Ann. Hum. Genet., Lond. (1972), **36**, 129
Printed in Great Britain

Fisher's 'fundamental theorem' made clear

BY GEORGE R. PRICE

The Galton Laboratory, University College London, London, W.C.1

1. INTRODUCTION

It has long been a mystery how Fisher (1930, 1941, 1958) derived his famous 'fundamental theorem of Natural Selection' and exactly what he meant by it. He stated the theorem in these words (1930, p. 35; 1958, p. 37): '*The rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time.*' And also in these words (1930, p. 46; 1958, p. 50): 'The rate of increase of fitness of any species is equal to the genetic variance in fitness.' He compared this result to the second law of thermodynamics, and described it as holding 'the supreme position among the biological sciences'. Also, he spoke of the 'rigour' of his derivation of the theorem and of 'the ease of its interpretation'. But others have variously described his derivation as 'recondite' (Crow & Kimura, 1970), 'very difficult' (Turner, 1970), or 'entirely obscure' (Kempthorne, 1957). And no one has ever found any other way to derive the result that Fisher seems to state. Hence, many authors (not reviewed here) have maintained that the theorem holds only under very special conditions, while only a few (e.g. Edwards, 1967) have thought that Fisher may have been correct – if only we could understand what he meant!

It will be shown here that this latter view is correct. Fisher's theorem does indeed hold with the generality that he claimed for it. The mystery and the controversy result from incomprehensibility rather than error.

2. THE MEANING OF THE THEOREM

This section will explain what Fisher's theorem states. The following section will give the evidence showing that the meaning explained here is indeed what Fisher meant.

Let M = the mean fitness in some population. (The precise definition of M will be given later.) Let dM = the change in M from time t to time $t+dt$. We can think of the change dM as made up of two components, one being the effect of natural selection, and the other being due to environment change. Therefore, in a purely formal way, let us write

$$dM = \partial_{\mathrm{NS}} M + \partial_{\mathrm{EC}} M. \tag{2.1}$$

Here $\partial_{\mathrm{NS}} M$ represents the change in M due to natural selection and $\partial_{\mathrm{EC}} M$ is the change in M due to environment change effects. (We are using partial differential and derivative notation in a slightly unconventional way, which will be made quite explicit below.) Fisher's 'fundamental theorem of Natural Selection' is

$$\partial_{\mathrm{NS}} M/\partial t = W, \tag{2.2}$$

where W is what Fisher termed the 'genetic variance in fitness' but which more commonly at present would be called the additive genetic variance in fitness. To make the theorem more explicit we can add subscripts indicating time:

$$\partial_{\mathrm{NS}} M_t/\partial t = W_t. \tag{2.3}$$

These subscript t's convey the sense of the repeated 'at that time' in Fisher's main statement of the theorem: 'The rate of increase in fitness of any organism *at any time* is equal to its genetic

variance in fitness *at that time*' (my italics). The main cause of misunderstanding about the theorem is that everyone has supposed that Fisher was talking about the total change dM/dt rather than just the fraction of this due to natural selection.

We next consider how $\partial_{\mathrm{NS}} M_t$ is to be defined. Let us write

$$M = c + \sum_{l,k} \beta_{m,lk} Q_{lk}. \tag{2.4}$$

Here l designates a gene locus, k designates a particular allele at the given locus, Q_{lk} is the population frequency of allele k of locus l, $\beta_{m,lk}$ is the partial linear regression of fitness on allele lk frequency in the members of the population (to be more precisely defined in a later section), and c is a constant. To simplify equation (2.4) we now replace the regression coefficients $\beta_{m,lk}$ by new regression coefficients $b_{m,lk}$ that incorporate the constant c:

$$M = \sum_{l,k} (B_{m,lk} + \lambda c) Q_{lk} = \sum_{l,k} b_{m,lk} Q_{lk}, \tag{2.5}$$

where λ = the reciprocal of the total number of gene loci. We further simplify by omitting the m, l and k subscripts, so that equation (2.4) becomes simply

$$M = \Sigma bQ.$$

Now we add subscripts to indicate time. Let $T = t + dt$. Obviously

$$M_t = \Sigma b_t Q_t, \quad M_T = \Sigma b_T Q_T,$$

and

$$dM_t = M_T - M_t = \Sigma b_T Q_T - \Sigma b_t Q_t. \tag{2.6}$$

We are now ready to explain $\partial_{\mathrm{NS}} M$ and $\partial_{\mathrm{EC}} M$ in equation (2.1). Though Fisher did not use the present partial differential notation nor the symbols b, Q and T, his point of view about the natural selection and environment change components of dM can be expressed as follows:

