with Theorem 5, as is also the case with Theorem 1, is that it is only a local result. There does not seem to be any obvious way by which one can show in general that when the Hardy-Weinberg system is stable, the non-Hardy-Weinberg system must be unstable. In any case we know from the examination of simple systems that when the Hardy-Weinberg equilibria are unstable, there may very well be stable non-Hardy-Weinberg equilibria.

It is difficult to imagine any evolutionary mechanism that will favour the development of stable Hardy-Weinberg equilibria. Here it is relevant to remember that we proved in Section E that in the case of multiplicative gene action between loci, a stable Hardy-Weinberg equilibrium does not present a local maximum of the mean fitness function. It is easily seen that the argument used in Section E extends naturally to the case of n loci. Lewontin's (1964) result on the mean fitnesses in the case of two loci, each with two alleles, suggests that there may very well be advantages to the non-Hardy-Weinberg equilibria, since they result in a higher mean fitness of the population. This is obviously a very deep problem which we will have to leave aside for the time being. The present state of knowledge on some aspects of these questions were discussed in Section E.

In the case of additive gene action between loci the proof of the theorem that the Hardy-Weinberg equilibrium is stable if and only if all the constituent single loci of the n-loci system are stable, does not appear to go through as smoothly as the proof for the stability conditions in the case of multiplicative gene action between loci. A proof for the additive case will, therefore, not be attempted here, especially since

there is very little reason to doubt the validity of the postulated n-loci result.

In all the results on additive and multiplicative gene action between loci the device of working with the 2-loci case was employed only to ease the problem of notation, and was in no way enforced by the logical or mathematical structure of the problem. Most of the results in Section E will, therefore, carry over to the case of n loci in a logical and obvious way. The speculation about the nature of the necessary and sufficient conditions for the stability of border points in the case of multiplicative gene action, which followed immediately on the derivation of Theorem 5, will for this reason carry over to the case of additive gene action between loci.

IV. LOAD CONCEPTS AND THEORY

A. The Single Locus Case

1. General introduction

Even a cursory glance at the literature is enough to convince one that the load theory was designed in single locus terms and that the generalization to the multi-loci situation was introduced without adequate theoretical justification. However, the historic approach to the multi-loci case via the single locus case will be followed here, because some of the results derived in this way will be needed later on, and because it is believed that to do so would be revealing.

A comparison between Chapters II and III will reveal that fitness was treated in essentially different ways. In Chapter II fitness was handled in the same way as a metric trait, that is, fitness was conceived to be the property of a single genotype. In Chapter III the fecundity aspect of fitness was described by a probability distribution associated with a pair of adult genotypes. However, it was shown that some meaning can still be given to the concept of a fitness value ascribed to a single genotype if the property, which we referred to as product fecundity, is assumed.

It is clear that the only observable quantity in an experimental context is the number of offspring per mating. It seems that the only rational way in which offspring can be attributed to a single parent is to assume that the offspring distribution of a pair of parents is equal to the product of two parental distributions, as follows from the discussion on Equation III.B.2. In other words, what is done here is to assume that

from the tu offspring of a mating between two parents of the genotypes (ij) and (rs), say, t can be ascribed to (ij) and u to (rs).

The problem here is that there is no way of knowing how to factor the tu offspring into the number ascribable to each parent, even if we are willing to make the assumption of independent offspring distributions, which is a more restrictive assumption than that of product fecundity. It follows that the best procedure is to work with pairs of genotypes in our derivations.

Let us, therefore, find the expectation of the number of infant offspring per adult mating pair. We will do so by finding conditional expectations. The expected number of infant offspring of the adult mating pair $(ij) \times (rs)$ is $\sum_{t=0}^{c} tb_t(ijrs) = b(ij)b(rs)$, as we defined it before in Section B of Chapter III. In the same way as before, let us denote the probability of survival from infancy to adulthood of the infant (ij) by $\ell(ij)$. It follows then from the theory in Section B, Chapter III that in a random mating population we can write

$$E(b(ij)b(rs)) = \frac{\sum_{ijrs} b(ij)b(rs)\ell(ij)\ell(rs)p(i)p(j)p(r)p(s)}{(\sum_{vw} \ell(vw)p(v)p(w))^2}$$

$$= \frac{(\sum_{ij} a(ij)p(i)p(j))^2}{(\sum_{vw} \ell(vw)p(v)p(w))^2} \quad ,$$

where, as before, we denote $b(ij)\ell(ij)$ by $a(ij)$.

It will be convenient for the purpose of the development of the load theory to define the mean fitness of a population by the expression $\sum_{ij} a(ij)p(ij)$, where $p(ij)$ is the frequency of the genotype (ij). This definition is in accord with our previous usage of the term mean fitness of

a population in Chapter III.

We see that the expected number of infant offspring per adult mating pair is a function of both the mean fitness and the mean viability of a population. It would, however, be convenient if we could find some quantity which may be observable and the expectation of which will be a function of the mean fitness alone. We will, therefore, investigate the expectation of the number of infant offspring per infant mating pair. For this purpose we clearly have to envisage an experimental procedure by which infants can be mated at random.

We first note that the event that a pair of adults are (ij) and (rs) implies that the pair of infants were (ij) and (rs), and vice versa. It follows that the expectation of the number of infant offspring from infant mating pairs can be written as

$$E(O) = \underset{\substack{\text{Identity}\\ \text{of infants}}}{E} \quad \underset{\substack{\text{Survival}\\ \text{of infants}}}{E} \quad E(O/\text{infants survived}/\text{infants were (ij) and (rs)}) .$$

The mean number of infant offspring of the adult mating type (ij) x (rs) is $\sum_{t=0}^{q} t b_t(ijrs) = b(ij)b(rs)$, as we defined it before in Section B of Chapter III. The expected number of infant offspring for the pair of infants (ij) and (rs) is $\ell(ij)b(ij)\ell(rs)b(rs) = a(ij)a(rs)$, where $\ell(ij)$ is the probability that the infant (ij) survives to adulthood. Hence, we have

$$E(a(ij)a(rs)) = \sum_{ij} \sum_{rs} a(ij)a(rs)p(i)p(j)p(r)p(s)$$

$$= (\sum_{ij} a(ij)p(i)p(j))^2 \quad . \qquad \text{(A.1)}$$

If we use the method of moments, it follows that the experimentally observed mean number of offspring per mating would be used as an estimator

of the square of the mean fitness of a population.

The mean number of infant offspring that one can ascribe to a specific infant genotype is also of some interest. We again assume random mating between adults. The expected number of infant offspring of a (ij) adult is, according to Kempthorne and Pollak (1969), equal to

$$b(ij)\frac{\sum_{rs} b(rs)\ell(rs)p(r)p(s)}{\sum_{uv}\ell(uv)p(u)p(v)}$$

$$= b(ij)\frac{\sum_{rs} a(rs)p(r)p(s)}{\sum_{uv}\ell(uv)p(u)p(v)} \quad .$$

Since the probability that a (ij) infant survives to adulthood is $\ell(ij)$, the expected number of infant offspring of a (ij) infant is

$$a(ij)\frac{\sum_{rs} a(rs)p(r)p(s)}{\sum_{uv}\ell(uv)p(u)p(v)} \quad .$$

It follows that the relative frequency of offspring attributable to the infants of the type (ij) will be

$$\frac{a(ij)p(i)p(j)\sum_{rs} a(rs)p(r)p(s)}{\sum_{uv} a(uv)p(u)p(v))^2}$$

$$= \frac{a(ij)p(i)p(j)}{\sum_{uv} a(uv)p(u)p(v)} \quad . \qquad \text{(A.2)}$$

It is clear that the Equation A.2 also holds for the case of random mating between infants.

The Equation A.2 shows clearly the reason why the approach given in Chapter III is operationally equivalent to the "conventional" way in which the change in gene frequencies is calculated. For a discussion of the "conventional" way of handling fitness Li(1967), or any standard textbook can be consulted.

In order to elucidate fully the extent to which fitness can be treated in the same way as a metric trait associated with selectively neutral genes, we now have to proceed from the previous discussion which assumed random mating, to a consideration of inbreeding. We shall assume that the selective forces are weak and that inbreeding takes place over a short period of time, so that the effect of selection on the inbred population structure can be considered to be negligible.

Let us first consider two individuals, X and Y, who are members of a population derived by some form of inbreeding with no selection from a random mating population with genotypic array $\sum_{ij} p(i)p(j)(ij)$. The nature of the relationship between X and Y and the nature of their inbreeding is specified by the values $P_1, P_2, \ldots, P_9$, which were defined by Harris (1964). Considering X and Y to be random members of the subpopulation composed of all pairs of individuals related as X and Y, i.e., who have the same values for $P_1, P_2, \ldots P_9$, the genotypic array of such pairs is

$$\begin{aligned} & P_1 \sum_i p(i)(ii)(ii) + P_2 \sum_{ij} p(i)p(j)(ii)(jj) + P_3 \sum_{ij} p(i)p(j)(ii)(ij) \\ & + P_4 \sum_{ij} p(i)p(j)(ij)(ii) + P_5 \sum_{ij} p(i)p(j)(ij)(ij) \\ & + P_6 \sum_{ijr} p(i)p(j)p(r)(ii)(jr) + P_7 \sum_{ijr} p(i)p(j)p(r)(ij)(rr) \qquad (A.3) \\ & + P_8 \sum_{ijr} p(i)p(j)p(r)(ij)(ir) + P_9 \sum_{ijrs} p(i)p(j)p(r)p(s)(ij)(rs) \ . \end{aligned}$$

We now compare this genotypic array of pairs of individuals on inbreeding with the one derived from

$$[F \sum_{i} p(i)(ii) + (1-F)\sum_{ij} p(i)p(j)(ij)]^2 \quad ,$$

which is equal to

$$F^2 \sum_{ij} p(i)p(j)(ii)(jj) + 2F(1-F) \sum_{ijr} p(i)p(j)p(r)(ij)(rr)$$
$$+ (1-F)^2 \sum_{ijrs} p(i)p(j)p(r)p(s)(ij)(rs) \quad . \quad (A.4)$$

The two genotypic arrays will be equal to each other if and only if

$$P_1 = 0,\ P_2 = F^2,\ P_3 = 0,\ P_4 = 0,\ P_5 = 0,\ P_6 + P_7 = 2F(1-F) \ ,$$
$$P_8 = 0,\ P_9 = (1-F)^2 \ .$$

A moment's reflection on the nature of mating pairs on inbreeding in a single population will convince one that it is rather unlikely that the arrays given as Equations A.3 and A.4 are going to be equal. P_2 will in our present setup be defined as $\text{Prob}(a=b\neq c=d)$, where the symbol = indicates identity by descent, and $\neq$ indicates non-identity by descent. The condition that $P_2 = \text{Prob}(a=b\neq c=d) = \text{Prob}(a=b)\text{Prob}(c=d) = F^2$, implies that the sources of inbreeding in the mating pair (ii) x (ii) must be independent in origin. This does not seem to be possible in any of the simpler mating systems, and can easily be verified not to be the case by following the inbreeding process for one generation when full-sibbing or selfing is practiced.

In the case of full-sibbing we can derive that $P_5 = \frac{1}{4}$, $P_8 = \frac{1}{2}$, $P_9 = \frac{1}{4}$, and that $P_1 = P_2 = P_3 = P_4 = P_6 = P_7 = 0$. These values can be verified from recurrence equations given by Gillois (1964). The case of full-sibbing with selection will be discussed in Chapter V, and since the foregoing values can easily be obtained from the formulation given there, we will delete their direct derivation. The value of F in the case of one generation of full-sibbing is $\frac{1}{4}$, as is well known.

Let us now define the mean fitness of a population derived from a random mating population with Hardy-Weinberg structure by a process of mating based purely on consanguinity until a degree of inbreeding F is reached, as

$$F \sum_i p(i)a(ii) + (1-F)\sum_i \sum_j p(i)p(j)a(ij) \; . \qquad (A.5)$$

The expectation of number of infant offspring in an infant subpopulation given by the genotypic array given as Equation A.3 can be obtained by substituting the symbols a(ij) for (ij). A formal derivation of the expected number of infant offspring in such an infant subpopulation can be obtained by the use of conditional expectations in the same way as in the random mating case.

It has been noted that the genotypic arrays given as Equations A.3 and A.4 will not be equal, except possibly in very special cases. It follows that the expected number of infant offspring derived from the infant array in Equation A.3 will not in general be equal to the square of the mean fitness of the inbred population given in Equation A.5. This implies that under inbreeding the mean number of infant offspring per infant mating pair will not be an estimate of the square of the mean fitness of the population, with the exceptions to this statement being in the cases where the arrays given as Equations A.3 and A.4 are equal.

We shall meet this problem of not having an estimate of the mean fitness of a population on inbreeding by constructing a population of inbred mating pairs with the desired genotypic array. The construction of a population with an array of mating pairs given as Equation A.4 is a relatively easy affair. The procedure is simply to obtain a number of independent

inbred lines from the same base population and to allow inbreeding to take place until each line reached the same degree of inbreeding as measured by the inbreeding coefficient F. The members of each line are then mated randomly to the members of the other lines. Our experimental procedure is then to mate infants at random within the required pattern, and to count the number of infant offspring per infant mating pair. No substitutions for mated infants who do not survive to adulthood are made. This procedure is important because it is only the mean number of infant offspring per infant mating pair which is an unbiased estimate of the square of the mean fitness of a sub-population with a specified degree of inbreeding.

In this context it is worthwhile to note that the expected number of offspring at infancy per adult mating pair for a population derived by the procedure given in the previous paragraph is

$$\frac{[F \sum_i p(i)a(ii) + (1-F)\sum_i \sum_j p(i)p(j)a(ij)]^2}{[F \sum_i p(i)\ell(ii) + (1-F)\sum_i \sum_j p(i)p(j)\ell(ij)]^2} , \qquad (A.6)$$

as follows easily by the procedures employed previously. It follows that if we want to work with adult mating pairs a correction will have to be made for the mortality between infancy and adulthood.

Note now that

$$E[\text{number of adults/the infant was } (ij)]$$

$$= 1\ \ell(ij) + 0[1-\ell(ij)]$$

$$= \ell(ij) \ .$$

It follows that E(number of adults per infant)

$$= F \sum_i p(i)\ell(ii) + (1-F)\sum_{ij} p(i)p(j)\ell(ij) \ . \qquad (A.7)$$

Since we are dealing with a binomial variable which takes only the values zero or one, it follows that the ratio of the number of adults to the number of infants is an unbiased estimator of

$$F \sum_i p(i)\ell(ii) + (1-F)\sum_{ij} p(i)p(j)\ell(ij) \quad .$$

We observe that Prob(a pair of infants survive to adulthood)

$$= \sum_{ijrs} p(ijrs)\ell(ij)\ell(rs) \ ,$$

where $p(ijrs)$ is the frequency of the pair of infants $(ij) \times (rs)$. The structure of the population induced by random mating between inbred lines causes $p(ijrs) = p(ij)p(rs)$. It follows that the probability of a pair of infants surviving to adulthood is equal to the product of the probabilities that each of them survives. It follows that

E(number of pairs of adults per infant pair)

$$= (\text{E(number of adults per infant)})^2 \ .$$

It is of some interest to observe that

Prob(a pair of infants survive $\cap$ t offspring from a pair of adults)

$$= \sum_{ijrs} p(ij)\ell(ij)b_t(ijrs)p(rs)\ell(rs)$$

= Prob(t offspring from a pair of adults) Prob(a pair of infants survive)

$$= \frac{\sum_{ijrs} p(ij)\ell(ij)b_t(ijrs)p(rs)\ell(rs)}{(\sum_{uv} p(uv)\ell(uv))^2} \sum_{defg} p(de)\ell(de)p(fg)\ell(fg) \ .$$

In this context the notation $\cap$ indicates the intersection of two events in the sample space.

It follows from the foregoing equation that we are dealing with independent events in our consideration of the viability and fertility aspects of fitness. Together with the results following from Equation A.7,

this indicates that it would be possible to consider a correction of the Equation A.6 for the purpose of the estimation of the square of the mean fitness.

We will, however, proceed by considering only the case in which the infant offspring per infant mating pair are counted, since it is theoretically the simplest and also the most direct procedure for the estimation of the square of the mean fitness of a population. In most cases where the foregoing theory is applicable there probably would not be any advantage to counting the number of infant offspring per adult mating pair, since knowledge of the mortality between infancy and adulthood is necessary for correction to render the estimate unbiased.

Before we move on to the consideration of other topics we note that the variance of the number of infant offspring per infant mating pair is given by

$$\sum_{ijrs} \sum_{t} t^2 b_t(ijrs)\ell(ij)\ell(rs)p(ij)p(rs)$$
$$- \left(\sum_{ijrs} \sum_{t} t b_t(ijrs)\ell(ij)\ell(rs)p(ij)p(rs)\right)^2$$
$$= \left(\sum_{ijrs} A(ijrs)p(ij)p(rs)\right) - \left(\sum_{ij} a(ij)p(ij)\right)^4 , \qquad (A.8)$$

where

$$A(ijrs) = \sum_{t} t^2 b_t(ijrs)\ell(ij)\ell(rs) .$$

We observe that this variance pertains to the procedure where infants are paired and their offspring are counted at infancey. We also note that the variance of the number of infant offspring per infant mating pair depends on the degree of inbreeding through the genotypic frequencies $p(ij)$. This result is, of course, only applicable to populations obtained in the manner indicated for the estimation of the square of the mean fitness of the population.

For the calculation of the variance of the mean number of infant offspring per infant mating pair, it is necessary to remember that the covariance between the number of offspring derived from different mating pairs may not be equal to zero. Let us consider a pair of infants from the same inbred line and denote their genotypes by (ij) and (de). Their frequency will be p(ijde), say, with the different cases of identity by descent described by the array given as the Equation A.3. Likewise let (rs) and (fg) be a pair of infants from another inbred line with the same degree of inbreeding as the first. According to the mating scheme that we described previously, the infant matings (ij) x (rs) and (de) x (fg) occur with frequency p(ijde)p(rsfg). The probability of obtaining t infant offspring from one mating and u infant offspring from the other mating is $b_t(ijrs)b_u(defg)\ell(ij)\ell(de)\ell(fg)$. The covariance between t and u is therefore derived to be

$$\begin{aligned}\mathrm{Cov}(t,u) = {} & \sum_{\substack{ijrs\\defg}} \sum_{tu} b_t(ijrs)b_u(defg)\ell(ij)\ell(rs)\ell(de)\ell(fg)p(ijde)p(rsfg) \\ & - \Big(\sum_{ijrs} \sum_{t} t b_t(ijrs)\ell(ij)\ell(rs)p(ij)p(rs)\Big) \\ & \times \Big(\sum_{defg} \sum_{u} u b_u(defg)\ell(de)\ell(fg)p(de)p(fg)\Big) \\ = {} & \Big(\sum_{ijde} a(ij)a(de)p(ijde)\Big)^2 - \Big(\sum_{ij} a(ij)p(ij)\Big)^4 . \qquad (A.9)\end{aligned}$$

The other case of interest in the consideration of covariances between the number of offspring from different infant mating pairs is where we have one pair of infants with genotypes (ij) and (de) from one inbred line mated to a pair of infants (rs) and (fg) from different inbred lines. In this case we obtain

$$\mathrm{Cov}(t,u) = (\Sigma_{ijde} a(ij)a(de)p(ijde))(\Sigma_{rs} a(rs)p(rs))^2 - (\Sigma_{ij} p(ij)a(ij))^4 . \quad (A.10)$$

We are now in a position to consider the possibilities of adapting the method of maximum likelihood to the estimation of the mean fitness of a population. Let us assume that we construct an array of mating pairs by random mating between members of equally inbred lines. Also assume that we take only one individual from each inbred line as a parent in our estimation population. We therefore denote the probability of obtaining t infant offspring per infant mating pair by

$$\Sigma_{ijrs} b_t(ijrs)p(ij)p(rs) = Q_t \text{ , say,}$$

where the $p(ij)$'s take on values according to whether i is identical by descent to j or not, in the same way as in given in Equation A.5. We note that the definition of the Q_t's is in no way dependent on the assumption of a specific number of alleles and that the definition of Q_0 needs to include the event with probability $1 - \Sigma_{ijrs} \ell(ij)\ell(rs)p(ij)p(rs)$ of zero offspring due to the death of one or both paired infants before adulthood.

