

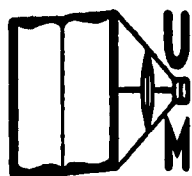
DOCTORAL DISSERTATION SERIES

TITLE THE PREDICTION OF QUANTITATIVE
CHARACTERISTICS IN POLYGENIC
SYSTEMS

AUTHOR ALLAN GEORGE ANDERSON

UNIVERSITY OF MICHIGAN DATE 1951

DEGREE Ph. D. PUBLICATION NO. 3577



UNIVERSITY MICROFILMS
ANN ARBOR . MICHIGAN

**THE PREDICTION OF QUANTITATIVE CHARACTERISTICS
IN POLYGENIC SYSTEMS**

by

Allan G. Anderson

**A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in the
University of Michigan
1951**

Committee in charge:

**Professor Paul S. Dwyer, Chairman
Professor Arthur H. Copeland
Associate Professor Charles W. Cotterman
Assistant Professor Donald A. Darling
Professor Sumner B. Myers**

PREFACE

The author wishes to take this opportunity to express his appreciation to Dr. Paul S. Dwyer, the chairman of his doctoral committee, and to Dr. Charles W. Cotterman and Dr. Donald A. Darling, who were always available for consultation and contributed far more in the way of time and advice than is customary in the case of committee members other than the chairman.

He also wishes to offer his thanks to Mr. G. H. Stringfield of the Ohio Agricultural Experiment Station, both for providing data to be used in demonstrating the methods of the present paper and for taking the time to discuss the general situation with the author during the formative stages of this work.

TABLE OF CONTENTS

Section	Page
1. Position in the field of mathematical genetics	1
2. General statement of purpose	3
3. Background and comparison of elements	4
4. The necessary modern algebra	5
5. A genetical model	9
6. Stock representations as algebraic elements	11
7. Inbred lines as a basing point	14
8. Symbolism and intermediate cases	17
9. General discussion	23
10. An essential manipulative operation	26
11. An example and the consequences of the development	28
12. Consideration of the product term	29
13. An example involving the second filial generation	32
14. Conclusion	35
Appendix	38
Bibliography	47

TABLES

Title	Page
1. Yields in bushels per acre of six inbred parent stocks and the fifteen hybrids resulting from all possible single crosses among them. (Data of Stringfield; <u>Zea mays</u>)	38
2. Values obtained from the data of Table 1 to be used in determining the coefficients in the expression for expected yield appearing in Theorem 10.1.	38
3. Estimated values of u_1 for the six inbred parent stocks of Table 1.	38
4. Comparison of the predicted and actual yields in bushels per acre for the fifteen hybrids of Table 1.	39
5. Analysis of the sums of squares obtained from Table 4 arranged in accordance with the outline given in Section 12.	40
6. Yields in bushels per acre of four inbred parent stocks and all possible crosses obtainable from them within two generations. (Data of Stringfield; <u>Zea mays</u>)	41
7. Values obtained from the data of Table 6 to be used in determining the coefficients in the expression for expected yield appearing in Theorem 10.1.	42
8. Stock representations associated with certain of the pedigrees appearing in Table 6.	42
9. Comparison of the predicted and actual yields in bushels per acre for the heterozygous stocks of Table 6.	43
10. Analysis of the sums of squares obtained from Table 9, arranged under each heading in accordance with the outline given in Section 12.	44
11. Analysis of the differences of observed and computed mean values of yield for certain types of crosses, obtained from the data of Table 9 and treated in accordance with the discussion given in Section 13.	46

1. Position in the field of mathematical genetics.

The field of mathematical genetics is a broad and rapidly expanding one, and consequently it seems desirable to indicate the general position of the present work within this framework. It is evident that the definition of mathematical genetics as the use of mathematical techniques in connection with problems which are primarily of a genetic nature renders it impossible to accomplish very much in any area of genetics without the use of mathematical genetics. In addition to the many appearances of mathematical genetics in diluted form in books and articles which are purely genetical, many individuals are more directly concerned with mathematical genetics as such than with mathematics merely as a tool to facilitate the treatment of problems in genetics. Interests of this type have given rise to entire books such as those of Hogben (17) and Malécot (23), as well as to the increasingly common practice among authors in the field of probability to include a chapter or at least several articles devoted to mathematical genetics along the same lines followed by Feller (11) and Neyman (26) in their recent books. There are also available in the literature a few research papers such as those of Geiringer (12) and Wright (36) which deal with rather large areas of mathematical genetics, although it is only natural that most papers are ordinarily concerned with more specific aspects of the field than are the books.

One of the many general and specific techniques which serve to make up the field of mathematical genetics is the use of genetic algebras. The present paper makes use of genetic algebras and thus is somehow related to works such as those of Etherington (9,10), Glivenko (13), Schafer (28), and Serebrovskij (29). However, these papers are

of a more theoretical nature and with the exception of those by Etherington are not directly concerned with any application to actual practice of the theory set down. Such application is the primary concern of the present work in which genetic algebras are used as a means to that end.

The problem treated here can also be approached in terms of the literature from another direction. Biologists in general and individuals such as agronomists in particular are interested in the problem of heterosis* and have thus become concerned with all aspects of the study of hybrid vigor. Since the present paper makes use of genetic algebras in order to deal with one part of the general heterosis problem, it is obviously related to other work done in this area without the use of the particular technique of genetic algebras. The problem of heterosis is treated "locally and in the large" throughout the recent literature in the field of genetics, and an attempt is made here to cite a few of the many articles which might prove of interest to the general reader of the present paper. No mention is made of any of the large number of papers concerned with the operation of heterozygosis in evolution rather than its operation in plant breeding. Due to its overall importance and the relative facility of obtaining data dealing with maize, corn has assumed a basic role in the study of heterozygosis in plant breeding. There are available in the literature several rather thorough surveys of the general problem of heterosis with particular attention to corn breeding. Among these are articles due to East (7), Jenkins (21), Kiesselbach (22), and Whaley (35). The papers by Jenkins and Kiesselbach also deal to some extent with the history of corn research in general, an aspect of the overall picture

*Heterosis and heterozygosity are here used interchangeably although such usage is not universal among geneticists.

which is treated in an article by Weatherwax (34). Included in the Bibliography without mention in the body of this text are references to several papers which deal with one or more specific aspects of the operation of heterosis in corn breeding. In the case of each of these articles it is clear from the title which area of the general problem is treated therein.

2. General statement of purpose.

The object of the present paper is to arrive at a satisfactory method for predicting the expected mean value of a given quantitative characteristic of a hybridized strain as a result of knowing the corresponding values for certain of the ancestors from which it is descended. A typical example of this general situation is yield of corn, and illustrations are given of applications to data of that type. In an earlier paper the author (1) made an attempt to get at the variance in first generation hybrid crosses, but the method of estimation was quite crude. The present paper leaves open the question of methods for the prediction of variance or higher moments as a result of pedigree.

The paper is intended to be largely self-contained for both the mathematician with little knowledge of genetics and the geneticist with little knowledge of mathematics, and as a result there will be occasions when for the benefit of one group detailed proofs will be given of statements which are obvious to members of the other group. The very basic concepts of modern algebra are not defined, but will be found discussed in detail in any standard text such as those by Birkhoff and Mac Lane (2) and Van der Waerden (33). The individual with

no knowledge at all of genetical terminology will be able to follow the discussion with the aid of the excellent glossary of these terms provided by Darlington and Mather (5) and occasional references to the genetical model which appears in section 5 of the present paper.

Hull (19) gives an expression for the expected value of a given quantitative characteristic for the offspring obtained by $n - 1$ generations of selfing of the hybrid resulting from the crossing of two distinct homozygous strains. In order to evaluate the coefficients found in this expression, it is necessary to have available the corresponding values for the two inbred parents and also for the hybrid of the F_1 generation resulting from the crossing of this pair of homozygotes. The chief purpose of this paper is to effect a considerable extension of these results under the same assumptions regarding the availability of data. A method is devised whereby the expected value of the given characteristic can be predicted for any hybrid having a pedigree showing it to be descended in any manner from only members of some original set of inbred strains.

3. Background and comparison of elements.

Genetic algebras as such seem to have had their beginnings with Serebrovskij (29) and Glivenko (13). Mendelian algebra was defined by Glivenko as a very general system involving three operations and having a base formed of genes. There has been a gradual evolution from this early work having only theoretical value until more specialized systems along the same general lines can now be used in practical applications to actual field problems in genetics.

The mendelian algebra of Glivenko has as elements genes, gametes,

zygotes, populations of each of these, and any other elements obtainable from the individuals and populations by means of the three operations. The gametic and zygotic algebras of Etherington (9,10) have as elements, respectively, populations which produce various types of gametes in given proportions and populations regarded as distributions of zygotic types. In the second of these papers four simple systems of mating within these populations are considered. The elements of the genotypic algebra of the present paper are individual strains having various proportions of different types of gene-pairs at the loci which affect the quantitative characteristic under consideration. The algebraic operation representing the crossing of two of these strains corresponds to the rather general situation of completely random mating between two populations in the algebras of Etherington.

