Sewall Wright: gene interaction and the Shifting Balance Theory

MICHAEL J. WADE

‘Context and interaction are of the essence’ (Lewontin 1974: 318.)

1. INTRODUCTION

1.1 The question of epistasis in evolution

The role of gene interactions in adaptive evolution is not well understood despite the frequent use of terms such as ‘co-adapted gene complexes’ and ‘genetic revolutions’ in evolutionary discussion. Although most evolutionary biologists and geneticists would acknowledge Sewall Wright’s principle of the ‘universality of gene interactions,’ there is no consensus regarding their role in evolution. Gene interactions and their statistical characterization as epistasis (the between-locus non-additive component of genetic variance) are not central concepts in current adaptation theory. The question might be put as follows: Does natural selection assemble gene combinations one gene at a time by operating on the average ‘additive’ effects of single genes or does natural selection operate to choose directly among different gene combinations or ‘systems of interacting genes?’ The Fisherian answer is that selection operates primarily on the average additive effects of single genes but, where loci are very tightly linked (e.g. inversions or supersgenes), there can be direct selection among combinations of linked loci. Wright would answer in a different way – namely, that the combination of local mass selection (operating on the ‘additive’ effects of single genes) and random genetic drift create different ‘adaptive peaks’ (each representing a unique gene combination) in different demes. Interdemic selection among the adaptive peaks occurs by differential dispersion. In this way interdemic selection operates to choose directly among different gene interaction systems or gene combinations. Fisher’s answer is more widely known and accepted whereas Wright’s extension has been largely ignored.
1.2 Definitions of epistasis

The term 'epistasis' is often used synonymously for the phrase 'gene interaction' but it has two very different meanings in genetics. In molecular and biochemical genetics, genes whose products function sequentially as substrates or catalysts in a common biochemical pathway are considered to be 'epistatic' to one another. Biochemical epistasis occurs when the expression of one gene wipes out the phenotypic effects of another gene' (Lewin 1990, p. 809). In this usage, one gene is 'epistatic' to another if the function of its gene product in a biochemical pathway is *conditional* on the success or failure of the other gene operating at a later step in the same pathway. This qualitative use of the term epistasis to indicate biochemical contingency corresponds to what might be called the epistatic effect, a linear measure of gene interaction. It is different from the meaning given to 'epistasis' by quantitative or statistical geneticists in reference to a component of genetic variance (a quadratic as opposed to linear measure). In statistical and quantitative genetics, epistasis is a populational concept describing the relationship between phenotypic variation and genetic variation (Fisher 1918; Cockerham 1954). Epistasis, along with dominance, is the 'non-additive' component of the genetic *variance* for a trait. This component of genetic variation measures the statistical effects of *variations* in gene combinations between individuals in relation to the total phenotypic variance, between individuals in a population. Epistasis between loci in statistical genetics requires genetic variation at each of at least two loci. Without genetic variation at two loci, the differences between individuals within a population cannot be attributed (in the statistical sense) to variations between individuals in gene combinations.

Genes can be epistatic to one another in the biochemical sense but not contribute to epistasis in the statistical sense. For example, using material from isogenic stocks of the fruit fly, *Drosophila melanogaster*, one can study the accumulation reaction products and thus characterize biochemical epistasis. However, there is no genetic variation within an isogenic stock by definition; all individuals share the same genetic constitution. One cannot partition components of genetic variance so there can be no statistical epistasis. Wright's principle of 'universality' of epistasis at the biochemical level (Wright 1969, pp. 59–60) does not guarantee an important or even a significant role for epistasis in evolution at the populational level. It is the statistical effects of gene interactions on the phenotypic variation that determine how gene combinations will evolve and whether or not selection can operate directly on differences between individuals in gene combinations.
1.3 Epistasis and Wright’s Shifting Balance Theory

Wright (1931) proposed his Shifting Balance Theory of evolution as a mechanism for selection to operate directly on the vast field of gene combinations. He postulated a three-phase process with all three phases acting simultaneously.

Phase 1: Random Genetic Drift where the gene frequencies at many loci in a finite subpopulation or deme drift at random about the set of gene frequencies corresponding to a local adaptive peak.

Phase 2: Mass Selection Within Demes which moves a local deme from one fitness peak to another whenever random genetic drift affects local gene frequencies sufficiently for the deme to cross over into the domain of attraction of another adaptive peak.

Phase 3: Interdemic Selection in which a deme at a high fitness peak systematically shifts the position of equilibrium of other demes to its own set of equilibrium gene frequencies by the differential dispersion of migrants away from the deme at the high peak and into neighboring demes at lower fitness peaks.

In Wright’s theory, random genetic drift and local individual selection in combination can achieve results unattainable by selection alone. This occurs as a result of different gene combinations arising between local demes in subdivided populations. The differences between demes in adaptive gene combinations would be associated with between-deme differences in average fitness. Interdemic selection, occurring by the differential dispersion of migrants out from demes of high local fitness and into demes of lower average fitness, is the process by which a favorable gene combination arising in a single locality could be introduced to other demes and thereby spread throughout a species. Individual selection operating solely in large panmictic populations was not viewed by Wright as a feasible method for selecting among gene combinations owing to recombination. Recombination breaks up gene combinations that are favored by individual selection and in this way it opposes within-deme selection. That is, gene combinations cannot be efficiently transmitted from parents to offspring because, owing to recombination, parents pass on genes and not genotypes (gene combinations) to their offspring. However, Wright proposed that locally fixed combinations causing a local adaptive peak in fitness could be exported to neighboring demes by migration over extended periods of time.

1.4 The Fisherian viewpoint

Although Wright’s major premise of the ubiquity of gene interactions and the complexity of the relationship between genes and character is shared by most evolutionary biologists, the importance of his Shifting
Balance Theory for adaptive evolution and selection among gene combinations is not similarly accepted. The alternative view, attributable in large part to R. A. Fisher (1918, 1958), is that the ultimate fate of a gene is determined by its average effect with respect to fitness within large populations and that co-dependent gene complexes arise largely as a by-product of the effects of selection on single genes. The average effect of a gene, $G$, on fitness is defined as the 'regression in the actual population of the genotypic measurement [fitness] on the number of $G$ genes' (Fisher 1941, p. 54). When there are gene interactions and small local breeding groups, the average effect of a gene on fitness can vary from group to group (see below). For this reason, Fisher insisted that:

\[ \ldots \text{the population used to determine its value comprises, not merely the whole of a species in any one generation attaining maturity, but is conceived to contain all the genetic combinations possible, with frequencies appropriate to their actual probabilities of occurrence and survival, whatever these may be, and if the average is based upon the statures attained by all these genotypes in all possible environmental circumstances, with frequencies appropriate to the actual probabilities of encountering these circumstances (Fisher 1958, pp. 30–31).} \]

The most important difference between Wright and Fisher in their views on evolution is the degree to which natural populations deviate from this ideal and the evolutionary consequences for single genes of such deviations. This may be the source of the 'intense controversy' (Provine 1986, p. 232) between Fisher and Wright over gene interactions and effective population size. Unless effective sizes are small and unless there are gene interactions, natural populations cannot significantly deviate from the Fisherian ideal.