$$\partial_{\mathrm{NS}} M_t = \Sigma b_t Q_T - \Sigma b_t Q_t, \tag{2.7}$$

$$\partial_{\mathrm{EC}} M_t = \Sigma b_T Q_T - \Sigma b_t Q_T. \tag{2.8}$$

It should be noted that (2.7) and (2.8) add together to give (2.6) (with the left sides of equations (2.7) and (2.8) adding in accordance with equation (2.1)). Fisher adopted the somewhat unusual point of view of regarding dominance and epistasis as being environment effects. For example, he writes (1941): 'A change in the proportion of any pair of genes itself constitutes a change in the environment in which individuals of the species find themselves.' Hence he regarded the natural selection effect on M as being limited to the additive or linear effects of changes in gene frequencies, while everything else – dominance, epistasis, population pressure, climate, and interactions with other species – he regarded as a matter of the environment. It can be seen that (2.7) expresses accurately the change in M due to additive effects of changes in gene frequencies. Thus it is consistent with what Fisher thought of as the change in mean fitness due to natural selection. Since (2.7) and (2.8) add to give (2.6), it therefore follows that (2.8) expresses all those other effects on mean fitness that Fisher thought of as constituting environment change effects. Changes in the environment (including the 'genic environment') of course cause changes in the regression coefficients b or β. Hence (2.8) expresses (according to Fisher's point of view) the effect of environment change without natural selection – for Fisher thought of changes in gene frequencies as being due to natural selection, and equation (2.8) is for constant gene frequencies. Therefore, surprising though it may seem to one upon first considering it, it is not un-

reasonable to think of (2.7) and (2.8) as describing, respectively, the effects of natural selection and environment change on mean population fitness.

It will be shown later that

$$(\Sigma b_t Q_T - \Sigma b_t Q_t)/dt = W_t. \tag{2.9}$$

Combining this result with (2.7) we obtain the 'fundamental theorem'. The geneticists who have published derivations of what they thought was Fisher's theorem have in most cases shown that the relation $dM_t/dt = W_t$ holds under special conditions. These conditions are exactly the conditions that eliminate all 'environment change' (in Fisher's extended sense) so that all $b_T = b_t$, with the consequence that (2.6) and (2.7) became equivalent. However, the matter of the constancy of the regression coefficients in (2.7) and (2.9) should not be misunderstood. Suppose that we use multiple primes to indicate successive infinitesimally later times. Thus let W be the variance at time t, W' = the variance at time $t+dt$, W'' = the variance at $t+2dt$, and so on; and correspondingly with M, b and Q. With this extended notation for indicating times, we write (2.3) and (2.9) combined as follows:

$$\partial_{\mathrm{NS}} M/\partial t = (\Sigma bQ' - \Sigma bQ)/dt = W,$$

$$\partial_{\mathrm{NS}} M'/\partial t = (\Sigma b'Q'' - \Sigma b'Q')/dt = W',$$

$$\partial_{\mathrm{NS}} M''/\partial t = (\Sigma b''Q''' - \Sigma b''Q'')/dt = W'',$$

and so on. From the form of these equations it can be seen that, even though the derivative must always be non-negative and in any real species will be positive (since the additive genetic variance cannot be negative and is not likely to be zero for any actual species), the fitness does not necessarily increase with time. This is because, with each change in the 'environment', there is a change in what constitutes 'fitness'. What Fisher's theorem tells us is that natural selection (in his restricted meaning involving only additive effects) at all times acts to increase the fitness of a species to live under the conditions that existed an instant earlier. But since the standard of 'fitness' changes from instant to instant, this constant improving tendency of natural selection does not necessarily get anywhere in terms of increasing 'fitness' as measured by any fixed standard, and in fact M is about as likely to decrease under natural selection as to increase.

Why Fisher nevertheless thought of his theorem as important will be explained in the next section.

3. EVIDENCE SUPPORTING THE EQUATION (2.3) INTERPRETATION

The evidence that equation (2.3), as interpreted in terms of equation (2.7), is what Fisher meant is, first, that this result is mathematically correct, secondly, that it holds with exactly the generality that Fisher claimed for his theorem, thirdly, that it agrees with his derivation, and fourthly, that it is in accordance with everything said in all his discussions of the theorem. Let us now turn to Fisher's writings in order to evaluate the details of the evidence. Here we will give our main attention to the last (1958) of his three publications on the theorem. (Readers possessing the 1930 edition of his book can manage fairly well by noting that the two editions are very similar up to the beginning of page 34, page 37 of the 1958 edition corresponds roughly to page 34 of the 1930 edition, and page 46 of the 1958 edition falls on pages 42–3 in the 1930 edition.)

The first point that needs to be explained is that Fisher never defined any notation corresponding to our $\partial_{NS}M$. Consequently, whenever he states a relation involving the change in M caused by natural selection he is forced to use words rather than symbols. This is why he states two main equations in his derivation partly in words and partly in symbols (in the sentence beginning with 'Moreover' near the middle of page 37, and where he writes 'the total increase in fitness is $\Sigma\alpha\, dp$' near the bottom of that page). On the other hand, when he states an equation involving M or dM rather than $\partial_{NS}M$ he commonly writes the equation entirely in symbols as in the three equations on page 46, for of course here he encounters no difficulty in symbolic representation. And as for the question why Fisher did not define notation to express what we are expressing by $\partial_{NS}M$, it should be noted that our $\partial_{NS}M$ and $\partial_{NS}M/\partial t$ are somewhat unconventional (as was pointed out earlier). No doubt that is why he did not employ partial differentials and derivatives. Presumably he could not think of any notation expressing the correct idea in a mathematically 'proper' way – hence his recourse to words. The result has been forty years of bewilderment about what he meant, whereas if he had been willing to make a slight sacrifice of strict mathematical propriety (as I have done) he could have expressed himself in a way that everyone would have understood.

