ON BIOMETRICAL GENETICS*

KENNETH MATHER
University of Southampton

1. ORIGIN OF BIOMETRICAL GENETICS

MANY things trace back to R. A. Fisher as their innovator in genetics as well as statistics. It is not for me to assess his great contribution to statistics, but among them must surely be numbered his conception of the science of experimentation itself. His notion of experimental design, with its emphasis on valid estimates of error, factorial arrangements and the use of confounding and other devices to increase precision, indeed the allied notion of the amount of information contained in a body of data as a means of assessing both the worth of a particular type of experiment and of estimates derived from it, are fundamental to modern biological practice. They are now so much a commonplace that it is difficult to imagine the situation before Fisher, and biologists of all kinds, not least among them geneticists, owe him an immense debt for this. Coupled with the principles and techniques of experimental design is the use of the analysis of variance as a powerful and indeed essential tool in the reduction and interpretation of the data the experiments yield. This idea of analysing variation goes back to his early work in genetics.

Among geneticists Fisher will again be remembered for many things; for this theory of the evolution of dominance, which has been persistently disputed in detail but never successfully challenged in principle; for his fundamental studies of the theory of natural selection upon which so much modern work is based; for his contribution to our understanding of the genetical determination of blood groups, especially his postulate of a trinitarian structure for the Rh gene; for his analysis of inbreeding; for his theory of polysomic inheritance; and, at any rate by those who worked with him, for the precision and ingenuity of his genetical experiments. This last was particularly well exemplified by the experiments which demonstrated the tetrasomic inheritance of the mid-style gene in *Lythrum*, in which incidentally not only was no illegitimate pollination ever involved but no artificial pollination of any kind was ever necessary. And not least he will be remembered for his first contribution to genetics—"The correlation between relatives in the supposition of Mendelian inheritance" in 1918—on a subject which he never took up experimentally, and to which he later returned only once (with Immer and Tedin in 1932), yet which virtually settled the controversy about the genetic determination of continuous variation and on which Biometrical Genetics is chiefly founded. Others had seen the problem; Fisher showed us how to solve it. Others had seen the possible reconciliation of biometry and Mendelism; Fisher brought them together and gave us a methodology which fundamentally we still

* Being the Fourth Fisher Memorial Lecture, delivered in London on 18th December 1970.

use. And this methodology brought in an analysis of variation for the first time.

The first decade of this century brought the first great generalisation of genetics: that Mendelian inheritance is ubiquitous throughout the living kingdom and that it is virtually exclusive. This generalisation is now so much taken for granted as a commonplace that it is easy to underestimate the stature of the men who established it and of whom I believe William Bateson (see especially 1909) to be the leader. Yet it cannot have been easy, involving as it did the interpretation of strange and at first sight non-Mendelian ratios in terms of genic interaction and lethal combinations as well as providing an at least formally satisfactory account of the first case of sex-linkage, which was observed by Doncaster (1908) in the moth *Abraxus grossulariata*. But all this applied to discontinuous variation of the kind which Mendel had used and whose significance in the wild Bateson had preached (some would say, over-preached) in the 1890's. There still remained continuous variation, the variation of small differences and smooth gradations that Darwin had seen as the material of adaptive change and that Galton had shown to be heritable in man. A flourishing school of biometry, led by Karl Pearson, had grown up to exploit Galton's approach and methods in the study of heredity. Biometricians and Mendelians appeared to agree on only one thing: that the determination of discontinuous and continuous variation could not follow the same principles.

Yule and others had pointed out that despite the apparent contrast with the sharp differences from which Mendelian inheritance was inferred, continuous variation need not be basically dissimilar in its mechanism provided that one assumed it to be dependent in expression on the simultaneous action of many genes whose effects supplemented one another. Pearson disputed even this, on the grounds that Mendelian factors, one of whose essential properties he took to be complete dominance, could not give a parent/offspring correlation greater than $\frac{1}{2}$ (which is, of course, true where dominance of one gene over its allele is complete) whereas correlations as high as a $\frac{3}{4}$ were commonly observed. Fisher approached the problem in a different and characteristic way by devising both the genetical framework and the statistical means of using the correlations observed among relatives to partition the variation not merely into its heritable and non-heritable fractions, but also of going further and dividing the heritable into further fractions ascribable to dominance and non-allelic interaction as well as simple additive effects, while making due allowance for the correlation observed between spouses. Not only were the correlations successfully interpreted in terms of established genetic phenomena; but the excess of the sib-correlation over that between parent and offspring was shown virtually to demand dominance, albeit incomplete dominance. After this there could be little doubt that continuous variation could also be brought into the Mendelian generalisation.

