



Sewall Wright

THE SHIFTING BALANCE THEORY AND MACROEVOLUTION¹

Sewall Wright

Department of Genetics, University of Wisconsin, Madison, Wisconsin 53706

CONTENTS

THE GAP IN DARWIN'S SYNTHESIS	1
KINDS OF MUTATIONAL CHANGES	2
THE ORIGIN OF THE SHIFTING BALANCE THEORY	4
THE SURFACE OF SELECTIVE VALUES	6
SPECIATION.....	9
NONADAPTIVE DIFFERENCES BETWEEN SPECIES	11
MACROEVOLUTION	13
AMBIGUITY OF THE CONCEPT OF CAUSATION.....	14
SUMMARY	15

THE GAP IN DARWIN'S SYNTHESIS

The extraordinary synthesis of data from all fields of biology achieved by Darwin (7, 8) in his theory of evolution by natural selection had one serious gap: the absence of a real understanding of heredity. His theory was open to Fleeming Jenkin's criticism (24) that the hereditary contribution of any new variation would be halved in each generation, and thus reduced to only one tenth of one percent in ten generations, according to the prevailing view that the heritage of an offspring is on the average an equal blend of those of its parents. Natural selection would have to operate with inconceiv-

¹*Editor's note:* This paper is based on a symposium held on June 15, 1981, during the annual meeting of the Genetics Society of America. Wright's shifting balance theory of evolution was first fully presented in his famous paper, "Evolution in Mendelian Populations" (45). The symposium was held 50 years later. It is particularly appropriate to present his thoughts on this subject half a century later.

able effectiveness to have an appreciable effect. Darwin was aware of this difficulty and in 1868 proposed his theory of pangenesis in an attempt to meet it.

The difficulty disappeared, however, soon after the rediscovery of Mendelian heredity in 1900. Yule (68) pointed out that unifactorial segregation in the 1 : 2 : 1 ratio in F_2 of a cross would persist unchanged thereafter under conditions of random mating in a large population. Castle (2) showed that the altered ratio resulting from selective elimination of recessives, generation after generation, would also persist unchanged with random mating after elimination ceased. Weinberg (35) and Hardy (22) independently expressed this idea in general form.

A big gap remained, however. Mendelian heredity applies to individuals, whereas evolution happens to populations. It was necessary to work out the statistical consequences of simultaneous operation of all of the processes involved in evolution: the recurrences of major and minor mutations, differential dispersion from localities, various kinds of selection, and the cumulative effects of accidents of sampling.

KINDS OF MUTATIONAL CHANGES

With respect to the kinds of variation utilized in natural selection, Darwin wrote, "Without variability, nothing can be effected. Slight individual differences, however, suffice and are probably the chief or only means in the production of species."

deVries (9) thought otherwise. He held that new species arise at a single step. His "mutation theory" had the merit of being based on observation: the finding of what seemed to be several new species of the American genus, *Oenothera*, among plants of *O. lamarckiana* that had escaped from cultivation in the Netherlands.

It turned out, however, that most of these plants were trisomics that merely involved changes in the proportions of existing heredities. Moreover, the extra chromosome was found not to be transmissible by pollen, so the new forms could exist only as segregants from the parent population.

One mutant type, *O. gigas*, that differed only slightly in phenotype from *O. lamarckiana*, turned out, however, to be a tetraploid, capable of reproducing itself but reproductively isolated because of the sterility of its triploid hybrids with the diploid parent. It thus behaved like a new species.

While no general conclusions about the origin of species could be drawn, it became clear from cytological studies that chromosome differences between related species were almost universal and that karyotypic change was very important in evolution. Nevertheless, the statistical consequences of

the processes under orderly Mendelian heredity were essential for character change.

Most early geneticists assumed that wild species are in general homallelic with respect to their array of genes, except for rare deleterious mutations and much rarer favorable mutations. Evolution was assumed to be due to the substitution of the latter, one at a time.

The course of elimination of a deleterious recessive was developed by Castle (2). H. T. J. Norton (28) worked out the course of fixation of favorable dominants and recessives. The evolutionary implications were spelled out by Chetverikov (6) in a paper that stimulated many studies in the USSR on the occurrence of mutant genes in wild populations of *Drosophila* (cf 66).

In a series of papers begun in 1924 (19) and summarized in 1932 (20), Haldane made the most systematic studies of the courses of substitution of favorable genes under diverse conditions. He did not consider this to be all that there was to evolution, but felt it was something that should be studied thoroughly.

Castle (3, 5) led a return to a more Darwinian view of evolution than that taken by most early geneticists. He challenged the prevailing view, tracing to deVries, that selection of quantitative variability has no permanent effects. He attempted to modify on a grand scale the black and white piebald pattern in a strain of rats by selection in both directions. After 20 generations, his selection lines were approaching self-black or self-white, when the experiments had to be discontinued because of low fecundity.

Meanwhile, Shull (33), Nilsson-Ehle (27), and East (10, 11) were demonstrating that many quantitative differences in plants were, after all, determined by multiple Mendelian genes. Castle at first attributed the changes in the piebald patterns of his rats to allelic mutations of the piebald factor itself, but after a crucial experiment (4) accepted the view that they were due largely to independent modifiers.