Now let us notice what words Fisher uses to express what we are expressing as $\partial_{NS}M$ or $\partial_{NS}M/\partial t$. In a few places he expresses these ideas clearly in words, for example: 'the rate of increase in the mean value of m produced by Natural Selection' (1958, pp. 45 f.); 'the rate of actual increase in fitness determined by natural selection' (1958, p. 46); 'the rate of increase in the average value of the Malthusian parameter [i.e. the fitness, which he represents by m or M] ascribable to natural selection' (1941, p. 57). Unfortunately, in the great majority of cases he leaves out the explanatory words 'produced by Natural Selection' (or 'determined by' or 'ascribable to') and leaves it to the reader to tell from context whether he is talking about M or dM or about the natural selection component in dM. For example, let us consider the verbal statements of the theorem quoted at the beginning of this paper. What Fisher should have written is something like this: 'In any species at any time, the rate of change of fitness ascribable to natural selection is equal to the [additive] genetic variance in fitness at that time.' Apparently he thought that his use of the word 'increase' made it obvious that he was referring to the natural selection effect; for he explains elsewhere that environment change must tend generally to decrease fitness (1958, pp. 41–5) and that the mean fitness 'cannot greatly exceed zero' (p. 46), which means, by elimination of these other two possibilities, that when he speaks of fitness as always *increasing* he must be referring to the natural selection effect. It was apparently a characteristic trait of Fisher's that often he would fail completely to put himself in the position of his hearers or readers, but instead assume that since his meaning was clear to himself it must be clear to others. There is probably no place where he has fallen into this error with more regrettable consequences than in his use of the word *fitness* in connexion with his 'fundamental theorem'.

A second main point that needs to be explained is Fisher's point of view concerning his theorem. Why was he more interested in an equation giving $\partial_{NS}M/\partial t$ than in one giving dM/dt? Of course the answer is that he realized that M has to remain near zero most of the time in every species or else the species would either become extinct or overwhelm the earth, and therefore dM/dt has to hover around zero, also. Consequently he felt that there was little of interest to be said about dM/dt, whereas he felt it was highly interesting and important that the natural selection component of dM/dt always equals W. His point of view was a dynamic one, a differential equation

point of view involving the forces acting on a species at a particular instant in time. One such 'force' is environment change, which Fisher shows will generally tend to decrease M. Opposed to this deterioration tendency is the force of natural selection, tending always to increase M. Concerning the balance between these two opposed forces, Fisher writes (1958, pp. 45 f.): 'The balance left over when from the rate of increase in the mean value of m produced by Natural Selection, is deducted the rate of decrease due to deterioration in the environment, results not in an increase in the average value of m, for this average value cannot greatly exceed zero, but principally in a steady increase in population.' He illustrates this by the following differential equation:

$$dM/dt + M/C = W - D. \tag{3.1}$$

Here D is the rate of 'deterioration of the environment', C is a constant, and the term M/C (which Kimura, 1958, explains more clearly than Fisher does) gives the effect of population pressure tending to decrease M. If we subtract M/C from both sides of (3.1) we obtain an explicit equation for dM/dt:

$$dM/dt = W - (D + M/C). \tag{3.2}$$

Here $-(D+M/C) = \partial_{\mathrm{EC}} M/\partial t$. It is interesting to note that Fisher felt such little interest in an equation for dM/dt that he does not even bother to show (3.1) solved for dM/dt as in our (3.2), nor does he distinguish (3.1) by calling it a 'theorem', but merely mentions it to illustrate one point that he makes in the course of discussing his 'fundamental theorem'. This should suffice to demonstrate that the 'increase in fitness' mentioned in his statement of the 'fundamental theorem' cannot possibly mean dM/dt.

Nor did Fisher think of his theorem as equivalent to equation (2.9). For one thing, he would hardly have compared such an equation to the second law of thermodynamics. Furthermore, he defined special notation (upper case roman letters in 1930, lower case Greek letters in 1958) to represent functions such as our ΣbQ. He speaks of such linear regression sums as 'expected values', 'genetic values', or 'the value of the genotype [phenotype] as best predicted from the genes present'. In no case does he use this notation or this terminology in connexion with the 'fitness' mentioned in his theorem. This shows that he did not think of the 'fundamental theorem' as involving linear regression sums. (Of course (2.9) and (2.3) are mathematically identical since (2.7) is true by definition. But this does not mean that Fisher would have accepted (2.9) as a statement of his theorem. Let us note that the equation $W_t = W_t$ also is mathematically identical with (2.3).)