Let us now classify the outcome of n independent matings from a population of mating pairs with respect to the number of offspring per mating, and let us denote the number of matings with t offspring by n_t. Hence, $\Sigma_{t=0}^{k} n_t = n$. It follows that the probability of obtaining $n_0, n_1, \ldots, n_k$ infant pairs with respectively $0,1, \ldots, k$ infant offspring per pair is

$$g(Q) = \frac{n!}{n_0!n_1!, \ldots, n_k!} (Q_0)^{n_0}(Q_1)^{n_1} \ldots (Q_k)^{n_k} , \quad (A.11)$$

where we assume that $Q_t > 0$, for $t = 0,1, \ldots, k$.

The maximum likelihood estimators of the Q_t are, of course, well known, and are simply given by $\frac{n_t}{n}$ for all t. From the definition of the

various quantities involved we have that

$$\sum_t tQ_t = \sum_{ijrs} \sum_t t b_t(ijrs)\ell(ij)\ell(rs)p(ij)p(rs) = \left(\sum_{ij} a(ij)p(ij)\right)^2 .$$

Hence, $\sum_t t \frac{n_t}{n}$ will be an unbiased estimate of the square of the mean fitness, since the $\frac{n_t}{n}$ are unbiased estimates of Q_t. $\sqrt{\sum_j t \frac{n_t}{n}}$ will be a biased estimate of $\sum_{ij} a(ij)p(ij)$. We observe that the foregoing estimates are valid for a population of mating pairs of an arbitrary degree of inbreeding, subject only to the limitations mentioned previously on the use of the inbreeding coefficient under the present circumstances.

The variance of the estimate of $\left(\sum_{ij} a(ij)p(ij)\right)^2$ follows easily from the variances and covariances of the estimates of $Q_0, Q_1, \ldots, Q_k$. Unfortunately the variance of the estimate of the mean fitness of the population cannot be derived by the foregoing procedure.

It is of some interest to note what will happen if we do not construct an estimation population of mating pairs described by the genotypic array given by Equation A.4, but instead proceed to take mating pairs in which both parents come from the same inbred line. In this case the probability of obtaining t infant offspring per infant mating pair is given by

$$\sum_{ijrs} b_t(ijrs)\ell(ij)\ell(rs)p(ijrs) \quad , \qquad (A.12)$$

where the $p(ijrs)$ take on different values according to the relationship of identity by descent between the genes (i), (j), (r) and (s) as given in Equation A.3. As in general $p(ijrs) \neq p(ij)p(rs)$, we see that if we proceed in the same way as in the case of mating between inbred lines we obtain in this case $\sum_t tQ_t \neq \left(\sum_{ij} a(ij)p(ij)\right)^2$, with the result that we are unable to obtain estimates of either the mean fitness or the square of the

mean fitness of a population.

Another angle from which one could try to approach the estimation of the mean fitness of a population would be to try and estimate the $a(ij)$'s, from which the mean fitness of the population under consideration can be calculated. A possible procedure here would be to use the function $g(Q)$ of the Equation A.11 and to differentiate with respect to the different $a(ij)$'s and $b_t(ijrs)$'s. However, for this approach at least the number of alleles should be known exactly. In the type of situation in which we are interested in fitness and loads, this would not in general be the case.

If the genotypes are identifiable one could consider repeating the mating $(ij) \times (rs)$ n times. In this case the probability of obtaining t infant offspring per adult mating pair is $b_t(ijrs)$. Let us write $A_t = b_t(ijrs)$ and let us denote the number of matings with t offspring by n_t. Hence, $\sum_t n_t = n$. It follows that the probability of obtaining $n_0, n_1, \ldots, n_k$ adult pairs with respectively $0,1, \ldots, k$ infant offspring per pair is

$$g(A) = \frac{n!}{n_0! n_1! \ldots n_k!} (A_0)^{n_0} (A_1)^{n_1} \ldots (A_k)^{n_k} . \qquad (A.13)$$

It then follows from the Equation A.13, by the same type of argument that we used on Equation A.11, that $\sum_t t \frac{n_t}{n}$ will be the unbiased estimator of $\sum_t t b_t(ijrs)$. With product fecundity it will be true that $\sum_t t b_t(ijrs) = b(ij)b(rs)$. Now, if different genotypes are included in the matings one will be able to see whether the assumption of product fecundity is tenable or not.

The problem of the estimation of fitness distributions certainly deserves more attention than it has been given here. However, at present

our main objective is to investigate the theory of genetic loads, and since we obtained enough information to enable us to give our inquiry into load theory a rational basis, we will leave matters as they stand.

The concept of two life phases of the individuals in a population was introduced because of its value in the discussion of population structure. For example, it was shown that in the single locus case the infants have Hardy-Weinberg structure of gene frequencies, and it is obvious that unless there is no viability selection, the array of adults will not possess the Hardy-Weinberg property. In any practical situation as many life phases can be distinguished as is desirable, but since the distinction between two life phases provides us with adequate conceptual tools and simplifies matters, the only distinction that we will make in this chapter is between infants and adults.

The difficulty involving the description of the genotypic array of the population by the use of Wright's coefficient of inbreeding has already been touched upon. Although it is conceptually bothersome, it is believed that the problem is more apparent than real in the case of genetic loads. Accordingly we will proceed to use the inbreeding coefficient for the description of the genotypic array and justify its use for the estimation of the load ratio in the sections dealing with this aspect of the problem.

2. The mutation load

We consider a single locus with alleles denoted by (i), $i = 1,2, \ldots ,m$. From Chapter III it can be seen that Hardy-Weinberg structure can be assumed for the population at infancy. The fitness of an infant genotype is defined in the same way as the previous section and in the same way as it was defined in Chapter III, except that from now on we

will write $x(ij)$ instead of $a(ij)$. We do so in order to emphasize that fitness is now treated in the same way as a metric trait associated with selectively neutral genes and that due care must therefore be exercised to remain within the bounds within which a legitimate correspondence exists between the two ways of looking at the characteristics of fitness. The bounds within which such a legitimate correspondence exists has been delineated in the previous section.

The justification for the calculation of the random mating load from the following description of a population will be come apparent in the subsequent development of our theme. It should be noted that in this section and the following one (on the segregation load) we are most of the time concerned with the development of functional or structural relationships. The problem of the estimation of the quantities involved will be touched on in Section 4.

In order to calculate the random mating load we write:

Genotype	(ij)	
Frequency	$p(i)p(j)$	(A.14)
Fitness	$x(ij)$,	

where $x(11) > x(ij)$ for $i \neq 1$ or $j \neq 1$ and where $p(1)$ is near to unity so that $p(i)p(j)$ for $i \neq 1$ and $j \neq 1$ can be considered to be of negligible magnitude. These assumptions are quite legitimate since from Chapter III, Section F.1.a we have that an equilibrium is stable in such a case. In Section F.1.c. of the same chapter we gave approximate solutions to the equilibrium equations in the two-allele case and showed that these equilibria will be stable under most conditions that one is likely to encounter. In the m-allele case approximations to the equilibrium gene

frequencies can be obtained from the derivations leading to Equation F.18 in Section F.1.b. of Chapter III.

We assume $x(11) = c$ and $x(1j) = c(1-h(j)s(jj))$, $j \neq 1$, where $s(jj) > 0$ and $0 < h(j) \leq 1$, and where c is a constant factor common to all genotypes. From the definition of the load (Chapter II, Equation 1) it follows that for the random mating load we have

$$L_0 = 1 - (1 - 2\,p(1) \sum_{j\neq 1} p(j)h(j)s(jj))$$

$$= 2p(1) \sum_{j\neq 1} p(j)h(j)s(jj) \quad , \tag{A.15}$$

since the $p(i)p(j)$ are very small for $i \neq 1$ and $j \neq 1$, and where L_0 denotes the random mating load.

To calculate the inbred load we assume that complete homozygosity is achieved without any change in gene frequencies. We write

Genotype	(jj)	
Frequency	p(j)	(A.16)
Fitness	x(jj) ,	

where $x(11) = c$ and $x(jj) = c(1 - s(jj))$ for $j \neq 1$. From Chapter II, Equation 1 the inbred load is then

$$L_1 = 1 - (1 - \sum_{j\neq 1} p(j)s(jj))$$

$$= \sum_{j\neq 1} p(j)s(jj) \quad . \tag{A.17}$$

It follows from Equations A.15 and A.17 that

$$\frac{L_1}{L_0} = \frac{\sum_{j\neq 1} p(j)s(jj)}{2 \sum_{j\neq 1} p(j)h(j)s(jj)} \quad , \tag{A.18}$$

since we assume $p(1)$ to be approximately equal to unity. It follows then

from the fact that the mean of positive numbers is always greater or equal to the smallest value and less than or equal to the largest that

$$\frac{1}{2 \max_{j \neq 1}(h(j))} \leq \frac{L_1}{L_0} \leq \frac{1}{2 \min_{j \neq 1}(h(j))} . \tag{A.19}$$

This result agrees with the two allele case given by Crow (1958) and which was also derived in Chapter II.

3. The segregation load

We assume that the fitness of the genotype (ij) is x(ij) and that its frequency is p(i)p(j) . The assumption of Hardy-Weinberg frequencies is legitimate in the context of Section 1 if the parents and offspring are counted in infancy. We also assume that we are in a system in which the mutation rates are so small that the approximation of a selection system with mutation by a system without mutation is satisfactory. This topic was covered in detail in Chapter III, Section F.1.b. For present purposes we assume that all the alleles (i), i=1, ... ,m are internal points of the genetic system as defined in Section III.C and in subsequent developments given in Chapter III.

We have

$$\begin{aligned} &\text{Genotype} && (ij) && i,j = 1,2, \ldots ,m \\ &\text{Frequency} && p(i)p(j) \\ &\text{Fitness} && x(ij) = c(1 - s(ij)),\ 0 \leq s(ij) \leq 1 . \end{aligned} \tag{A.20}$$

We denote

$$x(..) = \sum_i \sum_j p(i)p(j)x(ij)$$
$$x(i.) = \sum_j p(j)x(ij) .$$

The change in gene frequency is

$$\Delta p(i) = \frac{\sum_j p(i)p(j)x(ij) - p(i)x(..)}{x(..)}$$

$$= \frac{p(i)[x(i.) - x(..)]}{x(..)} ,$$

which is in agreement with Equation B.11 of Chapter III for reasons explained in Section A.1 in the derivation of Equation A.2.

Since we assume that the population is in equilibrium through selection we have that at equilibrium

$$x(i.) = x(..) \quad .$$

If it is assumed that the optimum genotype, one of the heterozygotes, has fitness c, then we define the fitness of the genotype (ij) as $c(1 - s(ij))$, where $s_{ij} = 0$ for the maximal genotype.

We then have that

$$c(1-s(i.)) = x(i.) \quad \text{and} \quad c(1-s(..)) = x(..) \quad .$$

It follows from the foregoing that at equilibrium

$$s(i.) = s(..) \quad . \tag{A.21}$$

In an infant population derived by random mating, the random mating load is

$$L_0 = \frac{c - c \sum_i \sum_j p(i)p(j)(1-s(ij))}{c}$$

$$= s(..) \tag{A.22}$$

$$= s(i.) \quad \text{at equilibrium.}$$

We have that $s(i.) \geq s(ii)p(i)$, so that at equilibrium $s(..) \geq s(ii)p(i)$.

It follows that

$$ms(..) \geq \sum_{i=1}^{m} p(i)s(ii) = L_1 \quad ,$$

where L_1 is the inbred genetic load.

It follows that

$$mL_0 \geq L_1$$

or

$$\frac{L_1}{L_0} \leq m \quad , \tag{A.23}$$

where m is the number of alleles in the genetic system at equilibrium.

It should be noted that for the totally inbred population the mean of the population is given by

$$c \sum_{i=1}^{m} p(i)(1-s(ii)) \quad .$$

It follows that if we calculate the load to be

$$L_1 = \sum_{i=1}^{m} p(i)s(ii) \quad ,$$

we have to assume that the maximal genotype in the inbred population is still the one with fitness c. The difficulty is of course that we assumed the maximal genotype to be a heterozygote, and that in a totally inbred population there will be no heterozygotes. The problem is now to reconcile the above calculation of the inbred load with the definition of the genetic load as

$$L = \frac{x_{max} - x(..)}{x_{max}} \quad .$$

Let us accept for the moment that the following representation of a population in terms of the inbreeding coefficient is valid:

Genotype	(ii)	(ij)
Frequency	$Fp(i)$	$(1-F)p(i)p(j)$
Fitness	$c(1-s(ii))$	$c(1-s(ij))$.

It follows that the mean of the population is

$$c[F \sum_i p(i)(1-s(ii)) + (1-F)\sum_i \sum_j p(i)p(j)(1-s(ij))]$$

$$= c[F-F \sum_i p(i)s(ii) + (1-F) \sum_i \sum_j p(i)p(j)(1-s(ij))] \ ,$$

and hence that

$$L_F = \frac{c[(1-F) + F \sum_i p(i)s(ii) - (1-F) \sum_i \sum_j p(i)p(j)(1-s(ij))]}{c} \ .$$

Then

$$\lim_{F \to 1} L_F = \sum_i p(i)s(ii) = L_1 \ .$$

This argument justifies the definition of the inbred load as it is given in the foregoing, because the use of the inbreeding coefficient in the present context is considered to be legitimate, as will become clear later.

4. The estimation of load ratios

We have now derived properties by which one can determine the nature of the load in a population if one has an estimate of the magnitude of the load ratio and some idea of how many genes per locus to expect and of the degree of dominance as measured by the $h(j)$'s of Section 2.

For the purpose of laying bare the arguments employed for the derivation of the estimation procedure in the literature (that of Morton, Crow and Muller (1956)), we will derive a multi-loci result by an essentially single locus procedure. In Section B of this chapter it will then be shown

that this result is a special case derived from a more general one on the assumption of no linkage between loci. We leave the justification of the description of the genotypic array by the inbreeding coefficient until the end of this section. Further, we assume that in a population with degree of inbreeding F, the genotypic array can be written as follows:

Genotype	$(i^a \; i^a)$	$(i^a \; j^a)$, $(i^a \neq j^a)$	
Frequency	$Fp(i^a) + (1-F)p(i^a)p(i^a)$	$(1-F)p(i^a)p(j^a)$	(A.24)
Fitness	$x(i^a i^a)$	$x(i^a j^a)$.	

The mean fitness of the population due to gene action at locus a is then:

(i) When $F = 0$

$$x_0^a = \sum_{i^a} \sum_{j^a} p(i^a)p(j^a)x(i^a j^a) \quad ,$$

(ii) For arbitrary F

$$x_F^a = F \sum_{i^a} p(i^a)x(i^a i^a) + (1-F) \sum_{i^a} \sum_{j^a} p(i^a)p(j^a)x(i^a j^a) \quad , \tag{A.25}$$

(iii) For $F = 1$

$$x_1^a = \sum_{i^a} p(i^a)x(i^a i^a) \quad .$$

Also we define the load at the a-th locus to be

$$L_F^a = \frac{x_{max}^a - x_F^a}{x_{max}^a}$$

so that

$$x_F^a = x_{max}^a(1-L_F^a) \quad . \tag{A.26}$$

We have then from Equations A.25 and A.26 that

$$x_F^a = Fx_1^a + (1-F)x_0^a$$

$$= x_0^a + F(x_1^a - x_0^a) \qquad (A.27)$$

$$= x_{max}^a[1-L_0^a + F(L_0^a-L_1^a)]$$

$$= x_{max}^a[1-\{L_0^a + F(L_1^a-L_0^a)\}] \quad . \qquad (A.28)$$

Now if the loci act multiplicatively and independently we have that the total fitness is

$$x_F = \prod_{a=1}^{n} x_F^a \quad .$$

Also, if the effect of a single locus on fitness is small, we have that

$$x_F^a = x_{max}^a \exp[-\{L_0^a + F(L_1^a-L_0^a)\}] \quad .$$

It follows that

$$x_F = (\prod_{a=1}^{n} x_{max}^a)\exp[-\{\sum_{a=1}^{n} L_0^a + F(\sum_{a=1}^{n} L_1^a - \sum_{a=1}^{n} L_0^a)\}] \qquad (A.29)$$

or

$$x_F = x_{max}\exp[-(A + BF)] \, , \qquad (A.30)$$

where

$$A = \sum_{a=1}^{n} L_0^a \quad \text{and} \quad B = \sum_{a=1}^{n} L_1^a - \sum_{a=1}^{n} L_0^a$$

If we postulate additive gene action between loci we have that the total fitness is

$$x_F = \sum_{a=1}^{n} x_F^a \, ,$$

so that in this case the Equation A.28 becomes

$$x_F = \Sigma x_{max}^a - \Sigma x_{max}^a L_0^a + F\{\Sigma x_{max}^a L_0^a - \Sigma x_{max}^a L_1^a\} \quad . \qquad (A.31)$$

If we assume zero correlation over loci between x^a_{max} and L^a_0 and between x^a_{max} and L^a_1 we can write Equation A.31 as

$$x_F = x_{max}\{1 + \frac{1}{n}\sum_a L^a_0 + F(\frac{1}{n}\sum_a L^a_1 - \frac{1}{n}\sum_a L^a_0)\} \quad , \qquad (A.32)$$

where n is the number of loci. Let us write $D = x_{max}(1 + \frac{1}{n}\sum_a L^a_0)$ and $E = x_{max}(\frac{1}{n}\sum_a L^a_1 - \frac{1}{n}\sum_a L^a_0)$, so that it follows that Equation A.32 can be written as

$$x_F = D + EF \quad . \qquad (A.33)$$

If we assume multiplicative gene action between loci and $x_{max} = 1$, we see from Equation A.30 that we will be able to construct estimates of

$$\frac{\sum_a L^a_1}{\sum_a L_0} \quad ,$$

since

$$\frac{A + B}{A} = \frac{\sum_a L^a_1}{\sum_a L^a_0} \quad .$$

Under the assumption of additive gene action between loci the same type of approach would apply to Equation A.33, since

$$\frac{D + E - 1}{D - 1} = \frac{\sum_a L^a_1}{\sum_a L^a_0} \quad .$$

It will be better to delay a full discussion about the estimation problem in the context of genetic loads until the derivation of the n-loci case has been given. However, some aspects of the problem pertinent in the present context will be noted.

The derivation of the properties of $\frac{A+B}{A}$ given here differs in two important respects from that given by Morton, Crow and Muller (1956).

First, these authors in effect assumed $x_{max}^{a} = 1$ for all a, and consequently they could write Equation A.30 as

$$x_F = \exp[-(A+BF)] \quad . \tag{A.34}$$

Second they assume that they have a estimate y_F of x_F, which they then substitute for x_F in the Equation A.34, from whence they then proceed by taking logarithms for the purpose of doing a weighted regression analysis of Log y_F on F.

In contrast to this, we showed for a population derived from inbred lines with the same inbreeding coefficient F, that $E(z_F) = x_F^2$, where z_F is the mean number of infant offspring per infant mating pair. The construction of the population from inbred lines for the purpose of estimation was discussed in Section 1.

It is apparent that the differences between the approach of Morton, Crow and Muller (1956) and the one followed here will in general lead to divergent consequences. However, it will be shown in the next chapter that when natural selection operates through differential mortality an unbiased estimate of the mean viability of a population can be obtained. This fact will allow theoretical justification of some aspects of the theory of Morton, Crow and Muller (1956).