Fisher (13), in contrast with Haldane, attempted to find a single general principle of evolution, comparable to the second law of thermodynamics in physics. He made certain simplifying assumptions: first, that the likelihood that a mutation would be favorable falls off so rapidly with the magnitude of its effect that only those with minor effects need be considered; second, that the local effects of accidents of sampling are negligible because they are overwhelmed by dispersion from neighboring localities and third that one-to-one relationships between gene and character are the norm. He considered epistasis unimportant and devised a special hypothesis to account for the prevalence of dominance: the existence of a host of modifiers with no effects in homozygous wild-type so that selection, in populations including

rare deleterious mutations, would shift the phenotypes of the heterozygotes toward that of the wild-type homozygotes. He arrived at what he called the “fundamental theorem of natural selection”—“the rate of increase in fitness of any organism at any time is equal to its (additive) genetic variance in fitness at that time” (13).

According to this theorem, dominance and epistasis merely slow down the rate of increase of fitness. While there are qualifications in connection with frequency-dependent selection, and with selection or linkage sufficiently strong that there is appreciable linkage disequilibrium, the theorem is a very useful approximation in effectively panmictic populations. Fisher essentially put Darwin’s theory into quantitative form.

THE ORIGIN OF THE SHIFTING BALANCE THEORY

Initially I leaned strongly toward Darwin’s theory, having read *The Origin of Species* long before I began academic studies of biology. Darwin’s view seemed to be confirmed by the results of Professor Castle’s selection experiments with rats, for which I was his assistant from 1912 to 1915. I was, however, somewhat disturbed by the termination of both experiments by low fecundity.

My own researches concerned the effects in combination of the factors affecting coat color and of those for other characteristics of the guinea pig. I was continually surprised by unexpected interaction effects. The systematic study of the relationships between genotype and phenotype became my principle experimental project from 1912 to 1954, apart from my concentrated study of the effects of close inbreeding and cross breeding on guinea pigs during 1915–1925 at the Animal Husbandry Division of the US Bureau of Animal Industry. The results of these two lines of research greatly influenced my ideas on evolution.

The objective of the studies of factor interaction was to devise hypothetical networks of gene-controlled processes that could account for them (63). The conclusions were that genotypes are in general related to phenotypes by a very complex network of biochemical and developmental reactions, such that (a) each character is usually affected by many gene substitutions; (b) each substitution usually has numerous pleiotropic effects; and (c) the intervening processes involve nonadditive interactions. This viewpoint is diametrically opposed to that implied by the common treatment of organisms as mosaics of unit characters (Fisher’s norm) in considering their evolution.

From this viewpoint, evolution becomes a much more intelligible process if based on natural selection among interaction systems rather than among

alleles at each locus separately. Unfortunately, only the latter is possible (unless linkage is very strong) under natural selection among individuals in a panmictic population, the sort of population to which Fisher's fundamental theorem applies. Deviations from random combinations at any two loci located in different chromosomes are halved in each generation under biparental reproduction. They are halved in about seven generations with 10% recombination and in 69 generations with 1% recombination [(Robbins (31))]. Even the latter process is rapid in terms of geologic time.

Resolution of this difficulty was suggested by observation of the profound differentiation among 23 closely inbred strains of guinea pigs in every character studied [Wright (41, 65)], and by study of how British breeds of livestock had been improved since the time of Robert Bakewell (1725–1795) (26, 39, 43).

Bakewell produced a herd of Longhorn cattle and a flock of Leicester sheep by careful selection and close inbreeding, that were generally considered the best. These breeds were made over by the selection by breeders in general of Bakewell's strains as the preferred sources of sires. Bakewell's methods were widely emulated in the founding of most of the other British breeds in spite of the use of inbreeding to a degree that had been considered deleterious.

I, joined later by H. C. McPhee, (26, 43) made an intensive study of the breeding history of the Shorthorn breed of cattle beginning with its foundation in the late 18th century. Inbreeding coefficients (40, 42) devised in order to measure the overall approach of neutral genes to fixation (F_{IT}), that which would be expected on starting random mating (F_{ST}) and an index of current inbreeding (F_{IS}). The last was significantly negative around 1850 when the hitherto most favored source of bull calves tended to be avoided because of low fecundity of the cows produced. This study clearly showed the importance of selection at two levels—selection by the more ambitious breeders in building up their herds, and by breeders in general among the herds as sources of sires.

To grasp the full significance of the two levels, one must consider the consequences of the Mendelian mechanisms. The selection among individuals in building up herds implies selection favoring the allele at each pertinent locus that gives the most favorable effect on the average of all combinations with such alleles at other loci. This is *genic* selection. The resulting patterns were fixed, more or less, by close inbreeding. The breeders' selection among such herds as sources of sires was a selection among the diverse interaction systems that happened to have been arrived at. This is *organismic* (or genotypic) selection [Wright (67)].

Recognition that the two-level process was much more efficient than mere individual selection led to consideration of whether an analogous two-level

process might not occur in nature. The first draft of a paper developing this idea was written immediately after the Shorthorn study in 1925 but was not published until 1931 (abstract 1929).

The detailed process in nature is necessarily very different from that under artificial selection. Random differentiation among local populations takes the place of the efforts of the leading breeders, which involved different degrees of skill as well as chance. Random differentiation in nature may be based on the cumulative effects of accidents in sampling (if numbers are very small), or on fluctuations in the conditions of selection [Wright (45, 58)], or in amount and quality of immigration into each local population, irrespective of numbers [Wright (58)].