If the 'rate of increase in fitness' in Fisher's statement of the theorem is neither dM/dt nor $(\Sigma b_t Q_T - \Sigma b_t Q_t)/dt$, there is nothing else that it can plausibly mean but $\partial_{\mathrm{NS}} M/\partial t$. This agrees with Fisher's words quoted earlier, and it agrees with his derivation, as will be presently shown.

And if any doubt still remains in the reader's mind, here – saved for last – is the strongest and clearest item of evidence. Immediately before stating his theorem in the words quoted earlier (1930, p. 35; 1958, p. 37), Fisher writes: 'the rate of increase in fitness due to all changes in gene ratio is exactly equal to the genetic variance in fitness W which the population exhibits'. His reason for speaking of 'gene ratio' is that his 1930 derivation is in terms of two alleles per locus, so that mention of a 'ratio' is equivalent to mentioning the two allele frequencies at a locus. In the next paragraph after the statement of the theorem, 'ratio' in 1930 becomes 'frequencies' in 1958, but Fisher overlooked making this change in the passage just quoted. If we

make this change now, then Fisher is saying that 'the rate of increase in fitness due to all changes in gene frequencies is exactly equal to the genetic variance in fitness W', which is an explicit statement of our equation (2.2) as interpreted with (2.7).

4. FISHER'S THREE PUBLICATIONS ON HIS THEOREM

In addition to the central confusion resulting from the use of the word *fitness* in two highly different senses, Fisher's three publications on his theorem contain an astonishing number of lesser obscurities, infelicities of expression, typographical errors, omissions of crucial explanations, and contradictions between different passages about the same point. It is necessary to clarify some of this confusion before explaining the derivation of the theorem.

We will look first at the 1930 edition of his book. (Much of what is said here applies to the 1958 edition also.) Chapter II begins with four sections about the 'Malthusian parameter of population increase', represented by m, and 'reproductive value', represented by v. This part of the chapter has been explained in a previous paper in this journal (Price & Smith, 1972). Then comes a section headed 'The genetic element in variance', where Fisher defines two variables, 'average excess', represented by a, and 'average effect', represented by α, and then derives an equation for W in terms of a and α. The definition of average effect on page 32 (in both editions) contains the most confusing published scientific writing I know of. What seems to have happened is that Fisher wrote an earlier version of this section in which he defined both a and α in terms of haploid 'half-individuals' or haploid chromosome sets. Presumably he explained this in the part about the first variable that he defines, a. Then he apparently decided it was clearer to define a in terms of diploid complete individuals, while he retained the half-individual definition of α. When he changed the first definition he removed the explanation about 'half-individuals' that I assume he must originally have provided. Unfortunately he did not realize that this made his definition of α almost impossible to understand, especially since on page 32 he refers to 'two moieties' which he claims to have mentioned earlier, so that the reader assumes that he refers to the 'two groups' of diploid individuals mentioned on page 30. The result is total confusion, since actually it can be shown that the 'moieties' must consist of haploid half individuals, and nothing on page 32 makes sense if the reader does not understand this.

In addition, these corrections are needed in the section on 'The genetic element in variance'. (i) On page 32, change 'term' to 'factor' in line 31, change 'individual' to 'half-individual' in line 30, and rewrite line 37 as: 'in the $2Np$ half-individuals of one moiety and $(-p\alpha)$ in the $2Nq$ half-individuals of'. (ii) Several expressions containing α must be multiplied by 2, to give $2pqa\alpha$ at the top of page 32, and $2pqa\alpha$ and $\Sigma(2pqa\alpha)$ at the top of page 33. (On page 37 of the 1958 edition, Fisher implies that the 1930 omission of the 2's was correct for the haploid treatment that he employed then. However, this is not so, as Fisher himself confirms by adding the 2's while leaving the rest of the haploid treatment unaltered in the 1958 edition.)

The derivation is completed in the following section, headed 'Natural Selection'. Here a few statements are made and some simple equations are shown, and not very much seems to be happening, and then suddenly the theorem is stated. The effect is like sleight-of-hand, and the reader goes back over those few lines again and again wondering where the theorem came from. The main part of the mystery is cleared up if one keeps in mind that 'increase in fitness' means 'increase in fitness ascribable to natural selection'. Other points that need explanation are these:

(i) In the first four sections m was defined as a population measure; at the beginning of the 'Natural Selection' section m changes into an individual or genotypic measure of fitness. (ii) Fisher states that he is using m as he used a variable x in defining and explaining average excess and average effect. This is not strictly correct. Actually his way of defining average excess for m is different from the page 30 definition. (iii) As before, 2's need to be added in several places, to give $2pq\alpha x$ near the middle of page 34, $2\alpha dp$ five lines below that, $2\alpha dp = 2pq\alpha x\,dt$ at the bottom of the page, and

$$\Sigma(2\alpha dp) = \Sigma(2pq\alpha x)\,dt = W\,dt$$

at the top of the next page. (iv) Weighting by reproductive value, v, is omitted in all equations. This means that most of the equations in this section are not strictly correct as they are written. Then at the end of the derivation Fisher suggests that v weighting should be used in the calculation of population gene frequencies p and q.

