It has long been known by breeders that sometimes a few offspring indicate an animal's real breeding value more accurately than the animal's own appearance or performance does. The rediscovery of Mendelism and the discovery a little later (by Bateson and others) of complementary genes and of other epistatic relations among non-allelomorphic genes, provided a simple explanation for this. Occasionally there are cases where a single offspring furnishes more convincing proof of its parent's genotype than could ever be obtained by studying the parent itself. So far as color is concerned, this would be the case with a purebred red calf born in one of the black breeds of cattle.

These are special cases, deserving attention to be sure, but perhaps they are more significant as exceptions to a general rule than as examples of one? Some writers on animal breeding topics have considered them to be examples. For instance we are sometimes told that a progeny test is "genotypic" and therefore dependable, whereas a cow's own record is "phenotypic" and therefore not dependable. But no fundamental distinction is really involved, since the progeny test itself consists of some kind of an average of "phenotypic" records of the offspring. To suppose that the process of averaging would purge these phenotypic records of all their error and would bring out only the pure gold of genotypic truth seems to require too much faith in the potency of arithmetic!

The present paper presents the results of an inquiry into the biometrical relations which determine the relative accuracy of progeny test and parent's own performance as indicators of the parent's real breeding value. The complete answer contains many variables but most of these have little influence. An answer accurate enough for practical purposes can be presented in usably simple form. For convenience the problem is phrased in terms of dairy cattle breeding thus: Is a cow's own record or the average record

Received for publication May 31, 1934.
Journal Paper, No. J173, from the Iowa Agricultural Experiment Station. Project No. 31.
of n of her daughters the more dependable indicator of the cow's own breeding value?

METHODS USED AND DEFINITIONS

These biometrical relations were explored by Wright's method of path coefficients (10). Figure 1 shows those relations. The small letters indicate path coefficients which measure (provided the cause and effect relationships are portrayed correctly in the diagram) the importance of the causes of variation in the effect. The path coefficient (h) from heredity (H) to observed characteristic or yield (C) is the ratio of the standard deviation of C due to H to the total standard deviation of C. The path coefficients h and e may be used to express the relative importance of heredity and environment on production records in a particular population of data. Because variation is measured by squares of deviations h and e will not add up to unity. Where there is no correlation between heredity and environment, the squared path coefficients, \( h^2 \) and \( e^2 \), will come near to expressing "the relative importance of heredity and environment" in the form in which that idea is ordinarily encountered in practical problems. A definite numerical value for \( h^2 \) is a description of the population from which it was derived and may not, without other knowledge, be safely applied to populations where the variations in environment or in heredity may be quite different. For prediction purposes, path coefficients are regression coefficients expressed on a scale in which each standard deviation is one unit.

Most of the terms in Figure 1 are self-explanatory. Phenotype is used here to mean what the outwardly observed record would be if it were possible to keep all environmental circumstances rigidly controlled. It differs from the actually observed record in that the latter has been modified by the variations in environment to which the various phenotypes were exposed. Non-additive effects of environment on the phenotype (such as concern Hogben so much in his recent "Nature and Nurture" (5)) may be minimized by expressing the records on scales which show in the most nearly additive way the effects of variations in environment and in heredity. For example there is some evidence to indicate (but perhaps not enough to prove conclusively?) that dairy production records should be expressed in logarithms of the yield instead of in pounds if the most efficient use is to be made of the information in those records. However, for the present question, non-additive combination effects of phenotype and environment may be considered as included in the environment, since they tend to decrease the correlation between records and genotypes.

"Genotype" is obtained by assigning each gene such a value that the simple sum of all the genes in each individual will give the "genotype" for that individual subject to the "least squares rule," that (1) the average
genotypic value of the population shall coincide with its average pheno-
typic value and (2) the sum of the squares of the individual deviations of "genotype" from "phenotype" shall be a minimum.* This "genotype" seems to be the same as Fisher's "expected value" of which he says (4, page 32), using human stature for illustration: "This expected value will not necessarily represent the real stature, though it may be a good approximation of it, but its statistical properties will be more intimately involved in the inheritance of real stature than the properties of that variate itself."