It has been mentioned that a serious objection to the part of the load theory as developed in this section lies with the use of the inbreeding coefficient in describing the genotypic array. The description of the genotypic array of a population by $F(\sum_i p(i)(ii)) + (1-F)(\sum_j p(i)(i))^2$ presupposes selectively neutral genes, no mutation and that the population was derived from an original population, which itself arose by random mating, by a process of mating based purely on consanquinity.

However, it is possible to argue that although the use of the inbreeding coefficient F weakens the logical cogency of the load argument, in practice no great error is introduced. The argument would be that with an organism having something like 10,000 genes, one cannot apply strong selection to each of them. If the effect of inbreeding on decreasing heterozygosity is, say, 25 percent per generation, this means that even intense selection cannot counteract the increased homozygosity of more than a small fraction of these in any one population.

There is also another slightly different way of looking at the effect of describing the genotypic array by the use of the inbreeding coefficient F for the purpose of estimation of the load ratio. Let us assume that the population can be placed in an experimental situation where all genes become selectively neutral. Suppose furthermore, that fitness is under these circumstances an observable characteristic in the same way as is, say, wool production in sheep. In other words, we assume that an experimental situation can be created in which the load theory holds exactly. The parameters estimated under such an idealized experimental situation would then be applicable to the description of populations under natural circumstances.

From this point of view, then, the errors in our estimates in practice derive from our inability to create a suitable experimental situation. One way in which the ideal experimental situation is often approximated in practice is to have all inbreeding to take place in one generation so as to give selection as little chance as possible to counteract the effects of inbreeding. It is clear that the effect of selection would be to introduce some divergence between the theoretical F and a "realized" F beyond that which can be expected on the basis of finite population size. This

question will be given further consideration in the next chapter.

B. The Two-loci Case

1. The description of the genotypic array on inbreeding in the case of selectively neutral genes

Following Shikata (1965) we define the following inbreeding coefficients in the two-loci case with the first locus denoted as the a-locus and the second as the b-locus

$$\begin{aligned} F_{11} &= \text{Prob}[a_r = a_s \,,\, b_t = b_u] \,, \\ F_{10} &= \text{Prob}[a_r = a_s \,,\, b_t \neq b_u] \,, \\ F_{01} &= \text{Prob}[a_r \neq a_s \,,\, b_t = b_u] \,, \\ \text{and} \qquad F_{00} &= \text{Prob}[a_r \neq a_s \,,\, b_t \neq b_u] \,, \end{aligned} \tag{B.1}$$

from which it follows that we have

$$F_{1.} = F_{11} + F_{10} = P[a_r = a_s] \,,$$

with obvious extensions to $F_{.1}$, $F_{0.}$ and $F_{.0}$. The equality relation $a_r = a_s$ means here that a_r and a_s are genes identical by descent. We note that straightforward correspondences exist between the functions employed here for the description of the inbreeding process and those used by Schnell (1961) for the same purpose. Schnell's work can therefore be consulted to develop an understanding about inbreeding in the case of more than one locus.

It is well known that without selection and mutation the genotypic array of a random mating population in equilibrium can be written as $[\sum_{i^1}\sum_{i^2} p(i^1)p(i^2)(i^1)(i^2)]^2$, where $p(i^1)$ is the frequency of the gene

(i^1) at the first locus and where $p(i^2)$ is the frequency of the gene (i^2) at the second locus.

By the type of argument which was employed by Kempthorne (1957) in the single locus case, we have by virtue of the Equations B.1 that, with mating based purely on consanquinity, the genotypic array resulting from inbreeding will be given by

$$F_{11} \sum_{i^1} \sum_{i^2} p(i^1)p(i^2)(i^1i^1i^2i^2) + F_{10} \sum_{i^1} {}_{i^2j^2} \sum p(i^1)p(i^2)p(j^2)(i^1i^1i^2j^2)$$

$$+ F_{01} {}_{i^1j^1} \sum \sum_{i^2} p(i^1)p(j^1)p(i^2)(i^1j^1i^2i^2) \qquad \text{(B.2)}$$

$$+ F_{00} {}_{i^1j^1} \sum {}_{i^2j^2} \sum p(i^1)p(j^1)p(i^2)p(j^2)(i^1j^1i^2j^2) \quad .$$

It is now easy to see that by random mating between members of independently derived inbred lines one would obtain a population with an array of mating pairs given by the square of the array given as Equation B.2. The process is entirely analogous to the case of a single locus which was spelled out in detail because of its simple explanation of the principles involved.

2. The relationship between the number of offspring per mating and mean fitness

The quantities involved in obtaining the expected number of offspring per mating pair are in the two-loci case entirely analogous to the single locus case, which was covered in extenso in Section A.1 of this chapter. In the two-loci case, the expected number of infant offspring for the infant mating pair $(i^1j^1i^2j^2) \times (r^1s^1r^2s^2)$ is $a(i^1j^1i^2j^2)a(r^1s^1r^2s^2)$. Hence, we have that $E(a(i^1j^1i^2j^2)a(r^1s^1r^2s^2))$

$$= \Big(\sum_{\substack{i^1 i^2 \\ j^1 j^2}} p(i^1 i^2)p(j^1 j^2)a(i^1 j^1 i^2 j^2)\Big)^2 \qquad \text{(B.3)}$$

for a random mating population and where $p(i^1 i^2) = p(i^1)p(i^2)$ in the case of Hardy-Weinberg structure at each of the constituent loci. From the comments in the previous section it is also clear that with random mating between infants from different independent inbred lines of the same degree of inbreeding, the expected number of infant offspring from a specific subpopulation of infant mating pairs would be given by the square of the genotypic array denoted by Equation B.2, where the symbols $a(i^1 j^1 i^2 j^2)$ are substituted for the symbols $(i^1 j^1 i^2 j^2)$. It is again of importance to note that this will be true only if the population from which the inbred lines were derived, had Hardy-Weinberg structure at each locus.

In the situation under consideration the variance of the number of infant offspring per infant mating pair is

$$\sum_{\substack{i^1 j^1 r^1 s^1 \\ i^2 j^2 r^2 s^2}} \sum_t t^2 b_t(i^1 j^1 i^2 j^2 r^1 s^1 r^2 s^2)\ell(i^1 j^1 i^2 j^2)\ell(r^1 s^1 r^2 s^2) \times p(i^1 j^1 i^2 j^2)p(r^1 s^1 r^2 s^2) - \Big(\sum_{i^1 j^1 i^2 j^2} a(i^1 j^1 i^2 j^2)p(i^1 j^1 i^2 j^2)\Big)^4 . \qquad \text{(B.4)}$$

The reason for the comparatively simple structure of the Equation B.4 is that because of our mating scheme of random mating between equally inbred lines, the frequency of the mating between the infant genotypes $(i^1 j^1 i^2 j^2)$ and $(r^1 s^1 r^2 s^2)$, $p(i^1 j^1 i^2 j^2 r^1 s^1 r^2 j^2)$ is equal to $p(i^1 j^1 i^2 j^2)p(r^1 s^1 r^2 s^2)$, where the $p(i^1 j^1 i^2 j^2)$ take on values according to the relationship of identity by descent between the genes i^1 and j^1 and between i^2 and j^2 as it is given in the array represented by the Equation B.2.

We note here that it is easy to obtain the recurrence relations for the calculation of F_{11}, F_{10} and F_{01} in the case of selfing with arbitrary linkage, but that thus far the only other case in which detailed derivations has been obtained is that of full-sibbing (Cockerham and Weir, 1968). These authors can also be consulted for a general survey of the topic.

The relationship $p(i^1_j i^1_i i^2_j i^2_r i^1_s i^1_r i^2_s i^2_s) = p(i^1_j i^1_i i^2_j i^2_j)p(r^1_s i^1_r i^2_s i^2_s)$ between the frequencies of individuals from different inbred lines causes the covariance between the infant offspring of the infant mating pairs A x B and C x D to be zero if A, B, C and D are all obtained from different inbred lines. The argument is entirely analogous to the case given in Equation A.9 where the covariance will be equal to zero if $p(ijde) = p(ij)p(de)$. In the case where A and C or B and D are derived from the same inbred line formulas analogous to the Equations A.9 and A.10 can easily be developed. However, the evaluation of $p(i^1_j i^1_i i^2_j i^2_r i^1_s i^1_r i^2_s i^2_s)$, representing the frequencies of pairs of individuals such as A and C or B and D, would present a severe problem. The problem of the relationship between two individuals, as defined by the relationship of identity by descent, is difficult enough in the single locus case as one has to deal with 15 mutually exclusive events. With two loci there will be 15 x 15 events that need to be considered. The case of n loci will encompass 15^n events.

3. <u>The relationship between population mean fitness and loads on inbreeding</u>

In the same way as in the single locus case we denote the fitness of a genotype $(i^1_j i^1_i i^2_j i^2_j)$ by $x(i^1_j i^1_i i^2_j i^2_j)$ instead of $a(i^1_j i^1_i i^2_j i^2_j)$ in order

to emphasize that in some ways these numbers are going to be treated in the same manner as measurements on a metric trait associated with selectively neutral genes and therefore that due care must be exercised in our thinking. The assumption of Hardy-Weinberg structure of gene frequencies in the population before inbreeding implies the assumption of multiplicative or additive gene action between loci as was shown in Section D of Chapter III. In the additive case we write $x(i^1j^1i^2j^2) = x(i^1j^1) + x(i^2j^2)$ and in the multiplicative case we write $x(i^1j^1i^2j^2) = x(i^1j^1)x(i^2j^2)$.

We now define the following means

$$
\begin{aligned}
x_{11}^{12} &= \sum_{i^1 i^2} p(i^1)p(i^2)x(i^1i^1i^2i^2) \quad , \\
x_{10}^{12} &= \sum_{i^1} \sum_{i^2 j^2} p(i^1)p(i^2)p(j^2)x(i^1i^1i^2j^2) \quad , \\
x_{01}^{12} &= \sum_{i^1 j^1} \sum_{i^2} p(i^1)p(j^1)p(i^2)x(i^1j^1i^2i^2) \quad , \\
\text{and} \quad x_{00}^{12} &= \sum_{i^1 j^1} \sum_{i^2 j^2} p(i^1)p(j^1)p(i^2)p(j^2)x(i^1j^1i^2j^2) \quad .
\end{aligned}
\tag{B.5}
$$

By an obvious extension of the Equations B.5, we write in the additive case

$$x_{11}^{12} = x_1^1 + x_1^2 \quad \text{and} \quad x_{10}^{12} = x_1^1 + x_0^2 \text{ , etc.}$$

In the multiplicative case we write

$$x_{11}^{12} = x_1^1 x_1^2 \quad \text{and} \quad x_{10}^{12} = x_1^1 x_0^2 \text{ , etc.,}$$

where x_1^1 , x_0^1 are mean fitnesses at the first locus and x_1^2 , x_0^2 are mean fitnesses at the second locus.

In the same way as in the single locus case it follows from Equations B.2 and B.3 that the population mean fitness, $x^{12}_{F_{11},F_{10},F_{01}}$, can be defined on inbreeding as being

$$x^{12}_{F_{11},F_{10},F_{01}} = F_{11}(x^{12}_{11} - x^{12}_{00}) + F_{10}(x^{12}_{10} - x^{12}_{00}) + F_{01}(x^{12}_{01} - x^{12}_{00}) + x^{12}_{00} , \tag{B.6}$$

since $F_{00} = 1 - F_{11} - F_{10} - F_{01}$.

We note that in the absence of linkage between loci the relationships $F_{11} = F^2$, $F_{10} = F(1-F)$ and $F_{01} = F(1-F)$ hold, where F is the single locus inbreeding coefficient. In the case of no linkage and of additive gene action between loci, Equation B.6 becomes by virtue of the extensions to the Equations B.5

$$x^{12}_{F_{11},F_{10},F_{01}} = [x^1_0 + F(x^1_1 - x^1_0)] + [x^2_0 + F(x^2_1 - x^2_0)] . \tag{B.7}$$

In the case of no linkage and multiplicative gene action between loci the extensions to the Equations B.5 cause Equation B.6 to become on simplification

$$x^{12}_{F_{11},F_{10},F_{01}} = [x^1_0 + F(x^1_1 - x^1_0)][x^2_0 + F(x^2_1 - x^2_0)] . \tag{B.8}$$

The Equations B.7 and B.8 are entirely equivalent to the Equation A.27 from which the single locus load formulas were developed. It is therefore clear that the loads in the multi-loci case will be equivalent to the loads in the single locus development if there is no linkage between loci.

There are also other ways by which equations similar to those obtained in the single locus case can be obtained. By substituting the additive form of the Equations B.5 into B.6, and by rearranging and summing the relevant terms we get

$$x^{12}_{F_{11},F_{10},F_{01}} = [x^1_0 + F_{1.}(x^1_1 - x^1_0)] + [x^2_0 + F_{.1}(x^2_1 - x^2_0)] \quad . \qquad (B.9)$$

Since $F_{1.} = F_{.1} = F$, say, Equation B.9 is precisely equivalent to Equation B.7 and hence the conclusion follows that linkage has no effect in the case of additive gene action between loci.

As is to be expected, the case of multiplicative gene action between loci is slightly more complicated than the additive case. Entirely analogous to the single locus case we now define the following loads:

$$L^{12}_{11} = \frac{x^{12}_{max} - x^{12}_{11}}{x^{12}_{max}} \quad ,$$

$$L^{12}_{00} = \frac{x^{12}_{max} - x^{12}_{00}}{x^{12}_{max}} \quad , \qquad (B.10)$$

$$L^{12}_{01} = \frac{x^{12}_{max} - x^{12}_{01}}{x^{12}_{max}} \quad ,$$

and

$$L^{12}_{10} = \frac{x^{12}_{max} - x^{12}_{10}}{x^{12}_{max}} \quad .$$

Remembering that $x^1_1 = x^1_{max}(1-L^1_1)$ and $x^2_1 = x^2_{max}(1-L^2_1)$ and that $x^{12}_{max} = x^1_{max}x^2_{max}$, it follows easily from the Equations B.10 that $L^{12}_{11} = L^1_1 + L^2_1 - L^1_1L^2_1$, etc., where L^1_1 and $L^{2'}_1$ are inbred loads at the first and second loci and where L^1_0 and L^2_0 are random mating loads at the first and second loci respectively. Hence, for small loads at each locus we have

$$L_{11}^{12} = L_1^1 + L_1^2 \ ,$$

$$L_{00}^{12} = L_0^1 + L_0^2 \ ,$$

$$L_{01}^{12} = L_0^1 + L_1^2 \ , \qquad \text{(B.11)}$$

and

$$L_{10}^{12} = L_1^1 + L_0^2 \ .$$

Let us now substitute loads for the means in Equation B.6 according to the relationships given by the Equations B.10 to get

$$x^{12}_{F_{11},F_{10},F_{01}} = x^{12}_{max}(1 - F_{11}L_{11}^{12} - F_{10}L_{10}^{12} - F_{01}L_{01}^{12} - F_{00}L_{00}^{12}) \ . \qquad \text{(B.12)}$$

Hence, on substituting the Equations B.11 in Equation B.12 and on summing over the necessary terms we obtain

$$x^{12}_{F_{11},F_{10},F_{01}} = x^{12}_{max}(1 - F_{1.}L_1^1 - F_{.1}L_1^2 - F_{0.}L_0^1 - F_{.0}L_0^2) \ , \qquad \text{(B.13)}$$

or

$$x_F^{12} = x^{12}_{max}[1 - L_0^1 - L_0^2 - F(L_1^1 + L_1^2 - L_0^1 - L_0^2)] \ , \qquad \text{(B.14)}$$

since $F_{1.} = F_{.1} = F$, say, and hence $F_{0.} = F_{.0} = 1\text{-}F$. Equation B.14 is the two-loci equivalent to Equation A.28 in the single locus case.

4. <u>The load ratios at each of the two loci</u>

From our work in Section F of Chapter III it is clear that the same limits can be attached to the range of values of

$$\frac{L_1^1}{L_0^1} \quad \text{and} \quad \frac{L_1^2}{L_0^2}$$

as in the single locus case which was given in Equations A.19 and A.23.

In Section D.3 we showed that at equilibrium in the case of additive gene action between loci, the same Hardy-Weinberg frequencies obtain at each of the two loci as that which would have followed from a consideration of each locus separately. The same property holds in the case of multiplicative gene action between loci, provided that the linkage between the loci is not too tight.

In Section III.F.2.c we showed in the discussion on Equations III.F.30 through III.F.33 and on Equations III.F.36 through III.F.40 that, if the mutation rates are small, a system with mutation can be reasonably well described by a system without mutation.

The foregoing results are of importance for the segregation load. The results of importance for the mutation load are that of Equations III.F.25 and III.F.26, and of Equations III.F.41 through III.F.44. In the transposition of these results to the consideration of the single locus loads, it is well to remember that in the derivation of Equation A.19 we were not concerned with the precise values of the gene frequencies, but rather with the assumption that some are small, and others are near to unity.

C. The n-Loci Case

1. The relationship between population mean fitness and loads on inbreeding

The theory of loads in the n-loci case is pretty much a straightforward extension of the 2-loci case. Again we have to assume Hardy-Weinberg structure of gene frequencies in the population. The conditions under which this is a possibility was given in Section III.G. Since we deduced from Theorem 5 that all possible two-loci Hardy-Weinberg systems must be stable in order for the n-loci system to be stable, we have from Theorem 1 in Section III.D

that in the case of multiplicative gene action between loci, the linkage between loci must not be too tight. Additive gene action between loci will, of course, be enough to guarantee the existence of Hardy-Weinberg equilibria, as follows from Chapter III. It follows from the discussion in Sections III.E and III.G that, in the case of the fitness matrix being of full rank on the internal points, the n-loci Hardy-Weinberg equilibrium will be the only possible stable equilibrium under the assumption of additive gene action between loci.

In the same way as in the Equations B.1 we denote the inbreeding coefficients here as $F_{y_1 y_2 \cdots y_n}$, where $y_a = 0,1$ for $a = 1,2, \ldots ,n$. If y_a takes the value unity for a particular a, it means that at that locus the status of identity by descent between genes holds. Similarly the value zero for y_a indicates that the status of not identical by descent is under consideration. It follows that the genotypic array of a population without selection, derived from a random mating population in equilibrium by a process of mating based purely on consanguinity, can be written in the n-loci case in a fashion precisely analogous to Equation B.2. In the same way as in the Equations B.5 we also define means denoted by $x^{12\ldots n}_{y_1 y_2 \cdots y_n}$, with $y_a = 0,1$ for $i = 1,2, \ldots ,n$. If $y_a = 1$ the genes at the a-th locus are identical by descent with an associated frequency of $p(k^a)$ for the k^a-th allele. If $y_a = 0$ the genes at the a-th locus are not identical by descent, and hence the frequency associated with the a-th locus is $p(r^a)p(s^a)$ for the r^a-th and s^a-th alleles respectively. It follows that at infancy the population mean fitness on inbreeding, which we denote by $x^{12\ldots n}$, can be written as

$$x^{12\ldots n} = \sum_{y_1, y_2, \cdots, y_n = 0,1} F_{y_1 y_2 \cdots y_n} x^{12\ldots n}_{y_1 y_2 \cdots y_n} \quad . \qquad (C.1)$$

In the case of additive gene action between loci we can write $x^{12\ldots n}_{y_1 y_2 \ldots y_n} = \sum_{a=1}^{n} x^{a}_{y_a}$. Hence, in this case Equation C.1 becomes

$$x^{12\ldots n} = F(\sum_a x^a_1) + (1-F)(\sum_a x^a_0)$$

$$= \sum_a x^a_0 + F(\sum_i x^a_1 - \sum_i x^a_0) \quad , \qquad (C.2)$$

since $F_{1\ldots} = F_{.1\ldots} = F_{..1\ldots} = \ldots = F$, say, and since $F_{0\ldots} = F_{.0\ldots} = F_{..0\ldots} = \ldots = 1 - F$, where a dot indicates that a subscript has been summed over. If we remember the relationship $x^a_{y_a} = x^a_{max}(1-L^a_{y_a})$ then it is easy to see that Equation C.2 leads to Equation A.31, and hence that with additive gene action between loci an essentially single locus approach led to the correct results.