2. MENDELIAN GENETICS AND BIOMETRICAL GENETICS

To some of us the origin of biometrical genetics has a historical interest in its own right. It is also essential if we are to understand what biometrical genetics was designed to achieve, in what its strength lies and what its limitations are. Fisher's problem was that of designing a method of analysis

which could be applied in circumstances that precluded use of the already established Mendelian method of recording and interpreting the ratios obtained when individuals were assigned to clearly distinguishable classes of established genetical significance, yet which allowed of clear and confident interpretation in terms of accepted genetic phenomena. He solved it by translating these genetical phenomena into biometrical terms, into terms of variances, covariances, and correlations (to which we would now add means) and then using appropriate statistical devices, which could be introduced where they did not already exist, to manipulate the statistical quantities and interpret them in their genetical relations. This is still the essential aim of biometrical genetics.

The techniques of biometrical genetics are sometimes represented, explicitly or implicitly, as rivals of the Mendelian method, and disputed, denied, denounced or denigrated accordingly. This is of course misrepresentation, for the two are not antagonistic but complementary in the situations to which they can be applied and in the coverage of variation that they offer. They explore the same genetic systems in different ways and lead to conclusions of different but complementary contents. The very phenomena with which they must both be concerned may require different representation and concepts which are readily utilised in the one emerge less clearly when applied in the other, with a consequent deepening of understanding if the reasons for the blurring are appreciated rather than the analysis itself denounced.

The biometrical method is derivative in the sense that it depends on Mendelian genetics for the phenomena in terms of which it attempts to interpret the variation it sets out to measure and analyse. No basic genetical phenomenon, except perhaps certain of the relations between heritable and non-heritable variation, have been discovered by it: all have been recognised first by Mendelian analysis. Nor indeed, apart from segregation with its implication of particulate inheritance, could any of them—linkage, crossing-over, the tie with the chromosomes, dominance, interaction in its many forms and the rest—have been recognised easily and with confidence, if at all, by biometrical analysis. But given the concepts, they can be applied over a much wider range of variation and in a much wider range of circumstances by the biometrical method, for all variation is always open to such attack while the success of Mendelian analysis depends on the choice of suitable differences for investigation, as Mendel himself clearly realised. As he says of his peas: "Some of the characters noted do not permit of a sharp and certain separation, since the difference is of a 'more or less' nature, which is often difficult to define. Such characters could not be utilised for the separate experiments; these could only be applied to characters which stand out clearly and definitely in the plants." (Translation in Bateson, 1909.)

The Mendelian approach depends on the successful recognition of clearly distinguishable phenotypic classes from which the relevant genetical constitution can be inferred. It is at its most powerful when there is a one-to-one correspondence of phenotype and genotype, though some ambiguity of the relationship, as when complete dominance results in heterozygote and one homozygote having the same phenotype, is acceptable. In such cases, of course, it is implicit that an appropriate test mating will effect the separation and so clear up the ambiguity should this be necessary.

In this way individual genes can be followed and their relations with other genes in both transmission and action investigated, so leading through the study of linkage to the tracing of chromosome behaviour on the one hand and through the study of interactions to the elucidation of biochemical relations on the other. Their mutations can also be followed with consequent information on internal genic organisation.

The approach can be refined and made more powerful by the use of special breeding techniques, test crosses and special stocks, and by the more careful definition of characters and the control of the environment to clarify the situation by the removal of phenotypic obscurities and overlaps. All of these methods can be and have been used successfully and most valuably in the analysis and interpretation of phenotypic variation which at first sight appeared to be so confused as to offer no prospect of success. But they require stocks which may not be available in the species under investigation, control of the environment which may not be practicable, or time which the investigator cannot afford with longer-lived species. And even when successful they can result in a fractionation of the genetical material into lines which must be put together again in ways which reintroduce the original obscurity and overlap if the purpose for which the analysis was undertaken is to be achieved. Furthermore, there always remains a residuum of variation which cannot be handled by these means. For many genetical purposes this residuum can be ignored—indeed, as Mendel himself appreciated, often must be ignored; but for other purposes, especially in applied genetics, all variation in a character is prospectively of interest and that part of it which is the most troublesome to analyse and manipulate can be the most valuable.

The biometrical approach is from a different direction starting with the character rather than the individual determinant. It makes no requirement that the determinants be traceable individually in either transmission or action. It seeks to measure all the variation in a character and then, by comparing individuals and families of varying relationships, to partition the differences observed into fractions ascribable to the various genetical (or for that matter non-genetical) phenomena which are to be taken into account, whose consequences can thus be measured and whose significance can be tested by comparison with the variation remaining unaccounted for.

Its applicability is not confined in principle to continuous variation: it could be used even where variation is discontinuous, arising from the segregation of individually traceable genes. Applied properly in such a case it would come up with the right answers, though it would be more tedious and cumbersome to use than Mendelian analysis which is thus to be preferred for that part of the variation it can handle.

Biometrical analysis characteristically comes into its own where the variation is continuous or quasi-continuous (to use Grüneberg's term) and where there is no ready means to hand of imposing genetically meaningful discontinuities on it. In so far as it is genetical in its determination, such variation commonly arises as the manifestation of polygenic systems; but the use of the techniques of biometrical genetics is neither confined to nor required by polygenic variation. On the one hand the segregation arising from a simple genetical system may be obscured by, for example, wide non-heritable variation and so require a biometrical approach, at any

rate initially; and, on the other, a polygenic system may be capable of at least partial analysis by Mendelian techniques where circumstances are favourable and the necessary special stocks and methods are available. Continuous variation will commonly imply polygenic determination and commonly require biometrical analysis; but to assume, as is sometimes done, that the three are so indissolubly bound together that each implies the others is naïvely to confuse the variation that we observe with the analytical techniques that we can apply and the causation that we may infer.