4. The necessary modern algebra.

A linear algebra over a field F is defined by Birkhoff and Mac Lane (2) as a set A which is a finite-dimensional vector space over F and which admits an associative and bilinear multiplication

$$x(yz) = (xy)z$$

$$x(ay + bz) = a(xy) + b(xz)$$

$$(ax + by)z = a(xz) + b(yz)$$

where these laws are to hold for all scalars a and b in F and for all x, y, z in A . The order of A is its dimension as a vector space. In the interests of the non-mathematician a breakdown of the above definition into its component parts due to Brauer (3) is presented. A is a linear algebra over F if for all a and b in F and for all x, y, z in A the following conditions are satisfied.

- 1) A sum $x + y$ is defined, an element of A .
 - 2) A product xy is defined, an element of A .
 - 3) A product ax is defined, an element of A .
 - 4) $x + (y + z) = (x + y) + z$.
 - 5) $x + y = y + x$.
 - 6) A contains an identity element, zero, with respect to addition.
 - 7) If x is an element of A , then $-x$ is also an element of A .
 - 8) $x(yz) = (xy)z$.
 - 9) $x(y + z) = xy + xz$.
 - 10) $(x + y)z = xz + yz$.
 - 11) $a(x + y) = ax + ay$.
 - 12) $(a + b)x = ax + bx$.
 - 13) $(ab)x = a(bx)$.
 - 14) $1x = x$.
 - 15) There exists a basis $x_1, \dots, x_n$ of elements of A such that there is a unique representation $y = a_1x_1 + \dots + a_nx_n$ where the a_i are elements of F .
 - 16) $a(xy) = (ax)y = x(ay)$.
- 4) through 10) are simply the conditions that A be a ring. If all the conditions with the exception of 8) are satisfied, A is said to be a linear non-associative algebra of order n over F .

Consider the system S of hypercomplex numbers of the form

$$aX + bY + cZ$$

subject to the multiplication table

$$XX = X$$

$$XY = X/2 + Y/2 = YX$$

$$XZ = Y = ZX$$

$$YX = X/4 + Y/2 + Z/4$$

$$YZ = Y/2 + Z/2 = ZY$$

$$ZZ = Z$$

and having coefficients in the field F . It should be noted that some authors including Birkhoff and Mac Lane (2) and Van der Waerden (33) restrict the use of the term hypercomplex system to cases where the multiplication table is such that multiplication is associative, thus making linear algebra and hypercomplex system synonymous.

Theorem 4.1 : S is a linear non-associative algebra of order three over F .

Proof: This proof consists of the application of the sixteen conditions for a linear algebra over a field and making the three observations that 8) is denied, all the other conditions are verified, and the basis consists of three elements.

$$1) (aX + bY + cZ) + (gX + hY + kZ)$$

$$= (a + g)X + (b + h)Y + (c + k)Z.$$

$$2) (aX + bY + cZ)(gX + hY + kZ) = (ag + ah/2 + bg/2 + bh/4)X$$

$$+ (ak + ah/2 + bg/2 + bh/2 + bk/2 + ch/2 + cg)Y$$

$$+ (bh/4 + bk/2 + ch/2 + ck)Z.$$

$$3) d(aX + bY + cZ) = (ad)X + (bd)Y + (cd)Z.$$

4) This follows directly from the associative property of addition.

5) This follows directly from the commutative property of addition.

6) $OX + OY + OZ$ is an element of S since O is an element of F .

$$7) -(aX + bY + cZ) = (-a)X + (-b)Y + (-c)Z.$$

8) This is clearly not satisfied since

$$X(YZ) = X(Y/2 + Z/2) = X/4 + 3Y/4 \neq 3Y/4 + Z/4 = (X/2 + Y/2)Z$$

$$= (XY)Z.$$

- 9) This follows directly from the associative and commutative properties of addition and the distributive property of the elements of F .
- 10) This follows directly from the commutative property of multiplication.
- 11)
$$\begin{aligned} d[(a + g)X + (b + h)Y + (c + k)Z] \\ &= (da + dg)X + (db + dh)Y + (dc + dk)Z \\ &= [(da)X + (db)Y + (dc)Z] + [(dg)X + (dh)Y + (dk)Z] \\ &= d(aX + bY + cZ) + d(gX + hY + kZ). \end{aligned}$$
- 12)
$$\begin{aligned} (d + g)(aX + bY + cZ) &= (da + ga)X + (db + gb)Y + (dc + gc)Z \\ &= [(da)X + (db)Y + (dc)Z] + [(ga)X + (gb)Y + (gc)Z] \\ &= d(aX + bY + cZ) + g(aX + bY + cZ). \end{aligned}$$
- 13)
$$\begin{aligned} dg(aX + bY + cZ) &= (dga)X + (dgb)Y + (dgc)Z \\ &= d[(ga)X + (gb)Y + (gc)Z] = d[g(aX + bY + cZ)]. \end{aligned}$$
- 14) $e(aX + bY + cZ) = (ea)X + (eb)Y + (ec)Z = aX + bY + cZ$, where e is the unit element of F .
- 15) The x_i are $eX + 0Y + 0Z$, $0X + eY + 0Z$, $0X + 0Y + eZ$ and $n = 3$.
- 16) This follows directly from the commutative property of multiplication among the elements of F .

The mathematician will observe that most of these verifications are simply applications of the fact that F is a field.

A basic algebra is defined by Etherington (9) as a linear algebra A , associative or not, which possesses a non-trivial scalar representation on the field F over which the algebra is defined. This means that to any element x of A there corresponds a scalar $f(x)$ of F , not identically zero, such that for all x any y in A and for all a in F the following conditions are satisfied.

- 1) $f(x + y) = f(x) + f(y)$.
- 2) $f(ax) = a f(x)$.
- 3) $f(xy) = f(x) f(y)$.

$f(x)$ is called the weight of x , or the weight function of A .

Theorem 4.2 : S is a baric algebra having weight function

$$a + b + c.$$

Proof: This proof consists of the verification of the three conditions for a baric algebra.

- 1)
$$\begin{aligned} f[(a + g)X + (b + h)Y + (c + k)Z] &= (a + g) + (b + h) + (c + k) \\ &= (a + b + c) + (g + h + k) \\ &= f(aX + bY + cZ) + f(gX + hY + kZ). \end{aligned}$$
- 2)
$$\begin{aligned} f[d(aX + bY + cZ)] &= f[(da)X + (db)Y + (dc)Z] = da + db + dc \\ &= d(a + b + c) = d[f(aX + bY + cZ)]. \end{aligned}$$
- 3)
$$\begin{aligned} f[(aX + bY + cZ)(gX + hY + kZ)] &= f[(ag + ah/2 + bg/2 + bh/4)X \\ &\quad + (ak + ah/2 + bg/2 + bh/2 + bk/2 + ch/2 + cg)Y \\ &\quad + (bh/4 + bk/2 + ch/2 + ck)Z] \\ &= (ag + ah/2 + bg/2 + bh/4) \\ &\quad + (ak + ah/2 + bg/2 + bh/2 + bk/2 + ch/2 + cg) \\ &\quad + (bh/4 + bk/2 + ch/2 + ck) \\ &= ag + ah + ak + bg + bh + bk + cg + ch + ck \\ &= (a + b + c)(g + h + k) = f(aX + bY + cZ) f(gX + hY + kZ). \end{aligned}$$

5. A genetical model.

Consider a situation in which the members of a population have the allelic genes A and a , B and b , C and c , and D and d at four distinct loci. The possible gene-pairs in each individual member at the first locus are of the forms AA , Aa , aa and the situations at the

other three loci are comparable. An individual is said to be homozygous at a given locus if the gene-pair at that locus consists of two identical genes. Examples of homozygosity at a given locus are pairs of the types AA, aa, bb, CC, etc. The individual is said to be homozygous only if it is homozygous at every locus, and is otherwise said to be heterozygous. Thus a member of the population is heterozygous if it exhibits heterozygosity at one or more loci. An example of a homozygous individual is one having the gene-pairs

AA
bb
CC
DD

at the four loci, while two individuals having the gene-pairs

Aa
bb
CC
DD

and

AA
Bb
cc
Dd

are both heterozygous.

In crossing two members of a population such as that described in the preceding paragraph the number of possible occurrences at a given locus is quite small, and these possibilities can be readily treated in terms of probability. For example, if both parents are homozygous at the locus only three situations in the offspring can arise there. If each parent has a gene-pair of the type AA at the first locus the offspring must also have a pair of this type, and similarly in the case of an aa pair in each parent. However, if one parent has an AA pair and the other an aa pair at the locus, the offspring must have an Aa pair. In the case where either or both of the parents are heterozygous at the locus, it is considered to be exactly as likely that that parent or parents will contribute an A gene as an a gene to the

make-up of that particular gene-pair in the offspring. Consequently, if one parent is of the form AA and the other Aa at the first locus, the probability is one-half that the offspring will be AA and one-half that it will be Aa at this locus. The case for Aa and aa pairs in the parents is comparable. When both parents are heterozygous at the locus, it is clear that the probabilities of the various types of gene-pairs at that locus in the offspring will be $1/4$ for AA, $1/2$ for Aa, and $1/4$ for aa. It should be noted that the operation of this type of probability situation leads to expected values rather than exact results when the make-up of the gene-pairs affecting a particular characteristic in a given individual is under consideration.