It is now necessary to define n-loci loads as

$$L^{12\ldots n}_{y_1 y_2 \ldots y_n} = \frac{x^{12\ldots n}_{max} - x^{12\ldots n}_{y_1 y_2 \ldots y_n}}{x^{12\ldots n}_{max}} \quad , \qquad (C.3)$$

where $x^{12\ldots n}_{max}$ is the value of the genotype with the maximum fitness. We also have under the assumption of multiplicativity that

$$x^{12\ldots n}_{max} = \prod_{a=1}^{n} x^a_{max}$$

and that

$$x^{12\ldots n}_{y_1 y_2 \ldots y_n} = \prod_{a=1}^{n} x^a_{y_a} \quad ,$$

in the same way as in the multiplicative case of the Equations B.5. Hence, if we recall that

$$x^{a}_{y_a} = x^{a}_{max}(1-L^{a}_{y_a}) \quad ,$$

where $L^{a}_{y_a}$ is the load at the a-th locus, we can write that

$$x^{12\ldots n}_{y_1\ldots y_n} = x^{1\ldots n}_{max}\left(1 - \sum_{a} L^{a}_{y_a} + \sum_{a<b} L^{a}_{y_a} L^{b}_{y_b} - \sum_{a<b<c} L^{a}_{y_a} L^{b}_{y_b} L^{c}_{y_c}\right.$$

$$\left. + \ldots + (-1)^{n} L^{1}_{y_1} \ldots\ldots L^{a}_{y_n}\right) \quad ,$$

from which it follows by the Equation C.3 that

$$L^{1\ldots n}_{y_1\ldots y_n} = \sum_{a} L^{a}_{y_a} - \sum_{a<b} L^{a}_{y_a} L^{b}_{y_b} + \sum_{a<b<c} L^{a}_{y_a} L^{b}_{y_b} L^{c}_{y_c}$$

$$- \ldots - (-1)^{n} L^{1}_{y_1} L^{2}_{y_2} \ldots\ldots L^{n}_{y_n} \quad .$$

We now assume the loads at each locus to be small enough so that their products with each other are of negligible magnitude. From the previous equation it is clear that although there are very many of these small terms of load products, they occur in a series with alternating signs to its terms. Hence, the equation

$$L^{12\ldots n}_{y_1 y_2 \ldots y_n} = \sum_{a=1}^{a} L^{a}_{y_a} \qquad (C.4)$$

will probably hold to a satisfactory degree if the loads at each locus are reasonably small.

By virtue of Equation C.3 it is now possible to write Equation C.1 as

$$x^{12\ldots n} = x^{12\ldots n}_{max}\left(1 - \sum_{y_1, y_2, \ldots, y_n = 0,1} F_{y_1 y_2 \ldots y_n} L^{12\ldots n}_{y_1 y_2 \ldots y_n}\right) . \qquad (C.5)$$

If we now substitute Equations C.4 into Equation C.5 we get

$$x_F^{12\ldots n} = x_{max}^{12\ldots n}(1 - F\sum_a L_1^a - (1-F)\sum_a L_0^a)$$

or

$$x_F^{12\ldots n} = x_{max}^{12\ldots n}[1 - \sum_a L_0^a + F(\sum_a L_0^a - \sum_a L_1^a)] \quad , \qquad (C.6)$$

since, as before, $F_{1\ldots} = F_{.1\ldots} = \ldots = F_{\ldots 1} = F$, say, and since $F_{0\ldots} = F_{.0\ldots} = F_{.0\ldots} = \ldots = F_{\ldots 0} = 1 - F$, where a dot indicates that a subscript has been summed over.

We now transform Equation C.6 to

$$x_F^{12\ldots n} = x_{max}^{12\ldots n} \exp - \{\sum_a L_0^a + F(\sum_a L_1^a - \sum_a L_0^a)\} \quad ,$$

which is equivalent to Equation A.29 in the single locus derivation. This involves not only the assumption that the loads at each locus are small, but also that their sums over all loci must be small. Such an assumption may be justifiable in a few cases, as will be noted in the next chapter on applications of the load theory, but in general such an assumption of small sums of loads seems not to be acceptable as there must be very many loci determining fitness. It follows that one can expect the estimation procedure pertinent to Equations A.29 or A.30 not to be applicable in all cases.

In order to ease our notational problem, let us write the Equation C.6 in the form

$$x_F^{12\ldots n} = c[(1-L_0) + (L_0 - L_1)F] \quad , \qquad (C.7)$$

where

$$c = x_{max}^{12\ldots n} ,$$

$$L_0 = \sum_{a=1}^{n} L_0^a ,$$

and

$$L_1 = \sum_{a=1}^{n} L_1^a .$$

2. The estimation of genetic loads

We now assume that we have a population of mating pairs which is divided into subpopulations derived from inbred lines according to the scheme outlined in Section A.1. Each subpopulation is constructed from independently derived inbred lines which are inbred to the same degree of inbreeding as measured by the inbreeding coefficient F. Each subpopulation is, therefore, characterized by the degree of inbreeding of its constituent inbred lines, and each of the different subpopulations is associated with a different inbreeding coefficient.

We presume that the number of infant offspring per infant mating pair is observable and denote by z_{ij} the number of infant offspring from the j-th infant mating pair of the i-th subpopulation. It follows by extension from the results given by Equation A.1 in the single locus case and by Equation B.3 in the case of two loci that

$$E(z_{ij}) = x_i^2 , \qquad (C.8)$$

or in words that the expected number of infant offspring per infant mating pair is equal to the square of the mean fitness, under the conditions as noted before. From these considerations and from the Equation C.7 it follows that a natural model for the situation under consideration is

$$z_{ij} = c^2[(1-L_0) + (L_0-L_1)w_i]^2 + e_{ij} \quad , \qquad (C.9)$$

where w_i denotes the degree of inbreeding of the subpopulation i, and where e_{ij} denotes the deviation of the number of infant offspring from the j-th infant mating pair from the mean of the i-th subpopulation. It follows from the description of the model that $E_j(e_{ij}) = 0$, where E_j indicates that the expectation is taken with respect to the variable indicated by the second subscript.

Recall now that we wish to estimate $\frac{L_1}{L_0}$. It is then easy to see from the relationship $E(z_{ij}) = c^2[(1-L_0) + (L_0-L_1)w_i]^2$, which follows from Equation C.9, that $\frac{L_1}{L_0}$ is non-estimable if c is unknown. This conclusion follows since there is no way by which $\frac{L_1}{L_0}$ can be expressed as a function of the $E(z_{ij})$ only.

It is, however, of some interest to proceed with the derivation of estimation procedures in order to get some idea of what may be achieved if estimates of c are available or if c is known.

The natural extension of Equation B.4 which gives the two-loci formula for the variances of the e_{ij} indicates that $Var(e_{ij}) = \sigma_i^2$, since the variance of the z_{ij} depend on the degree of inbreeding of the i-th subpopulation as it is described by the multivariate inbreeding coefficients $F_{y_1 y_2 \cdots y_n}$. The magnitude of these multivariate inbreeding coefficients depend on the degree of linkage which will in general be unknown. The σ_i^2 also depend on the unknown parameters which describe the fitness of the various genotypes. We note that $Cov(e_{ij}, e_{ij'}) \neq 0$, if the infants from the matings j and j' have parents from the same inbred line. This follows from Equations A.9 and A.10 in the single locus case and from the

discussion of the considerations following Equation B.4 in the case of two loci. The use of more than one individual from the same inbred line as a parent in the estimation of the mean fitness of a population may be impossible to avoid under some circumstances.

The Equation C.9 can be written in the form

$$z_{ij} = c^2[(1-L_0)^2 + 2(1-L_0)(L_0-L_1)w_i + (L_0-L_1)^2 w_i^2] + e_{ij} , \qquad (C.10)$$

which corresponds to a polynomial model

$$z_{ij} = f + gw_i + hw_i^2 + e_{ij} . \qquad (C.11)$$

If the relavant variances and covariances of the errors of the model of Equation C.10 can be estimated, a weighted regression analysis can then be performed on the basis of the model of Equation C.11. Then, if c is known, or an estimate of c is available, L_0 and L_1 can be estimated according to their correspondences with f, g and h.

We will also consider an adaptation of the maximum likelihood approach here, since it gives us the opportunity to extend the single locus results on the estimation of the mean fitness of Section A.1 to the multi-loci case.

For the moment, let us assume two loci to ease the notation. Also assume that we take only one individual from each inbred line as a parent in our estimation population. We know that in the case of the event of both infants surviving to adulthood the probability of obtaining t infant offspring per infant mating pair is given by

$$\sum_{\substack{i^1_j i^1_i i^2_j i^2 \\ r^1_s r^1_r r^2_s r^2}} b_t(i^1_j i^1_i i^2_j i^2 r^1_s r^1_r r^2_s r^2)\ell(i^1_j i^1_i i^2_j i^2)\ell(r^1_s r^1_r r^2_s r^2)p(i^1_j i^1_i i^2_j i^2)p(r^1_s r^1_r r^2_s r^2) .$$

It is now necessary to remember that the $p(i^1j^1i^2j^2)$ depend on the degree of inbreeding of the inbred lines that constitute the mating pairs of the subpopulation i. We therefore denote the foregoing expression by P_{it} in the case of the subpopulation i.

We note that we have to include the probability for the event of zero offspring due to the death of one or both members of an infant mating pair before adulthood in the probability for zero offspring as we defined it in the above. It is also clear that the definition of the P_{it}'s is in no way dependent on the assumption of a specific number of loci, or on the assumption of a specific number of alleles per locus.

Let us now classify the outcome of n_i independent matings from the subpopulation of mating pairs, i, with respect to the number of offspring per mating, and let us denote the number of matings with t offspring in the ith subpopulation by n_{it}. Hence, $\sum_{t=o}^{k} n_{it} = n_i$. It follows that the probability of obtaining $n_{i0}, n_{i1}, \ldots, n_{ik}$ infant pairs with respectively $0,1, \ldots, k$ offspring per pair is

$$\frac{n_i!}{n_{i0}!n_{i1}! \cdots n_{ik}!} (P_{i0})^{n_{i0}} (P_{i1})^{n_{i1}} \cdots (P_{ik})^{n_{ik}}, \qquad (C.12)$$

where we assume $P_{it} > 0$ for all i and for $t = 0,1, \ldots, k$.

The maximum likelihood estimators of the P_{it} are, of course, well known, and are simply given by $\frac{n_{it}}{n_i}$ for all t. However, we are at present not interested in the estimation of the P_{it}. From the definition of the various quantities involved we have that the expected number of infant offspring per infant mating pair in the subpopulation i is given by $\sum_t tP_{it} = x_i^2$, as follows from the extension to the Equation B.3 and where

x_i is the mean fitness of the i-th subpopulation. It follows from the Equation C.7 that

$$\sum_t tP_{it} = c^2[(1-L_0) + (L_0-L_1)w_i]^2 \quad , \qquad (C.13)$$

where w_i denotes the degree of inbreeding of the subpopulation i.

It is also necessary to note that the Equations C.13 will always be consistent, because the relationship given as Equation C.7 is an identity relationship and will hold good as long as all the subpopulations i were derived from the same original population.

Let us now assume we have three independent subpopulations, i.e. that i = 1,2,3. It is also clear that we can take square roots on both sides of the Equations C.13 without it being necessary to be concerned about the signs, since the left-hand side of each equation is always positive, and so is c on the right-hand side. By the Equation C.13 we can, therefore, write

$$\begin{aligned}
\sqrt{Q_1} &= \sqrt{\textstyle\sum_j jP_{1j}} = c[(1-L_0) + (L_0-L_1)w_1] \\
\sqrt{Q_2} &= \sqrt{\textstyle\sum_j jP_{2j}} = c[(1-L_0) + (L_0-L_1)w_2] \qquad (C.14) \\
\sqrt{Q_3} &= \sqrt{\textstyle\sum_j jP_{3j}} = c[(1-L_0) + (L_0-L_1)w_3] \quad .
\end{aligned}$$

We immediately note that we are essentially in the same trouble as we were in the case where we tried estimation by least squares. The Equations C.14 are of the form $y_i = a + bx_i$, which allows the estimation of only two parameters. In terms of maximum likelihood theory the trouble arises because a many to one relationship exists between $\sqrt{Q_1}$, $\sqrt{Q_2}$, $\sqrt{Q_3}$ and c, L_0 , L_2 .

The next step in our inquiry is to try and trace the reason for the non-estimability of $\frac{L_1}{L_0}$, or alternatively of c, L_0 and L_1. From the Equations C.14 it is clear that the problem is essentially that we have only two independent equations to solve for the three unknowns c, L_0, and L_1. This problem arises from the substitution of the two quantities defined in the Equation C.3 for the means $x^{12\ldots n}_{y_1 y_2 \ldots y_n}$ in the Equation C.1. For each mean two quantities are substituted, namely the corresponding load and the value of the maximal genotype. It can be seen from Equation C.1 that if we know all 2^n of the multiple loci inbreeding coefficients we would be able to estimate the 2^n means associated with them by the procedure set forth in the Equations C.14. If we look at the Equation C.1 in a multiple regression context, we observe that there are 2^n-1 independent multiple loci inbreeding coefficients which can be used as independent variables in the estimation of 2^n means. In contrast to this there are in the Equation C.5 2^n loads to be estimated together with the unknown value of the maximal fitness of the population. The subsequent simplification of the Equation C.5 did nothing to alleviate this problem. The fact that in the least squares procedure, which we employed in this section, both sides of the Equation C.1 has to be squared, only makes the estimation problem more complicated, but does not alter its substance.

The foregoing argument arose form the load theory in the case of multiplicative gene action between loci. It is, however, readily apparent that the problem also exists under the assumption of additive gene action between loci, as can be seen from the Equation C.2 and subsequent developments.

Since all the means that can be estimated from a population are functions of loads and the value of the maximal genotype, one possible way out of our dilemma would be to try to obtain an estimate of the value of the maximal genotype. Unfortunately the estimation of the value of the maximal genotype seems to be a practical impossibility. The maximal genotype is either a complete homozygote, or is completely heterozygous at all loci affecting fitness. In the case of many loci the probability of such a genotype occurring must be extremely small. Added to this difficulty would be the problem of identifying such an individual when it is present in the population. These points are well discussed by Levene (1963). It is unfortunate that the load theory should run into difficulty in this way.

These problems in relation to the estimation of the maximal genotype tie up with the only other conceivable approach to the load question. This alternative approach to the estimation of the genetic load would be to estimate all the fitnesses and to calculate the load directly from these estimates. In this case we would look at the likelihood function given as Equation C.12 as a function of all the probabilities for survival and offspring. However, it is apparent from our discussion on Equations A.11, A.12 and A.13 and from the definition of the quantities P_{it} that for this procedure at the very least we will have to know the number of loci and the number of alleles per locus. Such knowledge will only be available in very exceptional cases, so that we will not pursue the matter here.

Two other topics are also of interest. The first is the matter of error in the realized F's or the w_i's in the terminology employed here. We will, however, delay the discussion of the results that we were able to

obtain in this area to the next chapter, since it will fit in better there. The second topic pertains to the relationship between the theoretical and estimated load ratios. We recall from Section B.4 that the load theorems which distinguish between mutation and segregation loads pertain to ratios at a single locus, i.e., we proved results on $\frac{L_1^a}{L_0^a}$, where the superscript a indicates the a-th locus. It follows that estimates of $\sum_a \frac{L_1^a}{L_0^a}$, would be the easiest to interpret. However, it follows from the discussion on Equations C.7 and C.9 that if an estimate of the maximal genotype were available, we would be able to obtain estimates of

$$\frac{L_1}{L_0} = \frac{\sum_a L_1^a}{\sum_a L_0^a} .$$

About the only thing that can be said of this problem, it seems, is that the values of

$$\frac{\sum_a L_1^a}{\sum_a L_0^a} \quad \text{and} \quad \sum_a \frac{L_1^a}{L_0^a}$$

will not differ much if a relationship $L_1^a = kL_0^a$ holds approximately for all a, where k is a constant.

V. THE APPLICATION OF LOAD IDEAS

A. Loads in the Case of Metric Traits

1. General remarks

Loads can also be defined for metric traits such as mortality, morbidity due to specific causes and different sorts of abnormalities. For a more detailed description of these traits, Crow (1962) and especially Morton (1960, and references contained therein) can be consulted. In the present context it is sufficient to state that these metric traits may be components of fitness, and that thus far all applications of load theory in an experimental context have been to metric traits.

It is of great importance to note that we have to assume Hardy-Weinberg equilibria in order to use the inbreeding coefficient in the description of the genotypic array of a population. Of even greater importance, however, is to remember that the theorems which were designed to allow discrimination between the mutation and segregation loads on the basis of the ratio of the inbred load to the random mating load, only hold good when the trait under consideration is fitness itself, or when the genetic control of each component of fitness is totally independent from that of the others. The reason for this is that the equilibrium frequency of a gene is determined by its relationship to fitness.

Since it fits in well with our previous discussion, and since it is in one form or another the trait to which the load theory has been most often applied, we will modify the theory of the previous chapter to fit the necessary assumptions for metric traits in terms of the trait viability (or mortality) between infancy and adulthood. This simple division of the

life of an individual into two phases is considered to be adequate in conveying the nature, if not the precise magnitude, of the problem involved in the application of load theory.

2. Load theory in the case of metric traits

a. The single locus case We have from the Equation IV.A.7 that the expected number of adults per infant offspring is

$$F\sum_i p(i)\ell(ii) + (1-F)\sum_{ij} p(i)p(j)\ell(ij) \ , \qquad (A.1)$$

where all the terms are as defined previously. It follows that the ratio $\frac{\text{total number of adults}}{\text{total number of infants}}$ is an unbiased estimate of the mean viability of a population with degree of inbreeding F, if we define the mean viability of a population as being equal to the expression given as Equation A.1.

We see here that viability (or mortality) is a property of an individual, as opposed to fitness which is the property of a pair of individuals. It follows that it will be unnecessary to construct subpopulations for estimation purposes in the manner outlined in the previous chapter. It remains, however, necessary to be cautious about the use of the inbreeding coefficient F. The reasons for this are the same as those outlined in Section IV.A.1.

It is easy to see that the variance of the number of adults per infant offspring is equal to

$$\sum_{ij} \ell(ij)p(ij) - \left(\sum_{ij} \ell(ij)p(ij)\right)^2 \ , \qquad (A.2)$$

where $p(ii) = Fp(i)$ if the genes are identical by descent and $p(ij) = (1-F)p(i)p(j)$ otherwise. The covariance between the number of

adults per infant between infants from the same inbred line is

$$\sum_{ijrs} \ell(ij)\ell(rs)p(ijrs) - \left(\sum_{ij} \ell(ij)p(ij)\right)^2 , \qquad (A.3)$$

where the $p(ijrs)$ take on different values according to the relationships of identity by descent between the two individuals with genotypes (ij) and (rs), as was outlined in Equation IV.A.3. The covariance for the number of adults per infant between infants from independently derived inbred lines will be zero. It follows that the variance of the ratio

$\frac{\text{total number of adults}}{\text{total number of infants}}$ can be derived from the expressions given as Equations A.2 and A.3 if we know the relationship between different individuals in a population or a sample from a population.