![Diagram](image)

**Fig. 1. Path Coefficient Diagram Showing Biometric Relations Between Parent’s Records and Parent’s Breeding Value (G) And the Biometric Relations Between the Progeny Average and the Parent’s Breeding Value.**

These genotypic values will be perfectly correlated with Turner’s definition (9, page 45) of "transmitting abilities," in so far as dominance is the only cause of discrepancy between genotype and phenotype, but Turner’s definition makes no provision for non-additive combination effects of genes which are not allelomorphic to each other. For convenience in this article such non-additive combination effects of non-allelomorphic genes will be

*I am indebted to Dr. Sewall Wright for this method of expressing as much as possible of hereditary differences in terms which contribute to the correlations between relatives.*
called epistasis or epistatic effects, although this is a slightly broader meaning than that for which "epistatic" was first used in genetics. In the usage of practical animal breeders, "nicking" most nearly has the same meaning but many breeders also apply the term, "nicking," to phenomena which are nothing but unusual Mendelian recombinations of genes each of which may have nearly additive effects.

"Genotype" as used here is perfectly correlated with breeding value if the latter term is understood to express the average merit of the offspring produced by mating this animal to mates which are a truly representative sample of their breed. If the mates are not a truly representative sample and if there is much epistatic interaction or if there is much dominance and mating is far from random, it must sometimes happen that an individual will produce offspring of higher (or lower) merit in one kind of herd than it would in other herds where the most frequent gene-combinations are quite different. In such cases an animal's breeding value as measured by actual records of progeny, will deviate from what its breeding value would be if its mates were truly representative of their whole breed. However such offspring will not transmit such dominance deviations or epistatic deviations, since those deviations are functions of the gene combination rather than of the individual genes considered separately. The combination is not transmitted but segregates at the reduction division and the genes reassemble in other combinations in the next generation. Hence "genotype" as used here to express the sum of the average effects of the constituent genes is "breeding value" if one considers mating the animal to representative animals of its breed, though it may deviate from "breeding value" if the latter term is used to mean the contribution which this animal could make to descendants produced by breeding it to some isolated and non-typical sub-group of the breed.

"Phenotype" is used here as a convenient intermediate step to show the modification of the genotype by dominance or epistatic relations, which are usually considered genetic phenomena although really they are only consequences of the chemistry and physiology by which the genes produce their effects and are not involved in the mechanism by which genes are transmitted. By this device we can express separately the proportion \( h^2 \) of the observed variance which is hereditary in the broader sense of that word (\( i.e., \) including the effects of dominance and epistasis as "hereditary") and the proportion \( g^2 \) of the hereditary (in that broad sense) variance which can be expressed by this additive scheme for representing genotype. The proportion of the observed variance which is hereditary in the narrow sense of that word (\( i.e., \) which would contribute directly to parent-offspring correlations) would be \( h^2 g^2 \).

Non-additive effects of gene combinations are separated into those due to dominance \( (D) \) and all others \( (I) \). This is done merely because the sta-
Statistical consequences of dominance have been explored (2, 3, 12) and can be stated in definite formulae. The statistical consequences of epistatic interactions are not so well known and must be quite complex in some cases, but on the whole probably are much the same as those of dominance, i.e., to lower the correlation between genotype and outward characteristic and hence to lower the observed correlations between relatives, but to affect the correlations between some kinds of relatives more than those between other kinds. So far as dominance and epistasis merely lower the parent-offspring correlation (i.e., diminish the value of g) their effects nearly cancel out of the present problem since they weaken the correlation between daughter average and dam’s genotype almost as much as they do the correlation between the dam’s average and her genotype. However so far as dominance and epistasis lower the correlation between sisters less than they do that between parent and offspring, they actually prevent the progeny test from becoming relatively as accurate an indicator of the dam’s genotype as it would if there were no dominance or epistasis.

The importance of epistasis in practical breeding problems is still a matter of conjecture, with some workers thinking it very unimportant and others speculating that it may often be highly important, especially in herds already improved to a point considerably above the average level of the breed. Wherever the parent-offspring correlation is high under nearly random mating (after having been corrected for the effects of common environment), it seems likely that epistasis is unimportant. However a low parent-offspring correlation under those circumstances does not prove that epistasis is important, since such a low correlation might also result from h being small, i.e., from random variations in environment having important effects on the records.

The importance of dominance depends primarily on whether complete dominance is the rule, how abundant the dominant genes are in contrast to their recessive allelomorphs, and somewhat on the deviations from random mating. In a population mating at random with dominance complete and with no epistasis, \( g^2 \) for the effects of each gene pair = \( \frac{2(1-q)}{2-q} \) or conversely \( d^2 = \frac{q}{2-q} \) = the portion of variance in \( H \) due to dominance.