The theorem is discussed in the last half of the 'Natural Selection' section and in the remainder of the chapter. The following mistakes need correction. (i) At the bottom of page 35 add a coefficient of 2 to the expression $\alpha\delta p$ and a coefficient of 4 to the expression $pq\alpha^2$. (ii) Delete the 2 from the expression set on a separate line near the top of page 36. (iii) In the section headed 'The nature of adaptation' Fisher gives a correct relation at the top of page 39 for the case $n = 3$, but for higher n he is seriously in error (I suspect due to a mistake in integration) and greatly underestimates the probability that a random change will move closer to the centre of the hypersphere. (iv) The equations at the bottom of page 42 and near the top of page 43 should contain $C(W-D)$ in place of $(W-D)/C$.

Now we may briefly consider the 1941 paper. This is mainly about the variables average excess and average effect, which here are explained with much clarity. Then comes a brief summary of the main part of the derivation of the theorem. Unfortunately this is just as obscure as its 1930 counterpart. As far as I know, this paper contains no mathematical or typographical error.

Lastly we consider the 1958 revised edition of the book. Here something quite unusual has happened in the section on 'The genetic element in variance'. As was mentioned earlier, in the 1930 edition of the book (and also in the 1941 paper) the definitions of average excess and average effect are in terms of two alleles at every locus, with p being the population frequency of one allele and $q = 1-p$ being the frequency of the other. The 1958 version of this section starts out in the same way, giving the two-allele definitions only slightly changed from the 1930 wording. Then where the 1930 version ends on page 34, the 1958 version changes abruptly and without any word of introduction or explanation to entirely new definitions that apply with any number of alleles per locus. Some of the symbols are changed from those defined at the beginning of the section (so that, for example, on page 35 ξ is defined to have the meaning that is given to X in the first part of the section), and a new form of 'average excess' is defined that is not mathematically equivalent to the 'average excess' defined four pages earlier. Since Fisher refers at the bottom of page 36 to an 'expression obtained in the first edition of this book', whereas in fact this expression is also obtained in the 1958 edition (at the top of page 33), it seems clear that the 1958 edition must contain passages that Fisher intended to have deleted. A possible explanation for this is that he first prepared a slightly revised version of the 1930 treatment. Then, while he was reading the proofs for this, he decided to remove the limitation to two alleles per locus. Accordingly, he prepared an entirely new version for most of the 'Genetic element in variance' section and the beginning of the 'Natural Selection' section and mailed it to the publisher of the new edition.

Unfortunately he must have failed to make clear to the publisher that most of the earlier two-allele treatment was to be deleted, and both versions were printed, one after the other. Probably no later proof was sent to Fisher, for the added sections that describe the new multiple allele treatment contain an astonishingly large number of typographical errors.

The corrections needed to put the 1958 edition into the form that Fisher probably intended are as follows. (i) Everything from line 19 of page 31 to line 13 of page 34 should be deleted. (ii) The following sentence, which appears in the 1930 edition, should probably be added on page 37 at the beginning of the 'Natural Selection' section: 'The definitions given above may be applied to any characteristic whatever; it is of special interest to apply them to the special characteristic m which measures the relative rate of increase or decrease.' (iii) All subscripts shown on pages 34–6 as 'ik', 'lk', or '$1lk$' (eight such cases) should be '$1k$'. (iv) In the second equation on page 34 the numerator should contain $S(x_{11})$ instead of $S(n_{11})$ and $S(x_{1k})$ instead of $S(n_{lk})$. (v) Fisher uses Σ' for summation over all alleles at one locus, and Σ for summation over all loci; primes should be added to the two sigmas near the bottom of page 34 and to the sigmas on lines 5 and 15 of page 36. (vi) In line 4 of page 35, 'as' should be inserted after 'population'. (vii) The equation $\alpha_2 = -p_2\delta$ near the bottom of page 36 should be $\alpha_2 = -p_1\delta$. (viii) The expression $\Sigma\alpha dp$ near the bottom of page 37 should be $\Sigma\Sigma'(2\alpha dp)$.

In addition, the following should be noted. (i) The verbal definition of average effect given in the top half of page 35 does not agree with the definition in terms of 'normal equations' in the bottom half of the same page. Clearly the latter is the correct definition, and Fisher has erred in the verbal definition as a result of following too closely the wording he used in defining the two-allele form of average effect at the bottom of page 31. (ii) With the two-allele forms of a and α, $a = \alpha$ under Hardy–Weinberg conditions, but this is not true for the new multiple-allele forms defined in 1958. Consequently the remark about a and α being 'no longer distinct' with random mating, near the bottom of page 38, is not correct with the 1958 treatment though this argument is correct in the 1930 edition. (iii) The 1958 edition has 2's added where they are needed on pages 32, 33 and 37 and has the correct form $C(W-D)$ in the equations on page 46, but most of the other errors listed above for the 1930 edition remain uncorrected in the later edition.