The rest of the argument for loads in the case of metric traits proceeds now exactly according to the outline given in Section IV.A.4. The reason for this is that we can describe, for example, the mean viability in the fashion of Equation IV.A.25. As opposed to the case of fitness, we find that for the trait viability the approach of Morton, Crow and Muller (1956) will hold good. The reason for this is that the expectation of the proportion of infants surviving to adulthood is equal to the mean viability of the population. However, as was pointed out in Section IV.A.4, the assumptions of Morton, Crow and Muller's model require the limiting suppositions of no linkage between loci and that the maximal genotype has a numerical value of unity. The supposition of no linkage between loci can be relaxed if the sums of the loads are small, as follows from the discussion on Equation IV.C.6 referring to its possible equivalence to Equation IV.A.29.

The inclusion of an environmental effect is an additional feature of the model of Morton, Crow and Muller (1956) which deserves some attention. Let us denote the environmental effect by q and assume that the environment acts multiplicatively, so that we can modify the Equations IV.A.29 and IV.A.30 to write

$$z_F = qx_F = qx_{max} \exp [-(A+BF)] \quad . \qquad (A.4)$$

It is then obvious that, with a regression analysis of Log z_F on F in the case of x_{max} equal to unity, the environmental effect will be included in the estimate of A.

b. The multi-loci case To keep the notational problem within bounds which allow easy manipulation, we will again exemplify our general formulation by the two-loci case.

As we did before, we denote by $\ell(i^1j^1i^2j^2)$ the probability of survival from infancy to adulthood of an infant with genotype $(i^1j^1i^2j^2)$. We will consider only multiplicative gene action between loci, so that we write $\ell(i^1j^1i^2j^2) = \ell(i^1j^1)\ell(i^2j^2)$. Additive gene action between loci will not be considered here, since to write $\ell(i^1j^1i^2j^2) = \ell(i^1j^1) + \ell(i^2j^2)$, does not make sense as it implies that survival can be due to the effect of either the first or the second locus.

It follows easily by the technique of conditional expectation that the expected number of adults per infant is given by

$$\sum_{i^1j^1i^2j^2} \ell(i^1j^1i^2j^2)p(i^1j^1i^2j^2) \quad , \qquad (A.5)$$

where the $p(i^1j^1i^2j^2)$ will take on values as exemplified in the Equation IV.B.2. The variance of the number of adults per infant offspring is

$$\sum_{i^1_j{}^1 i^2_j{}^2} \ell(i^1_j{}^1 i^2_j{}^2)p(i^1_j{}^1 i^2_j{}^2) - \left(\sum_{i^1_j{}^1 i^2_j{}^2} \ell(i^1_j{}^1 i^2_j{}^2)p(i^1_j{}^1 i^2_j{}^2)\right)^2 \qquad \text{(A.6)}$$

The covariance between the number of adults per infant for two infants from the same inbred line is

$$\sum_{\substack{i^1_j{}^1 i^2_j{}^2 \\ r^1_s{}^1 r^2_s{}^2}} \ell(i^1_j{}^1 i^2_j{}^2)\ell(r^1_s{}^1 r^2_s{}^2)p(i^1_j{}^1 i^2_j{}^2 r^1_s{}^1 r^2_s{}^2) - \left(\sum_{i^1_j{}^1 i^2_j{}^2} \ell(i^1_j{}^1 i^2_j{}^2)p(i^1_j{}^1 i^2_j{}^2)\right)^2 , \qquad \text{(A.7)}$$

where the $p(i^1_j{}^1 i^2_j{}^2 r^1_s{}^1 r^2_s{}^2)$ take on values according to the two-loci analog of Equation IV.A.3. This matter was covered in Section IV.B.2 in the discussion following Equation IV.B.4. It follows from the Equation A.7 that the covariance between the number of adults per infant for two infants from independently derived inbred lines will be zero, since in this case it would be true that $p(i^1_j{}^1 i^2_j{}^2 r^1_s{}^1 r^2_s{}^2) = p(i^1_j{}^1 i^2_j{}^2)p(r^1_s{}^1 r^2_s{}^2)$.

In the case of two loci the relationship between loads and the mean viability follows exactly the pattern given in Section IV.B.3, and is exemplified by the Equations IV.B.5 and by the Equations ranging from IV.B.10 through IV.B.14. In the case of an arbitrary number of loci the relationship between mean viability and loads follows the pattern of Section IV.C.1.

In the context of this and the previous chapter the inclusion of an environmental effect in any sort of general way will cause considerable difficulty. The remedy to the situation would be to introduce a variable environment into the derivation of the equilibrium conditions which are

necessary for the load theory. This does not seem to be an easy undertaking and will not be attempted here. However, one can easily introduce a constant probability of death, or survival, affecting all individuals equally. Let, for instance, the probability of a random environmental death be (1-k). On the assumption that the environmental and genetic causes of death act independently, the probability of survival of an infant with genotype $(i^1j^1i^2j^2)$ is $k\,\ell(i^1j^1i^2j^2)$. It follows that in this case the expected number of adults per infant is equal to $k\,\Sigma\,\ell(i^1j^1i^2j^2)p(i^1j^1i^2j^2)$. It follows that in the case of n loci, the Equation IV.C.7 become

$$x_F^{12\ldots n} = kc[1-L_0 + (L_0 - L_1)F] \quad , \qquad (A.8)$$

where $x_F^{12\ldots n}$ in this case simply represents the mean viability of a population with degree of inbreeding F.

The introduction of a constant multiplicative environmental effect will not have any influence on the stability conditions derived in Chapter III, since in the formulas for the change of gamete frequencies the environmental effect will cancel. This remark is of importance in the cases where one is willing to assume that the genes which influence viability have no effects on other components of fitness, so that the load ratio theorems are applicable to them.

The estimation of genetic loads in the case of metric traits is a topic which can be best discussed in the next section on the application of load theory. The reason for this is that estimation procedures are dependent on the experimental situations in which they are employed.

3. The application of load theory in the case of metric traits

a. The estimation of loads in an experiment The application of load theory will be discussed with reference to the work of Malogolowkin-Cohen and associates (1964). This paper was chosen because of its clear display of the experimental set-up and the extensive reporting of its data on Drosophila willistoni.

The experimental flies were obtained from eight different localities and consisted of females enseminated in nature. The progeny of the different females were obtained in different cultures. To produce offspring with an inbreeding coefficient of $\frac{1}{4}$, groups of ten virgin females and ten males which were obtained from the same female progenitor were allowed to produce eggs, of which the viability to adulthood was studied. These matings were treated as follows: the ten pairs of flies were allowed to mature for 3 to 4 days at 25°C in the same creamer with culture medium. Then each group of flies was transferred without etherization to a clean spoon with cornmeal-molasses-agar medium, to which carbon black was added for easier visibility of the eggs. After approximately 12 hours, a portion of the culture medium with 50 eggs deposited on it was lifted from the spoon and transferred to a regular half-pint culture bottle with cream of wheat-molasses-agarless medium. Two such 50-egg samples (i.e., a total of 100 eggs) were taken from each group of parents. The cultures were allowed to develop at 25°C. Complete counts of the adult flies hatching in the cultures were made.

To obtain eggs with inbreeding coefficient zero, groups of ten virgin females from one progenitor were mated to males from the progeny of another wild collected mother. Yet another procedure was to mate a single male

with two females from different cultures, i.e., a male from progeny A was mated to a single female from progeny B, and four days later the same male was mated to a female from progeny C. The females were separated and allowed to produce offspring; ten females from one progenitor were mated to ten males from the other. A mating of ten males with ten females is called a cross, and in this case the resulting eggs have an inbreeding coefficient of .125.

For the purposes of analysis the data from the experiment was classified according to level of inbreeding, locality from which the wild collected progenitors was derived and according to the number of cultures with a certain percentage of survival from egg to adult stage. With a single exception all localities were represented in the class with survival of 97 to 100 percent. It is a fair assumption, therefore, that the value for the maximal genotype in this case is unity.

The authors then assumed that the data can be described by Equation A.4 with $x_{max} = 1$, and proceeded to take logs on both sides of the equation so as to do a regression analysis. The problem with this type of approach is that it ignores the complications that arise from the nature of an error term which should be incorporated on the right-hand side of Equation A.4 in an experimental situation.

Let us therefore proceed to consider the assumptions under which the models that we derived previously can be assumed to fit the given experimental circumstances. We first note that for the purpose of describing the relationship between the mean viability and the loads by the use of the inbreeding coefficient we had to assume Hardy-Weinberg frequencies at all loci. Under the assumptions of Section III.A we showed in the subsequent

sections of Chapter III that Hardy-Weinberg frequencies are possible with multiplicative gene action if the linkage between loci is not too tight. We will not assume additive gene action here, since we do not consider it applicable to viability as we pointed out previously in this chapter and in Section III.D. From the work reviewed by Li (1967) it is clear that tight linkage and deviations from multiplicativity may cause considerable deviations from Hardy-Weinberg equilibria. It is clear from Equation IV.B.2 and its derivation that deviations from Hardy-Weinberg equilibria might cause considerable difficulties. This matter will be referred to again in Section 4.

Epistasis and linkage are, however, not the only types of causes of deviation from Hardy-Weinberg equilibria that we have to keep in mind. It is clear, at least under the assumptions for the derivations in Section III.B, that product fecundity is necessary for the panmictic structure of gamete frequencies, and that the panmictic structure applies only to the moment of union between sperm and ova before any non-random mortality occur. In terms of the experimental situation under consideration it means that we have to assume that all unions between sperm and ova result in eggs with the same probability, regardless of the genotypes involved. The effect of differential mortality is relatively easy to keep track of, and will be examined in the next section. The effects of the other causes of deviations from Hardy-Weinberg equilibria are refractory.

Now, under these assumptions the expected number of adults per egg is

$$\sum_{i^1j^1i^2j^2} x(i^1j^1i^2j^2)p(i^1j^1i^2j^2) \quad , \qquad (A.9)$$

as was explained in the derivation of Equation A.5, and where we now use x's

instead of ℓ's in conformity to the notation of the previous chapter. Again we exemplify the case of n loci by the 2-loci case. Let us now denote by r_{ijk} the number of adults from the k-th egg derived from the j-th cross at the i-th level of inbreeding. Then we denote the proportion of eggs per cross surviving to adulthood by $z_{ij} = \frac{\Sigma r_{ijk}}{100}$, since the number of eggs used was always equal to 100.

It follows that

$$E(z_{ij}) = \sum_{i^1_j i^1_i i^2_j i^2_j} x(i^1_j i^1_i i^2_j i^2_j) p(i^1_j i^1_i i^2_j i^2_j) \quad ,$$

where the expectation is taken with reference to the third subscript in r_{ijk}, and where the right-hand side is a function of the i-th level of inbreeding through the genotypic frequencies. Hence,

$$z_{ij} = q \exp[-(A+Bw_i)] + e_{ij} \quad , \qquad \text{(A.10)}$$

if we assume the relationship between the mean viability and loads given by Equations A.4 to be justifiable under the circumstances of the experiment. In Equation A.10, w_i indicates the degree of inbreeding of the i-th level of inbreeding, and the term $e_{ij} = \Sigma e_{ijk}$ denotes a sampling error with $E(e_{ij}) = 0$. It should be noted that an environmental effect peculiar to the j-th cross at the i-th level of inbreeding is not included in Equation A.10. Under the present circumstances this omission is probably justifiable since the experiment was conducted under well-controlled circumstances. Also, we assume that the Equation A.10 will be applied on the basis of data within localities of origin of the progenitors, so that an extra subscript for localities on the z_{ij}'s would be unnecessary.

The standard procedure in the literature is to take logarithms for the estimation of the loads, which are functions of A and B. However, it can

easily be seen from the Equation A.10 that this is not a procedure for which adequate justification can be given. It is clear that in general

$$E(-\log z_{ij}) \neq \log q + A + Bw_i \quad ,$$

so that we are not in a Gauss-Markov set-up, even when $q = 1$ and under the assumption of homogeneous error variances. It is very difficult to say what sort of properties least squares estimates of A and B will have under the present circumstances.

The structure of the error variances in the model given as Equation A.10 also deserves some attention. From the experimental set-up and from the Equations A.3 and A.7 we see that $\text{cov}(e_{ijk}, e_{ijk'}) \neq 0$, where the magnitude of the covariance between e_{ijk} and $e_{ijk'}$ will depend on whether the eggs k and k' are full-sibs, half-sibs or from different pairs of parents altogether. It is clear that the structure of the relationships between the different eggs may vary from cross to cross, since we denoted by a cross a mass mating between 10 males and 10 females. For this reason, and since the variances in the Equations A.2 and A.6 depend on the degree of inbreeding of a population through the genotypic frequencies, we have to write $\text{var}(\sum_k e_{ijk}) = \sigma^2_{ij}$, where it is understood that $\sigma^2_{ij} \neq \sigma^2_{i'k'}$ if $i \neq i'$ or $j \neq j'$. If the parents from the different crosses are unrelated it will be true that $\text{cov}(e_{ijk}, e_{i'j'k'}) = 0$ if $i \neq i'$ or $j \neq j'$. It is not stated whether the parents in the different crosses were related or not, but for the purpose of this discussion we will assume that they are unrelated.

It is clear that we are in a situation with a very complicated error structure. For the purpose of examining the effect of the logarithmic

transformation on the error structure of the population, we will simplify matters a bit by assuming that a cross consists of a mating between a single male and a single female and that individuals from different mating pairs are unrelated. We also assume that only one egg per cross is taken at random, and that it is noted whether the egg gives rise to an adult fly. In this case $r_{ij} = 1$ if the j-th cross gives rise to an adult, and zero otherwise. Let $j = 1, \ldots, s$, so that in this case we write $z_i = \frac{\Sigma r_{ij}}{s}$. It follows for the two-loci case that $E(z_i)$ is equal to the expression given as Equation A.9, so that we can write $E(z_i) = Q_i$, where Q_i represents the Equation A.9, which is a function of the i-th level of inbreeding through the genotypic frequencies.

From our assumption of independent mating pairs it follows that $\text{cov}(r_{ij}, r_{ij'}) = 0$ if $j \neq j'$. Hence, from Equation A.6 it follows that

$$\text{Var}(z_i) = \frac{Q_i(1-Q_i)}{s} \quad . \tag{A.11}$$

The situation now corresponds to that of a binomial with a probability of success (or in this case, survival) Q_i. That we are indeed in a simple binomial situation follows easily when we consider that under our simplified assumption we have s independent events with probability of survival

$$\sum_{i^1 j^1 i^2 j^2} \ell(i^1 j^1 i^2 j^2) p(i^1 j^1 i^2 j^2) = Q_i \quad ,$$

in our present notation.

The point is that we are now in a situation for which it is well known that to homogenize the variances of the z_i the correct transformation is $\sin^{-1}\sqrt{z_i}$. In this case the homogenization of $\text{var}(z_i)$ will also

accomplish the homogenization of $\text{var}(e_i)$, since z_i and e_i differ only by an additive constant. We conclude then that the transformation $\log(z_i)$ will not be satisfactory with regard to the homogenization of variances. There is no reason why this conclusion whould not carry over to the more complex case of the experimental situation of Malogolowkin-Cohen and his associates (1964). This statement is in accord with approximate variances of the z_{ij} that can be derived by the use of Taylor's expansion of a differentiable function.

The foregoing argument is not meant to imply that a $\sin^{-1}\sqrt{z_{ij}}$ transformation is recommended. The emphasis is rather on the fact that the logarithmic transformation will not cause the error variances to be homogeneous. Thus, another objection against the use of the logarithmic transformation is added to the one already given in the discussion which followed on Equation A.10.

We are, therefore, in a position to state that the procedure of taking logarithms on both sides of Equation A.10 for the estimation of A and B does not have much, except computational convenience, to commend it. It will be a sounder procedure to derive a method of estimation directly from the Equation A.10. However, we will not do so here since we have shown that there is better justification for assuming Equation A.8 to be the correct relationship between the mean viability and the loads, than is the case with Equation A.4.

With the same justification from the experimental set-up as for Equation A.10 we write, by virtue of Equation A.8, our model as

$$z_{ij} = k[1-L_0 + (L_0 - L_1)w_i] + e_{ij}$$

or (A.12)

$$z_{ij} = a + bw_i + e_{ij} \quad ,$$

where, as we pointed out before, we have to write $\text{var}(e_{ij}) = \sigma^2_{ij}$, and where $\text{cov}(e_{ij}, e_{i'j'}) = 0$ if $i \neq i'$ or $j \neq j'$. This last statement follows from our assumption that the parents involved in the different crosses are unrelated.

Let us now assume $j = 1,2, \ldots, s_i$ and let us write $z_{i.} = \sum_j \frac{z_{ij}}{s_i}$ and $e_{i.} = \sum_j \frac{e_{ij}}{s_i}$. It follows that $\text{var}(e_{i.}) = \sum_j \frac{\sigma^2_{ij}}{s_i}$. Hence, we can estimate a and b from the model given as Equation A 12 by doing a weighted regression analysis of the $z_{i.}$ on the w_i , where the weights are $\frac{s_i}{\sum_j \sigma^2_{ij}}$.

Under the assumption of $k = 1$ the weighted least squares estimates of L_0 and L_1 are obtained from the equations $\hat{a} = 1 - \hat{L}_0$ and $\hat{b} = \hat{L}_0 - \hat{L}_1$, where the circumflex indicates that we are dealing with estimates of the parameters involved. The variances and covariances of $\hat{L}_0$ and $\hat{L}_1$ also follows easily from the variances and covariances of $\hat{a}$ and $\hat{b}$.

The reason why we have to assume that the probability of a random death, $1 - k$, is known (for covenience we took $k = 1$) is that we wish to estimate $\frac{L_1}{L_0}$. The only way in which this can be done from estimates of a and b is to obtain $\frac{\hat{L}_1}{\hat{L}_0}$ under the assumption of known k . With unknown k we are again in the position which we discussed at the end of the last chapter of trying to estimate three unknowns from two equations.

$\frac{\hat{L}_1}{\hat{L}_0}$ is, of course, going to be a biased estimate of $\frac{L_1}{L_0}$. To reduce

the bias a method first proposed by Quenouille (1956) can be used. Miller (1964) can be consulted for a discussion of this method which was dubbed the "jackknife" method by Tukey. Miller also discusses an approximate distribution for a jackknife estimator. The jackknife method was employed by Levene and associates (1965). Although it was not clearly stated that such is the case, it appears that Levene and his associates employed the jackknife to reduce the bias in their estimates of A and B which result from the taking of logarithms on both sides of Equation A.10.

An approximate formula for the variance of $\frac{\hat{L}_1}{\hat{L}_0}$ is

$$\mathrm{Var}\left(\frac{\hat{L}_1}{\hat{L}_0}\right) = \frac{\mathrm{Var}(\hat{L}_1)}{\hat{L}_0^2} - 2\,\frac{\hat{L}_1 \mathrm{Cov}(\hat{L}_1, \hat{L}_0)}{\hat{L}_0^3} + \frac{\hat{L}_1^2}{\hat{L}_0^4}\,\mathrm{Var}(\hat{L}_0) \qquad \text{(A.13)}$$

The problem with doing a weighted regression analysis on the model of Equation A.12 is of course that the weights will be unknown. The best that can be done in this situation is to estimate $\sum_j \frac{\sigma^2_{ij}}{s_i}$ from the mean squares between the z_{ij}'s for a given i. This estimator will at least be unbiased. This was also the method of dealing with the problem of heterogeneous variances recommended by Malogolowkin-Cohen and his associates (1964) for their analysis based on taking logarithms on both sides of Equation A.10.

For the simplified case in which only one individual from each inbred line is measured for estimation purposes, after the fashion which we explained for the derivation of Equation A.11, maximum likelihood estimators for L_0 and L_1 or $\frac{L_1}{L_0}$ can easily be constructed by following the same pattern as that of Section C.2 of Chapter IV. The reason for this being the case is that we are here in a simple binomial situation where it is easy

to derive the likelihood of a sample. In the case where related individuals are measured the likelihood of the sample will become unmanageably complex, especially if more than one locus is involved.