For Wright, population subdivision and gene interactions opened the important possibility of significant variation between local demes in the average effect of a gene on fitness. In the absence of population subdivision, this is not a likely possibility and Fisher's concept of population size is important to his views on the evolutionary process:

\[ \ldots \text{I believe that N must usually be the total population on the planet} \ldots \] (correspondence between Fisher and Wright, quoted in Provine 1986, p. 255). Natural populations in Fisher's view approach the ideal population fairly closely and he expects the evolutionary fate of a gene to be determined by its species-wide average effect on fitness. With this view, it is very unlikely that random genetic drift could lead to the fixation of a gene throughout the whole of a species, except with extremely weak selection. If a gene's average effect on fitness is positive, then natural selection will sweep that gene to fixation, and conversely if its average effect is negative. Gene interactions are of little consequence in this picture because in the idealized population, a gene is "tried" by
natural selection in all possible genotypic configurations and in all relevant environments exactly in the expected frequencies of each genotype, each environment, and each gene–environment combination. In contrast, Wright focused not on the $N$ of the whole species but on the local effective population number.

More recently, Williams (1966, p. 56) has expressed considerations similar to those of Fisher:

Obviously it is unrealistic to believe that a gene actually exists in its own world with no complications other than abstract selection coefficients and mutation rates. The unity of the genotype and the functional subordination of the individual genes to each other and to their surroundings would seem, at first sight, to invalidate the one-locus model of selection. Actually these considerations do not bear on the basic postulates of the theory. No matter how functionally dependent a gene may be, and no matter how complicated its interactions with other genes and environmental factors, it must always be true that a given gene substitution will have an arithmetic mean effect on fitness in any population. One allele can always be regarded as having a certain selection coefficient relative to another at the same locus at any given point in time. Such coefficients are numbers that can be treated algebraically, and conclusions inferred from one locus can be iterated over all loci. Adaptation can thus be attributed to the effect of selection acting independently at each locus.

Genetically complex adaptations are gradually built up by natural selection operating independently on single genes. The variations in the genetic background or environment that characterize real finite populations are of little long-term consequence because they do not significantly affect the relationship of a gene to fitness.

The definitions of gene effects and gene interactions are equivalent to the statistical concepts of main effects and interactions (Anderson and Kempthorne 1954). Fisher (1918, 1925) introduced both concepts. Within statistics, the limits to the utility of these concepts were challenged by Neyman (1935) (see Traxler 1976) and within evolutionary genetics by Wright (1930, 1931). Both Neyman and Wright used remarkably similar arguments, as I show below.

2. MAIN EFFECTS AND INTERACTIONS IN STATISTICS

2.1 Definitions of main effects and interactions

In biological systems, it is often the case that several different factors are important in producing a particular condition. For example, seed yield of a plot of land might depend on several factors such as (A) the variety of plant sown, (B) the amount of fertilizer used, and (C) the rainfall
during the growing season. Furthermore, the influence of a particular factor on seed yield might depend on the level of one or more of the other factors. For example, some plant varieties (levels of factor A) might be more sensitive to low levels of rainfall (factor C) than other varieties. There are many reasons, biological as well as economic, that make it desirable to understand the individual contribution of each factor to seed yield as well as the joint effects of the factor combinations. The analysis of variance in conjunction with replicated factorial experimental designs was introduced in order to achieve this understanding of individual and joint effects (Fisher 1937).

The individual contribution of a factor to a condition averaged over all contexts, i.e. over all combinations of the other factors, is defined as the main effect of that factor (Neyman 1935; Yates 1935; Fisher 1937). In Tables 1 and 2, Neyman’s model is presented. There are three factors, A, B, and C, with two levels of each factor, 1 (present) and 0 (absent). A fully factorial design would consist of eight treatments, each of which could be represented as a set of 0’s and 1’s. For example, (1, 0, 1) represents the treatment with factors A and C present but B absent and

### Table 1

**Simple effects, main effects, simple interactions, and interactions**

| Simple effects of A: | 1. \((1, 0, 0) - (0, 0, 0) = \delta_1\) |
| Main effect of A:     | \(a = (\sum \delta_k) = (\delta_1 + \delta_2 + \delta_3 + \delta_4)/4\) |
| Simple interactions of A and B: | 1. \((1, 1, 0) - (1, 0, 0) - (0, 1, 0) + (0, 0, 0))/2 = (\delta_2 - \delta_1)/2 = \beta_1\) |
|                        | 2. \((1, 1, 1) - (1, 0, 1) - (0, 1, 1) + (0, 0, 1))/2 = (\delta_4 - \delta_3)/2 = \beta_2\) |
| Interaction of A and B: | \(ab = (\sum \beta_k)/2 = (\delta_2 - \delta_1 + \delta_4 - \delta_3)/4\) |

### Table 2

**The Neyman model of nature**

<table>
<thead>
<tr>
<th>Factors</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>AB</th>
<th>AC</th>
<th>BC</th>
<th>ABC</th>
<th>~</th>
</tr>
</thead>
<tbody>
<tr>
<td>Code</td>
<td>100</td>
<td>010</td>
<td>001</td>
<td>110</td>
<td>101</td>
<td>011</td>
<td>111</td>
<td>000</td>
</tr>
<tr>
<td>Yield</td>
<td>98</td>
<td>91</td>
<td>91</td>
<td>109</td>
<td>109</td>
<td>91</td>
<td>109</td>
<td>102</td>
</tr>
</tbody>
</table>
(0, 0, 0) represents the mean of the control treatment with no factors present. The main effect of factor A is defined as the average of four treatment mean differences (cf. Table 1), which Neyman (1935) referred to as simple effects. A simple effect is the difference observed between two treatment means when the treatments differ by only the presence and absence of the factor in question. The simple effects are summed together and then averaged to calculate the main effect of a factor. In this example, four simple effects contribute to the average effect of factor A (cf. Table 1). Each simple effect contributing to the average can be considered a context-specific contribution of a factor to a main effect, where ‘context’ refers to a constant combination of all other factors.

The first order interaction effect is defined similarly as an average of ‘simple’ interactions. It is the contribution of a pair of factors to the condition under study averaged over all contexts in which the pair of factors appears. A ‘simple’ interaction is the observed difference in a factor’s simple effect when measured in the presence or absence of some other factor. A simple interaction is a linear combination of four treatment means; it is plus or minus one-half multiplied by (a) the mean of the treatment with both factors, (b) the mean of that treatment with the first but not the second factor, (c) the mean of that treatment with the second but not the first factor, and (d) the treatment mean with neither factor (cf. Table 1). In Neyman’s example, two such simple interactions are averaged to calculate the interaction effect of the AB pair of factors. The number of terms averaged to calculate a main effect can be greater than the number of terms averaged to calculate an interaction effect (four versus two, respectively, in the present example).