Selection among local populations is not, of course, according to the number of sires provided to the population in general, but according to rates of proliferation and dispersion into neighboring populations owing to the selective values of their genetic systems as wholes. The aspects of the total system that contribute most should continue to spread throughout the species. The spreading of two different favorable systems from different centers leads to overlapping and thus creates a new center from which the joint system spreads, and so on indefinitely.

THE SURFACE OF SELECTIVE VALUES

The distinction between the two modes of selection may be appreciated best from a geometric representation of the field of variability in which both operate. Assume orthogonal axes in a multidimensional space, a dimension for each allele other than the leading one at each locus, with the frequency in the specified population as the coordinate in locating this population. The populations considered are local random breeding populations with approximately random combination, both within and among loci. Add an additional dimension for the selective value for each population. This defines a multidimensional "surface" of selective values [Wright (46)].

At first sight, it may seem that this "surface" has only one maximum corresponding to the population in which each locus is represented by the allelic frequency that gives the highest average selective value in combination with other loci. If, however, there are nonadditive interactions with respect to selective value, there may be multiple maxima or "selective peaks."

This is true even in the case of the conventional pattern for quantitative variability, with several equivalent pairs of alleles with semidominance and additive effects on the character in question putting the optimum at an intermediate value of the character, close to the mean. It is obvious that this

optimum would be given by several different genetic combinations. Thus, there would be many different selective peaks connected by multidimensional saddles, and all at the same height in the ideal case. If the genes have different pleiotropic affects, however, the peaks would not be at the same height.

Of more interest are cases involving nonadditive interactions with respect to the character. A very rugged "surface" is to be expected with major peaks representing different methods of coping with the environment, as well as numerous subordinate peaks corresponding, as in the preceding case, to different genetic systems that give the same character.

As noted in the first paper in which the concept of a "selective surface" was presented [Wright (46)], "The problem of evolution is that of a mechanism by which the species may continually find its way from lower to higher peaks." For this to occur, random processes must from time to time carry the population against the pressure of selection. Saddles must be crossed. It is difficult to grasp what is meant by such a process in multidimensional space. As the conditions are sufficiently severe for simultaneous wide random drift against adverse selection pressures at only two loci, we need not consider the possibilities of crossing saddles of more than two dimensions. Cases in which a heterozygote is inferior to both homozygotes may be considered one-dimensional saddles, but these do not involve interaction effects. They are probably uncommon except in the case of chromosome rearrangements, the heterokaryons of which are selected against because of aneuploidy. We will come back to these in connection with speciation. Evolution within the species (microevolution) is concerned primarily with selection for two-factor interaction systems as a supplement to mass selection in relation to single favorable alleles.

For passage to a new two-factor selective peak, the equilibrium frequencies at both loci must be fairly high, say 10% or more, and the opposed pressures must be rather weak to permit the necessary wide stochastic deviations for the crossing of the saddle. Thus, for recurrent mutation at a rate of 10^{-5} per generation, balanced by adverse selection, the selection coefficient must be only 10^{-4} or less, assuming semidominance.

Replacement of individuals in a local population by immigration, representative of the species at ratio 10^{-3} per generation, must be opposed by locally adverse selection with coefficients 10^{-2} or less. Equilibrium due to opposing selection pressures such as those implied by rarity advantage in a heterogeneous environment [Wright (54)], by selective advantage in diversity itself [Wright (65), pp. 484-951], or in heterozygote advantage [Fisher (12), Wright (45)] must be such as to yield a similarly high equilibrium frequency to permit sufficiently wide stochastic variability at both loci to reach the saddle point.

The strongly held polymorphisms that have been described in many species are unlikely to contribute to the shifting balance process. They are to be interpreted as adaptations, end results of evolution, rather than material for further evolution. One should not, as some have done, take the very high selection coefficients in these cases as typical of selection in general. Very strong selection at any locus tends, indeed, to reduce that at all others since all selection must be taken out of a finite reproductive excess [Wright (45)]. This is Haldane's "cost" of selection (21), which he showed was about the same whether the effect was major or minor. He held that even geologic time might not be sufficient for many observed evolutionary changes to have been brought about by an accumulation of minor mutations. This, however, applies only to panmictic populations. "Cost" is not an important consideration if subdivided among many local populations, as occurs under the shifting balance process.

It is obvious that the probability that any particular two-locus saddle will be crossed at a given time is very slight. Higher organisms, however, have tens of thousands of loci. If there are 1000 loci that are strongly heterallelic with only weakly held equilibria, there are nearly 500,000 pairs among which there may be favorable interaction effects, nearly 50 million if there are 10,000 such loci. Peak-shifts are to be expected from time to time in a given local population. There may be thousands of more or less independent local populations, so that peak-shifts may be occurring fairly often somewhere within a species with a suitable population structure.

The conditions necessary for neighborhoods to be sufficiently independent within a continuous population have been discussed in several papers [Wright (48, 50, 55, 57, 61); summarized in (64), chapter 12]. At the opposite extreme from panmixia is a species in which there is a large region where the population is subdivided into small colonies, frequently subject to extinction and refounding by a few stray individuals from the superior colonies, perhaps from single fertilized females [Wright (50, 53)].