It is clear from the foregoing that the correct analysis of an experiment depends on the error structure induced by the experimenter. This aspect received rather scant attention in the load literature, and was therefore covered in considerable detail. Since we assumed thus far the error structure to have been completely of a sampling nature, we will investigate the possible nature of an environmental error very briefly.

In the same way as we introduced an environmental probability of survival in the derivation of Equation A.8, we can introduce an environmental probability of survival with reference to the j-th cross from the i-th level of inbreeding. This would make sense in a case where all the offspring from a specific cross shares a common environment.

Let us assume that the probability of survival of the k-th egg of the j-th cross, given that the k-th egg has genotype $(i^1j^1i^2j^2)$, is $g_{ij}\ell(i^1j^1i^2j^2)$, where g_{ij} is the environmental probability of survival in the j-th cross. It follows that the expected number of adults per egg of the j-th cross in the i-th level of inbreeding is

$$E(r_{ijk}) = g_{ij} \sum_{i^1j^1i^2j^2} \ell(i^1j^1i^2j^2)p(i^1j^1i^2j^2) . \qquad (A.14)$$

Hence, our model in this case can be written as

$$r_{ijk} = g_{ij}\{\sum_{i^1j^1i^2j^2} \ell(i^1j^1i^2j^2)p(i^1j^1i^2j^2)$$

$$+ [(\frac{r_{ijk}}{g_{ij}} - \sum_{i^1j^1i^2j^2} \ell(i^1j^1i^2j^2)p(i^1j^1i^2j^2)]\} ,$$

or, by Equation A.8, assuming $k = 1$, $c = 1$, as

$$r_{ijk} = g_{ij}[1 - L_0 + (L_0 - L_1)w_i + e_{ijk}] \quad , \tag{A.15}$$

by virtue of Equation A.14. By comparing the terms in Equation A.15 with those of the preceding equation we see that $E(e_{ijk}) = 0$ by reason of Equation A.14.

It is evident that the model represented by Equation A.15 would require a far different type of approach in estimation procedure than the one represented by the Equation A.12 or the one represented by Equation A.10. This again highlights the need for a careful consideration of the nature of the error structure in the analysis of load experiments.

b. A comparison between the alternative models Since all the estimates of loads in the literature have been obtained on the basis of the model represented by Equation A.10, it is of some interest to derive some sort of relationship between these load estimates and the ones that one would obtain by the use of the model represented by the Equation A.12.

From the derivation of these equations we have the correspondences A and L_0, B and $L_1 - L_0$, a and $1 - L_0$, and b and $L_0 - L_1$, all on the assumption of $k = 1$, $q = 1$ and $c = 1$. In order to avoid confusion, we will denote the load estimates on the basis of the model represented by Equation A.10 by $\hat{A}$ and $\hat{B}$ and the corresponding load estimates obtained from Equation A.12 by $\hat{L}_0$ and $(\hat{L}_1 - \hat{L}_0)$.

We will now compare the two models, represented by the Equations A.10 and A.12 respectively, by comparing the functional relationships from which they originated. We therefore recall the functional relationships given by the Equations IV.A.30 and IV.C.7 to write

$$x_F = \exp[-(A+BF)] \qquad (A.16)$$

and

$$x_F = 1 - L_0 - (L_1 - L_0)F \ , \qquad (A.17)$$

where x_F denotes the mean viability of a population with degree of inbreeding F.

It follows from the Equation A.16 that $A = -\log(x_0)$ and from the Equation A.17 that $L_0 = 1 - x_0$. An expansion in series form of $-\log(x_0)$ is

$$-\log(x_0) = 2[(\frac{1-x_0}{1+x_0}) - \frac{1}{3}(\frac{x_0-1}{1+x_0})^3 - \frac{1}{5}(\frac{x_0-1}{1+x_0})^5 \ . \ . \ . \ ] \ , \qquad (A.18)$$

which will always be convergent, since $0 < x_0 \leq 1$. Also all members of the right-hand side of Equation A.18 will be positive since $x_0 - 1 < 0$. We now substitute $A = -\log(x_0)$ in the Equation A.16 to get

$$BF = \log(x_0) - \log(x_F)$$

or

$$BF = \log(\frac{x_0}{x_F}) \ .$$

On expansion $\log(\frac{x_0}{x_F})$ yields

$$\log(\frac{x_0}{x_F}) = 2[\frac{x_0-x_F}{x_0+x_F} + \frac{1}{3}(\frac{x_0-x_F}{x_0+x_F})^3 + \frac{1}{5}(\frac{x_0-x_F}{x_0+x_F})^5 \ . \ . \ . \ . \ ] \ . \qquad (A.19)$$

Usually the terms on the right-hand side of Equation A.19 will be positive, since the mean fitness in general declines with inbreeding. Also $\frac{x_0-x_F}{x_0+x_F}$ will in general be small, so that the series will converge rapidly.

From Equation A.17 we now obtain $L_0 = 1 - x_0$ and $(L_1 - L_0)F = x_0 - x_F$. It follows from Equation A.18 that $A \geq L_0$, since $\frac{2}{1+x_0} \geq 1$, with equality in the case of $x_0 = 1$. Since $\frac{2}{x_0+x_F} \geq 1$ it follows from Equation A.19 that $B \geq (L_1 - L_0)$, with the equality holding in the case of $x_0 = x_F$.

In the load work it is usual to estimate the ratio $\frac{B}{A}$. Now if $\frac{1-x_0}{1+x_0}$ and $\frac{x_0-x_F}{x_0+x_F}$ are reasonably small, we use the first terms of the Equations A.18 and A.19 to give the relationships

$$L_1 - L_0 = \frac{x_0+x_F}{2} B$$

and

$$L_0 = \frac{1 + x_0}{2} A$$

and, hence

$$\frac{L_1 - L_0}{L_0} = \frac{(x_0+x_F)B}{(1 + x_0)A} , \qquad (A.20)$$

will hold to a fair degree of approximation.

We write, as before, z_F for the observed viability and indicate estimates with a circumflex. We then get a rough idea how the estimates of the various quantities will differ by considering the equation

$$\frac{\hat{L}_1 - \hat{L}_0}{\hat{L}_0} = \frac{(z_0+z_F)}{(1 + z_0)} \frac{\hat{B}}{\hat{A}} , \qquad (A.21)$$

The Equation A.21 is of more value than it will seem at first sight, because the inbred data are often concentrated at one or two points. In human data, for instance, the number of first-cousin marriages exceed the

other consanguineous marriages by far. In the data of Malogolowkin-Cohen and associates (1964) most of the data in the inbred class have a coefficient of inbreeding of one fourth. In other cases there may be other criteria for the choice of the z_F for use in the Equation A.21, for one might want to know what the largest or smallest values of $\frac{\hat{L}_1 - \hat{L}_0}{\hat{L}_0}$ are likely to be.

It is also clear that when the mean viabilities are high, there will not be much of a difference between the estimates from the two models. An instance of this being the case can be found in the human data used by Morton, Crow and Muller (1956), where the lowest mean viability from offspring of first cousin-marriages is near to 80 percent. The other use of Equation A.21 is simply to show that in general $\frac{\hat{L}_1 - \hat{L}_0}{\hat{L}_0} < \frac{\hat{B}}{\hat{A}}$.

For the sake of argument we will now assume that the genes which influence viability and the genes which influence the other components of fitness are independent from each other. The theorems on the load ratios given as Equations IV.A.19 and IV.A.23 are then applicable and we see that $\frac{\hat{B}}{\hat{A}}$ will give in general a higher estimate of $\frac{L_1 - L_0}{L_0}$ than $\frac{\hat{L}_1 - \hat{L}_0}{\hat{L}_0}$ and hence would tend to favor the mutation load hypothesis. However, an examination of the data from some of the most reliable load experiments with the aid of Equation A.21 would seem to indicate that the differences between the results that would have been obtained from an analysis based on the model of Equation A.12 and that which follows from the actually used model of Equation A.10, are not great enough to modify conclusions as to the possible nature of the origin of the loads.

There is one aspect related to the foregoing discussion that is still of some importance to consider, and this has to do with Morton, Crow and

Muller's (1956) concept of lethal equivalents. From Equation A.10 we have that for $F = 1$

$$E(\text{number of adults per egg}) = \exp[-(A+B)] \quad , \tag{A.22}$$

and from Equation A.12 we have that for $F = 1$

$$E(\text{number of adults per egg}) = 1 - L_1 \quad , \tag{A.23}$$

where in both of the above equations we assumed no random environmental cause of death. Under this assumption $A + B$ was designated by Morton and his associates (1956) as the number of lethal equivalents per gamete. The analog of $A + B$ in the alternative model given by the Equation A.12 is L_1. Since the left-hand sides of Equations A.22 and A.23 must always be positive we see that on the basis of the model represented by Equation A.10 there can be any number of lethal equivalents, but that on the basis of the model represented by Equation A.12, the number of lethal equivalents must always be less than unity.

It is now well to remember that the model of Equation A.10 was derived on the assumption of no linkage between loci or alternatively on the assumption that the loads are so small that $1 - L_0 - F(L_1 - L_0)$ can be represented by $\exp[-(L_0 + (L_1 - L_0)F)]$, as was pointed out in the discussion on Equation IV.C.6. In the case of $(L_1 - L_0)$ being relatively large, the foregoing approximation may still hold good for small values of F, but in such a situation the Equation A.22 represents an invalid extrapolation to the case of $F = 1$. We conclude, therefore, that estimates of the number of lethal equivalents which are greater than unity, fall outside the parameter space.

The data of Morton, Crow and Muller (1956) give too large a number of lethal equivalents in about all cases, on calculations based on the models of Equations A.10 and A.12. This is also the case with many of the estimates of Malogolowkin-Cohen and his associates (1964), and with quite a few estimates of Levene and associates (1965). A reason for this may be that the approximation represented by the Equation IV.C.4 does not hold good. The failure of Equation IV.C.4 to hold may be due to there being too many loci which have a relatively large influence on viability in such a way that the loads at these loci are not small. On the other hand, it is conceivable that other deviations from the basic assumptions of load theory may have the same effect.

Opposed to this we find that the human data of Neel and Schull (1962), the Drosophila data of Mettler and his associates (1966), the Drosophila pseudoobscura data of Stone and his associates (1963), of Dobzhansky and his associates (1963) and of Torroja (1964), all give estimates falling in the parameter space.

Obviously all sorts of considerations are important in an inquiry into the reasons for estimates not falling in the parameter space. Matters like the nature and origin of the populations are important, for they determine whether one can assume the population to be in Hardy-Weinberg equilibrium. A consideration of the possible nature of the error term deserves also some attention, as we pointed out in the discussion of Equation A.15. The list of possible considerations bearing on this topic is virtually endless, and for this reason matters will be left as they stand.

4. Some effects of departures from the basic assumptions

In this section we examine the possible effects of two departures from the basic assumptions in load theory. One is the effect of differential mortality during a previous life stage on the study of loads pertaining to a second life stage and the other is the possible effect of selection on the description of the inbred genotypic array by the coefficient of inbreeding.

The trouble with other possible sources of deviations from the basic assumptions is that they cannot be dealt with in the conceptual framework of load theory. In these cases the best that can be done at present is to point out why they undermine the structure of the theory. This will be done in Section b.

The papers in the literature which contain interesting information about matters which cannot be exhaustively discussed here, are those of Crow (1963), Levene (1963), Haldane and Jayakar (1965) and Schull and Neel (1965). However, it is felt that the material contained in this chapter will provide an adequate background for the evaluation of the force of the arguments presented by these authors.

a. The effect of differential mortality on loads For our study in this section is desirable to derive a few results on loads with reference to two life phases which may occur in some organism. Again we will exemplify the general case with a consideration of the case of two loci. Let us consider the survival from egg to adult in two stages, e.g. the survival from egg to larva and from larva to adulthood. Let the probability for the genotype $(i^1j^1i^2j^2)$ to survive from egg to adulthood be $\ell(i^1j^1i^2j^2) = \ell_1(i^1j^1i^2j^2)\ell_2(i^1j^1i^2j^2)$, where $\ell_1(i^1j^1i^2j^2)$ is the probability of survival from egg to larva and $\ell_2(i^1j^1i^2j^2)$ is the

probability of survival from larva to adulthood. Then further let $\ell_k(i^1j^1i^2j^2) = \ell_k(i^1j^1)\ell_k(i^2j^2)$ for $k = 1,2$. As a consequence of this it follows that $\ell(i^1j^1i^2j^2) = \ell(i^1j^1)\ell(i^2j^2)$. We also assume that $\max(\ell(i^1j^1i^2j^2)) = \max(\ell_k(i^1j^1)) = \max(\ell_k(i^2j^2)) = 1$, for $k = 1,2$.

In viability studies it makes sense to write $\ell_k(i^1j^1) = 1 - s_k(i^1j^1)$, where $s_k(i^1j^1)$ is the probability for a death during the k-th life phase. It follows that we can write

$$\begin{aligned}\ell(i^1j^1i^2j^2) &= 1 - s(i^1j^1i^2j^2) \\ &= [1 - s_1(i^1j^1i^2j^2)][1 - s_2(i^1j^1i^2j^2)] \\ &= 1 - s_1(i^1j^1i^2j^2) - s_2(i^1j^1i^2j^2) , \end{aligned} \tag{A.24}$$

if $s_1(i^1j^1i^2j^2)$ and $s_2(i^1j^1i^2j^2)$ are small enough so that their products can be neglected. In the same way we can write

$$\ell(i^1j^1i^2j^2) = 1 - s_1(i^1j^1) - s_1(i^2j^2) - s_2(i^1j^1) - s_2(i^2j^2) , \tag{A.25}$$

if again the different products are small enough to be negligible.

The expected number of adults per egg is

$$\begin{aligned}\sum_{i^1j^1i^2j^2} \ell(i^1j^1i^2j^2)p(i^1j^1i^2j^2) = 1 &- \sum_{i^1j^1} s_1(i^1j^1)p(i^1j^1) \\ &- \sum_{i^2j^2} s_1(i^2j^2)p(i^2j^2) \\ &- \sum_{i^1j^1} s_2(i^1j^1)p(i^1j^1) \\ &- \sum_{i^2j^2} s_2(i^2j^2)p(i^2j^2) , \end{aligned} \tag{A.26}$$

where the $p(i^1j^1i^2j^2)$ follow the pattern set forth in the Equation IV.B.2,

and where the $p(i^1j^1)$ and $p(i^2j^2)$ are assumed to follow the usual pattern of neutral genes on inbreeding, when the inbreeding process started off from Hardy-Weinberg frequencies.

By our usual definition of a load (Equation 1, Chapter I) we have

$$L_F^1(k) = 1 - \sum_{i^1j^1} p(i^1j^1)[1-s_k(i^1j^1)] = \sum_{i^1j^1} p(i^1j^1)s_k(i^1j^1) \quad (A.27)$$

and

$$L_F^2(k) = 1 - \sum_{i^2j^2} p(i^2j^2)[1 - s_k(i^2j^2)] = \sum_{i^2j^2} p(i^2j^2)s_k(i^2j^2) ,$$

for $k = 1,2$, and where F can take all values from 0 to 1 , endpoints included. We note again that in the present set-up we assume the maximal genotype to have a value of unity at all loci. It also follows from our definitions of the various quantities involved that we can write

$$\begin{aligned} L_F^a = L_F^a(12) &= 1 - \sum_{i^aj^a} p(i^aj^a)[1-s(i^aj^a)] \\ &= 1 - \sum_{i^aj^a} p(i^aj^a)[1-s_1(i^aj^a) - s_2(i^aj^a)] \\ &= L_F^a(1) + L_F^a(2) \quad , \end{aligned} \quad (A.28)$$

for $a = 1,2$. The notation which we introduced in the Equations A.27 and A.28 is self-explanatory, but none the less deserve some comments. First, it is important to note that the result of Equation A.28 holds only in the case where the maximal genotypes have a value of unity and where the selection coefficients are so small that we can assume their products of negligible magnitude in comparison with the other quantities involved. Second, we see that the load for the second life phase is computed on the assumption of Hardy-Weinberg gene frequencies. It is, of course, not true that

the population will be in Hardy-Weinberg equilibrium at the end of the first life phase. This matter will be further discussed later on.

In the conceptual framework given at the end of Section IV.A, the expected number of adults per egg in a population characterized by the inbreeding coefficient F, is by Equations A.26 and A.27

$$\sum_{i^1_j{}^1 i^2_j{}^2} \ell(i^1_j{}^1 i^2_j{}^2) p(i^1_j{}^1 i^2_j{}^2) = 1 - L^1_F(1) - L^2_F(1) - L^1_F(2) - L^2_F(2) , \quad (A.29)$$

or by Equation A.28

$$\sum_{i^1_j{}^1 i^2_j{}^2} \ell(i^1_j{}^1 i^2_j{}^2) p(i^1_j{}^1 i^2_j{}^2) = 1 - L_F(1) - L_F(2) . \quad (A.30)$$

The Equations A.29 and A.30 generalize immediately in the case of n loci to

$$E(\text{adults per egg}) = 1 - \sum_a L^a_F(1) - \sum_a L^a_F(2) = 1 - L_F(1) - L_F(2) \quad (A.31)$$

and

$$E(\text{adults per egg}) = 1 - L_F(12) = 1 - L_F . \quad (A.32)$$

We are now in a position to consider the effect of differential mortality in one life stage on the load estimates in a succeeding life stage. As before we will exemplify matters by the use of the case of two loci.

The expected number of adults per larva is easily seen to be of the form

$$E(\text{adults per larva}) = \frac{\sum \ell_2(i^1_j{}^1 i^2_j{}^2)\ell_1(i^1_j{}^1 i^2_j{}^2)p(i^1_j{}^1 i^2_j{}^2)}{\sum \ell_i(i^1_j{}^1 i^2_j{}^2)p(i^1_j{}^1 i^2_j{}^2)} , \quad (A.33)$$

where the $p(i^1_j{}^1 i^2_j{}^2)$ are in Hardy-Weinberg form at equilibrium with no

inbreeding, and where the $p(i^1j^1i^2j^2)$ follow the pattern of Equation IV. B.2 on inbreeding.

It should be noted that the Equation A.33 holds for any metric trait which should be considered to be the property of the larva rather than the eggs. If we let $x(i^1j^1i^2j^2)$ be the measurement of the trait on the larval genotype $(i^1j^1i^2j^2)$, we will get the equivalent of the Equation A.33 to be

$$E[x(i^1j^1i^2j^2)] = \frac{\Sigma x(i^1j^1i^2j^2)\ell_1(i^1j^1i^2j^2)p(i^1j^1i^2j^2)}{\Sigma \ell_1(i^1j^1i^2j^2)p(i^1j^1i^2j^2)} \quad . \tag{A.34}$$

In the steps that follow we will employ the type of notation which we used in Equation A.28. For instance, loads pertaining to measurements on the composite trait denoted by $\ell_1(i^1j^1i^2j^2)\ell_2(i^1j^1i^2j^2) = \ell(i^1j^1i^2j^2)$, will be denoted by $L(12)$. It then follows by the Equation IV.C.7, for the case of n loci and on the deletion of the random environmental effect from Equation A.8, that Equation A.33 or A.34 becomes

$$E(\text{adults per larva}) = \frac{c(12)\{1-L_0(12) + [L_0(12)-L_1(12)]w_i\}}{c(1)\{1-L_0(1) + [L_0(1)-L_1(1)]w_i\}} \tag{A.35}$$

where w_i is the degree of inbreeding of the subpopulation at the i-th level of inbreeding, and where $c(12)$ and $c(1)$ are the maximal genotypes on the composite and first life phase traits respectively.