In detecting a main effect of interaction, we need to calculate its standard error ($\sigma$). The variance of simple effects about the average effect for factor A is

$$\sigma^2 = \sigma^2 + \frac{1}{4} \left( (\delta_1 - \delta_4)^2 + (\delta_2 - \delta_3)^2 \right)$$ (1)

The variance of simple interactions about the interaction effect of A and B is similarly given by

$$\sigma^2 = (\delta_1 - \delta_2)^2$$ (2)

Despite having the same number of treatment means in the definitions of the main effect of A and its interaction with B ($a$ and $ab$ in Table 1), the standard error of a main effect can be smaller than the standard error of the interaction, making it more difficult to detect interaction effects than main effects. In Neyman’s example (Table 2), the main effect of A, $a$, is 12.5 with a standard error of 3.19 but the interaction $ab$ is equal to 5.5 with a standard error of 3.89 (Neyman 1935). It is this property of the definitions of main effects and interactions that can complicate the
interpretation of results in replicated, factorial experiments, as Neyman (1935) pointed out (Traxler 1976; Milliken and Johnson 1984).

2.2 The experiment as a questionnaire for Nature

Fisher considered the replicated, factorial design and analysis of variance (ANOVA) as a method for investigating and characterizing nature. The metaphor that he used to describe the statistical procedure was that of an experimenter submitting a questionnaire to Nature in the form of an experimental design (Fisher 1937). By interpreting Nature’s answers to the experimental questionnaire using ANOVA, the experimenter may learn about the characteristics of the natural world. When Nature is complex and many factors affect the condition under study, Fisher argued, the experimental study of one factor at a time is an inefficient method of investigating the causal relationships between factors and condition. He introduced the factorial design and the analysis of variance in part to address this inefficiency (Fisher 1937).

Factorial designs are more efficient than the sequential study of single factors in isolation for three reasons. First, we do not know a priori whether or not a particular factor produces its effect independent of all other possibly influential factors. The study of one factor at a time in isolation from other factors cannot enlighten us on this point. Secondly, in most biological systems, we cannot study one factor at a time. Instead, we study the effects of variations in one factor holding all other factors at a single, constant level or over a random sample of conditions unknown or uncontrollable with respect to the other factors. When factors interact, then single-factor effects estimated in this way can be confounded with the interaction effects unique to the special context defined by controlling the other factors constant. When we study a single factor and keep other possibly important factors constant, we are estimating that factor’s main effect by measuring only one of its simple effects. When we measure a factor’s performance in a single context, it may or may not be representative of the array of relevant biological contexts. Thirdly, if the magnitude of one factor’s effect is not independent of all other factors, the nature of the interdependence of factor effects may not be a simple one. The methods proposed by Fisher in the ANOVA are designed to accommodate arbitrarily complex interactions among factor effects.

The primary advantage of factorial experimental designs according to Fisher (1937) was that they permit the consistency of the effects of one factor to be explored across a field of possible factor combinations. Indeed, this is his motivation for defining the ‘main effect’ of a factor as the average of the simple effects (see above). For applied purposes, knowledge of this consistency is of considerable interest because an estimate of a factor’s main effect from a factorial design is expected to
have predictive value across a wider range of conditions than if only a single simple effect were measured. For the study of biological systems in nature, however, the lack of consistency, i.e. the variability of simple effects about the expected main effect, and the causes of such variability are also of concern. Fisher in his remarks on the additive effects of genes chose to emphasize the consistency that an estimate of 'genic main effects' offers for evolutionary prediction much in the same way that he argued for consistency of effects of any factor in a factorial design. In contrast, Wright emphasized the variability of a gene's simple effects when measured in different genetic backgrounds in his several discussions of gene interactions. The estimates of a gene's effect on fitness can vary not only in magnitude from deme to deme but also vary in sign. Wright emphasized the biological importance of the variation in the components of genic main effects, owing to gene interactions.

2.3 The problem that interactions cause for interpreting the questionnaire

Whenever there are interactions among factors, the detection and estimation of the main effects of single factors and the interaction effects of groups of factors becomes problematic (Milliken and Johnson 1984). The problem becomes more acute as the size of the experiment, i.e. the number of replicates per treatment, decreases (Neyman 1935). Because the standard error of an estimate of an interaction effect can be greater than the standard error of a main effect estimate, it can be more difficult to bound the estimate of an interaction away from zero than it is to bound the estimate of a main effect away from zero when the two kinds of effects are of equivalent magnitude. There can be a bias in the interpretation of Nature's answers to the experimental questionnaire that favors the discovery of main effects and the overlooking of interactions. One is more likely to make a Type II error (i.e. accept the null hypothesis of no effect when an effect is present) in the case of interactions than one is in the case of main effects because the power of the test for interaction can be weaker. For this reason, we are more likely to discover main effects than we are likely to discover interactions when we look for both using the ANOVA.

2.4 Neyman and Fisher

This problem was first brought to light by Neyman (1935). In reply to a paper by Yates (1935) on complex experiments and their interpretation, Neyman (1935) questioned the validity of inferences based on main effects in an ANOVA, especially when the numbers of replicates were small. Neyman's challenge was founded on the basic definitions of main effects
and interactions, their attendant measurement errors, and the problems of interpretation that can arise as a result of these definitions.

In order to illustrate the problems that could arise in the interpretation of an experiment by ANOVA, Neyman (1935) constructed a hypothetical model of nature (Table 2) and conducted 30 factorial 'experiments' using Monte Carlo simulation on this model of nature. That is, he submitted not one but 30 questionnaires to this model and used the recommended ANOVA methods to analyze nature's answers. He then compared the characterization of nature inferred from the ANOVA to the known underlying model.

He hypothesized a situation in which a factorial experiment was being employed to determine which one or combination of three fertilizers best increased crop yield. One of three factors, A, produced a true negative effect on yield of 4 per cent when added to the control (i.e. the simple effect of A alone was to decrease yield by 4 per cent). However, the simple interactions, ab, ac, and abc, involving A had a positive 7 per cent increase on yield. (The other main effects, b and c, and their simple interaction, bc, were zero.) Using this model of nature, he conducted Monte Carlo simulations of 30 factorial experiments, each with three replications per treatment, exactly analogous to the 'complex experiment' discussed by Yates (1935). Assuming a fixed rate of random errors of treatment means, he applied the ANOVA to the data from each of the 30 experiments.

In 28 of the 30 'experiments,' a significant effect was found for some factor or interaction of factors at the 0.01 level. In 27 of the 28 experiments with significant results (96.4 per cent of the time), a significant main effect of A was discovered, and in 61 per cent of the questionnaires it was the only significant effect. On the other hand, a significant interaction of A and B was discovered in five of the 28 cases (17.9 per cent of the time). These results illustrate the bias toward discovering main effects and overlooking interaction effects described above.