An effective shifting balance process involves three phases: first, extensive local differentiation, with wide stochastic variability in each locality; second, occasional crossing of a saddle leading to a higher selective peak under mass selection; and third, excess proliferation of, and dispersion from, those local populations in which a peak-shift has occurred, leading to occupation of the superior selective peak by the species as a whole. The process is not alternative to mass selection but supplementary, since such selection is involved in all three phases.

The surface of selective values changes with changes in environmental conditions. Progress under pure mass selection in a panmictic population comes to an end under constant environmental conditions, with firm establishment of the controlling peak. With changing conditions, the location of the species follows the movement of the peak if the change does not lead

to extinction. Old adaptations are lost as new ones are acquired. VanValen has compared the course of evolution to the running of Lewis Carroll's red queen to stay where she is.

Under the shifting balance process in a subdivided species, there tends to be continual progress from lower to higher selective peaks with constant environmental conditions, and there is more effective progress with changing conditions than if there were only pure mass selection.

SPECIATION

So far we have dealt with evolution only as based on substitutions of Mendelian alleles. Moreover, we have assumed that the phenotypic effects of these substitutions are slight. Obviously, this is not all there is to evolution. Cytological study has revealed that closely related species often differ visibly in chromosome pattern. This phenomenon must be accounted for. Moreover, major character changes sometimes occur apparently very rapidly, and some of these are of such a nature that seemingly they must have occurred at single steps. There is also the problem of branching, by which millions of species have become distinct.

We have already noted that deVries held that species arise as single steps, by a process that has nothing to do with the small changes that occur within species to produce races. When it turned out that most of his "mutations" were forms with extra chromosomes, incapable of perpetuation except as segregants from the parent species, most of the early geneticists turned to major Mendelian mutations as the material for evolution, but the evolutionary significance of the karyotypic differences between related species remained.

Goldschmidt was the most prominent exception. He had no concrete examples of the origin of new species and higher taxa from "hopeful monsters," but maintained that because there is such a "bridgeless gap" between the kinds of differences among related species and those among subspecies, an abrupt origin of the former must be postulated. Thus he held that "chromosome repatterning" is required for "macroevolution" as opposed to the microevolution within species [Goldschmidt (16), cf Wright (52)].

The botanist, Willis [(37, 38)], also believed that new species and higher taxa arise abruptly. He argued from certain inconclusive facts of geographic distribution (*Age and Area*), and especially from the "hollow curve" shown by the distribution of numbers of species within genera. It was shown, however, that such a distribution is to be expected, irrespective of the mode of origin [Wright (51)].

Schindewolf (32), a paleontologist, held that the morphological differences between many higher taxa, as seen in the fossil record, are such that they could only have arisen abruptly. Other paleontologists, such as Simp-

son, (34) did not agree. This issue has recently been raised again by Gould & Eldredge (18), and Gould (17).

Macroevolution obviously involves both major morphological and physiological change and reproductive isolation. There has been considerable confusion about the relationships between these aspects. The earlier taxonomists, classifying the forms of restricted regions, and paleontologists, studying the very incomplete fossil records, necessarily treated morphologic differences as the primary criteria in naming species. As taxonomic studies extended over larger areas, it became evident that many forms that had been considered distinct species intergraded and had to be considered subspecies [Osgood (29), Rensch (30)]. Reproductive isolation became the primary criterion, however awkward this might be in some cases.

It became clear that major morphological change and speciation in the new sense are wholly distinct phenomena, which, while usually associated, may occur separately. Intergrading subspecies may differ more at the centers of their ranges than do typical species. Moreover, the differences among breeds of livestock and especially of dogs testify to the very great morphological differences that may be brought about within species by artificial selection. On the other hand, studies of many genera, such as *Drosophila*, have revealed the existence of numerous "sibling" species, reproductively isolated in nature, but with few or no morphological differences.

While character change and speciation must be considered wholly distinct phenomena, the occurrence of either undoubtedly leads to the occurrence of the other. Reproductive isolation facilitates the establishment of different characters. A great morphological difference between adjacent subspecies due to different environmental conditions causes selection against wasteful hybridization and so leads to reproductive isolation.

Finally, the same population structure is favorable both for adaptative character change due to peak-shifts, and to incipient speciation from local fixation of a chromosome rearrangement [Wright (50, 53)]. This population structure is one in which the population is broken up into numerous small colonies, frequently subject to extinction and refounding by stray individuals (perhaps a single fertilized female) from the more flourishing colonies. This situation is obviously very favorable for a peak-shift. If one of the founders happens to be heterozygous for a rearrangement that has been kept at low frequency by selection against the heterozygotes because of high aneuploidy, there is an appreciable chance that the arrangement may drift past the barrier and become homozygous. Whether the peak-shift or the fixation of the rearrangement occurs first, occurrence in a colony gives a favorable start for a new species.

Not all speciation is, however, initiated or clinched by chromosomal change. Evidence comes from many chains of intergrading subspecies that

have returned on themselves, producing forms that coexist in the same region as if they were distinct species, requiring only extinction of the intermediates to become such [Osgood (29), Mayr (25)].