There does not seem to be any general way in which the Equation A.35 can be simplified. Some simplification is, however, possible if the assumptions hold which led to Equation A.28. These assumptions were that both traits have a maximal genotype with value unity and that the selection

coefficients pertaining to genes at each locus are small. This last assumption is more restrictive than the one that the load at each locus must be small, which we recall, was used in the derivation of Equation IV.C.7.

It follows then by the Equation A.28 that the Equation A.35 becomes

$$E(\text{adults per larva}) = \frac{1 - L_0(1) + [L_0(1)-L_1(1)]w_i - L_0(2) + [L_0(2)-L_1(2)]w_i}{1 - L_0(1) + [L_0(1) - L_1(1)]w_i},$$

which, if the different quantities involved are small, becomes

$$E(\text{adults per larva}) = \exp - \{L_0(2) + [L_1(2) - L_0(2)]w_i\} \quad . \qquad (A.36)$$

In the type of experimental set-up that we discussed previously the Equation A.36 will lead to the type of model exemplified by the Equation A.10. The results of Levene and his associates (1965) on data on Tribolium survival rates from eggs to larvae to adults, can be interpreted with the help of the results embodied in Equation A.36. These results show good agreement between the estimates of $L_0(12)$ and $L_1(12) - L_0(12)$ and the sums of the estimates of $L_0(1)$ and $L_0(2)$ and of $L_1(1) - L_0(1)$ and $L_1(2) - L_0(2)$. Exactly how much support this result gives to the theory encompassed in Equation A.36 is difficult to evaluate precisely. The reason for this statement will become evident after some more comments on Equation A.35.

The interpretation that can be given to the lethal equivalents of Morton, Crow and Muller (1956) in a study with two life phases ought now to be evident from a comparison of Equation A.31 with Equation A.23. The lethal equivalents of the composite trait would simply be equal to the sum

of the lethal equivalents of its components. This statement will, of course, hold only under the conditions which led to the derivation of Equation A.31.

Care should, however, be taken in the interpretation of the load ratios. Let us, for instance, assume that fertility differences do not complicate the picture. Then the equilibrium gene frequencies will depend on the total viability from egg to adult and the load ratio theorems will apply to the loads computed on survival data from egg to adult.

It is of some interest to inquire what will happen in an analysis of data in which the simplification of Equation A.35 to Equation A.36 does not hold. It is clear that the denominator of Equation A.35 can be written as $(1-y)^{-1}$, with $0 < y < 1$, so that it can be expanded in a convergent series. Hence, it follows that the Equation A.35 can be regarded as a polynomial in the w_i of degree $n \geq 2$. If a polynomial is fitted to data on this basis and it fits significantly a degree of more than two, one can probably conclude that the representation of the data by a model based on the Equation A.36 is invalid. From an experimental point of view, however, the problem would be that it could be difficult to obtain enough different F values in a population to determine what sort of polynomial fits the data best.

It is noteworthy that nonlinearity in the regression on the inbreeding coefficient can also be induced by linkage even if the mortality from egg to larva depends on a different set of genes than the mortality from larva to adulthood. Let us consider a 4-loci case, where the first two loci determine survival from larva to adult, and the second two loci determine survival from egg to larva. As before, we assume multiplicative gene

action between loci, so that we can write

$$E(\text{adults per larva}) = \frac{\Sigma \ell_2(i^1j^1)\ell_2(i^2j^2)\ell_1(i^3j^3)\ell_1(i^4j^4)p(i^1j^1i^2j^2i^3j^3i^4j^4)}{\Sigma \ell_1(i^3j^3)\ell_1(i^4j^4)p(i^3j^3i^4j^4)}, \tag{A.37}$$

where ℓ_2 refers to the survival from larva to adult and ℓ_1 to the survival from egg to larva. The property that gives rise to the trouble resulting from the Equation A.37 is that on inbreeding it would not be true that

$$p(i^1j^1i^2j^2i^3j^3i^4j^4) = p(i^1j^1i^2j^2)p(i^3j^3i^4j^4) ,$$

as follows easily if we consider the 4-loci analog of the Equation IV.B.2 in the case of linkage between loci (1,2) and loci (3,4). Hence, on applying the result of Equation IV.C.7 to the Equation A.37 we obtain

$$E(\text{adults per larva}) = \frac{c(12\bar{3}\bar{4})[1 - \sum_{a=1}^{4} L_0^a + (\sum_{a=1}^{4} L_0^a - \sum_{a=1}^{4} L_1^a)w_i]}{c(\bar{3}\bar{4})[1 - \sum_{a=3}^{4} L_0^a + (\sum_{a=3}^{4} L_0^a - \sum_{a=3}^{4} L_1^a)w_i]}, \tag{A.38}$$

where the maximal genotypes are denoted in the same way as in Equation A.35. The same type of argument then applies to A.38 as was applied to Equation A.35 when it was argued that its denominator can be expanded in a convergent series.

These results on differential mortality causing curvilinear regression equations to obtain, are important, since in the literature curvilinearity is usually ascribed to epistasis.

b. The consequences of departures from Hardy-Weinberg frequencies on load theory. The problem caused by epistasis in load theory is of a twofold nature. First, epistasis other than the kind generated by

multiplicative gene action between loci in fitness, or components of fitness, will cause the assumption of Hardy-Weinberg equilibria no longer to hold. Second, epistasis of the type mentioned causes the single locus loads not to be meaningful any longer, and the load ratio theorems not to hold.

The assumption of Hardy-Weinberg frequencies is necessary in order to describe the genotypic array with the coefficient of inbreeding F and its multivariate analogs, as it was outlined in Section IV.B. In the single locus case Jacquard (1968) derived a formula describing inbreeding in the absence of Hardy-Weinberg structure. It may be possible to extend this work to two loci, although it will not be easy.

The next step would then be to try to calculate multi-loci loads of the type defined in Equation IV.C.3. The problem here would be that a multivariate regression analysis will be involved, and that it will have to be a system of great complexity, depending at the very least on 2^n-1 multi-loci inbreeding coefficients, where n is the number of loci. The problem is furthermore compounded by the fact that we will in general not know the number of loci in a genetic system, or their linkage relationships. The knowledge of the linkage relationships is necessary for calculation of the multi-loci inbreeding coefficients. Also these inbreeding coefficients are difficult to calculate, and recurrence relations have been derived only for two special two-loci cases.

If all these problems do not seem to be prohibitive enough one might reflect on the fact that there are no theorems available in the case of non-multiplicative epistasis by which one can try to make deductions as to the nature of the origin of different types of loads. There does not seem

to be a general principle by which such theorems can be derived.

If such is the case, the only reason for the calculation of loads would be as a tool for the study of the fitness of populations. Here one should remember, as was pointed out in Chapter II, that loads from different genetic systems give information about the mean fitness of populations only if the maximal genotypes of the different genetic systems are of the same magnitude. There is no reason why such should be the case.

We conclude then that the only reason for the study of non-multiplicative epistasis in a load context would be to try to develop tests by which one can determine whether the load assumptions hold. From the foregoing considerations it follows that this will be no easy task which can be disposed of by an application of techniques which are already in existence. It will, therefore, not be attempted here.

The problems due to deviation from Hardy-Weinberg frequencies because of too tight linkage between loci will be pretty much the same as those due to epistasis. Here also the load theorems will not hold, since we had to assume Hardy-Weinberg frequencies in their derivation. The description of the genotypic array by the various inbreeding coefficients will be invalid, for the same reasons as in the case of epistasis.

That the occurrence of this problem is quite possible follows from the fact that studies on the fine structure of the gene showed that the unit of recombination is in all probability a nucleotide pair. Benzer (1967) compares the total linkage length and total DNA content of a T_4 virus particle and concludes that the ratio between the linkage distance and the molecular distance is of the order of 10^{-3} percent recombination per nucleotide pair. These are rather small values, so that one would judge

that the Equations III.D.17 or III.D.20 may often force a conclusion that the existence of Hardy-Weinberg frequencies is impossible. On the other hand, one might argue that with these small recombination values the last term of the right-hand side of the Equations III.B.8 will for adjacent nucleotide pairs be small compared to the other quantities involved, so that with the aid of the Equations III.F.24 one will be able to show that the alleles at a composite locus may have Hardy-Weinberg frequencies if the matrix $(a(i^1j^1i^2j^2))$ has the requisite properties. Stability at the composite locus implies, of course, that the matrix $[a(i^1j^1i^2j^2)]$ can not be the Kronecker product of the matrices $[a(i^1j^1)]$ and $[a(i^2j^2)]$. We thus have a pattern that requires, for Hardy-Weinberg frequencies to be possible, comparatively large recombination values between units of function that interact with each other in a multiplicative fashion, and small recombination values within such units of function (or cistrons).

It does not seem to be easy to judge whether the pattern of gene action and recombination values that would result in Hardy-Weinberg frequencies at all cistrons is probable or possible. This matter certainly deserves further inquiry, but since the requisite information does not appear to be readily available, we will leave matters here. In the case of additive gene action between cistrons linkage will, of course, not be a problem at all if the recombination within cistrons is small.

c. <u>The effect of selection on the description of the genotypic array by means of the coefficient of inbreeding</u> The effect of selection on inbreeding is difficult to handle mathematically. We will therefore consider the simple case of a single locus population. We assume that we start off with a population of infants with Hardy-Weinberg frequencies,

and subject it to one generation of full-sib mating.

Let us consider the mating between the offspring of two individuals with genotypes (ij) and (rs) respectively. We start off with the two infants (ij) and (rs), which survive to adulthood with probabilities $\ell(ij)$ and $\ell(rs)$ respectively. The mean number of offspring of the mating (ij) x (rs) we denote by $b(ij)b(rs)$, as we did before. The array of full-sib infants in the population is therefore

$$\frac{1}{D_1}\sum_{ij}\sum_{rs}p(i)p(j)p(r)p(s)a(ij)a(rs) \times [\tfrac{1}{4}(ir) + \tfrac{1}{4}(jr) + \tfrac{1}{4}(is) + \tfrac{1}{4}(js)] \quad , \tag{A.39}$$

where D_1 is a normalizing factor to make the probabilities or relative frequencies add up to unity. The array of full-sib adults in the population is given by

$$\frac{1}{D_2}\sum_{ij}\sum_{rs}p(i)p(j)p(r)p(s)a(ij)a(rs) \times [\tfrac{1}{4}\ell(ir)(ir) + \tfrac{1}{4}\ell(jr)(jr) + \tfrac{1}{4}\ell(is)(is) + \tfrac{1}{4}\ell(js)(js)] \tag{A.40}$$

where D_2 is again a normalizing factor, but different from the one in the previous case. Full-sib mating now takes place, and the array of mating pairs is given by

$$\frac{1}{D_3}\sum_{ij}\sum_{rs}p(i)p(j)p(r)p(s)a(ij)a(rs) \times [\tfrac{1}{4}\ell(ir)(ir) + \tfrac{1}{4}\ell(jr)(jr) + \tfrac{1}{4}\ell(is)(is) + \tfrac{1}{4}\ell(js)(js)]^2 \, . \tag{A.41}$$

Due to the fact that the mean number of offspring of a pair of adult genotypes, say, (ij) and (rs), is given by b(ij) x b(rs) in which b(ij) is ascribed to (ij) and b(rs) to (rs), the array of relative frequencies of infant offspring that can be ascribed to the pairs of infants in the parental generation is

$$\frac{1}{D}\sum_{ij}\sum_{rs}p(i)p(j)p(r)p(s)a(ij)a(rs)$$

$$\times\left[\tfrac{1}{4}a(ir)(ir)+\tfrac{1}{4}a(jr)(jr)+\tfrac{1}{4}a(is)(is)+\tfrac{1}{4}a(js)(js)\right]^2 . \qquad (A.42)$$

We now find it useful to classify the mating pairs according to whether they have two variables in common, like (ir) and (ir), or one variable in common, like (ir) and (jr), or no variables in common, like (ir) and (js). We find that of the 16 possible mating pairs there are four with two variables in common, eight with only one in common and four with no variables in common. Because of the symmetric way in which the variables enter in Equation A.42, we can write the array of relative frequencies of offspring from the different mating types as

$$\sum_{ij}\sum_{rs}\frac{Q(ij)Q(rs)}{D}\{\tfrac{1}{4}a^2(ir)[(ir)\times(ir)]+\tfrac{1}{2}a(ir)a(jr)[(ir)\times(jr)]$$

$$\times\tfrac{1}{4}a(ir)a(js)[(ir)\times(js)]\}\ , \qquad (A.43)$$

where $Q(ij)=a(ij)p(i)p(j)$, $a^2(ir)=a(ir)a(ir)$ and where D is a normalizing factor.

It follows that we can write the offspring array as

$$\sum_i \frac{1}{8D}\{\sum_{jrs} Q(jr)Q(is)[a^2(ij) + a(ir)a(ij)]\}\ (ii)$$

$$+ \sum_{ij} \frac{1}{8D}\{\sum_{rs} Q(ir)Q(js)[a^2(ij) + 2a(ij)a(ir) + a(is)a(rj)]$$

$$+ \sum_{rs} Q(ij)Q(rs)[a(ir)a(jr) + a(ir)a(js)]\}\ (ij) \qquad \text{(A.44)}$$

We see that if we assume $a(ij) =$ constant, for all i, j, we get in Equation A.44 the usual array of

$$\tfrac{1}{4}\Sigma p(i)(ii) + \tfrac{3}{4}\sum_{ij} p(i)p(j)(ij) \quad .$$

In the same way the Equation A.41 yields

$$\tfrac{1}{4}\sum_{ij} p(i)p(j)[(ij) \times (ij)] + \tfrac{1}{2}\sum_{ijr} p(i)p(j)p(r)\ [(ij) \times (ir)]$$

$$+ \tfrac{1}{4}\sum_{ijrs} p(i)p(j)p(r)p(s)[(ij) \times (rs)] \quad ,$$

which corresponds to the values given for full-sibbing in the discussion on Equation IV.A.3.

We see from the Equations A.41 and A.44 that there is no easy pattern for the description of the effect of selection on inbreeding. The only possible generalization seems to be that if there is a small difference in the fitnesses, the infant arrays will be reasonably well predicted in the short run by the usual inbreeding formulas.

It is also easy to see that the Equation A.44 cannot in general be written in the form

$$f\sum_i q(i)(ii) + (1-f)\sum_{ij} q(i)q(j)(ij) \quad ,$$

where f is a constant such that $0 \leq f \leq 1$ and $q(i)$ is the frequency of the i-th allele. As a consequence the work on equilibria under

selection and inbreeding, as reported by Li (1967), is invalid, since it depends on this sort of description of the genotypic array.

For these reasons the use of the usual concepts of inbreeding theory has to be sharply circumscribed in situations where selection is operating. In the development of the load theory we have emphasized frequently the approximate nature and short term validity of the use of the different inbreeding coefficients. The main arguments in this regard were given at the end of Section IV.A. The weakness in these arguments is, of course, that they are not of a precise quantitative nature, and in this section we have not been able to improve matters.

The alternative approach of working with loads in a generation matrix context, where selection can be accommodated, was rejected, because with this type of approach no use can be made of Crow's load ratio theorems. The reason for this being the case is that in the derivation of the load ratio theorems in Section IV.A we had to assume that the gene frequencies would not be influenced by the inbreeding process.

One would also like to try to determine the influence of selection in the estimation of a and b from the model

$$z_{ij} = a + bw_i + e_{ij} \quad ,$$

where w_i is degree of inbreeding at the i-th level of inbreeding in a population, and z_{ij} represents the measurement on the j-th individual at the i-th level of inbreeding. It is conceivable that for each level of inbreeding there might exist a F' which will give a better relationship between the mean viability and the loads. The problem would be that we will not be able to calculate such a F', and conquently we will be forced

to use the conventional F in the estimation of a and b. This will result in bias in our estimates. To try and investigate this bias we will have to consider other inbreeding systems that might be used for the purposes of the estimation of loads in the same way as we did with full-sibbing.

It is clear that this problem will involve heavy algebra, with a possiblility of poor returns relative to the time and effort that it might require. Although it might be well to keep the existence of this problem in mind, a solution will, therefore, not be attempted here.

B. The Possibility of Estimating Loads in the Case of Fitness

In the previous sections it became clear that deviations from the basic assumptions can cause many problems in the estimation of loads. We gave much attention to the effect of differential mortality before the life stage of interest, and the effect of such mortality on the estimation of loads in metric traits was pointed out.

We now revert to our terminology of Chapters III and IV, and will talk again about the possible meaning of our terms infant and adult in a practical situation. It is clear that the properties of our infant population will only occur at the moment of fertilization, before any chance of non-random deaths occur. In the case of genetic deaths, the population at any later life stage will exhibit the type of genotypic array characteristic of a population of adults.

In practice no population can be observed at the stage which we described as infancy. In humans, for instance, Stern (1963, p. 86) mentioned

that from 25 to 40 percent of all zygotes perish prior to delivery. Gowen (1963, p. 87) can be consulted for references on other mammals. He states that the more frequent values for loss between ovulation and birth are between 30 and 50 percent. These values are certainly high enough for their effects to be taken into serious consideration.

The type of problem that will result from differential mortality will be well illustrated by considering the expected number of adult offspring per adult mating pair. We will consider the case of two loci under the assumptions of Sections III.A and III.B. We also assume mating between individuals from different inbred lines in the fashion described in Section IV.A.

Consider now the adult mating pair $(r_0^1 r_1^1 r_0^2 r_1^2) \times (s_0^1 s_1^1 s_0^2 s_1^2)$. Under the assumption of no mutation we get the infant offspring array

$$\sum_{x_1,x_2=0,1} \sum_{y_1,y_2=0,1} \gamma_{x_1x_2}\gamma_{y_1y_2}(r_{x_1}^1 s_{y_1}^1 r_{x_2}^2 s_{y_2}^2) ,$$

where the $\gamma_{x_1x_2}$ are recombination values as defined in Section III.B. Let us assume that the mating under consideration gave t offspring. Then, since we assume a multinomial situation, the expected number of infants with the genotype $(r_{x_1}^1 s_{y_1}^1 r_{x_2}^2 s_{y_2}^2)$ will be $t\gamma_{x_1x_2}\gamma_{y_1y_2}$. It likewise follows that the expected number of adults for the mating pair under consideration, given t infant offspring, is

$$\sum_{x_1,x_2=0,1} \sum_{y_1,y_2=0,1} t\gamma_{x_1x_2}\gamma_{y_1y_2}\ell(r_{x_1}^1 s_{y_1}^1 r_{x_2}^2 s_{y_2}^2) ,$$

where $\ell(r_{x_1}^1 s_{y_1}^1 r_{x_2}^2 s_{y_2}^2)$ is the probability of survival from infancy to adulthood of the genotype under consideration. By the same type of

argument as was used in the derivation of Equations IV.A.1 and IV.B.3 it follows that E(number of adult offspring per adult mating pair)

$$= \frac{1}{D} \sum_{\substack{r_0^1 r_1^1 \\ r_0^2 r_1^2}} \sum_{\substack{s_0^1 s_1^1 \\ s_0^2 s_1^2}} \sum_{x_1,x_2=0,1} \gamma_{x_1 x_2} \sum_{y_1,y_2=0,1} \ell(r_{x_1}^1 s_{y_1}^1 r_{x_2}^2 s_{y_2}^2) a(r_0^1 r_1^1 r_0^2 r_1^2) a(s_0^1 s_1^1 s_0^2 s_1^2)$$

$$\times\, p(r_0^1 r_1^1 r_0^2 r_1^2) p(s_0^1 s_1^1 s_0^2 s_1^2) ,$$

where

$$D = \Big[\sum_{\substack{i_0^1 i_1^1 \\ i_0^2 i_1^2}} \ell(i_0^1 i_1^1 i_0^2 i_1^2) p(i_0^1 i_1^1 i_0^2 i_1^2) \Big]^2 \qquad \text{(B.1)}$$

As before, the $p(r_0^1 r_1^1 r_0^2 r_1^2)$ will take on values according to Equation IV.B.2 in an population of mating pairs constructed from inbred lines. Since $\sum_{x_1 x_2} \gamma_{x_1 x_2} = 1$, it follows that with $\ell(r_{x_1}^1 s_{y_1}^1 r_{x_2}^2 s_{y_2}^2) = 1$ for all variables involved, the Equation B.1 reduces in the case of random mating to the Equation IV.B.3.