This was not the only problem with nature's answers to the questionnaire of the ANOVA, however. The main effect of A was estimated to be positive in all of the 28 cases where it was detected when in fact the simple effect of A alone was negative. The inference regarding the sign of the main effect of A was the opposite of its simple effect when B and C were missing.

Neyman concluded that '... the frequency of cases in which practical conclusions concerning the properties of treatment a would be entirely false is very considerable and is certainly much greater than the level of significance 0.01 ...' (Neyman 1935, p. 241). The cause of the problem was two-fold: (1) the effect that the simple interactions had when averaged to calculate the main effect of A according to definition; and, (2) the inability of the experiment to detect interactions when they were present
owing to the larger standard error of estimates of the interaction effect in ‘small’ experiments (i.e. experiments with few replications). There are ways to address these problems as outlined in Milliken and Johnson (1984).

An important remaining question is how representative of nature Neyman's example might be. The hypothetical model of Nature explored by Neyman's simulation of small experiments is ‘frivolous,’ in that the simple interactions of A with B and C are all of opposite sign to the simple effect of A alone. Factor A alone is bad for yield although in combinations with other factors it is good for yield. When these several simple effects are averaged, the sign of the main effect is the opposite of the sign of the simple effect. Indeed, the only way for the sign of a factor's main effect to be the opposite of that factor's simple effect is for the simple interactions involving one factor to be of opposite sign to the simple effect of one factor alone. In this respect, Neyman's model of nature is a special case and its results cannot be generalized. It illustrates what could happen and not what will necessarily happen. Below, I will argue not only that the main effects of genes on fitness can be of opposite sign to their interaction effects on fitness but also that they should be. Thus, Neyman's 'frivolous' special case of nature may be representative of the case of gene interactions and fitness: the case on which Wright based his Shifting Balance Theory of Evolution.

3. GENE INTERACTIONS IN ADAPTIVE EVOLUTION

3.1 Wright and Fisher

The arguments of Wright and Fisher regarding gene interactions and adaptive evolution parallel those of Neyman and Fisher regarding factor interactions. Wright did not challenge Fisher's definitions of a gene's average effect or average excess but rather questioned the utility of these concepts for evolutionary biology. He discussed the inadequacy of such definitions for describing gene action and evolution on many occasions.

In recognizing the multiplicity of genic effects or pleiotropy as a universal property of living systems (Wright 1969, pp. 59–60), he distinguished the effect of a gene on fitness from the effects of that gene on other phenotypic traits. By 'fitness', Wright meant the relative numbers of progeny contributed to the next generation by a particular genotypic or phenotypic class. Wright also considered traits for which maximum fitness was associated with an intermediate value of a quantitative phenotype. In his comments on the 'inadequacy of the simple additive concept of gene effect' (Wright 1969, p. 419), he concluded that '... all genes that approach additivity in their effects on quantitatively varying characters
will be favorable in some combinations and unfavorable in others in terms of natural selective value (fitness) and, thus, exhibit interaction effects of the most extreme sort in the latter respect' (Wright 1969, p. 420). Thus, Wright identifies fitness as a unique trait with respect to the predominance of genetic interaction effects. He stresses the point that genetic effects may be additive with respect to many characters but that fitness is a fundamentally different kind of character with strong interaction effects. This point is also made by Falconer (1981, p. 394): 'Abundant evidence proves that virtually all metric characters are genetically variable in populations that are more or less in equilibrium, including characters that affect fitness. There must therefore be genetic variance of fitness. But, since selection for fitness produces no response, there can be no additive genetic variance for fitness; so all the genetic variance for fitness must be non-additive, i.e., variance due to dominance and epistatic interactions.' Genotype-by-environment interactions could further complicate the evaluation of fitness.

Wright (1969, p. 420) elaborated on the multiplicity of the effects of single genes on the phenotype (universal pleiotropy): 'This again insures that natural selective value (fitness) is a function of the system of genes as a whole rather than something that can be assigned individual genes.' Here, Wright questions the proposition that a multivariate analysis of gene effects can result in a strictly additive partitioning of phenotypic variation among genetic factors. A partitioning adequate for describing the effects of a gene on a single character becomes inadequate when characters are combined into a whole system, a single unitary phenotype.

In his review of The Genetical Theory of Natural Selection (Fisher 1930), Wright (1930) discussed this difference between himself and Fisher with respect to gene effects. 'He [Fisher] assumes that each gene is assigned a constant value, measuring its contribution to the character of the individual (here fitness) in such a way that the sums of the contributions of all genes will equal as closely as possible the actual values of measures of the character in the individuals of the population' (Wright 1930, p. 353). (The phrase 'additive gene effects' means that 'the sums of the contributions of all genes will equal' the value of the character.) However, Wright (1930, p. 353) continues, 'Genes favorable in one combination, are, for example, extremely likely to be unfavorable in another.' That is, the additive concept of gene effect for Wright is not adequate for characterizing the relationship between a gene and fitness because the sign of a gene's contribution to fitness is not constant, but changes, depending on the genetic background.

The position taken by Fisher (1937, p. 108) is the opposite of that taken by Wright: 'In studying the properties of a system of interaction factors it has been shown (Fisher 1918) that departures from the simple additive law of interaction will usually have effects somewhat similar to
non-heritable modifications. We may therefore be confident that, even if a strictly additive interaction is not exactly realized, the mass effects of segregation in a large number of factors will closely simulate those of simple cumulative systems. Fisher equated gene interaction effects with 'non-heritable' variation and advocated the additive model of a 'simple cumulative system.' The disparity of opinion on the importance of gene interactions is so extreme as to merit emphasis here. Fisher, on the one hand, dismisses gene interactions as comparable in their effect on the evolutionary process to the transient influence of 'non-heritable factors' (a view common to some of the current models of the evolution of continuous characters, e.g. Lande 1988). Wright, on the other hand, advocated the primacy of interaction effects in evolution and developed his Shifting Balance Theory as the most effective mechanism for selecting directly on gene interactions to insure creative adaptive evolution.

3.2 The parallels between Neyman and Wright regarding interactions

Just as Neyman and Fisher disagreed on the usefulness of the ANOVA as a method for estimating and interpreting main effects when interactions were present, Wright and Fisher also disagreed about the role of gene interactions in experimental genetics and evolutionary theory. There are several interesting parallels between the examples used by Neyman and Wright in their separate arguments with Fisher emphasizing the importance of interaction effects. The first is that the definition of additive genic effects and genic interactions depends on the application of ANOVA to specific breeding designs in experimental statistical genetics. As pointed out by Anderson and Kempthorne (1954, p. 897), 'The factorial gene model . . . is an adaptation of the factorial model used in the design of experiments.' From the statistical perspective, Neyman and Wright are discussing quite similar phenomena in their respective disagreements with Fisher. The argument of Neyman (1935), although not the example, is general and concerns the interpretation of the results of any factorial experimental design in terms of main effects and interactions (Milliken and Johnson 1984). The argument of Wright (1969, ch. 15) is necessarily more specific and concerns the application of such designs to the biological problem of the study of gene action.