As a consequence of the above discussion it is possible to consider in detail the mating of the individuals X and Y having gene-pairs as shown at the four loci.

<u>X</u>	<u>Y</u>
AA	AA
bb	Bb
cc	CC
Dd	Dd

An offspring of this particular cross will obviously receive an A gene from each parent at the first locus, and will thus have an AA pair there. Similarly, it must have a Cc pair at the third locus. However, at the second locus Bb and bb pairs are equally likely and at the fourth locus the probabilities of DD, Dd, and dd pairs are $1/4$, $1/2$, and $1/4$ respectively.

6. Stock representations as algebraic elements.

It is possible to account for continuous variation on the basis of the simultaneous operation of a number of genes. These genes are

inherited in the mendelian way, but their differences have effects which are small in relation to the total variation, similar to one another, and supplementary to one another. Such a set of genes is said by Darlington and Mather (5) to constitute a polygenic system, and its individual members are termed polygenes.

Consider now a situation in which a given quantitative characteristic is affected by contributions from N gene-pairs having identical effects and such that gene action is additive. It is further assumed that inheritance is symmetrical; a condition which is generally accepted as being satisfied in problems of the type with which the present paper is concerned. This assumption is implicit in the formulas given by Hull (18,19) for F_1 and F_n in terms of the parents from which they are descended. Etherington (9) states that inheritance is symmetrical if the sex of a parent does not affect the distribution of gametic types produced. Symmetry of inheritance may be disturbed by unequal crossing over in male and female or by sex linkage, but Darlington and Mather (5) state that in a polygenic system the results of reciprocal crosses are alike. This statement is verified in the case of hybrid corn by a recent communication from G. H. Stringfield of the Ohio Agricultural Experiment Station (31) to the effect that in data taken over a period of twelve years the differences shown between reciprocal crosses have been so slight as to be negligible.

Under the above assumptions it is possible to represent a given stock as $n_1AA + n_2Aa + n_3aa$, where A and a denote two allelic genes at any one of N different loci, A stands for the effective allele which adds something over the contribution of a , the n_j represent the numbers of gene-pairs of the three possible types, and $n_1 + n_2 + n_3 = N$.

The system of all such stock representations will be denoted by S^* .

Theorem 6.1 : The elements of S^* are in 1-1 correspondence with the subset of S for which $a + b + c = N$ and a, b, c are non-negative integers.

Proof: The conditions on a, b, c are exactly those satisfied by the n_j . It remains to show that AA, Aa, aa satisfy the multiplication table of S .

- 1) $(AA)(AA) = AA.$
- 2) $(Aa)(AA) = (1/2)AA + (1/2)Aa = (AA)(Aa).$
- 3) $(AA)(aa) = Aa = (aa)(AA).$
- 4) $(Aa)(Aa) = (1/4)AA + (1/2)Aa + (1/4)aa.$
- 5) $(Aa)(aa) = (1/2)Aa + (1/2)aa = (aa)(Aa).$
- 6) $(aa)(aa) = aa.$

It is evident that 2), 4), and 5) actually give expected values.

An element x of a baric algebra A is said by Etherington (9) to be normalized when it is replaced by the element $x/f(x)$ of unit weight. It is immediately apparent that all elements of non-zero weight can be normalized. The elements of S^* , in particular, can be normalized by dividing through by N to obtain a new system S^{**} of normalized stock representations $uAA + vAa + waa$ such that $u + v + w = 1$ and u, v, w are non-negative.

Theorem 6.2 : S^{**} is closed under multiplication.

Proof: Since AA, Aa, aa obey the multiplication table of S and $u + v + w = 1$, S^{**} is isomorphic to a subset of the normalized elements of S for which a, b, c are non-negative. It is clear that the product of two elements of S having non-negative coefficients will also have non-negative coefficients, and it follows from the multiplica-

tive property of the weight in a baric algebra that the product of two elements will be automatically normalized if both elements are.

Theorem 6.3 : The expected values of the proportions of loci of the three types AA, Aa, aa in a stock are readily obtainable as functions of the proportions in the two stocks crossed to produce it, provided that these stocks are unrelated.

Proof: This follows directly from Theorem 6.2 and the definition of the product of two elements of S which is given in Section 4.

It should be noted that this theorem is not applicable in the case of selfing or the crossing of related strains, as the implicit assumption of a random distribution over the N loci of the gene-pairs of the various types in one parent with respect to the distribution of those of the other parent would be violated. Theorems 7.1, 8.2, and 8.3 deal with situations of this sort and in doing so bring out the fact that the primary concern should be with crosses between unrelated stocks.

7. Inbred lines as a basing point.

It is now desirable to find a suitable starting point which is both useful in theory and physically practical. It would take a great deal of information concerning various crosses involving given strains to enable one to estimate with any accuracy their proportions of loci of the three different types. Thus Theorem 6.3, while perfectly valid, is of very little practical value at the present point in our discussion. For purposes of the present paper the selfing of a stock is taken to mean either self-fertilization of an individual or the mating of two individuals of the same F_1 stock. The representation approach-

ed by the S_n resulting from successive generations of selfing of a stock P and its descendants is called the limiting stock representation of P by selfing. At this point it becomes useful to consider specific formulations of two of the classical theorems of genetics.

Theorem 7.1 : The number of Aa pairs in any given stock is halved by selfing.

Proof: This follows directly from the multiplication table, since in a cross of this kind each gene-pair is opposite another of the same type.

- 1) $(AA)(AA) = AA.$
- 2) $(Aa)(Aa) = (1/4)AA + (1/2)Aa + (1/4)aa.$
- 3) $(aa)(aa) = aa.$

The selfing of $P = uAA + vAa + waa$ thus produces

$$S_1 = (u + v/4)AA + (v/2)Aa + (v/4 + w)aa.$$

It is evident that the coefficients in the expression for S_1 are actually expected values.

Theorem 7.2 : For any given stock the limiting stock representation by selfing is homozygous.

Proof: By Theorem 7.1, $S_1 = (u + v/4)AA + (v/2)Aa + (v/4 + w)aa$ for $P = uAA + vAa + waa$. If the n^{th} filial generation is denoted by

$$S_n = u_n AA + v_n Aa + w_n aa,$$

then $v_1 = v/2$, $v_2 = v/4$, $v_3 = v/8$, ... , $v_n = v/2^n$.

$$S_1 - P = (v/2) \left[(1/2)AA - Aa + (1/2)aa \right],$$

$$S_2 - S_1 = (v_1/2) \left[(1/2)AA - Aa + (1/2)aa \right] = (v/4) \left[(1/2)AA - Aa + (1/2)aa \right],$$

$$S_3 - S_2 = (v_2/2) \left[(1/2)AA - Aa + (1/2)aa \right] = (v/8) \left[(1/2)AA - Aa + (1/2)aa \right],$$

...

$$S_n - S_{n-1} = (v/2^n) \left[(1/2)AA - Aa + (1/2)aa \right].$$

The total increment in n generations is therefore

$$S_n - P = (1/2 + 1/4 + 1/8 + \dots + 1/2^n) v \left[(1/2)AA - Aa + (1/2)aa \right].$$

Hence the n^{th} filial generation is

$$\begin{aligned} S_n &= uAA + vAa + waa + (1 - 1/2^n) v \left[(1/2)AA - Aa + (1/2)aa \right] \\ &= (u + v/2 - v/2^{n+1})AA + (v/2^n)Aa + (v/2 - v/2^{n+1} + w)aa. \end{aligned}$$

As the number of generations increases, this quickly approaches the limiting stock representation $(u + v/2)AA + (v/2 + w)aa$.

Theorem 7.1 is employed in the work of Hull (18,19), and Theorem 7.2 corresponds exactly to a result utilized in connection with the first type of mating system discussed by Etherington (10) for the elements of his zygotic algebra. As a result of these two theorems it is clear that stocks can be obtained which are essentially homozygous. For purposes of the present paper a stock will be called inbred if the selfing process has been carried out sufficiently so that its representation differs from a limiting stock representation so slightly as to justify its being treated as homozygous in theoretical considerations. These inbred lines can be represented as

$$P_i = u_i AA + (1 - u_i)aa,$$

and they constitute the desired starting point for actual field experiments. It is now possible to state the following basic principle as an immediate consequence of Theorem 7.1 and the repeated application of Theorem 6.3.

The coefficients in the normalized representation of a stock descended from inbred parent lines in any way such that all crosses are between unrelated stocks are readily obtainable as functions of the u_i of the inbred stocks to which its pedigree can be traced.

It is actually possible to obtain the coefficients for a stock

descended in any known way from inbred parent lines for which the u_i are known. However, the computations involved become rather tedious when crosses of related strains are included. Theorems 8.2 and 8.3 treat two such cases and serve to show that these computations are unnecessary if one is interested in those strains for which the expected value of the quantitative characteristic under consideration is most favorable.