(The letter \( q \) here represents the frequency of the dominant genes among all the genes occupying that locus in the whole breed.) This \( d^2 \) which Fisher calls “the dominance ratio” (3) is thought by him to be typically about \( \frac{1}{2} \) for most characteristics. Wright (12, pages 137–139) for other reasons, which appear to him more convincing, believes that this ratio is more typically about \( \frac{1}{6} \), even if complete dominance is the almost universal rule among all genes. Moreover he questions (13) whether such complete dominance is the rule, especially among modifying genes each of which has
a comparatively small effect. Such modifying genes probably furnish much of the variance in quantitative characteristics which, like milk production, are vitally necessary in some degree but which may vary widely among different individuals in the breed.

Whether complete dominance is the rule or not, it is obvious that dominance will not often be the major obstacle in estimating an animal's genotype from its own appearance or performance. Even if dominance is universal and Fisher's estimate of $d^2$ is correct, $g$ will generally have a value of about .82, while if Wright's view is correct $g$ will usually be nearer .90. Both these figures are based on the assumption that epistasis is too small to be important. If epistasis is important, $d^2$ and $g^2$ would still retain about the same relation to each other, though both would be smaller.

There will be occasional characteristics (of which lethals are notable examples) where dominance is complete and the recessive gene is so rare that the "dominance ratio" approaches close to unity. In those cases dominance may be the major source of the discrepancies between genotype and phenotype. However in such traits the recessive individuals are rare and each such trait will contribute but little to the total variance of the population. For example, in the United States in black breeds of cattle like the Angus, Galloway, or Holstein-Friesian, the gene for black must be around ten to fifteen times as frequent as its recessive allelomorph. Hence the "dominance ratio" for this particular trait will be well above .80. But the gene for red is so rare that less than 1% of the purebred calves of these breeds are born red. Hence this trait is a comparatively unimportant source of loss to breeders. Dominance has been discussed here more than its importance justifies, largely because it has received more attention than it deserves in the writings on applied genetics.

**SOLUTION UNDER SIMPLEST CONDITIONS**

Figure 1 shows the relations and correlations as they would be under random mating. The correlation between one record of the dam and the dam's genotype will be $h_p g_p$, while the correlation between one record of one daughter and the dam's genotype will be $h_gab$. When $m$ unselected records of the dam are averaged together the correlation between that average and the dam's genotype becomes

$$h_p g_p \sqrt{\frac{m}{1 + (m-1) (h_p^2 + e_p^2 r_{EpEp})}}$$

which approaches

$$g_p \sqrt{\frac{h_p^2}{h_p^2 + e_p^2 r_{EpEp}}}$$

as a limit as $m$ becomes indefinitely large. If $e_p^2 r_{EpEp} = zero$ (the most favorable condition for high values of $r_{MG}$), this approaches $r_{MG} = g_p$ as a
limit. Increasing the number of records observed thus tends to eliminate errors arising from random uncorrected variations in environment but does not tend to eliminate errors arising from dominance or epistasis or consistently biased and uncorrected environment.

On the other hand, the average record of n daughters each tested once is correlated with the dam’s genotype thus:

$$\text{ hgab} \sqrt{\frac{n}{1+(n-1)(h^2g^2a^2b^2 + e^2r_{EE})}}$$

which approaches

$$\sqrt{\frac{h^2g^2a^2b^2}{h^2g^2a^2b^2 + e^2r_{EE}}}$$

as n becomes indefinitely large. This approaches unity when $e^2r_{EE}$ is zero. In other words, increasing n tends to cancel errors arising from the imperfections of both h and g, whereas increasing m tends to cancel the imperfections of $h_p$ but does nothing to correct for the fact that $g_p$ is less than unity. However the progeny test involves ab which arises from the intervening generation of Mendelian segregation and appears not at all in the correlation between dam’s record and dam’s genotype.

Thus a daughter’s record usually is a less accurate (on account of ab) indicator of her dam’s genotype than the dam’s own record is, but the errors in daughters’ records tend to be more completely cancelled by the averaging process than is the case with errors in the dam’s record. Hence if $g$ and $g_p$ were quite small, it might be possible for n to be so large that its cancellation of errors arising from dominance or epistasis would more than make up for the Mendelian sampling errors which affect the progeny average but not the dam’s own records.

The problem of the relative accuracy of progeny test and dam’s own record as indicators of the dam’s genotype therefore becomes the quantitative one of balancing the errors introduced by ab against the greater effectiveness of the progeny average in eliminating errors introduced by dominance and epistasis. Unfortunately for the simplicity of the answer, $e^2r_{EE}$ and $e_p^2r_{EpEp}$ cannot safely be considered zero and their existence sets serious limits on the effectiveness of averaging as a means of eliminating errors, either from the daughter average or from the average of the dam’s own records.