The two techniques of analysis, Mendelian and biometrical, are not mutually exclusive, but may be combinable with advantage as we can see in two ways. In the first place a given population, or set of families, may show both discontinuous and continuous variation simultaneously in respect of a character. Thus, for example, human stature affords one of the classical examples of continuous variation for which the biometrical approach was first developed. Yet human populations regularly contain individuals readily distinguishable from their fellows in respect of stature in ways making it clear that they do so because of a major genic difference establishable by the Mendelian method. Achondroplastic dwarfs are an example of such a difference in respect of human stature, and we must expect differences of this kind to be found in respect of virtually any character in any species, for our experience suggests that few if any characters are not subject to both major genic and polygenic effects.

In the second place, even where the variation in the material under analysis is wholly continuous, the analysis can often be furthered by the introduction of marker genes, where these are available, through whose use a polygenic system can be at least partly fractionated by tracing the degree of association of the continuous variation with the pieces of chromosome recognisable by virtue of the markers they carry. The use of such markers early provided evidence that at least some of the determinants of continuous variation must be carried on the chromosomes, and later, through the now highly developed technique of chromosome assay, that where the test could be applied all, or virtually all, of these determinants must be nuclear (see Mather and Harrison, 1949). In recent years, Thoday (1961) and his colleagues have gone further, and using the special stocks available in *Drosophila melanogaster* have shown how, under favourable circumstances, components of a polygenic system can be related to specific loci in the chromosomes. Again drawing on the unrivalled resources of *Drosophila melanogaster*, Cooke and Mather (1962) have been able to compare the conclusions reached by purely biometrical analysis of continuous variation with expectation synthesised from measurements obtained by using marker genes to follow chromosomes in the classical way. The results agreed satisfactorily within the limitations of comparability to be expected of the two techniques, and in doing so strengthened to this extent our confidence both in the genetical interpretation of the results of biometrical analysis and in the justification of making the analysis and interpreting its results in terms of phenomena that we recognise from classical genetical studies, yet could hardly derive from biometrical considerations alone. In short, we can feel confident that we are following the same basic mechanisms of heredity by techniques each appropriate to its own circumstances and yielding comparable information of its own particular kind.

3. THE NATURE OF BIOMETRICAL GENETICS

When individual genes are identified, their properties of action and their interactions one with another can be followed in detail. In biometrical genetics genes are not, and in general cannot be, identified, and the analysis must therefore be in terms of phenomena, additive action, dominance, non-allelic interaction, linkage, non-random mating and so on, the contribution of any given gene, or set of genes, to which is inseparable from the bulk effect. Thus even where we have an idea of the number of genes or other units contributing to the gross effect we can at best arrive at no more than an estimate of the average contribution. The information we can gain is therefore limited and the absence of an overall component relatable to a particular phenomenon is no guarantee that the individual genes do not display it. With dominance, for example, a pair of gene differences may fail to show an overall effect as well because their individual effects are equal but opposite in sign, say 1 and -1 on a suitable scale, as because both are 0: the sum, and average, of 1 and -1 are the same as those of 0 and 0. Equally if the two contributions are, say, 1 and $-\frac{1}{2}$ the overall effect is the same as where they are $\frac{1}{2}$ and 0 or both $\frac{1}{4}$, though here the overall positive contribution shows that dominance is at least preponderantly in that direction. The failure to appreciate this prospective balancing effect can lead and has led to misconception, such as the assumption that heterosis implies over-dominance at individual loci, or that the coincidence of the expression of a character in an F_1 with the average expression in the two parents implies an absence of dominance. In both cases the observations could be interpreted in this way; but in neither case is the evidence compelling or even good.

Such examples could be multiplied endlessly, but only one further will be given. Some years ago a situation was described in which the expression of the character in F_1 fell outside the range set by the two parents but where it was possible to extract from F_2 and later generations a line breeding true for an expression more extreme even than that of F_1 . The question was then asked as to whether the F_1 should be described as showing heterosis, the comparison with the parents being held to suggest that it should, but comparison with the derived line that it should not. The interpretation is, of course, simple. The range delimited by the parents must have been restricted by dispersion and partial balancing of $+$ and $-$ alleles in them, the dispersion being removed or at least reduced in the true-breeding line later derived from their cross. The average dominance (or h in Mather's notation) taking sign into account, was thus less than the average additive effect (or d) of the genes. The question about heterosis thus becomes meaningless or at least irrelevant in this context.