8. Symbolism and intermediate cases.

F_1 is ordinarily defined as stated by Darlington and Mather (5) as the first generation of the cross between two homozygous individuals, and F_n as the result of successive selfing of an F_1 and its descendants. This meaning of F_n can be extended, and the notation used by Etherington for a pedigree will be altered to conform to this extension. Inbred stocks are represented as P_i , and $F_{1:ij} = P_i \times P_j$. This manner of definition is extended to

$$F_{2:ijk} = F_{1:ij} \times F_{1:km},$$

$$F_{3:ijkmpqrt} = F_{2:ijk} \times F_{2:pqrt}, \text{ etc.}$$

Etherington (8) defines the shape of a non-associative product as the manner of combination of its factors without regard to their identity, and a non-associative product or shape may be visualized as a pedigree. The shapes generated by repeated squaring of an element, and products having such a shape, are referred to by Etherington as plenary. To eliminate brackets in writing non-associative products, Etherington employs a notation due to Peano in which groups of dots are used to separate the factors when necessary and such that fewness of dots implies precedence in multiplication. $AB.CD:EF.GH$ is an example of a

plenary shape.

Theorem 8.1 : The pedigree representing the descent of a stock in any known way from inbred parent lines may be written as a plenary shape.

Proof: From Theorem 7.1 it is evident that $P_1 \times P_1 = P_1$. Thus in any pedigree a P_1 can be replaced by the same P_1 taken any desired number of times. In a pedigree demonstrating ultimate ancestry in terms of only inbred lines, in particular, any P_1 which appears as crossed with a stock n generations removed from homozygosity for all its ancestors can be replaced by the same P_1 taken 2^n times. It is thus clear that any pedigree given completely in terms of inbred stocks can be converted to a plenary shape.

Plenary shapes lend themselves to convenient representation as F_n involving the 2^n appropriate subscripts. The pedigree of $F_{3:12334555}$ in the notation of Etherington would be $P_1P_2 \cdot P_3P_4P_5 \cdot P_5$. The plenary shape corresponding to this pedigree is $P_1P_2 \cdot P_3P_3 \cdot P_4P_5 \cdot P_5P_5$.

From the commutativity of the crossing operation it follows that $F_{1:ij} = F_{1:ji}$. It is also obvious that $F_{2:ijklm} = F_{2:ijmk} = F_{2:mkiij}$. Similar expressions of a necessarily more complicated nature can be obtained for stocks which are a greater number of generations removed from homozygosity of all ancestors. As a result of the basic principle of Section 7 and using the multiplication table, the proportions of loci of the various types in $F_{1:ij}$, $F_{2:ijklm}$, $F_{3:ijkmpqrt}$, etc., for the cases when all subscripts are different are readily obtainable in terms of the proportions of AA pairs in each of the 2^n inbred stocks from which they are descended. The table is not applicable for all steps in the cases $F_{1:ii}$, $F_{2:ijij}$, $F_{3:ijijijij}$, $F_{3:ijkmiijkm}$, etc., but

$F_{1:ii} = P_i$ and Theorem 7.1 also takes care of the other steps where use of the table is excluded as a result of selfing. When $F_{3:ijijijij}$ is obtained by the crossing of two distinct individuals rather than the selfing of one, it is noted by Darlington and Mather (5) that the result is a biparental progeny of the third generation and not an F_3 in the usual sense. The intermediate cases such as $F_{2:ijik}$, F_3 stocks in which from one to three of the first four subscripts appear again among the last four, $F_{3:ijkmikjm}$ also remain to be considered. The computations become quite involved for pedigrees of these types, and only $F_{2:ijik}$ and $F_{3:ijkmiijpq}$ are considered in detail in the present paper. It is seen that the intermediate cases involving crosses of related stocks lead to undesirable values of the quantitative characteristic, so that further consideration of them is unnecessary.

Theorem 8.2 : Under the assumption that $u_1 = u$ is the same for all the inbred stocks involved, the expected values of the proportions of loci of the three types in $F_{2:1213}$ are midway between those in $F_{2:1234}$ and those in $F_{2:1212}$.

Proof: There are four possible arrangements of the different types of gene-pairs opposite each other in the cross $F_{1:12} = P_1 \times P_2$. These arrangements and the resulting situations in $F_{1:12}$ are:

$$(AA)(AA) = AA$$

$$(AA)(aa) = Aa$$

$$(aa)(AA) = Aa$$

$$(aa)(aa) = aa.$$

There is no loss in generality in assuming $u = 1/2$. The representation of this cross is thus $F_{1:12} = (1/4)AA + (1/2)Aa + (1/4)aa$, and by Theorem 7.1 $F_{2:1212} = (3/8)AA + (1/4)Aa + (3/8)aa$. By application of

the multiplication table $F_{2:1234} = (1/4)AA + (1/2)Aa + (1/4)aa$. Either type of gene-pair in P_3 can be opposite each of the four arrangements in P_1 and P_2 . The resulting situations in $F_{1:13}$ are AA and Aa opposite each of the first two in $F_{1:12}$, and Aa and aa opposite each of the last two. Thus $F_{2:1213}$ leads to the eight situations:

$$(AA)(AA) = AA$$

$$(AA)(Aa) = (1/2)AA + (1/2)Aa$$

$$(Aa)(AA) = (1/2)AA + (1/2)Aa$$

$$(Aa)(Aa) = (1/4)AA + (1/2)Aa + (1/4)aa$$

$$(aa)(AA) = (1/4)AA + (1/2)Aa + (1/4)aa$$

$$(Aa)(aa) = (1/2)Aa + (1/2)aa$$

$$(aa)(Aa) = (1/2)Aa + (1/2)aa$$

$$(aa)(aa) = aa.$$

The corresponding representation is

$$F_{2:1213} = (5/16)AA + (3/8)Aa + (5/16)aa.$$

The proof of the statement similar to Theorem 8.2 which can be made for the $F_{3:12341256}$ is quite long and some of the details are omitted.

Theorem 8.3 : Under the assumption that the proportions of loci of the three types are the same for all first generation hybrids involved, the expected values of the proportions of loci of the various types in $F_{3:12341256}$ are midway between those in $F_{3:12345678}$ and those in $F_{3:12341234}$, where the latter is obtained by the crossing of two distinct individuals rather than by the selfing of one.

Proof: There are twenty-seven possible arrangements of the various types of gene-pairs opposite each other in $F_{1:12}$, $F_{1:34}$, and $F_{1:56}$. The $F_{2:1234}$ situations resulting from the nine arrangements in the

first two are:

$$\begin{array}{c}
 AA \\
 (1/2)AA + (1/2)Aa \\
 Aa \\
 (1/2)AA + (1/2)Aa \\
 (1/4)AA + (1/2)Aa + (1/4)aa \\
 (1/2)Aa + (1/2)aa \\
 Aa \\
 (1/2)Aa + (1/2)aa \\
 aa.
 \end{array}$$

From these the $F_{3:12341234}$ situations are obtained as:

$$\begin{array}{c}
 AA \\
 (9/16)AA + (3/8)Aa + (1/16)aa \\
 (1/4)AA + (1/2)Aa + (1/4)aa \\
 (9/16)AA + (3/8)Aa + (1/16)aa \\
 (1/4)AA + (1/2)Aa + (1/4)aa \\
 (1/16)AA + (3/8)Aa + (9/16)aa \\
 (1/4)AA + (1/2)Aa + (1/4)aa \\
 (1/16)AA + (3/8)Aa + (9/16)aa \\
 aa.
 \end{array}$$

There is no loss in generality in assuming

$$F_{1:1j} = (1/3)AA + (1/3)Aa + (1/3)aa$$

for all the first generation hybrids involved. The representation of the earlier cross is $F_{2:1234} = (1/4)AA + (1/2)Aa + (1/4)aa$, and by application of the multiplication table $(1/4)AA + (1/2)Aa + (1/4)aa = F_{3:12345678}$. The respective pairs in the twenty-seven arrangements for $F_{1:12}$, $F_{1:34}$, $F_{1:56}$ and the resulting situations in $F_{3:12341256}$ are:

AA, AA, AA :	AA
AA, AA, Aa :	$(3/4)AA + (1/4)Aa$
AA, AA, aa :	$(1/2)AA + (1/2)Aa$
AA, Aa, AA :	$(3/4)AA + (1/4)Aa$
AA, Aa, Aa :	$(9/16)AA + (3/8)Aa + (1/16)aa$
AA, Aa, aa :	$(3/8)AA + (1/2)Aa + (1/8)aa$
AA, aa, AA :	$(1/2)AA + (1/2)Aa$
AA, aa, Aa :	$(3/8)AA + (1/2)Aa + (1/8)aa$
AA, aa, aa :	$(1/4)AA + (1/2)Aa + (1/4)aa$
Aa, AA, AA :	$(9/16)AA + (3/8)Aa + (1/16)aa$
Aa, AA, Aa :	$(3/8)AA + (1/2)Aa + (1/8)aa$
Aa, AA, aa :	$(3/16)AA + (5/8)Aa + (3/16)aa$
Aa, Aa, AA :	$(3/8)AA + (1/2)Aa + (1/8)aa$
Aa, Aa, Aa :	$(1/4)AA + (1/2)Aa + (1/4)aa$
Aa, Aa, aa :	$(1/8)AA + (1/2)Aa + (3/8)aa$
Aa, aa, AA :	$(3/16)AA + (5/8)Aa + (3/16)aa$
Aa, aa, Aa :	$(1/8)AA + (1/2)Aa + (3/8)aa$
Aa, aa, aa :	$(1/16)AA + (3/8)Aa + (9/16)aa$
aa, AA, AA :	$(1/4)AA + (1/2)Aa + (1/4)aa$
aa, AA, Aa :	$(1/8)AA + (1/2)Aa + (3/8)aa$
aa, AA, aa :	$(1/2)Aa + (1/2)aa$
aa, Aa, AA :	$(1/8)AA + (1/2)Aa + (3/8)aa$
aa, Aa, Aa :	$(1/16)AA + (3/8)Aa + (9/16)aa$
aa, Aa, aa :	$(1/4)Aa + (3/4)aa$
aa, aa, AA :	$(1/2)Aa + (1/2)aa$
aa, aa, Aa :	$(1/4)Aa + (3/4)aa$
aa, aa, aa :	aa