As to the importance of different modes of speciation, White (36) states that: "Over 90% and perhaps 98% of all speciation events are accompanied by karyotypic change and in the majority of these cases the structured chromosomal rearrangements have played a primary role in initiating the divergence."

NONADAPTIVE DIFFERENCES BETWEEN SPECIES

While an association between the processes of speciation and of major adaptive change is both expected and observed, apparently nonadaptive differences have also been observed between related species and may constitute the most clearcut criteria for identification. In many cases, however, such differences have turned out upon careful study to have adaptive significance. Where they did not, such differences may be supposed to be pleiotropic effects of gene substitutions that are, overall, adaptive.

Before dismissing all cases as one or other of those discussed above, one should consider whether there is any theoretical way by which a genuinely nonadaptive gene substitution may occur at the species level. This possibility seems very unlikely in the case of the splitting of a species into two daughter species, both of such great effective size that the stochastic distributions of neutral gene frequencies are i-shaped ($4Nv$ greater than 1, where N is effective population number, and v is the rate per generation of origin of mutation).

If the effective population size of a species is persistently so small, over a long period, that the stochastic distributions of neutral genes are strongly U-shaped, slightly deleterious as well as neutral alleles are expected to reach fixation or near-fixation and accumulate, leading to serious inbreeding depression and ultimate extinction of the species.

I may have somewhat overestimated the likelihood of these alternatives in my early papers (45–46). In populations of intermediate size ($4Nv$ only slightly less than 1), nearly neutral genes may remain close to fixation for long periods without a serious accumulation of deleterious genes at other loci.

The most favorable situation for a nonadaptive differentiation of species, however, seems to be the passage of one of them through a bottleneck of small population size during which one or more nearly neutral genes happen to become fixed, unassociated with fixation of any seriously deleterious genes at other loci, with subsequent expansion of population number to such an extent that further fixation by random drift becomes unlikely. The

bottleneck may occur at the time of speciation, whether by fixation of a chromosome rearrangement or as a result of complete geographic isolation of a few stray individuals, or it may come later in the history of the species.

The discussion above concerns random drift from accidents of sampling (1). There is also the possibility of a period of extensive random drift from fluctuations in selection or other systematic factor, followed by a long period in which such fluctuations are absent.

Although I touched on the possibility of nonadaptive differentiation of species by sampling drift in a paper published in 1932, I did not consider this phenomenon to contribute more than some unimportant “noise” in the process of adaptive evolution, which was the main subject of this and earlier papers. Nevertheless, certain authors thought nonadaptive differentiation of species my primary concern, and that I considered it an alternative to progress by natural selection [Huxley (23), Fisher & Ford (13a, 14) (cf Wright (58, 62) the former approvingly, the latter not)]. Unfortunately, a great many others have followed their lead during the past 50 years, without checking on what I actually wrote in 1929, 1931, 1932 and later (44–46). I emphasize here that while I have attributed great importance to random drift in small local populations as providing material for natural selection among interaction systems, I have never attributed importance to nonadaptive differentiation of species. There have been even more extreme misrepresentations of the shifting balance theory than that discussed above, but these have not been taken up by others so extensively [(65), chapter 13; (67)].

There may have been some confusion because the qualitative theory proposed in 1929, 1931, 1932 far outran its mathematical treatment (in 1931). The latter was limited to two kinds of balance: that among all of the evolutionary pressures in determining an equilibrium frequency for pairs of alleles; and that between movement of such frequencies toward the equilibrium value and movement from them by the cumulative effects of sampling drift determining a certain stochastic distribution about equilibrium (not, in general, fixation). Moreover, only the simplest mode of selection was considered.

More general treatments of selection came later [Wright (47, 48, 54)]. Formulae for the stochastic distributions of multiple alleles and multiple interacting genes were presented in 1937 and 1949 (47, 59). The only population structure discussed in 1931 was the mathematically simple but rather unrealistic “island model” in which the influx into a local population in each generation is taken as representative of the whole species. A cluster model in which the influx comes only from neighboring localities was presented in 1951 (61). In the meantime, the possibilities of local differentiation within a continuum (area or linear) because of restricted dispersion was explored in several papers [Wright (50, 55, 57, 61, 62)]. Random drift from

causes other than accidents of sampling (referred to in 1931) was treated mathematically in 1948 (58). The last phase of the shifting balance process, the spreading of a favorable interaction system from its center of origin throughout the whole species by excess proliferation and dispersion, still awaits full mathematical treatment. Such spreading will obviously occur in the absence of any opposing process. These matters are reviewed in Wright (65).

MACROEVOLUTION

Most species are restricted to a single ecological niche by the occupation of all closely related niches by other species. Their evolution is usually restricted to increasing the efficiencies of their physiological processes and very slowly improving their adaptations to their particular niches. This is microevolution.

From time to time, however, a species is presented with vacant niches. This precipitates evolutionary changes of a different order of magnitude, those that constitute macroevolution [Wright (51, 52, 56, 59, 60)]. Individuals of the species may have moved into relatively unoccupied territory. Darwin was much impressed, on visiting the Galapagos Islands, by the presence of an endemic family of birds, the Geospizidae, which had several genera and many species. There was a strong suggestion that the Geospizidae had evolved from stray individuals of a mainland species. A similar example is presented by another family of birds, the Drepanididae, of the Hawaiian Islands. Such cases are not restricted to islands. Studies of geographic distribution in conjunction with paleontological data provide numerous examples from the continents. Darwin was impressed by such indications in South America.