The value of ab will be nearly .5 even with moderately large departures from random mating (11, pages 118–119). Substituting that value for ab, the ratio (B) of $r_{Ag}$ to $r_{Mg}$ becomes

$$B = \frac{h g \sqrt{\frac{n}{h_p g_p \sqrt{\frac{1+(m-1)(h_p^2 + e_p^2 r_{EpEp})}{4+(n-1)(h^2 g^2 a^2 b^2 + 4e^2 r_{EE})}}}}}}$$

(1)
If dam and daughters are produced by the same kind of a breeding system \( g = g_p \) and if dams and daughters are equally typical samples of their generations and if the records used to represent the productiveness of dam and of daughters are chosen by a method equally fair to both, \( h = h_p \) and the first part of this expression cancels. Assuming (as the case most favorable to the progeny test) that \( m = 1 \) and squaring to simplify the expression we have

\[
B^2 = \frac{n}{4 + (n-1)(h^2g^2 + 4e^2r_{EE})} \quad \text{(2)}
\]

For the progeny test to be more accurate than the dam's own record \( B^2 \) must be larger than 1.0. This condition is satisfied when \( n > 4 + (n-1)(h^2g^2 + 4e^2r_{EE}) \). For this to be true, \( n \) must be larger than 4, even when \( h^2g^2 \) and \( e^2r_{EE} \) are extremely small. If \( e^2r_{EE} \) is as large as .25, \( n \) cannot possibly be large enough to make the progeny test average as reliable as the dam's own performance.

Equation 2 is pictured graphically in figures 2 and 3. Since four variables are involved (treating \( h^2g^2 \) as a single variable and \( e^2r_{EE} \) as another single variable, since each behaves as such in this statement of the problem) it seems necessary to select certain reasonable or limiting values of one of the four and for this constant value of one variable to show the interrelations of the other three as a curved surface whose height above a base is \( B^2 \).

Figure 2 is designed to show how \( B^2 \) varies with \( n \) and with \( h^2g^2 \). The upper surface shows the limiting condition \( (e^2r_{EE} = \text{zero}) \) most favorable to the progeny test. Such a condition would be encountered in actual practice when half sisters resemble each other only because they have the same dam. Any correlation between half sisters because they are kept under similar environment, or are related through their other parent, or because \( r_{DD} \) or \( r_{HH} \) are not zero would have the same kind of effect as giving \( e^2r_{EE} \) a positive value. Some such effects are almost certain to occur in any set of actual data, even when such data have been obtained under experimental conditions controlled as rigidly as possible. The lower surface in figure 2 shows \( B^2 \) when \( e^2r_{EE} = .20 \). Such a condition would arise whenever unrelated cows in the same herd showed a correlation of +.20 just because they were managed and tested in the same herd in a population of data coming from many herds kept under varying management. Such a value for \( e^2r_{EE} \) is roughly consistent with much of the C.T.A. and Official data yet analyzed in a way which permits any estimate of \( e^2r_{EE} \). However the value of \( e^2r_{EE} \) naturally will vary from population to population, since it is a description of the heterogeneity of the conditions under which different sets of half sisters are kept.
The double line in figure 2 shows a plane parallel to the base at the level \( B^2 = 1.0 \). Points above this level denote conditions under which the progeny test is more accurate than the dam’s own record. When \( e^{2r_{EE}} = \) zero, and \( h^2g^2 \) is as low as .10, the progeny test surpasses the cow’s own record when \( n = 5 \), but \( n \) must exceed 6 when \( h^2g^2 = .4 \) and must exceed 11 when \( h^2g^2 = .7 \). This brings out the important points that the progeny test is not very dependable where the daughters resemble each other for other reasons than that they are out of the same dam\(^1\) and that the superiority of the progeny test over the dam’s own record is most marked where \( h^2g^2 \) is smallest, i.e., where the characteristic is least hereditary in the narrow sense.

\(^1\)This same point was shown in more detail by the present writer in an article (8) directed primarily at how \( r_{AD} \) varies with increasing \( n \). In that article \( e^2 \) was used to include all the causes for half sisters resembling each other except that they are by the same parent.
of that word, and therefore where neither indicator is highly accurate. Values of $h^2g^2$ less than .10 are not shown in either figure 2 or figure 3. For such values $B^2$ would become very large if $e^2r_{EE}$ were also very small.