The corresponding representations for the two crosses of related stocks are

$$F_{3:1234|234} = (1/3)AA + (1/3)Aa + (1/3)aa$$

and

$$F_{3:1234|256} = (7/24)AA + (5/12)Aa + (7/24)aa.$$

It is clear that Theorems 8.2 and 8.3 will no longer be valid as such if the proportions of loci of the various types in the P_i and $F_{1:ij}$, respectively, are not all the same. However, one is not concerned with exactly where the intermediate case lies on the scale between the cross of unrelated stocks and the cross of identical stocks, but only with the fact that it actually is intermediate between the two with regard to the quantitative characteristic under consideration. Thus these two theorems are seen to be simply two specific formulations of the general statement that the degree of heterozygosity present in the result of a cross is inversely related to the degree of relationship between the stocks crossed.

9. General discussion.

The expressions previously obtained are actually expected values rather than exact results, although this was not stated explicitly in every case. In the proof of Theorem 6.1 it was pointed out that the genotypic multiplication table leads to these expected values, and further mention of the situation was made only when not to do so might conceivably have caused confusion. For unrelated stocks the assumption implicit in the use of the multiplication table, that the gene-pairs of the various types in the one parent are distributed randomly with respect to the distribution of those of the other parent over the

N loci, is not unwarranted, but those results obtained through its use are clearly expected values rather than exact expressions.

The gap between use of the multiplication table and Theorem 7.1 is not as great as it may appear to be. It is interesting to note that the crosses with which one is most concerned in the practical applications are among those for which the proportions of the various types of loci are obtainable in a straightforward manner. The theorems dealing with selfing are important, as they justify the assumption that inbred lines can be used as a starting point and thus bring about a considerable simplification of the entire discussion. As a result of the indication of the presence of an overdominance factor, to be discussed in the next section, it is seen that crosses of unrelated strains will in general give the most desirable values for the quantitative characteristics in which one is interested. Thus the difficulty involved in obtaining the proportions of loci of the various types in crosses of related strains is of very little significance, as the expected values of the characteristics in question for crosses of this kind will always be intermediate between the most desirable values and the poorer values resulting from selfing.

Stringfield (30) brings out the striking relationships between degree of heterozygosis and three different quantitative characteristics in maize. It is seen that height of ear node and acre grain yield are directly related to degree of heterozygosity, while the length of the period from planting to silking is inversely related. The 1948 report of cooperative tests of corn hybrids (32) contains two specific instances which tend to show this same relationship. In 34 crosses among ten inbred lines, the only one involving related parent

stocks has a yield considerably smaller than that of any other cross and it is also observable that none of the other 33 crosses requires a longer period to silk. The second case is perhaps even more convincing, as the one cross involving related inbred lines has a yield which differs from the next smallest yield of any of the 36 possible crosses of nine inbred lines by more than the maximum difference in yield for any two of the other 35 crosses. The period to silk is also longer for this stock than for any of the 35 crosses involving unrelated parents.

It is certain that the assumption of N gene-pairs having identical effects upon a given quantitative characteristic is not ideally realized in actual physical situations. However, it is clear that studies such as that of Hull (18) made on this basis give results which do not differ significantly from those obtained in practice. It is seen in the last section of this paper that expressions based on the assumption that a polygenic system of this type is operating give a satisfactory fit to data from actual field experiments. Such experiments must, of course, be carefully organized and standardized to minimize the effects of non-heritable agencies upon the criteria under consideration.

From the definition of the system S^{**} it is evident that several distinct strains may be represented by the same element of the system. This situation arises when the stocks in question have the same proportions of gene-pairs of the three types distributed differently over the N loci. With inbred parent lines as a starting point no difficulty is caused by this seemingly paradoxical situation, since it is always clear in crossing two stocks having the same representation

whether or not it is a case of selfing. It is immediately apparent whether a cross of inbred stocks breeds true or not, and all other stocks considered have pedigrees which can be traced back to sets of inbred lines which either are or are not identical. Thus one invariably knows whether to apply Theorem 6.3 or Theorem 7.1 to a given cross.

10. An essential manipulative operation.

It now remains to be able to estimate for inbred stocks the proportions of AA and aa gene-pairs at the N loci affecting a given quantitative characteristic. For simplicity the specific characteristic yield is considered. The yields of the inbred stocks and of the hybrids of the F_1 generation resulting from crosses within the group of homozygous stocks are assumed known.

For the N loci let the basic effect of a gene substitution A for a be d, dominance deviation of any single gene-pair kd, proportions of loci AA in P_1 and P_2 respectively u and v, yield of the bottom recessive T, and gene action additive. It is evident that the total interval for inbred stocks is $2Nd$. $k = 0$ if there is no dominance effect, $k = 1$ represents complete dominance, and k is greater than 1 in the case of overdominance. The results of Hull (18) indicate that overdominance is to be expected in situations of the type to which the methods of the present paper are applicable. Several field experiments comparable to those treated in the next sections of this paper are recorded there, and in each case a value of k greater than 1 is obtained.

In the notation of S**:

$$P_1 = uAA + (1 - u)aa$$

$$P_2 = vAA + (1 - v)aa$$

$$F_{1:12} = uvAA + (u + v - 2uv)Aa + (1 - u - v + uv)aa.$$

Thus in terms of yield

$$P_1 = u(T + 2Nd) + (1 - u)T = 2uNd + T$$

$$P_2 = v(T + 2Nd) + (1 - v)T = 2vNd + T$$

$$\begin{aligned} F_{1:12} &= uv(T + 2Nd) + (u + v - 2uv)(T + Nd + kNd) + (1 - u - v + uv)T \\ &= (u + v)(1 + k)Nd - 2uvkNd + T. \end{aligned}$$

$$\bar{P} = (P_1 + P_2)/2 = (u + v)Nd + T.$$

$$P_1 P_2 = 4uvN^2d^2 + 2(u + v)NdT + T^2.$$

It can now be seen that

$$F_{1:12} = (1 + k + kT/Nd)\bar{P} - (k/2Nd)P_1 P_2 - kT^2/2Nd - kT,$$

a result stated by Hull (18). From this expression for expected yield it is clear that the following theorem holds.

Theorem 10.1 : The expression for expected yield of a first generation hybrid can be written as $z_2 = ax + by + c$, where z_2 is the expected yield, x is the sum of the two parent yields, y is the product of the parent yields, and a, b, c are real numbers.

The non-mathematician will find an excellent treatment of the use of normal equations in connection with multiple regression given by Rider (27). One can estimate z from x and y by means of a multiple regression equation $z_2 = ax + by + c$ in which a and b are called the partial regression coefficients of z on x and y respectively. Geometrically this equation represents a plane, and the ordinary procedure is to determine the coefficients so as to minimize the sum of squares of vertical deviations (assuming the z -axis to be vertical) from the plane. Employing the usual methods of the calculus partial derivatives

of $\sum (z - z_2)^2 = \sum (z - ax - by - c)^2$ with respect to a , b , and c are taken and set equal to zero to give the normal equations required to solve for a , b , c as

$$\sum z = a \sum x + b \sum y + cm$$

$$\sum xz = a \sum x^2 + b \sum xy + c \sum x$$

$$\sum yz = a \sum xy + b \sum y^2 + c \sum y,$$

where m is the number of crosses involved. The values obtained for a , b , c lead to corresponding values for the unknowns k , T , N_d and the last two of these in turn make possible the estimation of the proportions of AA loci in the inbred parent stocks. In the next section a specific example is employed to show how these results are obtained.

11. An example and the consequences of the development.

This example makes use of unpublished data obtained by G. H. Stringfield in making all possible crosses among six inbred lines of corn. The resulting data, which were also used by Hull (18) in a different connection, appear in Table 1 of the Appendix. The parental yields occupy the cells of the diagonal, and those cells to the left of the diagonal are blank since the data showed the differences between reciprocal crosses to be negligible.