Another way in which a species may be presented with vacant niches is by surviving a catastrophe that has destroyed forms in niches related to its own. The world underwent extinction of many forms, including the dinosaurs at the end of the Mesozoic period, which opened the way for the enormous expansion of the mammals during the Paleocene and later. The mammals had existed as a relatively insignificant group for some hundred million years.

Of greatest importance, perhaps, are those cases in which a species, gradually perfecting its adaptation to a particular niche, has reached a point at which some previously nonexistent niches have opened up. Most of the higher taxa probably originated this way.

Occupation of a vacant niche is a process that differs markedly from perfection in a single niche. It generally involves a much more drastic change in physiology and morphology. Most importantly, there is little or

no competition. A mutation that is only very imperfectly adapted and has rather serious deleterious side effects may be the best available mutation to occupy the new niche. There is no reason to suppose that it occurs any more frequently than before the niche opened up. Having occupied the niche after a fashion, it should be subject to very rapid improvement.

In many, perhaps most, cases, the initial occupation may, however, be by a minor allelic substitution. Drastic change can then occur very rapidly in the absence of competition by a succession of such substitutions, which is essentially a continuation of microevolution except in rate. The process would be facilitated by subdivision of the sparse population into randomly differentiated colonies and the shifting balance process.

There are cases, however, in which a single rather complicated mutational change seems to have been required for the species to occupy the new niche at all [as maintained by deVries (9), Goldschmidt (16), Willis (37, 38), and Schindewolf (32)]. Such a change by a single mutation is highly improbable, but becomes more plausible as a consequence of selection among interaction systems by the shifting balance process. The kind of interaction system most pertinent here is that in which a drastic mutation is associated with an array of nearly neutral modifiers that tend to alleviate the inevitable deleterious side effects. It may be assumed that all possible combinations of alleles of such modifiers occur among local populations of the parent species and that the drastic mutation is carried at a low frequency owing to recurrent mutation. The system consisting of the favorable combination of modifiers plus the drastic mutation becomes firmly established in one of the colonies by mass selection and spreads throughout the region by excess proliferation and dispersion [Wright (62)].

AMBIGUITY OF THE CONCEPT OF CAUSATION

Most of those who have written on the subject, including those mentioned above, have postulated the occurrence of a drastic mutation as the "cause" of macroevolutionary change. This may be true if "cause" refers to a necessary condition, but not if by "cause" we mean the change in conditions that precipitates the phenomenon in question at a particular time and place. The occurrence of a particular sort of drastic mutation at a rate of, say, 10^{-6} per generation during a period of millions of years wherever the species exists is obviously not a precipitating cause. The ultimate dynamic factor in evolution is to be sought in the universal tendency of living things to persist, if possible, and multiply. With respect to the nature of the change at any time, the most important consideration is the state of organization already attained. Next is the character of the mutations to which the genetic material is at the moment subject. The natures of the physiological and

morphological changes brought about by these mutations are of most interest to those concerned with the course of evolutionary change. The word "cause" may be used in all of these cases in this connection, but it is only the presentation of an adequate ecological opportunity, a vacant niche, that is the "cause" that precipitates a very rapid evolutionary change at a particular time and place and thus the "cause" in the sense used here. There may, indeed, be some delay in the assemblage of a favorable array of modifiers with the pertinent drastic mutation. The local sampling event or other form of random drift, and local occurrences of the drastic mutation, may thus be intermediary precipitating cues, but these will be of no avail without the vacant niche. The latter will usually be effective in inducing a macroevolutionary step even though the exact time and place of its first establishment depend on the former.

SUMMARY

Because of persistent misunderstandings, the shifting balance theory proposed half a century ago, primarily in connection with microevolution, is reviewed here at some length before discussing its relation to macroevolution. A basic assumption was that so complex a network of biochemical and developmental processes intervenes between the primary effects of genes and observed characters that typically many genes interact, often nonadditively, to affect each character, and, moreover, each gene substitution typically affects many characters.

It is also assumed that there are always a great many loci at which second alleles are maintained at equilibrium at fairly high frequencies, say 10% or more, by opposing pressures, usually weak, from recurrent mutation, influx from neighboring populations under somewhat different conditions, and diverse sorts of selection.

If all local populations are treated as being located in a multidimensional space, with a dimension for each allele other than the leading allele, coordinates according to the allelic frequencies, and a dimension added for selective value, the resulting topography of selective values is expected to be very rugged, with innumerable major and subordinate selective peaks connected by saddles. A given population tends to remain close to the peak that for historical reasons has come to control it. The peak itself, followed by the population, tends to move as conditions change, but even under constant conditions the population tends to wander continually in the neighborhood of its controlling peak because of changes in its set of gene frequencies due to accidents of sampling (if the effective population number is small, about 100 or less) and to fluctuations in the systematic pressures. The shifting balance process consists of occasional shifts locally across a saddle, from

control by a lower selective peak to control by a higher one. The process requires the occurrence of exceptionally wide local stochastic deviations, typically at two loci, with reversals of the direction of selection at these loci as the saddle is crossed.