The Equation B.1 illustrates precisely the problem which will result from an attempt to apply load theory to fitness in practice, for as we pointed out at the beginning of this section, our inability to count offspring at conception forces us into a situation equivalent to adult to adult observations on parent and offspring in terms of the model employed in Chapters III and IV. We note that even if we have an estimate of the mean viability, which appears in the denominator of Equation B.1, we still would not be able to construct an estimate of the square of the mean fitness of the population. Hence, we are in no position to derive a model

like Equation IV.C.9 for the estimation of the genetic loads.

A moment's reflection will convince one that a maximum likelihood approach in the manner of Equations IV.C.13 and IV.C.14 will be of no avail here. The reason for this is that an adult to adult analog of the expression $\sum_t tP_{it}$ will not be equal to the square of the mean fitness of the population, but will have to be a function similar to Equation B.1, of the various quantities involved.

It is clear, therefore, that we are here in a situation in which we will have to be able to estimate the coefficient of recombination, the viabilities and mean fertilities involved in Equation B.1 in order to be able to apply the load theory.

In the light of our work on estimation in Chapter IV it is fair to assume that in order to estimate all these parameters one will have to know things like the number of loci and the number of alleles per locus. Such knowledge is in general not available, and indeed if we already knew so much about the genetic structure of a population, it will be unlikely that loads will add much to our information.

The comments in the previous section about the disturbance caused in load theory by non-multiplicative epistasis and too tight linkage in the case of multiplicative gene action between loci carry over to the application of load theory to fitness. However, these sources of disturbance are really not too worrisome, because it is conceivable that one could evolve tests to detect deviations from Hardy-Weinberg equilibria and the load theory will then simply not be applicable. The real shipwreck of the load theory is on the rocks of differential mortality and, as follows from our discussion at the end of Chapter IV, lack of information about the

value of the maximal genotype. This unfortunate conclusion follows, for even though the population might have Hardy-Weinberg structure at the moment of conception, these factors will cause the load theory to yield the wrong answers.

VI. SUMMARY AND CONCLUSIONS

The key assumption underlying all of the theory of genetic loads is that of the existence of Hardy-Weinberg equilibria at all loci that influence fitness. The conditions under which such equilibria can be expected to exist in an infinite population were, therefore, investigated in some detail for the case of two loci with an arbitrary number of alleles per locus. This procedure had the advantage of simplifying the notation to a considerable extent, while nevertheless retaining enough generality in many circumstances to allow straightforward generalization to the case of n loci.

To keep the mathematics workable, two life phases, infant and adult, non-overlapping generations and a uniform environment were assumed. In order to ensure the existence of Hardy-Weinberg equilibria, it was further necessary to assume no sex difference in fitness, ramdom mating and that the expected number of offspring of a mating pair is the product of two means, one corresponding to each parent. This last assumption was referred to by the term product fecundity. It seems that in general the only modes of gene action that lead to the existence of Hardy-Weinberg equilibria at all loci are those of additive and multiplicative gene action between loci.

In the case of no mutation and additive gene action between two loci, the necessary and sufficient condition for the existence of a stable Hardy-Weinberg equilibrium is that both the constituent single locus cases are in stable Hardy-Weinberg equilibria. In the case of no mutation and multiplicative gene action between two loci, both single locus cases must also be in stable Hardy-Weinberg equilibria, but in addition the linkage

between the two loci must not be too tight, as indicated by a simple function of the characteristic roots of a transformation of the single locus fitness matrices.

In the case of additive gene action between loci there is no reason why the two-loci result should not generalize to the case of n loci, although no rigorous proof was derived. The necessary and sufficient conditions for the stability of a Hardy-Weinberg equilibrium in the case of n loci with multiplicative gene action between loci was shown to be determined by the degree of recombination between loci as well as the characteristic roots of a simple transformation of the single locus fitness matrices. It was shown that the modulus of a linear combination of these characteristic roots weighted by the recombination values must be less than or equal to unity in order to ensure Hardy-Weinberg stability.

The foregoing results refer only to local equilibria, i.e., they do not allow any statement about possible non-Hardy-Weinberg equilibria. However, in the case of additive gene action between two loci it was shown that the mean fitness is nondecreasing from one generation to the next. In the case where the two-loci fitness matrix is of full rank, this statement allows the deduction that the stable Hardy-Weinberg equilibrium will be a global equilibrium, that is, in this case no other stable equilibria exist. This statement seems to be applicable to the case of n loci, although the development of a formal proof may be quite difficult.

Another topic of interest in the study of stable equilibria is the relationship between maximum mean fitness and the fitness at Hardy-Weinberg equilibrium. With additive gene action between two loci it was shown that a Hardy-Weinberg equilibrium is stable if and only if it represents a local

maximum of the mean fitness function. In the case of multiplicative gene action between loci the Hardy-Weinberg equilibrium does not give a maximum for the mean fitness function.

All of the foregoing discussion refers only to internal points on the equilibria or maxima, that is, only to points where the equilibrium frequencies are greater than zero and less than unity. The possibility of new genes, which may arise through mutation, becoming established in a population with Hardy-Weinberg structure was dicussed in conjunction with the stability of border points. Clearly, if a gene is associated with an unstable border point it will increase, and there may be a chance that it will be established in the population. The stability of border points in a Hardy-Weinberg system depends on the non-zero equilibrium gene frequencies, the equilibrium mean fitness and the fitnesses of the new gene in heterozygotic combination with the established genes.

The foregoing work was based on the assumption of no mutation. The problem with mutation is that it causes the equilibrium equations to become systems of cubic equations to which literal solutions can only be obtained in the simplest cases. In this respect the effect of mutation is the same as that of recombination between loci with arbitrary gene action between loci. In the case of a single locus with two alleles the effect of mutation was considered in detail by the analysis of the single cubic equilibrium equation. In both the single locus and two-loci systems it was shown that with small mutation rates, especially in the case of mutation to the border points, a selection system with mutation can be locally approximated by a selection system without mutation. This result is of importance for the segregation load, since the assumption of a selection system with-

out mutation is unrealistic. For the present purposes only cases were considered where Hardy-Weinberg equilibria are possible on the internal points of a selection system.

In the cases where a selection system by itself could exist only on the border points, heterozygosity can be maintained by a balance between mutation and selection, a situation which, in load terms, would result in a mutation load.

In this study load theory was considered from the point of view of its utility as a tool for the study of the genetic architecture of a population. Hence the whole development was aimed at the estimation of the ratio of the inbred load to the random mating load in the cases of the mutation load and the segregation load. A concise discussion of these matters was given in Chapter II.

In the present approach to the estimation of the load ratios it is necessary to assume that the requisite inbreeding is accomplished rapidly, so that selection would not interfere too much with the description of the genotypic array by the inbreeding coefficients.

In this study the basic premise is that fitness is the property of a pair of genotypes. This implies in general that the only observable connected to fitness in an experimental context is the number of offspring per mating. Hence it is necessary to construct populations for estimation purposes by random mating of infants derived from different equally inbred lines. The single locus inbreeding coefficients that characterize the different sub-populations derived in this way, are then used for the purpose of doing a regression analysis.

From the way in which fitness is defined, it is also possible to show that the expected number of infant offspring per infant mating pair is equal to the square of the mean fitness. This will also be true on inbreeding with random mating between inbred lines, if one takes care to stay within the limits enforced by selection on the description of the genotypic array by multi-loci inbreeding coefficients. An alternative approach, suggested by the method of maximum likelihood, for the estimation of population mean fitness was also explored.

In the context of the assumptions that would result in Hardy-Weinberg frequencies it was found that the regression of the number of infant offspring per infant mating pair on the inbreeding coefficients of the different estimation subpopulations will allow the estimation of the genetic loads for fitness in the case where the value of the maximal genotype is known.

There are two problems here. The first is that the value of the maximal genotype will not be known and that it will be impossible to estimate in most cases. The other problem is that the stage designated in this study by the term infancy would correspond in practice to the moment of conception before any differential mortality occurs. In an experimental situation it is often only possible to observe the number of births. It is shown that if there is a significant degree of mortality before birth, the load theory will no longer apply. The same conclusions apply in the context of the approach suggested by the method of maximum likelihood.

The load theory in the case of metric traits, as exemplified by the trait viability, was also examined. In most cases where the load theory has been applied in practice, it was to the trait viability. Here the

model that follows from the present investigation is somewhat different from the one in the literature. Reasons are given why this alternative model is considered to be more realistic. However, it is concluded that in most cases the two models will lead to the same conclusions.

It is axiomatic that the correct analysis of data depends on the error structure that derives from an experimental situation. An attempt has therefore been made to expose the assumptions which underlie the usual analysis of load data by looking at data from an actual experiment. An indication as to the correct analysis in the case of errors of a sampling nature is given. It is also shown that an environmental effect peculiar to individuals from a specific cross at a certain level of inbreeding would give rise to an error structure entirely different from the one arising from random sampling.

The problem with the application of load theory to traits other than fitness is that the load ratio theorems that distinguish between mutation and segregation loads are in general no longer valid. In the case of viability these theorems would only be applicable if the genes that influence viability are independent in action from the genes that influence fecundity.

In the case of metric traits it was furthermore shown that differential mortality before the time of measurement will cause the regression on the inbreeding coefficient to be curvilinear. With respect to the disturbance caused by non-multiplicative epistasis it is concluded that such epistasis presents a situation to which the concepts and theory of genetic loads are inapplicable.

In the case of the application of load ideas to the trait viability, only multiplicative gene action was considered, since it is felt that the assumption of additive gene action between loci would be unrealistic. In the case of multiplicative gene action between loci, deviations from Hardy-Weinberg equilibrium can occur if the linkage between the different loci is too tight. The deviation from Hardy-Weinberg equilibrium will invalidate the description of the genotypic array by the use of multi-loci inbreeding coefficients, so that the effect of too tight linkage on the load theory will be similar to that of non-multiplicative epistasis.

The effect of selection on the genotypic array was investigated in the simple case of a single locus with one generation of full-sib mating. It is concluded that the usual description of the population by the use of Wright's inbreeding coefficient would be a reasonable approximation only if the differences between the fitnesses of the different genotypes are small. The genotypic array produced under the present circumstances cannot be described in such a way as to support the procedure of deriving the conditions for equilibrium gene frequencies under selection by the use of Wright's coefficient of inbreeding.

VII. BIBLIOGRAPHY

Atkinson, F. V., Watterson, G. A. and Moran, P. A. P. 1960. A matrix inequality. Quarterly Journal of Mathematics 11: 137-140.

Bellman, R. 1960. Introduction to matrix analysis. New York, New York, McGraw-Hill Book Company, Incorporated.

Benzer, S. 1962. The structure of the gene. Scientific American 206: 70-84.

Benzer, S. 1967. Genetic fine structure and its relation to the DNA molecule. Gene Theory. Belmont, California, Dickenson Publishing Company.

Bodmer, W. F. and Felsenstein, J. 1967. Linkage and selection: theoretical analysis of the deterministic two locus random mating model. Genetics 57: 237-265.

Cockerham, C. C. and Weir, B. S. 1968. Sib mating with two linked loci. Genetics 60: 629-640.

Crow, J. F. 1958. Some possibilities for measuring selection intensities in man. Human Biology 30: 1-13.

Crow, J. F. 1962. Population genetics: selection. Methodology In Human Genetics 1962: 53-75. San Francisco, Holden-Day, Incorporated.

Crow, J. F. 1963. The concept of genetic load: a reply. American Journal of Human Genetics 15: 310-315.

Crow, J. F. and Kimura, M. 1964. The theory of genetic loads. Genetics Today. 11th International Congress of Genetics, The Hague, Netherlands, September 1963, Proceedings: 495-506. London, England, Pergamon Press.

Dobzhansky, T., Spassky, B. and Tidwell T. 1963. Genetics of natural populations. XXXII. Inbreeding and the mutational and balanced genetic loads in natural populations of Drosophila pseudoobscura. Genetics 48: 361-373.

Fisher, R. A. 1930. The genetical theory of natural selection. Oxford University Press, Oxford, England.

Gillois, M. 1964. La relation d'identité en génétique. Unpublished Ph.D. thesis. Paris, France, Library, University of Paris, France.

Gowen, J. W. 1963. Genetic selection in man. Third Macy Conference on Genetics. Ann Arbor, Michigan, University of Michigan Press.

Griffing, B. 1960. Theoretical consequences of truncation selection based on the individual phenotype. Australian Journal of Biological Science 13: 307-343.

Haldane, J. B. S. 1927a. A mathematical theory of natural and artificial selection. Part 4. Cambridge Philosophical Society Proceedings 23: 607-615.

Haldane, J. B. S. 1927b. A mathematical theory of natural and artificial selection. Part 5: Selection and mutation. Cambridge Philosophical Society Proceedings 23: 838-844.

Haldane, J. B. S. 1937. The effect of variation on fitness. American Naturalist 71: 337-349.

Haldane, J. B. S. and Jayakar, S. D. 1965. The nature of human genetic loads. Journal of Genetics 59: 143-149.

Harris, D. L. 1964. Genotypic covariances between inbred relatives. Genetics 50: 1319-1348.

Hiraizumi, Y. and Crow J. F. 1960. Heterozygous effects on viability, fertility, rate of development and longevity of Drosophila chromosomes that are lethal when homozygous. Genetics 45: 1071-1083.

Jacquard, A. 1968. Évolution des populations d'effectif limité. Population 23: 279-300.

Kempthorne, O. 1957. An introduction to genetic statistics. New York, New York, John Wiley and Sons, Incorporated.

Kempthorne, O. and Pollak, E. 1966. Fisher's fundamental theorem of natural selection. Unpublished seminar given at Ames, Iowa State University of Science and Technology. Ames, Iowa, Statistical Laboratory, Iowa State University.

Kempthorne, O. and Pollak, E. ca. 1969. Concepts of fitness in Mendelian populations. Submitted for publication in Genetics ca. 1969.

Kimura, M. 1955. Random genetic drift in a multi-allelic locus. Evolution 9: 419-435.

Kimura, M. and Crow, J. F. 1964. The number of alleles that can be maintained in a finite population. Genetics 49: 725-738.

Kimura, M., Maruyama, T. and Crow, J. F. 1963. The mutation load in small populations. Genetics 48: 1303-1312.

Kingman, J. F. C. 1961a. A mathematical problem in population genetics. Cambridge Philosophical Society Proceedings, Part 3, 57: 574-582.

Kingman, J. F. C. 1961b. On an inequality in partial averages. Quarterly Journal of Mathematics 12: 78-80.

Kojima, K. and Kelleher, T. M. 1961. Changes of mean fitness in random mating populations when epistasis and linkage are present. Genetics 46: 527-540.

Levene, H. 1963. Inbred genetic loads and the determination of population structure. National Academy of Sciences Proceedings 50: 587-592.

Levene, H., Lerner, I. M., Sokoloff, A., Ho, F. K. and Franklin, J. R. 1965. Genetic load in Tribolium. National Academy of Sciences Proceedings 53: 1042-1050.

Lewontin, R. C. 1964. The role of linkage in natural selection. Genetics Today. 11th International Congress of Genetics, The Hague, Netherlands, September 1963, Proceedings 3: 517-525. London, England, Pergamon Press.

Lewontin, R. C. and Kojima, K. 1960. The evolutionary dynamics of complex polymorphisms. Evolution 14: 458-472.

Li, C. C. 1967. Genetic equilibrium under selection. Biometrics 23: 397-484.

Malogolowkin-Cohen, CH., Levene, H., Dobzhansky, N. P. and Simmons, A. S. 1964. Inbreeding and the mutational and balanced loads in natural populations of Drosophila willistoni. Genetics 50: 1299-1311.

Mettler, L. E., Moyer, J. E. and Kojima, K. 1966. Genetic loads in large populations of Drosophila. Genetics 54: 887-898.

Miller, R. G. 1964. A trustworthy jackknife. Annals of Mathematical Statistics 35: 1594-1605.

Moran, P. A. P. 1962. Statistical processes of evolutionary theory. Oxford, England, Clarendon Press.

Moran, P. A. P. 1964. On the nonexistence of adaptive topographies. Annals of Human Genetics, London 27: 383-393.

Moran, P. A. P. 1967. Unsolved problems in evolutionary theory. Berkeley Symposium on Mathematical Statistics and Probability, 5th, Biology and Problems of Health, Proceedings 4: 457-480.

Morton, N. E. 1960. The mutational load due to detrimental genes in man. American Journal of Human Genetics 12: 348-364.

Morton, N. E., Crow, J. F. and Muller, H. J. 1956. An estimate of the mutational damage in man from data on consanguineous marriages. National Academy of Sciences Proceedings 42: 855-863.

Mulholland, H. P. and Smith, C. A. B. 1959. An inequality arising in genetical theory. American Mathematical Monthly 66: 673-683.

Neel, J. V. and Schull, W. J. 1962. The effect of inbreeding on mortality and morbidity in two Japanese cities. National Academy of Sciences Proceedings 48: 573-582.

Nei, M. 1964. Effects of linkage and epistasis on the equilibrium frequencies of lethal genes. II. Numerical Solutions. Japan Journal of Genetics 39: 7-25.

Norton, H. T. J. 1928. Natural selection and Mendelian variation. London Mathematical Society Series 2 Proceedings 28: 1-45.

Penrose, R. 1955. A generalized inverse for matrices. Cambridge Philosophical Society Proceedings 51: 406-413.

Quenouille, M. H. 1956. Notes on bias in estimation. Biometrika 43: 353-360.

Scheuer, P. A. G. and Mandel, S P. H. 1959. An inequality in population genetics. Heredity 13: 519-524.

Schnell, F. W. 1961. Some general formulations of linkage effects in inbreeding. Genetics 46: 947-957.

Schull, W. J. and Neel, J. V. 1965. The effects of inbreeding on Japanese children. New York, New York, Harper and Row, Publishers.

Shikata, M. 1965. A generalization of the inbreeding coefficient. Biometrics 21: 665-681.

Stern, C. 1963. Genetic selection in man. Third Macy Conference On Genetics. Ann Arbor, Michigan, University of Michigan Press.

Stone, W. S., Wilson, F. D. and Gerstenberg, V. L. 1963. Genetic studies of natural populations of Drosophila: Drosophila pseudoobscura, a large dominant population. Genetics 48: 1089-1106.

Torroja, E. 1964. Genetic loads in irradiated experimental populations of Drosophila pseudoobscura. Genetics 50: 1289-1298.

van Aarde, I. M. R. 1963. Covariances of relatives in random mating populations with linkage. Unpublished Ph.D. thesis. Ames, Iowa, Library, Iowa State University of Science and Technology.

VIII. ACKNOWLEDGEMENTS

A special word of thanks is due to Professor Oscar Kempthorne for the opportunity of being apprenticed to him in learning the craft of population genetics. Thanks also to Professor L. N. Hazel and his department for teaching and stimulation in the field of population genetics and animal breeding. Much thanks is also due to Dr. Edward Pollak for his discussion and teaching on genetic statistics.

The author also wishes to thank Mrs. Shirley Saveraid for her patience and skill in preparing this manuscript.

The author is grateful to the Department of Agriculture of the Republic of South Africa for support during graduate study and the initial work on this thesis.

Support from the National Institute of Health under grant number G-13827 is also gratefully acknowledged.