Finally, it may be useful to mention some matters of terminology that apply to all three publications. Fisher uses 'genotypic' for 'phenotypic', uses 'factor' where I would say 'locus', and writes 'loci' where I would write 'chromosome sets' or 'haploid half-individuals'. Perhaps his most surprising terminological usage is that he sometimes uses 'actual' to mean 'theoretical' (which one would think is the opposite of 'actual'). For example, he refers to 'the actual increase' in certain measures that would occur in an imaginary, ideal experiment that could not *actually* be performed (1941, p. 53).

Nevertheless, despite the many errors that he made, Fisher still arrives at a correct result with his theorem. Evidently he knew intuitively the result that he wanted to obtain, so that he was able to arrive at the correct result no matter how many mistakes he made *en route*.

5. THE 1958 DERIVATION

The 1958 multiple-allele derivation will now be explained. The derivation will first be given without reproductive value weighting; then a few words will be added about how to obtain the theorem in the full form that Fisher intended with weighting by reproductive value.

As was mentioned earlier, some of Fisher's equations are incorrect in the unweighted forms he shows. The cause of the difficulty is that either all variables and functions in these equations should be weighted by v or none should be weighted, and his m includes v weighting while his other variables are unweighted. To avoid this error we will define an unweighted form of m, which later will be replaced by a weighted form. Let x = age and let subscript γ designate particular genotypes. Let $r_{x\gamma}$ = the mean reproduction rate of individuals of age x and genotype γ (crediting each parent with half of each offspring), and let $\mu_{x\gamma}$ = the mean death-rate for genotype γ individuals at age x. If individual I_i is of age x and genotype γ, then m_i, the 'fitness' of I_i, is defined as

$$m_i = r_{x\gamma} - \mu_{x\gamma}. \tag{5.1}$$

Then M is the population mean of the m_i, and W_m is the additive genetic variance of m in the given population.

To clarify Fisher's definitions and use of a and α, we will add subscripts. Let ϕ designate some quantitative character and let ϕ_i = the value of character ϕ in individual I_i. Let $a_{\phi,lk}$ represent the average excess of allele lk for character ϕ. Fisher's 1958 definition of average excess (page 34) is equivalent to

$$a_{\phi,lk} = \operatorname{cov}(\phi, q_{lk})/Q_{lk}, \tag{5.2}$$

where Q_{lk} is the population frequency of allele lk, as defined at the beginning of this paper, q_{lki} is the individual frequency of allele lk in individual I_i (defined in Price, 1970), and $\operatorname{cov}(\phi, q_{lk})$ is the covariance of ϕ_i and q_{lki}. Let $\alpha_{\phi,lk}$ represent the average effect of allele lk on character ϕ. Fisher's 1958 definition of average effect (lower half of page 35) is equivalent to

$$\alpha_{\phi,lk} = \frac{1}{2}\left[\beta_{\phi,lk} - \sum_{k=1}^{s(l)} (\beta_{\phi,lk} Q_{lk})\right], \tag{5.3}$$

where $\beta_{\phi,lk}$ is the partial linear regression of ϕ_i on q_{lki}, and $s(l)$ is the number of different alleles for locus l present in the population. (Of course at each locus only $s(l)-1$ regression coefficients can be independently determined. The equation (5.3) definition holds with any consistent set of regression coefficients.)

Now we will derive the relation

$$W_\phi = \sum_{l,k} (2Q_{lk} a_{\phi,lk} \alpha_{\phi,lk}), \tag{5.4}$$

which Fisher states a little below the middle of page 36 (but using p instead of Q, omitting subscripts, and without using the symbol W). Substituting from (5.2) and (5.3) into (5.4) we obtain

$$\sum_{l,k} (2Q_{lk} a_{\phi,lk} \alpha_{\phi,lk}) = \sum_{l,k} [\beta_{\phi,lk} \operatorname{cov}(\phi, q_{lk})] - \sum_{l,k} [\operatorname{cov}(\phi, q_{lk}) \sum_{k} \beta_{\phi,lk} Q_{lk}]. \tag{5.5}$$

The final term of (5.5) can be shown to equal zero since

$$\sum_{k=1}^{s(l)} [\operatorname{cov}(\phi, q_{lk})] = \sum_{k=1}^{s(l)} \sum_{i=1}^{N} [(\phi_i - \bar{\phi})(q_{lki} - Q_{lk})]/N,$$

which obviously equals zero since q_{lki} and Q_{lk} summed over all k both equal 1. This leaves the first term on the right side of equation (5.5). To put this into a more convenient form we will make use of the following circumflex notation for linear regression estimated values (or what Fisher on page 35 terms 'the value of the genotype as best predicted from the genes present'):

$$\hat{\phi}_i = c_\phi + \sum_{l,k} b_{\phi,lk} q_{lki}. \tag{5.6}$$

Now we write

$$\sum_{l,k}[\beta_{\phi,lk}\,\mathrm{cov}\,(\phi, q_{lk})] = \mathrm{cov}\,[\phi, \sum_{l,k}(\beta_{\phi,lk} q_{lk})] = \mathrm{cov}\,(\phi, \hat{\phi}). \tag{5.7}$$

As Fisher shows in the first part of page 36, and as can easily be verified, $\mathrm{cov}\,(\phi, \hat{\phi}) = \sigma^2(\hat{\phi}) = W_\phi$ for any variable. Hence (5.4) follows from (5.5) and (5.7).