Natural selection occurs at two levels: (*a*) selection among individuals, tending toward the establishment of the allele at each locus that has, on the average, the most favorable interaction effects with all such alleles at other loci, a conservative process (genic selection) that tends to bind the population ever more firmly to the selective peak that happens to control it, and (*b*) selection among differentiated local populations consisting in excess proliferation and dispersion from those that happen to have reached the higher selective peaks (organismic or genotypic selection). This process may be expected to continue until the whole species comes under control of the highest selective peak.

The theory was suggested by an analogous two-level process of artificial selection that has been employed in the improvement of livestock, involving (*a*) the development of superior herds by the more ambitious breeders by individual selection supplemented with close inbreeding together with good fortune in escaping serious inbreeding depression. This is followed by (*b*) selection among such herds as sources of breeding stock, especially males, by the breeders in general.

Most species in nature are restricted to a single niche by the occupation of all closely related niches by other species. Their evolution is restricted to very gradual improvement in physiology and in adaptation to their niche largely by the shifting balance process and relatively rapid adjustments to changing conditions, largely by individual selection. These processes constitute microevolution.

Occasionally, one or more vacant niches may be presented: (*a*) the species may enter relatively uninhibited territory; (*b*) it may survive a catastrophe that has eliminated other species in related niches; (*c*) it may achieve a breakthrough in the slow process of improvement that opens up previously non-existent niches.

The filling of a vacant niche may be expected to be an enormously more rapid process than further improvement in a long occupied niche. This is in part because of the premium on a major adaptive change but especially because of the lack of competition.

Such a major change may be accomplished by a rapid succession of small changes, but there is an opportunity, and sometimes a necessity, for utilization of a major mutation that has been recurring but has been kept at low frequency by adverse selection. Such a mutation may occupy the new niche in spite of the inevitable deleterious side effects because of the absence of competition, but occupation is facilitated if the population structure permits

operation of the shifting balance process with respect to nearly neutral modifiers that tend to remove the deleterious side effects, and thus favor establishment of the interaction system consisting of the major mutation and its modifiers at some locality.

A macroevolutionary step depends not only on establishment of a major character change but also on reproductive isolation (speciation). These are, in general, wholly distinct processes, but the occurrence of either favors establishment of the other. Moreover, the same extreme population structure that especially favors the shifting balance process also favors the fixation of a new chromosome rearrangement that gives incipient speciation, because of the aneuploidy of heterozygotes. This population structure is the presence in the species range of an extensive region inhabited by numerous small colonies frequently subject to extinction and refounding by stray individuals from the more successful colonies. A large number of such foundings, especially by single fertilized females, gives an appreciable chance that a rare chromosome rearrangement may pass the barrier presented by the aneuploidy of the heterozygotes.

The most likely precipitating cause of the origin of a macroevolutionary step is thus presentation of a vacant ecological niche to a species with a population structure that is favorable both for incipient speciation and for operation of the shifting balance process.

Literature Cited

1. Cain, A. J., Currey, J. D. 1963. Area effects in *Cepaea*. *Philos. Trans. R. Soc. London, Ser. B* 256:1-81
2. Castle, W. E. 1903. The laws of Galton and Mendel and some laws governing race improvement by selection. *Proc. Am. Acad. Arts Sci.* 39:233-42
3. Castle, W. E. 1916. Further studies of piebald rats and selection with observations on genetic coupling. *Carnegie Inst. Washington Publ.* 241:163-98
4. Castle, W. E. 1919. Piebald rats and selection: a correction. *Am. Nat.* 53: 370-76
5. Castle, W. E., Phillips, J. C. 1914. Piebald rats and selection. *Carnegie Inst. Washington Publ.* No. 195
6. Chetverikov, S. S. 1926. On certain aspects of the evolutionary process from the standpoint of modern genetics. *Zh. Eksp. Zool.* 2:3-54 (In Russian); *Proc. Am. Philos. Soc.* 105:167-95 (Trans M. Barker)
7. Darwin, C. 1859. *The Origin of Species by Means of Natural Selection*. London: John Murray. 432 pp.
8. Darwin, C. 1868. *The Variation of Animals and Plants Under Domestication*. Vols. 1, 2. London: John Murray. 473 pp., 495 pp.
9. de Vries, Hugo. 1901-1903. *Die Mutations Theorie*. Vols. 1, 2. Leipzig: Viet. 648 pp.
10. East, E. M. 1910. A Mendelian interpretation of variation that is apparently continuous. *Am. Nat.* 44:65-82
11. East, E. M. 1916. Studies on size inheritance in *Nicotiana*. *Genetics* 1:164-76
12. Fisher, R. A. 1922. On the dominance ratio. *Proc. R. Soc. Edinburgh* 42: 321-41
13. Fisher, R. A. 1930. *The Genetical Theory of Natural Selection*. Oxford: Clarendon. 272 pp.
- 13a. Fisher, R. A., Ford, E. B. 1947. The spread of a gene in natural populations in a colony of the moth, *Panexia dominula* L. *Heredity* 1:143-74
14. Fisher, R. A., Ford, E. B. 1950. The "Sewall Wright" effect. *Heredity* 4:47-49
15. Ford, E. B. 1964. *Ecological Genetics*. London: Methuen. 335 pp.
16. Goldschmidt, R. 1940. *The Material Basis of Evolution*. New Haven, Conn., Yale Univ. Press. 436 pp.