In Mendelian analysis as many classes can be recognised, at any rate in principle, in a segregating family such as F_2 as there are genic combinations, and the comparisons among these classes and the frequencies of their occurrence provide the means for a detailed description of the genetic situation. In biometrical genetics, on the other hand, one family can give of itself only two pieces of information, its mean and its variance, supplemented of course by its covariances with other families or groups of families where these are appropriate. All our conclusions must be drawn therefore from the comparisons available among these means, variances

and covariances of different types of family. Unless therefore the experimental data are very extensive, springing from a complex range of different types of family, the number of useful comparisons will be limited and the number of phenomena capable of being investigated correspondingly restricted. This is, however, a restriction of practice rather than of principle, for any phenomenon displayed by nuclear genes or indeed by non-nuclear determinants is capable of being expressed in biometrical terms (see Mather and Jinks, 1971) and therefore of being investigated in appropriately designed biometrical experiments: the limitation is that only a restricted number can be investigated at one time if an impossible experimental complexity is to be avoided. This need not, however, vitiate in any way the conclusions to be drawn given that, in accordance with good statistical practice, provision is made for a valid estimate of error to which the phenomena, nuclear, non-nuclear or environmental, not under investigation can make their due and proper contributions.

At the same time, just as any genetical phenomenon can be represented in biometrical terms and hence investigated biometrically, the representation and analysis can be extended to any type of genetical material. Though much of biometrical genetics has been developed from the study of crosses among true-breeding lines, in the same way as Mendel's principles were derived from studies of such crosses, it can be and indeed has been extended to randomly breeding and other types of populations, just as Mendel's principles have been. Indeed, it should be recalled that Fisher first developed his approach in 1918 for the study of human populations displaying not merely random mating but assortative mating too, and only in 1932 did he, in collaboration with Immer and Tedin, turn to apply his approach to crosses between true-breeding lines.

The contributions made by the various phenomena to the components of means on the one hand and second degree statistics, variances and covariances on the other take different forms and so supply potentially complementary information. The properties of each gene, or set of genes, must of course be represented in such a way as to facilitate summation over all genes or relevant sets of genes. This makes no problems in respect of the overall or additive effect of a gene pair, which can be represented by d , half the difference between the two homozygotes in expression of the character, or dominance which can be represented by h , the departure of the heterozygotes from the mid-point between the two homozygotes, it being clear that h takes sign according to the direction of the departure. Taking the mid-point as the natural origin the gene pair $A-a$ will contribute thus d_a or $-d_a$ to the departure from the origin of a homozygote, according to whether it is AA or aa , and h_a to the departure of the heterozygote; but it will contribute to variances terms in d_a^2 and h_a^2 . Then taking all relevant genes into account the expression for the mean of the character will include, in addition to the midparent; terms in $[d]$ and $[h]$ where the square brackets indicate summation with sign taken into account, so recognising that when genes at different loci pulling in opposite directions come together in homozygotes they will tend to balance out one another's effects, and in heterozygotes dominance in opposite directions at different loci will also tend to balance out. Variances and covariances will, however, characteristically include terms in $D = S(d^2)$ and $H = S(h^2)$, there now being no balancing effect due to sign.

Turning to non-allelic interaction, the situation becomes more complicated in two ways. First, the types of interaction, complementary, duplicate, etc., recognised in classical genetics cannot as such be directly represented or combined over different pair of genes. But all types of interaction between pairs of genes at two loci can be represented in terms of four parameters; viz. the interaction of d_a and d_b , which we denote as i ; the interaction of d_a and h_b , denoted as j_a ; the interaction of h_a and d_b , denoted as j_b ; and the interaction of h_a and h_b , denoted as l . All the classical types of interaction can therefore be defined in terms of d_a , d_b , h_a , h_b , i , j_a , j_b and l . Thus when all eight quantities are equal and of the same sign the interaction is complementary and with all equal but i , j 's and l of like sign to one another and opposite sign to the d 's and h 's it is duplicate (see Mather, 1967). Clearly strengths as well as types of interaction can be represented by the relations of these quantities and all digenic interactions, no matter how diverse in type or strength, can be combined in terms of i , j and l . Since these quantities can take sign, the expressions for means will include terms in $[i]$, $[j]$, and $[l]$ where interactions are present. Obviously trigenic and higher order interactions can be accommodated by an extension of this notation.

The second way in which the situation is complicated by interaction is that the interactions characteristically contribute to variances and covariances not just by the addition of terms in i^2 , j^2 and l^2 , though these may indeed appear, but by becoming confounded in characteristic ways with the d 's and h 's. Thus the variance of an F_2 characteristically includes terms not in d^2 and h^2 , but in $(d + \frac{1}{2}j)^2$ and $(h + \frac{1}{2}l)^2$. The same general types of term appear in F_3 , but with the coefficients of j and l reduced to $\frac{1}{4}$. Thus interaction expresses itself differently in first- and second-degree statistics: in the analysis of means it is recognised by components specifically related to it; but in variances and covariances it is detected by changes in the D and H components which are no longer dependent purely on items of the respective types d^2 and h^2 but include contributions from the interactions, with coefficients varying over the generations. Incidentally, this serves to emphasise the conditional nature of the values assigned to the primary parameters, d and h , when interactions are present; and indeed their values must be conditional for they will depend on the genetic background against which they are measured.