Now we proceed to the final part of the derivation, on page 37. Here the opening sentences tell us that we should replace ϕ by m in the expressions just given. Next Fisher tells us that $\Sigma'(2pa\alpha)$, which is

$$\sum_{k}(2Q_{lk} a_{m,lk}\alpha_{m,lk})$$

in our notation, is the contribution of a single locus to W_m, and that this summed over all loci equals W_m. This is of course the relation that we just derived, our equation (5.4). In the next sentence Fisher speaks of the change in the 'average fitness of the species' due to a change dp in the frequency of a single gene; what he means here can perhaps be expressed as

$$\partial_{\mathrm{NS}} M/\partial Q_{lk} = 2\alpha_{m,lk}. \tag{5.8}$$

Next he gives a simple equation, which by (5.2) can be rewritten as

$$d\log_e Q_{lk}/dt = \mathrm{cov}\,(m, q_{lk})/Q_{lk}. \tag{5.9}$$

This equation can be obtained from equation (A 24) of Price (1972) by removing the weighting variables, replacing r by m, and dividing both sides of the equation by Q_{lk}. Multiplying both sides of (5.9) by $2Q_{lk}\alpha_{m,lk}dt$, we obtain Fisher's equation $(2\alpha)\,dp = (2pa\alpha)\,dt$, which we rewrite in our more explicit notation as

$$2\alpha_{m,lk} dQ_{lk} = 2\alpha_{m,lk}\,\mathrm{cov}\,(m, q_{lk})\,dt. \tag{5.10}$$

If we multiply both sides of (5.8) by ∂Q_{lk} and substitute into (5.10) we obtain

$$\partial_{\mathrm{NS}} M_{(lk)} = 2\alpha_{m,lk}\,\mathrm{cov}\,(m, q_{lk})\,dt. \tag{5.11}$$

Then Fisher speaks of summing over all alleles, first at one locus and then at all loci. This gives the following set of equations, which I give with the correction previously mentioned, and with $\partial_{\mathrm{NS}} M$ substituted for Fisher's equivalent verbal expression:

$$\partial_{\mathrm{NS}} M = \Sigma\Sigma'(2\alpha\,dp) = dt\,\Sigma\Sigma'(2pa\alpha) = W\,dt. \tag{5.12}$$

Or, more explicitly:

$$\partial_{\mathrm{NS}} M = \sum_{l,k}(2\alpha_{m,lk} dQ_{lk}) = dt \sum_{l,k}(2Q_{lk} a_{m,lk}\alpha_{m,lk}) = W_m dt. \tag{5.13}$$

Now we substitute in (5.13) from (5.2) and (5.3). Here we can simplify by omitting the irrelevant summation term of (5.3), which has zero effect in the expressions of (5.13) just as it has zero effect in (5.5). Thus we replace $\alpha_{m,lk}$ by $\frac{1}{2}\beta_{m,lk}$. With these substitutions, (5.13) becomes

$$\partial_{\mathrm{NS}} M = \sum_{l,k}(\beta_{m,lk} dQ_{lk}) = dt \sum_{l,k}[\beta_{m,lk}\,\mathrm{cov}\,(m, q_{lk})] = W_m dt. \tag{5.14}$$

This gives an overall view of the derivation of the theorem. The first equality is equation (2.7) with β substituted for b. (Note that this substitution is proper in (2.7), though not in (2.8).) The second equality in (5.14) follows easily from (5.10). And the last equality in (5.14) follows from (5.7). If we divide both sides of (5.14) by dt, the result is (2.2), the 'fundamental theorem'. Thus the derivation is complete. (It may be noted that if our aim were merely to derive the theorem rather than to explain how Fisher derived it, the derivation can be accomplished far more simply if we work entirely with regression coefficients and covariances without using Fisher's special 'average excess' and 'average effect' variables.)

Lastly we consider how to add weighting by v. Here it will be convenient to make use of the 'weighted statistical function' notation defined in Price (1972). Population gene frequencies weighted by v are defined in equation (19) of Price & Smith (1972). For the weighted form of m_i we will not follow the procedure Fisher seems to have had in mind, of defining m for a genotype by applying his page 26 equation to a population consisting of all individuals of that genotype – for it can be shown that this procedure leads to difficulties and can sometimes give paradoxical results. Instead, we can define m for a particular genotype after the manner of equation (14) of Price & Smith:

$$m_{cx\gamma} = \left(\frac{\partial \log v_{cx}}{\partial_x}\right)_c + r_{cx\gamma}\frac{v_0}{v_{cx}} - \mu_{cx\gamma}. \tag{5.15}$$