The data in Table 2 of the Appendix may be used to evaluate the coefficients in the expression given in Theorem 10.1 on the basis of the observed yields of the fifteen crosses. Substituting in the normal equations from the table and simultaneous solution of the resulting expressions gives $a = 0.812213$, $b = -0.014102$, $c = 55.180$. The expression for expected yield in crosses involving the particular non-heritable factors present in this group of crosses is then

$$F_{1;12} = 0.812213 (P_1 + P_2) - 0.014102 P_1 P_2 + 55.180.$$

Comparison of this with the expression immediately preceding Theorem 10.1 gives $k = 1.87$, $T = -44.2$, $Nd = 66.35$. The relation stating that $P_1 = 2u_1Nd + T$ then leads to the values of u_1 given in Table 3 of the Appendix. The general consequence of the example treated in this section is stated in the following theorem.

Theorem 11.1 : The u_i can be estimated if the P_i ($i = 1, \dots, n$) and $F_{1;ij}$ ($i = 1, \dots, n-1$; $j = 2, \dots, n$; i less than j) are known for the n inbred stocks involved.

From the definitions of T , d , and k it is evident that the contributions to yield by each gene-pair of the three types aa , Aa , AA are respectively T/N , $(T/N) + (1 + k)d$, $(T/N) + 2d$. Thus, as a result of the basic principle of Section 7 and Theorem 11.1, it is possible to state the theorem which expresses the chief contribution of the present paper.

Theorem 11.2 : Any stock having a pedigree traceable in any way to only inbred stocks has a yield function which can be estimated when the yields of the inbred ancestors and the hybrids of the F_1 crosses among them are known.

12. Consideration of the product term.

It now seems desirable to examine the role played by the product term in the process of estimation of expected values of quantitative characteristics. It should be recalled that the underlying assumptions involved in the overall situation led to its inclusion, and this examination is entirely apart from that consideration. The data of Table 2 of the Appendix can be used in connection with the normal

equations

$$\begin{aligned}\sum z &= a \sum x + bm \\ \sum xs &= a \sum x^2 + b \sum x\end{aligned}$$

obtained from the regression equation $z_1 = ax + b$ in a manner comparable to that of Section 10 to give $z_1 = 0.26974 x + 76.643$ as the most suitable linear expression in x for purposes of estimation. Table 4 of the Appendix compares the observed yields of the fifteen hybrids with the estimates obtained both with and without the use of the product term. These estimates are designated respectively as z_2 and z_1 .

A test of a statistical hypothesis is a procedure for deciding whether to accept or reject the hypothesis. The level of significance of such a test is largely a subjective matter and depends upon the chances that an individual is willing to take that his judgment may be wrong. Many investigators are willing to regard as significant any deviation (or difference) for which the probability of a greater deviation (or difference) is .05, and as highly significant any deviation (or difference) for which this probability is .01 or less. For purposes of the present paper a 5% level of significance is employed in connection with statistical tests.

It is desirable to test the empirical data of Table 4 against the two sets of theoretical values. The sums of squares of deviations are considered in order to observe improvements in fit as the estimates $\bar{z}$, z_1 , z_2 are employed progressively to bring about reductions in the sum of squares. Goulden (14) outlines the following procedure for testing the significance of the variance corresponding to each degree of freedom used in fitting. It should be noted that the number of degrees of freedom for estimating the total variance is $m-1$, one having been de-

ducted because the mean was calculated from the set of values.

	SS	DF
Regression function	$\sum(z_1 - \bar{z})^2$	1
First residual	$\sum(z - z_1)^2$	$n - 2$
Total	$\sum(z - \bar{z})^2$	$n - 1$

	SS	DF
Difference, regression functions	$\sum(z_2 - \bar{z})^2 - \sum(z_1 - \bar{z})^2$	1
Residual	$\sum(z - z_2)^2$	$n - 3$
First residual	$\sum(z - z_1)^2$	$n - 2$

The significance of the variance corresponding to the two degrees of freedom used in fitting the quadratic expression can be tested directly in accordance with the usual procedure as follows.

	SS	DF
Regression function	$\sum(z_2 - \bar{z})^2$	2
Residual	$\sum(z - z_2)^2$	$n - 3$
Total	$\sum(z - \bar{z})^2$	$n - 1$

The non-mathematician should note that the variance is the sum of squares per degree of freedom, and that the value of F obtained is the quotient of the variance due to the regression function by that due to the residual. This value is then compared with the 5% point of F for the appropriate numbers of degrees of freedom.

In Table 5 of the Appendix an analysis of the reductions in the sum of squares brought about by the use of the linear and quadratic expressions is made for the data of Table 4. The first part deals with the reduction due to the linear expression, the second part considers the further reduction brought about by the product term, and the third part treats the quadratic expression as a whole. The con-

tribution of the linear expression is seen to be barely significant at the 5% level, and the additional contribution made by the product term is not statistically significant. The reduction due to the quadratic expression as a whole falls slightly short of the 5% level.

13. An example involving the second filial generation.

The example of Section 11 does not actually make use of the main result of the present paper, as only single crosses are involved and these can be dealt with through the direct application of the previously known Theorem 10.1. In this section an example involving all possible crosses obtainable within two generations from four inbred parent lines is treated in considerable detail. Results obtained by the methods of the present paper are compared with those previously available, with particular emphasis on the predictions of the yields of those crosses which result in stocks exhibiting 100% heterozygosis. The original unpublished data due to G. H. Stringfield appear in Table 6 of the Appendix showing the yields of the four inbred lines, the six resulting single crosses, and the 45 additional crosses obtainable from these ten stocks in the next generation.

The data in Table 7 of the Appendix are used in conjunction with the normal equations appearing in Section 10 to obtain the expression $z_2 = 0.50020 x - 0.0097238 y + 59.815$ for expected yield of the single crosses when the same set of non-heritable factors is present. It is thus clear from the expression for $F_{1:12}$ preceding Theorem 10.1 that $k = 1.525$, $T = -78.41$, and $Nd = 78.43$. These values lead to the representations given for the inbred stocks in Table 8 of the Appendix, since $P_1 = 2u_1Nd + T$ where u_1 is the proportion of AA pairs among

those affecting yield in the inbred strain. The representations appearing for the single, three-way, and double crosses then follow from the multiplication table and those for the F_2 crosses from Theorem 7.1. It is evident from Section 8 that a modification of the general method is essential in order to estimate the yields of the less important intermediate cases, and the method used in obtaining estimates for the back-crosses and those of the AB.AC type is described in the next paragraph.

Once again it is desirable to examine the contribution made by the product term to the accuracy of estimation. The data of Table 7 of the Appendix together with the normal equations of Section 12 lead to $z_1 = 0.08083 x + 78.174$ as the best linear function of x for predicting yield of single crosses. These estimates appear in Table 9 of the Appendix. The theoretical performances of the double crosses are obtained in the manner described by Kiesselbach (22) as presently accepted, using the mean performance of the four single crosses resulting from a combination of each of the lines of one single cross parent with each of the lines in the other parent. Similarly the estimated yields of the three-way crosses in the second column of Table 9 are each obtained as the mean of two single crosses, and the extension of this technique to the F_2 crosses leads to agreement with the result of Theorem 7.1. Since this is the case, it seems reasonable to apply this same method to the prediction of the back-crosses and those of the AB.AC type for both the linear data and that involving the product term as well. This clearly leads to the required 50% and 75% levels of heterozygosity, respectively. The complete set of observed and predicted yields for all 51 crosses appears in Table 9.

The analysis of the sums of squares which is based on Table 9 appears in Table 10 of the Appendix. Under each heading there are three parts to the analysis, and these are comparable to the parts of Table 5 which are described in Section 12. It should be noted that under each heading other than back-crosses the number of degrees of freedom for estimating the total variance is $N-1$, where N is the number of crosses involved. This is due to the fact that each mean was computed from the data. The discrepancy in the case of back-crosses is discussed in the last paragraph of this section.

At this point it becomes necessary to consider a second specific statistical test. Rider (27) gives a complete treatment of the problem of testing the significance of the difference between two means. The present paper is concerned only with the case where the samples are of the same size, as it deals with the means of corresponding sets of observed and computed values. This results in the consideration of

$$t = \frac{|\bar{z}_c - \bar{z}_o|}{\sqrt{(N-1)/(s_c^2 + s_o^2)}},$$

in which N is the number of pairs of values involved and s_c^2 and s_o^2 are the variances of z_c and z_o respectively. This t is distributed in Student's distribution with $2(N-1)$ degrees of freedom, one degree having been used up in calculating $\bar{z}_c$ and another in calculating $\bar{z}_o$. The value of t obtained is then compared with the 5% value of t for the appropriate number of degrees of freedom.

It should be noted that for those groups of crosses other than the single crosses, to which functions were fitted directly, there is no necessity for the mean of the estimated values to coincide with the mean of the observed yields. For this reason the significance of the difference between the two means is tested in Table 11 of the Appendix

under those headings for which differences were found. The significance of the difference at the 5% level in the case of the back-crosses is responsible for the reduction by one of the degrees of freedom throughout that portion of the analysis carried out in Table 10 which involves the back-crosses, as the deviations of the observed and computed values were each taken about their own means.