We can see even more clearly the difference in the effects of a phenomenon on first- and second-degree statistics when we turn to linkage. Of itself linkage has no effect whatsoever on the mean expression of families even where these are of segregating generations like F_2 , though it will distort the effects of interaction, if the interacting genes are linked. Thus the comparisons among means cannot generally reflect or be affected by the linkage relation of the genes except where the linked genes are also interacting with one another. Linkage does, however, affect variances and covariances by the introduction into the D and H components of terms in $d_a d_b$ and $h_a h_b$ respectively (where $A-a$ and $B-b$ are the linked loci) these terms having coefficients which depend on the linkage values of the genes. The coefficients of the terms in $d_a d_b$ and $h_a h_b$ vary not with the generations, as do the coefficients in D and H of interaction parameters, but with the rank of the statistic, i.e. with the number of rounds of meiosis which are reflected by the segregation which the variance or covariance in question is measuring. So linkage, like inter-

action, is detected by changes in D and H , this time however by changes with rank rather than generation. Changes with rank and generation are not always easy to separate because a variance of rank r cannot be obtained before generation r , though generation r will always be capable in principle of providing variances of rank $r-1$, $r-2$, ...1 also.

These ways of denoting, detecting and delimiting the effects of interaction and linkage are well known in biometrical genetics and could hardly have justified mention at such length here if they did not serve to emphasise the differences in structure of first- and second-degree statistics and hence of the information that can be gained from analyses. In principle, the analysis of variances and covariances is more informative than the analysis of means: comparison of means cannot reveal linkage whereas comparison of variances can, and $\sqrt{\frac{H}{D}}$ is a much more reliable measure of the average

dominance $\frac{h}{d}$ than is $\frac{[h]}{[d]}$ to mention but two illustrative examples. If it is to be efficient, however, the analysis of second-degree statistics involves much more elaborate and complex calculations than does that of means, and even when efficiency is achieved the sampling variances to which estimates of the quadratic quantities like D and H are subject are much greater than those of $[d]$, $[h]$, etc., derived from means. For this reason it has been argued, for example by Gilbert (1961), that analysis of second-degree statistics is a waste of time: only means should be considered in biometrical genetics, at any rate as it is put to use in applied genetics. Such a view is, of course, myopic in that it ignores the fact that second-degree statistics provide information unobtainable from means, and prospectively of greater genetical value, albeit of lesser statistical precision. The need is not for argument as to which type of analysis should be undertaken, for clearly both should be used; rather it is for even more investigation than has yet been given to it, of the ways in which experiments can be designed, both genetically and statistically, to improve the precision of the tests and estimates that second-degree statistics and only second-degree statistics can provide.

4. SCALES

On the whole, geneticists have paid little attention to the problems of the metric or scale on which a character is measured. This is not perhaps surprising since, with the discontinuous variation from which genes are individually inferred, scale matters relatively little: no reasonable change of scale will remove the discontinuity or remove overlap where this occurs. True, the degree and in extreme cases even the apparent direction of incomplete dominance can be altered by a change of scale, and the properties of incomplete interaction can also be modified; but this has seldom been of great importance in classical genetical situations and in any case neither complete dominance nor full interaction is affected.

With continuous variation and biometrical analysis, however, the metric used for representing the character becomes much more important, partly because of the metrical nature of the analysis into components of variation, with the emphasis it places on the measurement of partial manifestation of phenomena and the conditional nature of the phenomena it reveals,

partly because of the simplification of the analysis to which re-scaling can lead, and partly because of difficulties and apparent paradoxes of interpretation that can stem entirely from choice of scale. None of these problems is in essence unique to the study of continuous variation: all can also arise in principle in Mendelian analysis, but they seldom cause trouble in the latter, whereas by its very nature, biometrical analysis can hardly avoid them. And by forcing us to face them biometrical genetics can play its part in clarifying the concepts which it derives from Mendelian genetics and upon whose application its success depends.

In genetics it must be very rare for the metric we use in measuring the expression of a character to be based on biological considerations or to have any deep-seated biological significance. The metrics we use are taken because they are familiar, obvious, or happen to be convenient for some other reason. There is thus no reason why we should not transform the results on to a different scale if, by doing so, analysis is made simpler or its results made more useful or meaningful in some way. Such transformations are familiar in statistics, as witnessed by, for example, the angular and probit transformations. Many biologists, however, retain a suspicion of transformations, believing, it would appear, that they are used to conceal difficulties or to produce unjustifiable conclusions by artificial means. In point of fact an appropriate choice of metric may be necessary if artificial difficulties are to be avoided.

Many times over the years it has been observed that a character, which is itself the product of two sub-characters, can show heterosis when the sub-characters show dominance, or preponderances of dominance, in the opposite direction. Thus, for example, a hybrid may yield a greater total weight of fruit than either parent, even though the number of fruit it carries is less than the midparent number and the average weight of its individual fruit is less than the midparent average. It is then asserted or implied that heterosis cannot be understood in terms of dominance and that observations on dominance are useless in breeding for heterotic behaviour. It is of course true that heterosis may well involve more than just dominance, cases being well established where interaction of a complementary type must be involved; but the apparent contradiction between the direction of dominance in sub-characters and heterosis in the character itself can arise solely by the use of the linear scale in measuring the latter instead of the logarithmic scale which would be suggested by its being the product of the two sub-characters.