The second parallel is that both Neyman and Wright selected the same example when interested in emphasizing the potentially important role of interactions. Neyman (1935) illustrated the problems of interpretation that arise in the ANOVA by using the example (Table 2, discussed above) in which the effects of a factor changed sign according to the other factors present. He showed how the average main effect of a factor could be found to be of opposite sign to one simple effect and thus lead to 'very wrong' inferences or conclusions. Similarly, Wright (1930, 1931,
1969) argued that the effect of a gene on fitness was dependent on the entire genetic system in which it was embedded and that this was important in large populations with small subdivisions. In generating gene interactions for fitness from additive effects on a quantitative character, Wright considered maximum fitness to be associated with an intermediate optimum value of the character. The phrase most often used by Wright was that, in such small subpopulations or demes, a gene's effect on fitness might be 'favorable in one genetic background and unfavorable in another.' The sign of a genic main effect (or any average main effect) can be changed by varying the constellation of other interacting factors only if at least some of the interaction effects are of opposite sign to the gene's simple effect. Without such a conflict in sign between simple main and interaction effects, the sign of the average genic effect must remain constant. Thus, both Neyman and Wright appealed to the same model with conflicting direct and interaction effects, when emphasizing the importance of interactions. The example used by Neyman was designed to illustrate the potential frequency of 'very wrong conclusions' from the ANOVA for the specific case when Nature was 'frivolous.' However, Wright argued that, with regard to fitness, this model of gene action is expected to be a very general one. That is, genic main effects and interactions on fitness are expected to be similar to those illustrated in Neyman's example. Neyman's specific example is representative of the relationship between genes and fitness discussed by Wright.

3.3 When Nature does small experiments

Lastly, Neyman argued that the problems of interpretation become most extreme in the analysis of the results of small experiments. As I have shown, he illustrated this point with a Monte Carlo simulation of 30 small experiments applied to the same idealized model of Nature. In his example, the repeated empirical conclusion from small experiments was that factor A had a beneficial main effect on yield when in fact its simple direct effect was to decrease yield. In an analogous manner, Wright argued that the evolutionary effects of gene interactions would be most important in large populations with small subpopulations. In such small populations, random genetic drift could create variation between populations in the system of gene interactions. Differently put, Wright believed that Nature does indeed do 'small experiments' in the ongoing process of adaptive evolution in subdivided populations.

Furthermore, just as Neyman illustrated how the interactions among factors can determine our inference regarding the nature of main effects, Wright argued that it is the interactions among genetic factors that determines their effect on fitness and ultimately their evolutionary fate. That is, owing to finite population size, it is a gene's interactions with
other genes that determine its relationship to fitness. A population subdivided into more or less isolated breeding groups is analogous to several small experiments being conducted on the same genetic system. The evolutionary fate of a gene in one deme within the subdivided population will be determined by the system of interacting genes characteristic of that small 'experiment.' Because the multiplicity of genotypes is extremely large in relation to population size, the genetic results of natural selection operating in local demes (Nature's small experiments) can be diverse if gene interactions for fitness are important.

3.4 Falconer's fitness component argument applied to gene interactions

In this section, I apply the arguments of Falconer (1981) regarding interactions among fitness components to the interactions expected among genes affecting fitness. I argue that genic main effects and interactions are expected to be of opposite sign, for those cases where selection maintains genetic variation and where genes interact to affect fitness. That is, genic main effects and interactions will usually have properties similar to Neyman's (1935) model of Nature.

The "character" that natural selection selects for is fitness' (Falconer 1981, p. 301) and the distributions of all other characters change in direct proportion to their correlation with fitness (Robertson 1966; Wade and Arnold 1980; Lande and Arnold 1983; Arnold and Wade 1984; Wade 1987). If two correlated characters are both subjected to directional selection, as would be any two components of fitness, then the genetic correlation between these two characters is expected to become negative at equilibrium. This happens because pleiotropic genes with positive effects on both characters experience strong directional selection toward fixation. Those pleiotropic genes with negative effects on both characters will experience strong directional selection toward loss from the population. Those genes, however, with positive effects on one fitness component and negative effects on another will experience much weaker selection and remain at intermediate frequencies for a longer period of time. It is these genes that will remain segregating in the equilibrium population and contribute to the genetic covariance among components of fitness. It is these genes that cause the genetic covariance between fitness components to be negative (Falconer 1981, p. 300).

This same argument can be extended to gene interactions. If alleles at two different loci each contribute positively to fitness and their interaction also has a positive effect on fitness, then these alleles will experience strong directional selection toward fixation. This will occur no matter what the effective population size because, even in small populations, the main effects of such alleles on fitness remain positive. When simple
main effects and simple interactions are of the same sign (positive in this case), sampling cannot change the sign of the estimated main effect. Similarly, consider the fate of alleles at two loci with negative main effects on fitness and negative interactions with respect to fitness. These alleles will experience strong directional selection toward loss from the population, again independent of population size. (Random genetic drift and selection will still operate according to classic population genetic theory and the outcome of selection will be less determinate in small populations. However, the direction of selection on such alleles cannot change sign for the reasons discussed above.)

In the population near equilibrium, only those alleles at the two loci whose main effects are of opposite sign to their interaction effects on fitness will remain segregating in the population (or be fixed more slowly). For the same reason that we expect fitness components to be negatively genetically correlated, we also expect that the genes segregating at intermediate frequencies in the population will be those that exhibit main and interaction effects on fitness of conflicting sign. Thus, the special case that Neyman used to illustrate the problems of interpretation expected in ANOVA when there are interaction effects is also the expected case with respect to gene action and fitness. For this reason and because the number of interacting factors is large, I argue that it is likely that small populations would differ sufficiently in genetic background that the effect of a gene on fitness in one deme would be of opposite sign to its effect on fitness in some other deme (Fig. 1). Each deme may be viewed as a small sample of some of the myriad of simple interactions possible (Table 56, p. 283, Lewontin 1974) and a gene’s effect on fitness can be estimated within the limits of the local deme. Indeed, this is how we expect local adaptation by natural selection to operate. It remains an open empirical question, however, what proportion of new mutations have beneficial (or detrimental) main and interaction effects on fitness. The immediate fate of a new mutation must be closely tied to its ‘simple interaction’ effect in the genome in which it arises.

4. PREDICTIONS FOR EXPERIMENTALLY INVESTIGATING THE EFFECTS OF EPISTASIS ON ADAPTIVE EVOLUTION

The above arguments can be interpreted as making some testable predictions in natural or laboratory populations. There are three suggestions for empirical research that can be derived. (1) We should expect that the additive genetic variance for fitness could increase when equilibrium populations are inbred or subdivided. (2) We should expect that the average additive effect on fitness of single genes will vary when measured in different genetic backgrounds in different local demes. (3)
Fig. 1. A schematic illustration of the average effect on fitness and its variance among subpopulations. The abscissa is the value of the average additive effect, \( a \), of an allele and the ordinate is the frequency of subpopulations or demes with different local values of \( a \) owing to differences among demes in a genetic background. In a large population at equilibrium, we expect the distribution to be centered at or very near zero, i.e. at equilibrium a gene should have little or no direct effect on fitness, positive or negative. A distribution about zero is created by subdivision of the population into more or less isolated demes so that the effect of the gene on fitness in some demes is positive but in other demes it is negative. The variance of this distribution should be a function of the degree of population subdivision, the extent of gene interactions for fitness, and the multilocus genotype frequencies. See text for further discussion.