Figure 3 shows how $h^2g^2$ and $e^2r_{EE}$ affect $B^2$ for three selected values of $n$ (5, 10, and 20). The shading shows the regions where $B^2$ rises above 1.0.
The limiting effect of $e^2r_{EE}$ on the accuracy of the progeny test is vividly shown in this figure. Only rarely can the accuracy of the progeny test in a whole population average as high as that of the parent’s own record, if the correlation between herd mates due to herd environment is as high as .10 and if $h^2g^2$ is as large as .10.

**CONDITIONS WHICH MAY AFFECT THE VALIDITY OF THE SIMPLE SOLUTION**

Figures 2 and 3 show the situation when $m$ is one, each daughter is represented by one record only, breeding is at random, the sires are of unknown breeding value, dams and daughters are produced by the same kind of breeding system, and dams are just as representative of their generation as daughters are of theirs. Now these conditions will never be fulfilled exactly, not even in experimental data. It is important to glance for a moment at the way deviations from these conditions will affect figures 2 and 3, although only deviations from the last condition can alter those figures materially in the direction of making the progeny test more reliable.

When $m$ is more than one—and it usually will be if $n$ is as large as two—the dam’s record will be more reliable, in accordance with the formula in equation (1). When each daughter is represented by an average of several records, the effect is that of increasing $h$ and making the progeny test more reliable but this is usually more than offset by the increase in $m$. Any dam who has several daughters, each with more than one record, is almost certain to have had many records of her own, so that the average number of records per daughter will nearly always be less than $m$. Hence the use of an average of all available records will tend to make both indicators more reliable but will usually increase that of the dam more than that of the progeny.

Deviations from random mating may be of two kinds so far as they affect this problem. First there may be a correlation between the mates of a cow, irrespective of whether they are like or unlike her. A positive correlation of this kind has the same effect as an increase in $e^2r_{EE}$ and therefore lowers the reliability of the progeny test. A negative correlation of this kind (i.e., some degree of success in deliberately seeing to it that each cow was mated both to very good bulls and to an equal number of very poor bulls) would increase the accuracy of the progeny test slightly since it would eliminate some of the chance existing under truly random breeding, that one cow would happen to be mated only to good bulls, another only to poor bulls, etc. However such a negative correlation cannot be high if $n$ is large, because if a cow is mated to three or more bulls those bulls cannot each one be exactly opposite to every other one! Moreover it is hard to imagine that such a general negative correlation could exist under any circumstances except a breeding experiment where it was made
a cardinal point to get from each cow daughters by poor bulls and daughters by good bulls.

The second kind of deviation from random mating would be a general correlation between dams and sires so that it would generally be true through the whole breed that good cows were usually mated to good bulls, poor cows to poor bulls, etc. This could come about either through inbreeding or through deliberate assortative mating based on individual performance or the performance of near relatives. It is not unreasonable, either from the genetic principles involved or from the published studies of dairy data to suppose that this correlation within dairy breeds might well be as large as $\pm 0.15$ or $\pm 0.20$. So long as all breeders are working for high production, this correlation could hardly rise much above $\pm 0.25$ or $\pm 0.30$ without the practice of extreme inbreeding. The effect of such a general resemblance between sire and dam would be to increase the accuracy of the progeny test slightly, since each offspring would then tend to be like the dam, not only because of receiving half its inheritance from her but also because the other half of its inheritance came from a sire whose inheritance was more like that of the dam than if he had been chosen at random from all sires of the breed. A general correlation of $\pm 0.25$ between the genotypes of sires and dams would modify equation 2 enough that the progeny test might possibly be more accurate than the dam's own record while $n$ is as small as 3, but this would be true only if $e^2 r_{ee}$ were zero and $h^2 g^2$ were very small, and there were no other correlation between sires. Since it is not often that all three of these conditions will be fulfilled, it seems unlikely that such deviations from random mating as exist in actual dairy cattle populations will often alter figures 2 and 3 enough to make the progeny test in a whole population average as accurate as the cow's own record when the cow has as few as four daughters. It should be mentioned that in plants where self-fertilization is possible, the resultant perfect correlation between genotypes of sire and dam is quite attainable and makes it possible for a progeny test based on as few as two offspring to be as accurate as the parent's own record. This is doubtless the most important fundamental reason that plant breeders make so much use of the progeny test, particularly when working with self-fertilized organisms like wheat.