Let us consider two gene differences $A-a$ affecting fruit number and $B-b$ affecting average fruit weight. Where large number is associated with low weight in one parent and the reverse combination in the other, the two true breeding parents and their F_1 will have mean numbers of fruit, average weights and total yields as shown in table 1, which also sets out the midparent values. The upper part of the table shows the result of using direct measure and the lower part the result of using log measure when, of course, the sub-characters add instead of multiplying to give the full character. With log measure the F_1 clearly cannot exceed the midparent in total yield unless $h'_a + h'_b$ is +ve which means that at least one of the sub-characters must show dominance in the same direction as the heterosis and it must be the sub-character with the greatest degree of dominance, for $h' = \log(h/d)$. The F_1 will further show heterosis of total yield only if in addition to being

+ve, $h'_a + h'_b > |d'_a - d'_b|$. Using direct measure, however, the situation is more complicated. In the absence of any dominance ($h_a = h_b = 0$) the F_1 will always exceed the midparent in total yield by $d_a d_b$ and it will show heterosis if $d_a d_a > |m_a d_b - m_b d_a|$. If there is dominance for low expression

TABLE I

Dominance in two sub-characters and heterosis in the character which is their product

Direct measure

	1—Parents—2		Midparent	F_1
Number of fruit (A)	$m_a + d_a$	$m_a - d_a$	m_a	$m_a + h_a$
Average weight (B)	$m_b - d_b$	$m_b + d_b$	m_b	$m_b + h_b$
Total yield ($A \times B$)	$(m_a + d_a)(m_b - d_b)$	$(m_a - d_a)(m_b + d_b)$	$m_a m_b - d_a d_b$	$(m_a + h_a)(m_b + h_b)$

$$\begin{aligned}
 F_1 - \text{Midparent} &= d_a d_b + h_a h_b + m_a h_b + m_b h_a \\
 F_1 - \text{Parent 1} &= d_a d_b + h_a h_b + m_a h_b + m_b h_a + (m_a d_b - m_b d_a) \\
 F_1 - \text{Parent 2} &= d_a d_b + h_a h_b + m_a h_b + m_b h_a - (m_a d_b - m_b d_a)
 \end{aligned}$$

Log measure

	1—Parents—2		Midparent	F_1
Log number of fruit (A')	$m'_a + d'_a$	$m'_a - d'_a$	m'_a	$m'_a + h'_a$
Log average weight (B')	$m'_b - d'_b$	$m'_b + d'_b$	m'_b	$m'_b + h'_b$
Log total yield ($A' + B'$)	$m'_a + m'_b + d'_a - d'_b$	$m'_a + m'_b - d'_a + d'_b$	$m'_a + m'_b$	$m'_a + m'_b + h'_a + h'_b$

$$\begin{aligned}
 F_1 - \text{Midparent} &= h'_a + h'_b \\
 F_1 - \text{Parent 1} &= h'_a + h'_b - (d'_a - d'_b) \\
 F_1 - \text{Parent 2} &= h'_a + h'_b + (d'_a - d'_b)
 \end{aligned}$$

A prime is used to distinguish a parameter found in log measure from its counterpart found in direct measure.

in both sub-characters, h_a and h_b will both be -ve and $h_a h_b$ will be +ve. The F_1 will then exceed the midparent in total yield if

$$d_a d_b + h_a h_b > m_a h_b + m_b h_a$$

and will show heterosis if $d_a d_b + h_a h_b > m_a h_b + m_b h_a + |m_a d_b - m_b d_a|$. Thus apparent positive dominance and even heterosis for the full character may well be associated with negative dominance for both sub-characters, depending on the magnitude of the d 's and h 's not merely in relation to one another but also in relation to the average expressions, m_a and m_b , of the two sub-characters. Clearly in such a case the log transformation is essential if reasonably interpretable results are to be obtained: to retain the direct measure is to retain a confusion which would in practice be even greater than appears in the table, from which further complications due to non-heritable variation have been omitted.