Following from (2), we should expect that the genetic basis of the response to selection in small populations might vary from deme to deme even for uniformly imposed selection. I discuss each of these predictions in more detail below.

4.1 The change in additive genetic variance with population subdivision

The first prediction is not unique to epistatic systems. When rare deleterious recessives are made homozygous by inbreeding, an increase in the additive genetic variance for fitness can occur (Robertson 1952; Phillips and Wade 1992). A similar prediction arises from the following considerations for epistasis. Consider a large population of a haploid organism in which \( N \) loci affecting fitness are segregating for alternative alleles. I choose a haploid population in order to emphasize the effects of gene interactions between loci rather than allelic interactions or dominance effects within loci, which have received considerably more theoretical attention. There are \( N \) main effects and \( N (N - 1)/2 \) pairwise interaction effects. Both kinds of effects contribute to the total genetic variance. In a large population at mutation-selection equilibrium, we would further expect the additive variance for fitness attributable to main
effects to be nearly zero and owing to deleterious mutations according to Fisher’s Fundamental Theorem.

If this population were subdivided to a value of Wright’s inbreeding coefficient, \( F = 0.50 \), then half of the genes would be fixed and half of the genes would remain segregating in the average deme. Demes would differ from one another in the identity of the genes fixed, lost, and segregating. The fixation of \( N/2 \) genes implies that the numbers of main effects contributing to the genetic variance within demes is halved. However, fixation at half the loci also changes the within-deme constellation of simple interactions between the fixed loci and the \( N/2 \) loci still segregating. At each of these segregating loci, the local or within-deme average effect may be changed. For example, consider a building block model in which the phenotype of an individual is equal to the sum of the separate linear and quadratic contributions of only two component loci, \( X \) and \( Y \), each with many alleles. We can represent a haploid two-locus genotype generally as \( X_iY_j \), where the subscripts indicate different alleles at \( X \) and \( Y \). Now, using the heuristic of the building block approach, the absolute fitness of a genotype, \( W \), is defined as

\[
W(X_iY_j) = 1 + a(X_i) + b(Y_j) + c(X_iY_j) .
\]

The genetic variance in fitness is then given as

\[
\sigma_w^2 = a^2(\sigma_X^2) + b^2(\sigma_Y^2) + c^2(\sigma_{XY}^2) .
\]

(4)

Suppose, in one deme, that the allele at the \( Y \) locus is fixed, \( Y_c \), without changing allele frequencies at the \( X \) locus. Then the fitness of a genotype in this deme is given by

\[
W(X_iY_c) = (1 + bY_c) + (a + cY_c)(X_i)
\]

and the genetic variance in fitness within this deme is

\[
\sigma_w^2 = (a + cY_c)^2(\sigma_X^2) .
\]

(6)

The coefficient determining the contribution of the \( X \) locus to the genetic variance in fitness in eqn (6) can exceed that in eqn (4) depending on the relative values of \( a \), \( c \), and \( Y_c \). That is, interaction effects in a large outbred population can contribute to main effects in a subpopulation derived from it by random genetic drift or inbreeding. Goodnight (1988) has referred to this phenomenon as the ‘conversion’ of epistatic genetic variance into additive genetic variance by random genetic drift. For this reason, we might expect the additive genetic variance for fitness to increase with inbreeding and population subdivision.

The magnitude of this effect depends on a number of different elements, including the number of loci, \( N \), the relative strength of the linear and
interaction effects (here, a and c), and the gene frequencies. (In my example, I changed only the frequency at the Y locus and not those at the X locus to illustrate the effect. See Goodnight (1988) for a much more general diploid treatment.) If gene interactions for fitness are as important as suggested by Wright, then we might expect to see a measurable effect of inbreeding or random genetic drift on the additive genetic variance for fitness. Dominance variation can have a similar effect (Robertson 1952; Phillips and Wade 1992). The theoretical criteria for efficiently distinguishing the separate effects of dominance and epistasis on the additive genetic variance with inbreeding have not yet been elucidated (Goodnight, personal communication) but dominance effects on the within-deme variance must decrease with inbreeding beyond \( F = 0.50 \). For inbreeding with \( F \) values near 1, the among-deme genetic variance will be twice the original within-population genic variance plus four times its additive-by-additive epistatic variance. At complete fixation, the dominance variance does not contribute to the among subpopulation genetic variance. Thus, large increases in the genetic variance for fitness among demes with extreme inbreeding might reasonably be attributed to epistatic rather than to dominance effects.

We investigated the effects of population subdivision with laboratory populations of flour beetles, *Tribolium spp.* (Wade 1977, 1982, 1985, 1990, 1991; McCauley and Wade 1980, 1981; Wade and McCauley 1980, 1984; Wade and Goodnight 1991). We created experimental arrays of demes, each with a different population structure, using different values of \( N \) and different amounts and patterns of migration (\( m \)). In almost all population structures, we found large heritable differences in fitness arising among demes within a period of 10 to 15 generations. The population structures we studied correspond to values of \( F \) across the range of 0.02 and 0.33 (breeding adults per deme: \( 6 < N < 96 \); migration rates: 0.00 < \( M \) < 0.25). This rapid differentiation among demes occurred despite finding little or no additive genetic variance for fitness (Wade 1985) and, for the range of \( F \) values studied, a large increase in among-deme variance is not expected on the basis of dominance effects (Crow and Kimura 1970, pp. 339–345). However, approximately \( 4F(1 - F) \) of the original epistatic variance for fitness is converted into additive genetic variance within demes (Whitlock, personal communication).

The most relevant experimental study is that of Wade and Goodnight (1991) in which we created a laboratory model of Wright's process of interdemic dispersion. In our study, each subdivided population consisted of 50 demes and each deme was founded with 20 breeding adults. After a 62-day period of population growth, Wright's process of interdemic selection was imposed on the entire array of demes. Migrants were chosen from the most productive demes, those with the highest values of fitness, and distributed into the least productive demes. Our protocol weighted each deme by its fitness relative to the average demic fitness across the
array: the largest demes contributed the most migrants and the lowest demes received the most migrants. Each experimental array had its own paired control array of 50 demes which experienced the same amount of random migration. We observed a significant increase in mean fitness over time in all three population structures (0.025 < F < 0.050) in which we imposed Wright's process.