Sometimes the merits of the different sires will be partly known. If their genotypes were perfectly known and there were no correlation between sires and dams, formula 2 would become:

$$B^2 = \frac{n}{4 + (n-2)h^2g^2 + 4(n-1)e^2 r_{ee}}$$  \hspace{1cm} (3)$$

Thus even such extreme knowledge as this would have but a trifling effect on figures 2 and 3. If mating is not at random, such information about the sires is quite useful in that it would cancel and prevent errors which would
otherwise be introduced by any such correlation between sires as that dis-
cussed in the next to the last paragraph. Knowledge of the sires therefore
prevents a correlation between sires from causing the accuracy of the
progeny test to fall far below the levels pictured in figures 2 and 3, but can-
not of itself raise those levels much higher than is shown. If there is a
general correlation between sires and dams and the merit of the sires is
known from other evidence, then such knowledge will help to predict the
dam’s genotype but such help is because of that direct correlation between
the genotypes of sire and dam, and not because of evidence from the prog-
eny themselves.

Where dam and daughters have been produced by different breeding
systems, the values of $g$ for dam and daughter may not be identical and
hence will not cancel out of equation 1 as has been assumed so far. The
most extreme example of this, so far as concerns dominance, would be the
case where the progeny test is conducted by breeding the cows to homozy-
gous recessive bulls. In such a case the deceiving effects of dominance
would become zero for the daughters and a correspondingly higher percent-
age of the total variance among them than among the dams would be truly
 genetic in the narrow sense of that word. Naturally such a test could not
be made for a cow on account of her short life and limited reproductive-
ness, but a corresponding test of a bull might be practical if the informa-
tion gained were worth the cost of getting it. However dominance is
usually only a minor cause for actual records deviating from breeding
value. The assembling and maintaining of a herd of recessive cows for such
tests would be expensive, even if possible at all. Moreover such a plan does
not appear to offer any help for avoiding deception by epistatic effects.
These considerations make it seem quite unlikely that such a testing plan
would often yield information worth the cost of getting it. If the sire and
dam belonged to different breeds, there would be increased probability of
important misleading epistatic effects which would probably preclude using
for testers individuals of a breed developed for other purposes, as for exam-
ple the use of beef-bred cows for testing the breeding value of a dairy bull
for milk and fat production.

Some of the most brilliant research in genetics, particularly in the
genetics of Drosophila and of maize, owes its success to the devising and
use of tester strains so designed as to eliminate (except for Mendelian sam-
ping errors) the effects not only of dominance but also of epistatic inter-
actions, thus making the truly genetic portion of the variance ($h^2g^2$) among
the offspring much higher than was the case among their parents. How-
ever the devising of such a tester strain has for its prerequisite a reasonably
accurate hypothesis as to the number and kind of genes involved and (at
least so far as concerns epistasis) of the precise way in which the effect of
the gene combination differs from the sum of the average separate effects
of the genes concerned. The limited reproductive rate of farm animals, their comparatively long interval between generations and their expensive
ness make it well nigh a lifetime's work to create such a tester strain, even if the investigator were wise enough or lucky enough to have chosen the correct working hypothesis at his first attempt and even if the genetic situa-
tion were simple. Hence it appears unlikely that much use can be made of such tester strains in the breeding of farm animals, however useful and even necessary such tester strains may be in research on fundamental prin-
ciples of genetics.

The assumption that daughter and dam are equally representative of their generation and that their records were selected by methods equally
fair to both will often be untrue especially in semi-commercial data, and is the only one of these assumptions likely ever to be important enough in a whole population to make much change in figures 2 and 3. The effect of such differences between daughters and dams is to make h in figure 1 une-
qual to h.<sub>p</sub>. If they are unequal these two terms cannot be cancelled out of equation 1 as has been inferred hitherto.

If but one record is used to represent each animal and if that record is
selected on some such ex post facto basis as the choice of the highest record after having seen them all, such a selection deliberately picks out that record made under the most favorable combination of unknown conditions of environment and health. Such selection automatically tends to give to the cow having many lactations a figure less truly representative of her phenotypic producing ability than would be the case with a cow which had but one or two lactations from among which to select. Since the dam usually
has more records than the daughter, the net result is to make the dam's record usually less typical of her than the daughter's record is typical of her. This tends to make the progeny test more accurate than is pictured in figures 2 and 3. It should be pointed out that this comes not from any intrinsic superiority of the progeny test, but rather from the bias intro-
duced by the choice of records where variable numbers are available. Studies of the actual correlations between records of the same cow in differ-
tent lactations in the same herd have generally given values ranging from about +.30 to +.60. Correlations as low as this give ample room for selec-
tions of the largest record for each cow to cause such selected records to deviate far from each cow's phenotype, especially if some cows have only one or two records and others have 4 or 5, from among which to select.