The trouble stems, of course, from seeking to use an analysis into additive

components of variation on a character which is made up multiplicatively. When additivity is restored by taking logarithms the trouble vanishes. In this case simply noting that the character is the product of two sub-character points to both the prospective trouble and the way to avoid it. In most cases, however, there is likely to be no obvious guide to a desirable treatment. It is nevertheless always worth looking round to see whether there are difficulties which a suitable transformation might overcome. In general, interactions whether between genes or between genes and environment are complications without which analysis is easier and conclusions clearer if they can be scaled out by a suitable transformation, and it is generally worth looking for such a transformation if interactions exist. The successful reduction of interactions by transformation may also affect

TABLE 2
Number of ocelli in bar and ultra-bar eyed Drosophila at two temperatures
(Krafka's data quoted by Hogben, 1933)

Temperature	Strain	
	B	U
15° C.	198 (2.30)	52 (1.72)
25° C.	74 (1.87)	25 (1.40)

The upper figures are in direct measure and the lower bracketed figures are after transformation into logarithms.

the components of variation ascribable to other genetic phenomena. The dominance component, for example, may be changed and it may be either increased or decreased according to the situation. Mather and Jinks (1971) quote an artificial example where the untransformed data show interaction but no dominance, whereas after a log transformation which removes the interaction, dominance appears. This emphasises once again the conditional nature of genetic phenomena as they can appear in biometrical genetics, and by derivation presumably in Mendelian genetics too, though of course, as already noted, complete dominance and interaction can never be scaled out.

Interaction between genotypic and environmental differences can also sometimes be scaled out. The data in table 2 are taken from Hogben (1933) who was concerned to argue that the effects of genetic and environmental change are not simply additive in development. Seen on the direct scale, the data appears to bear out his contention. On a log scale, however, the interaction of genetic and environmental differences virtually vanishes and we see that his conclusion is not inescapable. Whatever may be the legitimacy of his argument in other cases it cannot be held to be generally valid, as in this particular instance we are fully entitled to treat the effects of the two types of change as additive if we so wish. The analysis becomes correspondingly simpler and the measurement of the genetic and environmental effects are no longer conditional when we make this simple transformation.

This is not, however, to say that inter-allelic interactions or genotype) environment interactions can always be scaled out. Cases are known where a transformation which removes genic interactions in one cross merely serves to introduce them into another which previously had shown no evidence of them (see Jinks, 1955). Evidently the genes segregating in the different crosses had not been the same, and had not been interacting in the same way. Equally we know that F_1 's, though displaying a character to some extent intermediate between those of its parents, can at the same time be less subject than the parents are to the influence of environmental changes. This is a genotype/environment interaction which cannot be removed by any reasonable transformation. Thus we must always be prepared to bring interaction explicitly into an analysis and, as we have seen, the theoretical framework is available for us to do so.

Interactions of all kinds are indeed worthy of even more study than they have yet received, for they are of immense prospective importance in applied genetics. To take but one example, Buccio Alanis (1966) found with plant height and flowering time in two inbred lines of *Nicotiana rustica*, that in environments which enhanced the average expressions of the characters, not only were the differences between the lines enhanced, so demonstrating a genotype $\times$ environment interaction, but that the changes of the differences were related in a simple linear way to the changes in the average expressions. Furthermore, Buccio Alanis and Hill (1966) found that the excess of the F_1 over the midparent value was related to the effect of the environment in just the same linear way though the slope of the line was not the same as for the parental difference.

These findings raise several important questions. Why is the relation between interaction and environmental effect linear? Why is the slope of the line different in parents and F_1 ? How general a property is this of genotype/environment interaction? And so on. They also have prospectively important consequences in applied genetics, for if the linear relation is general it means that the consequences of changing a variety's environment could be predicted much more readily. Furthermore, since the F_1 and parental lines do not show the same slope, it follows that heterosis could be displayed in some environments but not in others. In other words, heterosis becomes a conditional phenomenon. It is very much to be hoped that these investigations of interactions will be followed up for their prospective practical importance as well as their theoretical significance.

5. CONCLUSION

As we have seen, biometrical analysis in genetics differs from the classical Mendelian method in that on the one hand it does not lead to the recognition of individual genetic differences and so cannot lead to a detailed knowledge of the genotype, but on the other is not limited in its application to variation from which individual genetic differences can be established and, in fact, takes into account all the variation in respect of a character that is displayed by a family, group of families or population. It proceeds not by the comparison of the frequencies and phenotypes of establishable genetic classes but by the comparison of means, variances and covariances of families and groups whose genetic relations are known only in the broad sense. These means, variances and covariances are broken down into components

representing the effects of genetical phenomena whose existence and properties could hardly be established from biometrical studies themselves, but stem from the detailed studies of the genotype made possible by the classical methods of genetics. At the same time, because these phenomena are being examined in a different and broader context, they must be defined more clearly so as to permit the combination of contributions from all the genes that are acting over all the environments that are encountered. In doing so there emerges a prospective interchange of one phenomenon with another, when represented in detail on a metrical scale. In this way, biometrical genetics, while borrowing its basic genetical structure from classical studies, makes in turn its contribution by assessing the contribution of these phenomena to variation otherwise incapable of analysis, by requiring us to define and examine the phenomena and their effects in a more precisely quantitative way, and by making us recognise their frequently conditional nature and their capacities for mimicking and balancing one another in their effects on the overall variation of a given character.

Linkage and other types of association of genes in the derivation of a family or population on the one hand, and interaction of genes in producing their effects on the other, are both forms of correlation of the genes in making their contributions to the variation. They cannot merely reinforce, but in some measure mimic one another in their effects on the variation and consequently on the response of the family or population to any selection that may be applied. Yet in the long run their consequences are different, for linkage is resolvable by recombination, whereas interaction is not. So it is important to separate their effects in assessing variation and assigning it to its sources. In principle, this separation of their effects can be achieved biometrically, but in practice it may be difficult. Not all the ways of doing so have, however, yet been examined closely and further study might be well repaid.