14. Conclusion.

Estimation is useful only if it is reasonable to expect that the estimates obtained are sufficiently accurate to be worthwhile. It appears from Table 10 that on the basis of the data treated in the present paper the contribution of the product term in reducing the sum of squares is in general not significant. However, it also appears that the presently available methods are in general not significantly superior to chance in forming estimates. In the cases of those crosses which demonstrate 100% heterozygosity it is apparent that the incorporation of the product term brings about reductions in the sums of squares which, while not statistically significant, seem to indicate a trend when considered together. These are the crosses with which corn breeders and other biologists are concerned, and the fact that any additional reduction can be made is of interest since the reductions made by present methods over chance are in general not statistically significant although they seem to be considered as sufficient to justify the use of the methods.

It should be pointed out that the actual values of the estimates also depend upon the particular non-heritable factors in operation. Thus it is evident that numerical results obtainable as a consequence

of the methods of the present paper are valid only with the same environmental conditions as those under which the P_1 and $F_{1;ij}$ values were obtained. Haldane (15,16) discusses cases in which interactions of heredity and environment occur, and in the book treats the four possible interactions involving two races in two environments with considerable detail. However, the Haldane examples deal with environmental situations which are radically different, and it is clear that field crosses will be made under a set of non-heritable conditions comparable to those expected to prevail in the situations to which the resulting data are to be applied. It is to be expected that information as to the relative merits of various crosses will be applicable in an environment essentially similar to that in which the crosses are made. For this reason it is felt that the present discussion will make it possible to predict more accurately in advance which stock pedigrees are to give the best results and thus effect a further reduction beyond that made possible by present methods in the number of actual crosses necessary to obtain this information.

The influence of non-heritable factors on observable results cannot be overemphasized. The rather small accomplishments of both the present and proposed methods of prediction as compared with what remains to be accomplished indicate considerable room for further improvement. It is possible that such improvement might be brought about by more effective control of the non-heritable factors, a modification of the underlying assumptions involved in the idea of a polygenic system, or a combination of these two things. The relatively poor results obtained through the use of a method of prediction based firmly on the presently accepted notion of a polygenic system would

seem to indicate these as logical directions for a next step in the process of improving estimation of quantitative characteristics.

APPENDIX

Table 1: Yields in bushels per acre of six inbred parent stocks and the fifteen hybrids resulting from all possible single crosses among them. (Data of Stringfield; Zea mays)

<u>Stocks</u>	<u>4 - 8</u>	<u>90</u>	<u>Hy</u>	<u>02</u>	<u>WF9</u>	<u>51</u>
4 - 8	13.6	76.7	96.3	91.0	100.7	106.1
90		28.2	81.4	94.2	97.9	86.4
Hy			29.8	108.9	109.8	94.7
02				46.1	104.0	100.8
WF9					51.4	103.4
51						55.3

Table 2: Values obtained from the data of Table 1 to be used in determining the coefficients in the expression for expected yield appearing in Theorem 10.1.

$n = 15$	$\sum xy = 1,720,935.4$
$\sum x = 1122.0$	$\sum xz = 110,035.7$
$\sum y = 20,330.9$	$\sum yz = 2,014,353.8$
$\sum z = 1452.3$	$\sum y^2 = 35,830,162.8$
$\sum x^2 = 89,129.3$	

Table 3: Estimated values of u_1 for the six inbred parent stocks of Table 1.

4 - 8	0.436	02	0.680
90	0.546	WF9	0.720
Hy	0.558	51	0.750

Table 4: Comparison of the predicted and actual yields in bushels per acre for the fifteen hybrids of Table 1.

Cross	Observed	Estimated	Estimated
	$\bar{z}_0$	$\bar{z}_1$	$\bar{z}_2$
4-8 x 90	76.7	87.92	83.72
4-8 x Hy	96.3	88.35	84.72
4-8 x 02	91.0	92.75	94.83
4-8 x WF9	100.7	94.18	98.12
4-8 x 51	106.1	95.23	100.54
90 x Hy	81.4	92.29	90.44
90 x 02	94.2	96.69	97.19
90 x WF9	97.9	98.11	99.39
90 x 51	86.4	99.17	101.01
Hy x 02	108.9	97.12	97.45
Hy x WF9	109.8	98.55	99.53
Hy x 51	94.7	99.60	101.06
02 x WF9	104.0	102.94	100.96
02 x 51	100.8	104.00	101.59
WF9 x 51	103.4	105.42	101.76

Table 5: Analysis of the sums of squares obtained from Table 4 arranged in accordance with the outline given in Section 12.

<u>Sums of Squares</u>	<u>Degrees of Freedom</u>	<u>Variance</u>	<u>F</u>	<u>5% F</u>
378.54	1	378.54	5.20	4.67
<u>945.51</u>	<u>13</u>	72.73		
1324.05	14			
113.98	1	113.98	1.64	4.75
<u>831.53</u>	<u>12</u>	69.29		
945.51	13			
492.52	2	246.26	3.55	3.88
<u>831.53</u>	<u>12</u>			
1324.05	14			

Table 6: Yields in bushels per acre of four inbred parent stocks and all possible crosses obtainable from them within two generations. (Data of Stringfield; Zea mays)

A: 28 B: 51A C: Hy D: 40B

<u>Pedigree</u>	<u>Yield</u>	<u>Pedigree</u>	<u>Yield</u>
Inbred Parents		Single	
A	39.5	AB	86.0
B	53.3	AC	85.2
C	27.2	AD	86.8
D	46.0	BC	88.1
		BD	84.6
		CD	78.6
F₂		Three-Way	
AB.AB	62.9	AB.C	94.3
AC.AC	64.8	AB.D	72.7
AD.AD	61.7	AC.B	85.1
BC.BC	60.6	AC.D	83.9
BD.BD	61.8	AD.B	87.5
CD.CD	61.5	AD.C	82.9
		BC.A	86.9
Back		BC.D	76.4
AB.A	66.5	BD.A	90.4
AB.B	74.3	BD.C	82.3
AC.A	70.7	CD.A	92.5
AC.C	66.8	CD.B	83.5
AD.A	71.6		
AD.D	65.9	AB.AC Type	
BC.B	69.4	AB.AC	81.4
BC.C	60.5	AB.AD	76.9
BD.B	74.6	AC.AD	82.0
BD.D	65.9	AB.BC	76.9
CD.C	58.7	AB.BD	75.6
CD.D	69.9	BC.BD	77.4
		AC.BC	66.9
Double		AC.CD	74.8
AB.CD	85.2	BC.CD	74.8
AC.BD	82.1	AD.BD	69.7
AD.BC	78.4	AD.CD	76.2
		BD.CD	70.0

Table 7: Values obtained from data of Table 6 to be used in determining the coefficients in the expression for expected yield appearing in Theorem 10.1.

$n = 6$	$\sum xy = 874,149.720$
$\sum x = 498.0$	$\sum xz = 42,331.39$
$\sum y = 10,149.51$	$\sum yz = 863,805.036$
$\sum x^2 = 42,069.96$	$\sum y^2 = 18,566,951.7201$
$\sum z = 509.3$	

Table 8: Stock representations associated with certain of the pedigrees appearing in Table 6.

Pedigree	Proportions of loci which are:		
	AA	Aa	aa
A	.758613	.000000	.248317
B	.839659	.000000	.160341
C	.673270	.000000	.326730
D	.793121	.000000	.206879
AB	.63115740	.32902720	.03981540
AC	.50608561	.41278178	.08113261
AD	.59617557	.35245286	.05137157
BC	.56531721	.38229458	.05238821
BD	.66595119	.30087762	.03317119
CD	.53398458	.39842184	.06759358
AB.AB	.71341420	.16451360	.12207220
AC.AC	.60928105	.20639090	.18432805
AD.AD	.68428878	.17622644	.13948478
BC.BC	.66089085	.19114730	.14796185
BD.BD	.74117060	.15043880	.10839060
CD.CD	.63359004	.19921092	.16719904
AB.C	.535702	.397538	.066760
AB.D	.631064	.326665	.042271
AC.B	.598237	.355661	.046102
AC.D	.565080	.375437	.059483
AD.B	.648555	.314952	.036493
AD.C	.520035	.405602	.074363
BC.A	.568622	.370904	.060474
BC.D	.599968	.349650	.050382
BD.A	.613667	.340740	.045593
BD.C	.549651	.390358	.059991
CD.A	.551131	.382617	.066252
CD.B	.615634	.341586	.042780
AB.CD	.583382	.362102	.054516
AC.BD	.581659	.365549	.052792
AD.BC	.584295	.360277	.055428

Table 9: Comparison of the predicted and actual yields in bushels per acre for the heterozygous stocks of Table 6.