Probably more generally important is the fact that the dams are often a more highly selected group than the daughters are. Usually all of the progeny which were tested at all are included in the progeny test. If they were very poor producers, they may have been culled shortly afterward but their records remain for use in the progeny test. The conditions may be quite otherwise among the dams. Not all cows which are tested succeed
themselves in becoming dams of tested daughters. The dams being studied in any particular set of data are to some extent only the more favored survivors from among the group tested contemporarily with them. In studying correlations between the records of those dams and any measure of the subsequent performance of their progeny, one is by no means making a comparison between phenotypic and genotypic manifestations of production. Instead he is comparing the utility of those phenotypic differences still remaining among the dams after this selection has been practiced, with the utility of various averages of phenotypic differences among the less highly selected group of daughters. This seems to be a major part of the explanation for the practical findings of Jull (6, 7) and Copeland (1) that the ancestor’s production record is of relatively limited value in breeding selections. In Jull’s data all birds which laid less than 200 eggs were culled before they had a chance to become dams. What would have happened, if all hens or a strictly random sample of those hatched in each generation had been progeny-tested, of course can only be surmised. Copeland’s data came from many herds and at various dates, so that selection was not by such uniform rules as in Jull’s experimental data. Nevertheless Copeland’s tables 1 and 5 show that there had been stronger selection among the dams than among either their daughters or granddaughters, else the extreme regression (about 50 to 70 pounds of fat) from the dams’ average to the daughters’ and granddaughters’ averages would not have occurred.

Naturally such studies as those of Jull and Copeland (many others might be mentioned but these two are selected as typical of the more careful among recent ones) correspond closely to the conditions practical breeders actually face. Depending on the amount of selection involved and on how highly hereditary the trait really is, it may sometimes be true that such variation as remains among a highly selected group of dams may be less useful as an indication of the hereditary differences between those dams than the variation which exists among the relatively unselected daughters of those dams, even when there are as few as two daughters per dam. The point to be made here is that when this actually is the case, it is not because of any inherent biological superiority of the progeny test but rather it is because in the case of the dams we are comparing what further selection could do among a group of survivors on which considerable selection has already been practiced, with all which selection could do among a group of daughters where very little selection has yet been practiced. The practical utility of such comparisons is beyond argument for the breeder who must actually use, in some combination or other, some data from the more heavily selected population and some data from the less selected population. There is, however, some danger of confusion as to the utility of the initial selections in the populations from which the dams first come, if such studies are interpreted as meaning that an individual’s own record is in-
herently less representative of its breeding value than the average records of as few as two or three of its offspring.

GENERAL CONSIDERATIONS AND DISCUSSION

The general conclusion to be drawn from all these considerations is that only under rare and unlikely combinations of conditions would a progeny test based on as few as four-daughters-average in an unselected population as accurate an indicator of a dam's breeding value as the dam's own performance. If the trait is highly hereditary, in the narrow sense of that word, or if the progeny resemble each other very much for reasons of having been exposed to a common environment, many more than four daughters may be required or it may even be quite impossible for the progeny test to average more accurate in a whole population than the dam's own performance, no matter how many daughters there are.

The superiority of the progeny test is greatest for traits which are least hereditary in the narrow sense of that word. For traits which are very faintly hereditary and for which the offspring do not resemble each other for reasons of having been under common environment, the progeny test can become several times as accurate as the dam's own record, but for that to be true the trait must be so slightly hereditary that neither indicator can be highly accurate!

The progeny test is needed most where one sex cannot express the trait, as in milk and fat production in dairy cattle, egg production in poultry, prolificacy in swine, etc. Since the biometrical relations between sire's genotype and progeny average are the same as those between dam's genotype and progeny average (except for sex-linked inheritance which must be a small part of the total inheritance in mammals) the foregoing considerations apply also to the progeny test of the sire. However, for traits which the sire cannot express himself, there is nothing but a pedigree estimate of the sire against which to compare the accuracy of a progeny test of him. Since a progeny test surpasses even the best pedigree estimate when there are more than three progeny (provided the offspring do not resemble each other very much for any other reason except that they are by the same sire), the progeny test in such cases is much more useful and more urgently needed for the sire than for the dam. However such progeny tests will rarely tell as much about the sire as the available information correctly used will tell about a dam of nearly equal age. Enthusiasm over some such slogan as the current one that "The next best thing to a proved sire is the son of a proved sire" should not cause us to forget that such a son is one generation of Mendelian segregation away from this sire and that half of his inheritance (a little more than half if the probably small amount of sex-linked inheritance in mammals is also considered) comes from his dam.
AN ANIMAL’S BREEDING VALUE

whose breeding value may usually be estimated more closely than that of his proved sire, if the available information is fully and fairly used.