In the main, interactions have in the past appeared more as sources of distraction and inflators of error variance in the biometrical measurement of the main genetic components. Anything that can be done to mitigate these consequences is obviously desirable and, as we have seen, it is always worth looking for a scale on which to represent the variation, such that the effects of interaction are minimised. Not all interactions are, however, metrical in this sense and their properties can represent the consequences of the past action of selection in such a way as at any rate partly to determine the outcome of future selection (Mather, 1966; Mather and Jinks, 1971). At the same time the components of interaction are difficult to measure biometrically and here again further study of the experimental and analytical techniques for doing so would bring a most worthwhile reward. Such a study of genotype/environment interactions has already begun to pay dividends and if the rectilinear relation which Buccio Alanis and Hill have found between the magnitude of such interaction and the direct effects of the environment proves to be at all general, it will be of considerable practical importance in applied genetics as well as of theoretical interest.

One last point remains to be emphasised. Since individual determinants are not followed separately in biometrical analyses, the properties of dominance, interaction, linkage and so on are revealed as over-all properties, pooled or averaged over all the genes contributing to the variation. These may be enhanced by reinforcing effects stemming from different genes or

diminished by opposing or balancing effects. The amount of information they can give about the properties of the individual determinants is thus limited. Furthermore, when we are dealing with such aggregates of genes distinctions which may be valuable in their application to understanding the properties of individual genes lose their value and become virtually meaningless. Even when by the introduction of appropriate techniques based on the use of, for example, marker genes, we have partially fractionated the genotype in respect of its contribution to continuous variation we are still in general dealing with aggregates of genes carried on individual chromosomes or pieces of chromosome. In such a case the distinction between mutation and recombination as the basis of change in these units (or effective factors as we may call them) is not helpful since both will result in an alteration of the effective factor's properties in ways that are incapable of being distinguished by any but the most elaborate and time-consuming experiments. At the same time, since recombination is so much more common a phenomenon than mutation in all but the smallest pieces of chromosome, it must dominate the scene and we must expect the changes we observe to reflect the properties of recombination—an expectation that experiment has consistently borne out. Equally, concepts like overdominance and pleiotropy are of little value when applied to the aggregates of genes beyond which the nature of the variation precludes us from pushing the analysis. To observe them as properties of the aggregates in no way requires that they are properties of the individual genes: they can appear in an aggregate without being shown by any individual member of it and they can be altered, enhanced, or removed by recombination of the constituting genes without any change in the properties of the genes themselves.

Though ultimately the same phenomena are involved they may appear in different ways and may require different interpretation at different levels of analysis, and this we should remember if we are to make proper use of the power of biometrical genetics in extending analysis to variation beyond the scope of the classical methods and to situations which it is important to understand but would otherwise be incapable of attack.

6. REFERENCES

- BATESON, W. 1909. *Mendel's Principles of Heredity*. University Press, Cambridge.
- BUCIO ALANIS, L. 1966. Environmental and genotype-environmental components of variability. I. Inbred lines. *Heredity*, 21, 387-397.
- BUCIO ALANIS, L., AND HILL, J. 1966. Environmental and genotype-environmental components of variability. II. Heterozygotes. *Heredity*, 21, 399-405.
- COOKE, P., AND MATHER, K. 1962. Estimating the components of continuous variation. II. Genetical. *Heredity*, 17, 211-236.
- DONCASTER, L. 1908. On sex-inheritance in the moth, *Abraxis grossulariator* and its var. *lacticolor*. 4th Rep. *Evol. Comm., R. Soc. London*.
- FISHER, R. A. 1918. The correlations between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edin.*, 52, 399-433.
- FISHER, R. A., IMMER, F. R., AND TEDIN, O. 1932. The genetical interpretation of statistics of the third degree in the study of quantitative inheritance. *Genetics*, 17, 107-124.
- GILBERT, N. 1961. Polygene analysis. *Genetic Res.*, 2, 96-105.
- HOGGEN, L. 1933. *Nature and Nurture*. Allen and Unwin, London.
- JINKS, J. L. 1955. A survey of the genetical basis of heterosis in a variety of diallel crosses. *Heredity*, 9, 223-238.
- MATHER, K. 1966. Variability and selection. *Proc. Roy. Soc. B.*, 164, 328-340.

- MATHER, K. 1967. Complementary and duplicate gene interaction in biometrical genetics. *Heredity*, 22, 97-103.
- MATHER, K., AND HARRISON, B. J. 1949. The manifold effect of selection. *Heredity*, 3, 1-52 and 131-162.
- MATHER, K., AND JINKS, J. L. 1971. *Biometrical Genetics*. 2nd edition. Chapman and Hall, London.
- THODAY, J. M. 1961. Location of polygenes. *Nature, London*, 191, 368-370.