Pedigree	$\bar{x}_0$	$\bar{x}_1$	$\bar{x}_2$
AB.AB	62.9	66.04	66.08
AC.AC	64.8	58.46	58.04
AD.AD	61.7	63.92	63.83
BC.BC	60.6	62.46	63.12
ED.ED	61.8	67.92	67.64
CD.CD	61.5	60.34	60.43
AB.A	66.5	62.59	62.63
AB.B	74.3	69.49	69.53
AC.A	70.7	61.54	61.12
AC.C	66.8	55.38	54.96
AD.A	71.6	62.29	62.20
AD.D	65.9	65.54	65.46
BC.B	69.4	68.99	69.64
BC.C	60.5	55.94	56.59
ED.B	74.6	69.75	69.47
ED.D	65.9	66.10	65.82
CD.C	58.7	55.64	55.73
CD.D	69.9	65.04	65.13
AB.AC	81.4	73.36	73.49
AB.AD	76.9	74.12	73.95
AC.AD	82.0	73.06	72.85
AB.BC	76.9	76.81	76.94
AB.ED	75.6	77.56	77.40
BC.ED	77.4	77.07	77.30
AC.BC	66.9	70.28	70.42
AC.CD	74.8	69.98	69.78
BC.CD	74.8	70.54	70.77
AD.ED	69.7	75.74	75.58
AD.CD	76.2	74.68	74.48
ED.CD	70.0	75.24	75.47
AB	86.0	85.68	85.76
AC	85.2	83.57	82.73
AD	86.8	85.08	84.91
BC	88.1	84.68	85.98
ED	84.6	86.20	85.64
CD	78.6	84.09	84.26
AB.C	94.3	84.12	84.36
AB.D	72.7	85.64	85.28
AC.B	85.1	85.18	85.87
AC.D	83.9	84.58	84.59
AD.B	87.5	85.94	85.70
AD.C	82.9	83.83	83.50
BC.A	86.9	84.62	84.25
BC.D	76.4	85.14	84.95
ED.A	90.4	85.38	85.34

Table 9, continued:

<u>Pedigree</u>	<u>z_0</u>	<u>z_1</u>	<u>z_2</u>
BD.C	82.3	84.38	85.12
CD.A	92.5	84.32	83.82
CD.B	83.5	85.44	85.81
AB.CD	85.2	84.88	84.82
AC.BD	82.1	84.88	85.23
AD.BC	78.4	84.88	84.60

Table 10: Analysis of the sums of squares obtained from Table 9, arranged under each heading in accordance with the outline given in Section 12.

<u>Sums of Squares</u>	<u>Degrees of Freedom</u>	<u>Variance</u>	<u>F</u>	<u>5% F</u>
Single crosses:				
4.82	1	4.82	-	7.71
<u>50.11</u>	<u>4</u>	12.53		
54.93	5			
2.77	1	2.77	-	10.13
<u>47.34</u>	<u>3</u>	15.78		
50.11	4			
7.59	2	3.80	-	9.55
<u>47.34</u>	<u>3</u>	15.78		
54.93	5			
Three-Way crosses:				
No reduction				
.54	1	.54	-	5.12
<u>456.09</u>	<u>9</u>	50.68		
456.63	10			

No reduction

Table 10, continued:

<u>Sums of Squares</u>	<u>Degrees of Freedom</u>	<u>Variance</u>	<u>F</u>	<u>5% F</u>
Double crosses:				
.00	1	.00	-	161
<u>49.82</u>	<u>1</u>	<u>49.82</u>		
49.82	2			
1.44	1	1.44	-	-
<u>48.38</u>	<u>0</u>	-		
49.82	1			
1.44	2	.72	-	-
<u>48.38</u>	<u>0</u>	-		
49.82	2			
Selfing:				
	No reduction			
	No reduction			
	No reduction			
Back-Crosses:				
113.35	1	113.35	2.45	5.12
<u>417.06</u>	<u>9</u>	<u>46.34</u>		
530.41	10			
	No reduction			
98.92	2	49.46	-	4.46
<u>431.49</u>	<u>8</u>	<u>53.94</u>		
530.41	10			
Crosses of the type AB.AC:				
	No reduction			
	No reduction			
	No reduction			

Table 10, continued:

<u>Sums of Squares</u>	<u>Degrees of Freedom</u>	<u>Variance</u>	<u>F</u>	<u>5% F</u>
All 100% heterozygous crosses:				
No reduction				
4.76	1	4.76	-	4.41
<u>551.81</u>	<u>18</u>	30.66		
556.57	19			
No reduction				
All crosses:				
3166.21	1	3166.21	115.26	4.04
<u>1346.18</u>	<u>49</u>	27.47		
4512.39	50			
No reduction				
3147.60	2	1573.80	55.36	3.19
<u>1364.79</u>	<u>48</u>	28.43		
4512.39	50			

Table 11: Analysis of the differences of observed and computed mean values of yield for certain types of crosses, obtained from the data of Table 9 and treated in accordance with the discussion given in Section 13.

	<u>$\bar{x}_c$</u>	<u>$\bar{x}_o$</u>	<u>s_c^2</u>	<u>s_o^2</u>	<u>N</u>	<u>DF</u>	<u>t</u>	<u>5% t</u>
Double	84.88	81.9	0.07	7.73	3	4	1.51	2.78
Selfing	63.19	62.2	10.45	1.78	6	10	.62	2.23
Back	63.19	67.9	25.80	22.02	12	22	2.26	2.07
AB.AC	74.04	75.2	6.55	18.58	12	22	.78	2.07
All	74.67	75.8	98.67	87.28	51	100	.57	1.98

BIBLIOGRAPHY

- 1) Anderson, A. G., 1947: Evidences of an overdominance factor in hybrid corn crosses, thesis, Univ. of Florida
- 2) Birkhoff, G. and Mac Lane, S., 1941: A survey of modern algebra, Macmillan, New York
- 3) Brauer, R. D., 1950: Structure of rings, hectographed notes, Univ. of Michigan
- 4) Collins, G. N., 1921: Dominance and the vigor of first generation hybrids, Amer. Nat., 55: 116-133
- 5) Darlington, C. D. and Mather, K., 1949: The elements of genetics, Allen and Unwin, London
- 6) Dextator, C. W. and Johnson, I. J., 1936: Prediction of double cross yields in corn, Jour. Amer. Soc. Agron., 28: 460-462
- 7) East, E. M., 1936: Heterosis, Genetics, 21: 375-397
- 8) Etherington, I. M. H., 1939: On non-associative combinations, Proc. Royal Soc. Edinburgh, 59: 153-162
- 9) Etherington, I. M. H., 1939: Genetic algebras, Proc. Royal Soc. Edinburgh, 59: 242-258
- 10) Etherington, I. M. H., 1941: Non-associative algebra and the symbolism of genetics, Proc. Royal Soc. Edinburgh, 61: 24-42
- 11) Feller, W., 1950: An introduction to probability theory and its applications, Wiley, New York
- 12) Geiringer, H., 1949: On some mathematical problems arising in the development of mendelian genetics, Jour. Amer. Stat. Assoc., 44: 526-547
- 13) Glivenko, V., 1936: Algèbre mendélienne, C. R. (Doklady) de l'Académie des Sciences de l'URSS, 4: 385-386
- 14) Goulden, C. H., 1939: Methods of statistical analysis, Wiley, New York
- 15) Haldane, J. B. S., 1938: Heredity and politics, Norton, New York
- 16) Haldane, J. B. S., 1946: The interaction of nature and nurture, Ann. Eugen., 13: 197-205

- 17) Hogben, L., 1946: An introduction to mathematical genetics, Norton, New York
- 18) Hull, F. H., 1946: Regression analyses of yields of hybrid corn and inbred parent lines, Maize Genetics News Letter
- 19) Hull, F. H., 1947: Theoretical regression of F_N on homozygous parents with additive or complementary gene action, Genetics, 32: 90-91
- 20) Jenkins, M. T., 1934: Methods of estimating the performance of double crosses in corn, Jour. Amer. Soc. Agron., 26: 199-204
- 21) Jenkins, M. T., 1936: Corn improvement, U. S. Yearbook of Agriculture, 455-522
- 22) Kiesselbach, T. A., 1951: A half-century of corn research, Am. Scientist, 39: 629-655
- 23) Malécot, G., 1948: Les mathématiques de l'hérédité, Masson, Paris
- 24) Meyers, M. T., 1930: Determining the date of silking in experiments with corn, Jour. Amer. Soc. Agron., 22: 280-283
- 25) Neal, N. P., 1935: The decrease in yielding capacity in advanced generations of hybrid corn, Jour. Amer. Soc. Agron., 8: 666-670
- 26) Neyman, J., 1950: First course in probability and statistics, Holt, New York
- 27) Rider, P. R., 1939: An introduction to modern statistical methods, Wiley, New York
- 28) Schafer, R. D., 1949: Structure of genetic algebras, Amer. Jour. Math., 71: 121-135
- 29) Serebrovskij, A., 1934: On the properties of the mendelian equations, C. R. (Doklady) de l' Académie des Sciences de l'URSS, 2:
- 30) Stringfield, G. H., 1950: Heterozygosis and hybrid vigor in maize, Agron. Jour., 42: 145-152
- 31) Stringfield, G. H., 1950: Letter to the author, dated Dec. 30
- 32) United States Department of Agriculture, 1949: Report of the 1948 cooperative uniform comparisons of corn hybrids, mimeographed report, Beltsville, Md.

- 33) Van der Waerden, B. L., 1943: Moderne algebra, Ungar, New York
(also Springer, Berlin)
- 34) Weatherwax, P., 1950: The history of corn, Sci. Monthly, 71:
50-60
- 35) Whaley, W. G., 1944: Heterosis, Bot. Review, 10: 461-498
- 36) Wright, S., 1942: Statistical genetics and evolution, Bull.
Am. Math. Soc., 48: 223-246