Although the progeny test will not often be superior to the dam’s own record (where no differences of selection are involved), it will be noticed from equation 2 that $B = .5$ even when $n$ and $m$ are only one. Therefore even the most fragmentary progeny test is worth much as an indicator of the dam’s genotype. Naturally, as in any other prediction, both indicators should be used where both are available. The principles of multiple correlation govern the amount of attention to be paid to each indicator where both are used. However the formulae for the relative amount of attention to be given to the two records are not only complex but also involve a term (which is usually indeterminable) for the degree to which the dam’s record and her daughters’ records resemble each other because of being made under common environment. The ratio between the standard regression coefficients will be in the same direction from unity as is the ratio between the primary correlations ($B$ in the earlier algebraic discussion) but will be more extreme. Perhaps as good an approximate rule as any would be to give a daughter’s production nearly half as much attention as the cow’s own record where there is only one record for each, but to give between $\frac{2}{3}$ and $\frac{3}{4}$ as much attention to the daughter average as to the cow’s own record where there are at least two or three offspring. This rule would be approximately correct if daughters and dams had been equally selected from their generations. Where the dams have already been highly selected, more importance than this should be given to the progeny test in making further selections but it does not seem possible to develop any simple general rule for this, since such a rule would include terms for the intensity of selection among dams and among daughters and also a term for the extent to which the trait is really hereditary.

In actual practice selection will also be based in part on pedigree. Chronologically pedigree becomes available first, then the individual’s own performance and the progeny test comes latest. As some selection is practiced on each basis, the possibilities of further gains by additional selections on the same basis rapidly diminish unless new information becomes available. The sampling nature of Mendelian inheritance sets a lower limit on the usefulness of pedigree than is inherently necessary for either of the other two bases of selection. Hence the very early selections on the basis of pedigree come near exhausting the possibilities in that direction, although increased knowledge of the performance of parents and other ancestors or collateral relatives does make it profitable occasionally to revise an earlier pedigree estimate. On the other hand, knowledge of the individual’s own performance can continue to increase at a practically important rate as long as it lives and knowledge of its progeny can increase
as long as previously untested progeny come on test and (at a slower rate) as long as progeny already tested continue to be tested at later ages.

This leads to the general picture that pedigree occupies first place only for such selections as must be made before the other two criteria are available. It is very distinctly in third place after a few such early selections have been made. Individual performance occupies first place after it becomes available and until considerable use has been made of it. Then progeny test is the most useful basis for further selections but when it is much used for such selections, the possibilities in that tend to be exhausted and the individual's own performance might again assume first place. Among mature animals under practical conditions individual performance and progeny test therefore will vary or alternate as most useful for further selections, according to the amount of new knowledge becoming available about each and according to how nearly the possibilities for selection on the basis of the existing knowledge about each have already been exhausted.

SUMMARY

1. The biometrical relations governing the relative accuracy of progeny test and of the parent's own performance as indicators of the parent's breeding value are presented and discussed.

2. A solution under the simplest conditions is presented algebraically and graphically. Under those conditions there must be at least five offspring before the progeny test in a whole population will usually be a more accurate indicator of the parent's breeding value than the parent's own performance.

3. Most deviations from those simplest conditions have only slight effects on the solution. However any general resemblance between the offspring for any other reason than that they are half-sibs through the parent in question sets serious limits on the accuracy of the progeny test. On the other hand if the parents or the records used to represent them are more highly selected than the offspring or their records, the progeny test may become relatively more accurate than under the simple conditions for which the algebraic solution is given.

4. The progeny test is needed most for traits which cannot be expressed in one sex and for traits which are but slightly hereditary.

5. The bases for estimating breeding value are pedigree, own performance, and progeny test. As fast as some selection is practiced on one of these bases, the possibilities for further progress by additional selection on the same basis rapidly diminish and correspondingly increased attention should be given to one of the other bases.

REFERENCES