The empirical evidence necessary to separate dominance effects from epistatic effects in these experiments is not decisive, but it tends to implicate epistasis for fitness rather than dominance. For small values of Wright's inbreeding coefficient F and variance in fitness owing only to recessive deleterious genes in low frequency in the stock populations, we would not expect to observe a large increase in among-deme genetic variance for fitness or a large response to artificial interdemic selection (Wade and McCauley 1980, 1984; Wade 1982, 1990, 1991). For an F of 0.10, less than 3 per cent of the total genetic variance is expected to be among demes. In some studies, we were able to detect a correlation between population mean fitness and expected heterozygosity (McCauley and Wade 1981; Wade and McCauley 1984), even for weak levels of random genetic drift (0.02 < FST < 0.10). If this were due to an increase in frequency of rare deleterious recessives with inbreeding, then we would expect to observe an increase in mean deme fitness on outbreeding. However, direct and extensive experimental tests comparing inbred and outbred subpopulations at different densities did not support this expectation for the low values of F (McCauley and Wade 1980; Wade and McCauley 1980). In summary, the rapid genetic differentiation among demes for fitness for low levels of inbreeding (0.02 < F < 0.10) and the lack of a direct response of mean fitness to outbreeding argue against dominance variance for fitness (owing to recessive deleterious alleles) as the sole cause of the genetic effects we observed in subdivided laboratory populations of flour beetles. Direct measures of the change in the within-deme additive genetic variance for fitness with inbreeding are underway and may assist in discriminating the effects of dominance from epistatic genetic variance.

A second line of empirical evidence suggesting that epistasis for fitness is important in these experiments comes from the Shifting Balance Experiment (Wade and Goodnight 1991). We found that interdemic selection at every second generation produced nearly a two-fold larger increase in demic fitness than selection every generation or every third generation. In an additive world, one would expect the response to selection to be proportional to the intensity of selection (the selection differential). We did not find this. It is as though a store were 10 blocks away and we arrived at the store sooner by walking every other block than by running every block.
4.2 Change in the average additive effect on fitness with inbreeding and random genetic drift

Wright argued that an allele might have a positive effect on fitness in one genetic background but a negative effect on fitness in a different genetic background (see above). This is especially true for an allele that contributes additively to a continuously distributed trait for which an intermediate value has the highest or optimum fitness. If gene interactions for fitness are often or usually of opposite sign to genic simple effects, then we ought to be able to detect among-deme variations in the average additive effect on fitness of single genes. The degree of population subdivision necessary to detect such variance in average additive effects among demes will depend on several different factors. I will use a simple two-locus model from Crow and Kimura (1970, p. 79) to illustrate this dependence.

The model presented in Table 3 (from Crow and Kimura 1970, p. 79) includes additive effects (A and B), dominance effects (DA and DB) and epistatic interactions (I, J, K, and L) between two diallelic loci. The average effect, \( a_1 \), of the A1 allele is defined as the linear regression of genotypic value on the number of A1 alleles in the genotype. Let \( p_{A1} \) and \( p_{A2} \) be the frequencies of the A1 and A2 alleles at the A locus and, similarly, \( v_{B1} \) and \( v_{B2} \), be the frequencies of the B1 and B2 alleles at the B locus. The \( a_1 \) is given by

\[
a_1 = A + D_A(p_{A2} - p_{A2}) + I(v_{B1} - v_{B2}) + J2v_{B1}v_{B2}(p_{A2} - p_{A1}) \\
+ K2v_{B1}v_{B2} + L(v_{B1} - v_{B2})(p_{A2} - p_{A1}).
\]

<table>
<thead>
<tr>
<th>Genotype at the A locus</th>
<th>A1A1</th>
<th>A1A2</th>
<th>A2A2</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1B1</td>
<td>A+B+I</td>
<td>DA+B+L</td>
<td>-A+B+I</td>
</tr>
<tr>
<td>B1B2</td>
<td>A+Db+K</td>
<td>DA+D_b+J</td>
<td>-A+D_b-K</td>
</tr>
</tbody>
</table>

A, B = additive effect of the A1 and B1 alleles, respectively.
DA, Db = dominance effects at A and B loci, respectively.
I = additive-by-additive interaction.
J = dominance-by-dominance interaction.
K = additive(A)-by-dominance(B) interaction.
L = additive(B)-by-dominance(A) interaction.
When there is no dominance ($D_A = D_B = 0$) and no epistasis ($I = J = K = L = 0$), then $a_1$ is simply equal to $A$ and independent of gene frequency. To see the effects of epistasis in an array of subpopulations derived from a larger randomly mating population, set $D_A = D_B = 0$. In those demes where $B_1$ is fixed ($v_{B_1} = 1.0$) but gene frequencies are unchanged at the $A$ locus, we have

$$a_1 = A + I + L(p_{A2} - p_{A1}).$$  \hfill (7)

But, in those demes where the alternative allele, $B_2$, is fixed, we have

$$a_1 = A - I - L(p_{A2} - p_{A1}).$$  \hfill (8)

Clearly, the value $a_1$ can change in magnitude and sign depending on gene frequencies at both loci as well as on the relative magnitudes of the additive ($A$) and epistatic ($I$ and $L$) effects. In a strictly additive world, the average effect could not change sign.

Figure 1 illustrates the possible consequences of population subdivision on the average additive effect of a gene on fitness. In a large, randomly mating population, $a_1$ is expected to be very near zero for a gene contributing to fitness. In a subdivided population, averaged over the entire array of demes, the value of $a_1$ is also very near zero but it may deviate from this expectation in either the positive or negative direction in different demes (Dempster 1963). Figure 1 is schematic and illustrates what could happen. The specific distribution for any particular gene across an array of demes in a subdivided population is unknown. Several important empirical issues remain unanswered, including: (1) is there significant variation in the average additive effect on fitness of a specific allele in different genetic backgrounds?; (2) can random genetic drift lead to differences among demes in genetic background sufficient to change the average effect of a gene in the manner suggested by Wright?; (3) to what degree and for how long must an array of subdivided populations be isolated before significant variance among demes in the average additive effect of a gene on fitness can be measured?; (4) how does the variance about the average additive effect of a gene on fitness (Fig. 1) change with time and degree of population subdivision?; and (5) do natural populations exhibit a sufficient degree of population subdivision for these effects to be evolutionarily important? The picture in Fig. 1 is a static one and for any real array of subpopulations the variance of $a$ should change with time and continued subdivision if the arguments of Wright are correct.
4.3 The heterogeneous genetic effect of homogeneous selection

The third prediction concerns the expectation of a heterogeneous genetic response among demes to homogeneous directional selection. If artificial selection for increased body size were imposed on a series of replicate small populations, derived independently by inbreeding from a common outbred stock, the considerations discussed above imply that the response to selection in different demes could involve different gene interaction systems (Dempster 1963). Even if the separate demes responded to selection to the same degree phenotypically, the genetic basis of the response might differ among lineages because the same gene could contribute to the local demic response in opposite ways given sufficient differences among demes in genetic background. Again, the important empirical issues are to what degree must the lineages be inbred before such effects can be detected and, if detected, do the levels of inbreeding correspond to those that might occur in natural populations owing to random genetic drift? This is a possibility and we need an empirical evaluation of its probability in natural populations.