Here r replaces the b used in Price & Smith, γ designates a genotype, and the other variables on the right side of (5.15) are as defined in Price & Smith. We can define m_i, the fitness of I_i at time t, in terms of $m_{cx\gamma}$ given by (5.15). Alternatively, we might define m_i by

$$m_i = (v_{i,T} + \tfrac{1}{2}n_{i,t}v_0 - v_{i,t})/(v_{i,t}dt). \tag{5.16}$$

Here v_0 = the reproductive value of newly conceived individuals, $v_{i,t}$ = the reproductive value of I_i at time t, $v_{i,T}$ = the reproductive value of I_i at time $T = t+dt$ (defined to equal zero if I_i dies during the interval dt), and $n_{i,t}$ = the number of offspring conceived by I_i during the interval dt. Of course some conventions of smoothing must be applied with the (5.16) definition. If this is done, the (5.15) and (5.16) definitions are equivalent. M is defined as $\text{ave}_v m$, the weighted population mean of m_i. Weighted regression coefficients are determined by weighted 'normal equations' that take the form

$$\sum_i [v_i(q_{\lambda\kappa i} - Q_{\lambda\kappa,v})(\phi_i - \text{ave}_v\phi)] = \sum_i [v_i(q_{\lambda\kappa i} - Q_{\lambda\kappa,v})\sum_{l,k}(\beta_{\phi,lk}(q_{lki} - Q_{lk,v}))]. \tag{5.17}$$

Weighted additive genetic variance is defined by $W_{\phi,v} = \text{var}_v\phi$.

With these weighted variables (5.14) becomes

$$\partial_{\text{NS}}M = \sum_{l,k}(\beta_{m,lk}dQ_{lk,v}) = dt\sum_{l,k}[\beta_{m,lk}\,\text{cov}_v(m, q_{lk})] = W_{m,v}dt, \tag{5.18}$$

from which the weighted form of the 'fundamental theorem' is easily obtained. Here it is to be understood that we are using the weighted forms of M, m and b though this is not shown explicitly in (5.18).

6. DISCUSSION

Let us briefly consider what Fisher did – and did not – accomplish.

First of all, the generality of his theorem is very great since it depends only on statistical smoothing through large population size and on assumptions of absence of meiotic and gametic selection that are involved in the derivation of (5.9). We may next note that the 'fundamental theorem' is very probably the most that anyone has yet been able to say *correctly* about evolutionary increase in fitness under general and realistic natural conditions. Thus the theorem is by no means a trivial, uninteresting result.

Still one feels disappointed that it does not say more. One defect is the device of treating non-additive gene effects as 'environment'. (Kimura, 1958, treats these non-additive effects in an interesting way in a paper based on the usual interpretation of Fisher's theorem as an equation for dM/dt.) A much graver defect is the matter of the shifting standard of 'fitness' that gives the paradox of M tending always to increase and yet staying generally close to zero. Much more

interesting would be a theorem telling of increase in 'fitness' defined in terms of some fixed standard. Thus there is a challenge here to find a deeper definition of this elusive concept 'fitness' and to give a deeper and sharper explanation of why it increases and under what conditions.

Meanwhile I trust that the present paper corrects any diminution in Fisher's mathematical reputation resulting from the common belief that he was seriously mistaken about his theorem. Doubtless this paper also adds considerably to his reputation for incomprehensibility.

SUMMARY

Fisher's 'fundamental theorem of Natural Selection' is mathematically correct but less important than he thought it to be. It concerns the natural selection component in dM, the change in population fitness. Fisher's explanations of his theorem are afflicted by a truly astonishing number of obscurities, infelicities of expression, typographical errors, omissions of crucial explanations, and contradictions between different passages about the same point. The theorem is derived here in the full form that Fisher intended (continuous time model, with weighting by reproductive value).

I thank Professor Cedric A. B. Smith for much help throughout the long period of time during which work on the problem of understanding Fisher was intermittently pursued, and I thank the Science Research Council for financial support.

REFERENCES

CROW, J. F. & KIMURA, M. (1970). *An Introduction to Population Genetics Theory*. New York: Harper and Row.
EDWARDS, A. W. F. (1967). Fundamental theorem of natural selection. *Nature, London* **215**, 537–8.
FISHER, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press.
FISHER, R. A. (1941). Average excess and average effect of a gene substitution. *Annals of Eugenics* **11**, 53–63.
FISHER, R. A. (1958). *The Genetical Theory of Natural Selection*, 2nd ed. New York: Dover Publications.
KEMPTHORNE, O. (1957). *An Introduction to Genetical Statistics*. London: Chapman and Hall.
KIMURA, M. (1958). On the change of population fitness by natural selection. *Heredity* **12**, 145–67.
PRICE, G. R. (1970). Selection and covariance. *Nature, London* **227**, 520–1.
PRICE, G. R. (1972). Extension of covariance selection mathematics. *Annals of Human Genetics* **35**, 485–490.
PRICE, G. R. & SMITH, C. A. B. (1972). Fisher's Malthusian parameter and reproductive value. *Annals of Human Genetics* **36**, 1–7.
TURNER, J. R. (1970). Changes in mean fitness under natural selection. In *Mathematical Topics in Population Genetics* ed. K. Kojima. Berlin: Springer-Verlag.