Several experimental studies have been conducted to investigate whether the limits to individual selection can be enhanced or diminished by a combination of selection within and among demes. Wright believed that this kind of selection might permit the selection of epistatic gene combinations. The majority of these studies, however, involve artificial selection on primarily additive traits (e.g. Katz and Young 1975; Madalena and Robertson 1975; Rathie and Barker 1968) where among-line selection is always expected to reduce the total response owing to the loss of alleles in lineages extinguished during selection. Only the study by Katz and Young (1975) found population subdivision to enhance significantly the selection response and they did not report an investigation of the genetic basis of the among-lineage genetic variation. Furthermore, a significant fraction of the total response to selection in both the other studies was due to a small number of genes of large effect, initially at low frequency in the base population. In general, strong artificial selection on an additive trait is unlikely to lead to significant among-line variance in gene interactions for the selected trait but may lead to among-line variance in fitness owing to linked loci.

4.4 Epistasis, adaptation, and the Shifting Balance Theory

It is important to emphasize that the three predictions made above, even if confirmed experimentally, do not necessarily guarantee the existence of multiple adaptive peaks or the efficiency of their export by differential migration among demes, the latter phases of Wright's Shifting Balance Theory. Their confirmation would suggest that the genetic basis of local
adaptations is heterogeneous among demes, even for uniformly imposed selection, and that the process of local adaptation is itself inherently variable in genetic outcome owing to gene interactions. However, Wright's theory not only requires the existence of such interactions and their differentiation among demes but also that they must produce significant differences among demes in local mean fitness. Furthermore, the interaction systems underlying local adaptations and local mean fitness must be spread to neighboring demes with different interaction systems and at lower fitness peaks by interdemic selection (differential migration). That is, variation among demes in systems of interacting genes must lead to variation among demes in mean fitness, and this, in turn, must lead to the differential dispersion of migrants away from high peaks and into demes at lower peaks. The latter two effects do not follow necessarily from the existence of demic variation in interaction systems.

In fact, the conditions for genetic differentiation among demes may be contrary to those necessary for the efficient export of adaptive peaks. The genetic differentiation of local demes requires restricted migration but the efficient export of adaptive peaks appears to require somewhat more extensive migration. Recent theoretical research (Crow et al. 1990) indicates that perhaps these conditions are not as restrictive or contradictory as they appear (Nei 1987). If the domain of attraction of a high fitness peak is large, then small amounts of interdemic migration may be sufficient to initiate a 'peak shift,' the genetic transformation of a local deme from the gene interaction system characteristic of a low fitness peak to the interaction system characteristic of a higher peak. A peak shift is achieved by the joint action of interdemic dispersal and individual selection within the receiving deme and not solely by interdemic dispersal. Crow et al. (1990) investigated what they referred to as the 'critical migration rate,' the minimum rate of interdemic migration necessary to effect a peak shift. This can be viewed as a situation in which a peak shift occurs primarily as a result of individual selection subsequent to migration rather than to the interdemic selection process itself. (Clearly, both are required for a peak shift.) They found, first, that a peak shift would occur after a number of generations of differential migration even if further migration was halted completely. This clearly demonstrates that selection within demes can be responsible for much of the shift from one interaction system to another and that the peak shift process does not depend solely on interdemic selection by differential dispersion. All else being equal, the broader the domain of attraction of the higher fitness peak, the more the shift will depend on selection within demes rather than among demes. Secondly, they found that peak shifts could occur even when the rate of differential migration from the high peak to the low peak was an order of magnitude less than the reverse migration. Again, this demonstrates the potentially important role of
individual selection within demes for the peak shift and indicates that the
first and third phases of Wright's process need not be in conflict. The
small migration rates necessary for genetic differentiation may also be
sufficient for exporting adaptive fitness peaks. (Remember that, for
Wright, mass selection within demes is necessary for a 'fitness' peak.)

In a recent study of Wright's Shifting Balance Theory, using experimental
demonstrated that significant increases in mean fitness can occur with
interdemic migration rates less than 1 individual per deme per generation
with an effective deme size of approximately 20 breeding adults. This is
a level of migration that is often associated with significant genetic
differentiation of demes but not with significant effects on the evolution
of mean fitness. Additional experiments indicate that there is a non-
additive genetic basis for this response (Wade, in preparation).

The relationship between epistasis, adaptation, and Wright's theory
depends on a number of additional genetic and ecological features of
natural populations. However, despite these caveats, the role of gene
interactions in adaptive evolution and the origin and spread of adaptations
involving the coordinated action of many genes would be better understood
if the kinds of empirical data discussed above could be obtained.
Furthermore, the empirical predictions made here appear to be related
directly to the evolutionary inferences regarding gene interactions drawn
by Wright.

5. CONCLUSIONS

Neyman in statistics and Wright in evolutionary genetics both challenged
Fisher on the role of interactions in the interpretation of Nature and they
did so using arguments of remarkable similarity. By investigating the
issues in these disputes, some general implications can be drawn and
some novel experimental predictions for evaluating the role of gene
interactions can be made. The experimental investigation of these
predictions may permit a more complete understanding of the evolutionary
importance of gene interactions and the robustness of current theory that
by and large ignores them. The ultimate test of Wright's Shifting Balance
Theory depends on empirical demonstration of the heterogeneity of
genetic response to local selective pressures, the existence, density, and
size of adaptive peaks, and the competence of interdemic selection by
differential migration for exporting a favorable gene combination from
one deme to the next. Although Wright's Shifting Balance Theory has
been referred to as a 'cornerstone of modern evolutionary thought'
(Mettler *et al.*, 1988) and 'the dominant theory of evolution in the 20th
century' (Collias 1991), a number of its requirements and predictions
await empirical investigation.
ACKNOWLEDGMENTS

I would like to acknowledge gratefully the thoughtful and critical comments of Drs Jim Crow, Jerry Coyne, Patrick Phillips, John Willis, Felix Breiden, Jim Griesemer, Joe Travis, Dave McCauley, Charles Goodnight, Lori Stevens, Chris Boake, Alan Templeton, Robert Boyd, Thomas Park, Michael Whitlock, Doug Futuyma, and Alexander Gimelfarb. Drs Coyne and Crow, in particular, generously provided thoughtful and extensive comments on the manuscript and helped me greatly to correct and clarify my ideas. Many of the conceptual aspects of this work started to develop during and after a 6-week visit to the University of Edinburgh in 1984, which was supported by a PHS grant K04 HD00431, NIH grant GM22523, and a Burroughs-Wellcome travel grant. I am grateful for the open and challenging intellectual environment provided by Drs Alan Robertson, Bill Hill, Douglas Falconer, Trudy MacKay, Linda Partridge, and Jerry Wilkinson during that visit.

REFERENCES