17. Gould, S. J. 1980. Is a new and general theory of evolution emerging? *Paleontology* 6(1):119-30
18. Gould, S. J., Eldredge, N. 1977. Punctuated evolution: The tempo and mode of evolution reconsidered. *Paleontology* 3:115-51
19. Haldane, J. B. S. 1924. A mathematical theory of natural and artificial selection. Part 1. *Cambridge Philos. Soc. Trans.* 23:19-41
20. Haldane, J. B. S. 1932. *The Causes of Evolution*. New York: Harper. 235 pp.
21. Haldane, J. B. S. 1957. The cost of natural selection. *J. Genet.* 55:511-24
22. Hardy, G. H. 1908. Mendelian proportions in a mixed population. *Science* 28:49-50
23. Huxley, J. S. 1942. *Evolution: The Modern Synthesis*. New York: Harper. 645 pp.
24. Jenkin, F. 1867. Origin of species. *North Br. Review* 46:272-318
25. Mayr, E. 1963. *Animal Species and Evolution*. Cambridge, Mass: Harvard Univ. Press. 797 pp.
26. McPhee, H. C., Wright, S. 1925. Mendelian analysis of the pure breeds of livestock. III. The Shorthorns. *J. Hered.* 16:205-15
27. Nilsson-Ehle, H. 1909. Kreuzungsuntersuchungen an Häfer und Weizen. Lunds, Universitat. *Aerskr. N. F.* 5: 2:1-22
28. Norton, H. T. J. 1915. In *Mimicry in Butterflies*, by R. C. Punnett. Cambridge: Cambridge Univ. Press. 188 pp.
29. Osgood, W. H. 1909. Revision of the mice of the American genus *Peromyscus*. *North Am. Fauna No.* 28. Washington DC: Bur. Biol. Survey, US Dept. Agric. 285 pp.
30. Rensch, B. 1929. *Das Prinzip geographische Rassenkreise und das Probleme der artbildung*. Berlin: Borntraeger. 206 pp.
31. Robbins, R. B. 1918. Application of mathematics to breeding problems. II. *Genetics* 3:73-92
32. Schindewolf, O. H. 1950. *Grundfragen der Paläontologie*. Stuttgart: Schweitzerbart. 505 pp.
33. Shull, G. H. 1908. The composition of a field of maize. *Am. Breeders Assoc.* 4:296-301
34. Simpson, G. G. 1944. *Tempo and Mode in Evolution*. New York: Columbia Univ. Press. 237 pp.
35. Weinberg, W. 1908. Ueber den Nachweis der Vererbung beim Menschen. *Jahresch. Ver. Naturk. Wuerntemb.* 64:368-82
36. White, M. J. D. 1978. *Modes of Speciation*. San Francisco: Freeman. 455 pp.
37. Willis, J. C. 1922. *Age and Area*. Cambridge: Cambridge Univ. Press. 258 pp.
38. Willis, J. C. 1940. *The Course of Evolution*. Cambridge: Cambridge Univ. Press. 207 pp.
39. Wright, S. 1920. Principles of livestock breeding. *US Dept. Agric. Bull.* 905.
40. Wright, S. 1921. Systems of mating. *Genetics* 6:111-78
41. Wright, S. 1922. The effects of inbreeding and crossbreeding on guinea pigs. *US Dept. Agric. Bull.* 1090.
42. Wright, S. 1922. Coefficients of inbreeding and relationship. *Am. Nat.* 56: 330-38
43. Wright, S. 1923. Mendelian analysis of the pure breeds of livestock. II. The Duchess family of Shorthorns as bred by Thomas Bates. *J. Hered.* 14:405-22
44. Wright, S. 1929. Evolution in a Mendelian population. *Anat. Rec.* 44:287
45. Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16:97-159
46. Wright, S. 1932. The roles of mutation, inbreeding, cross-breeding and selection in evolution. *Proc. 6th Int. Congr. Genet.* 1:356-66
47. Wright, S. 1937. The distribution of gene frequencies in populations. *Proc. Natl. Acad. Sci. USA* 23:307-20
48. Wright, S. 1938. The distribution of gene frequencies under irreversible mutation. *Proc. Natl. Acad. Sci. USA* 24:253-59
49. Wright, S. 1940. The statistical consequences of Mendelian heredity in relation to speciation. In *The New Systematics*, ed. J. S. Huxley, pp. 161-83. Oxford: Clarendon
50. Wright, S. 1940. Breeding structure of populations in relation to speciation. *Am. Nat.* 74:232-48
51. Wright, S. 1941. The "Age and Area" concept extended. Review of book by J. C. Willis. *Ecology* 22:345-47
52. Wright, S. 1941. "The Material Basis of Evolution" by R. Goldschmidt: a review. *Sci. Mon.* 53:165-70
53. Wright, S. 1941. On the probability of fixation of reciprocal translations. *Am. Nat.* 75:513-22
54. Wright, S. 1942. Statistical genetics and evolution. *Bull. Am. Math. Soc.* 48: 223-46
55. Wright, S. 1943. Isolation by distance. *Genetics* 28:114-38

56. Wright, S. 1945. "Tempo and Mode in Evolution," a critical review. *Ecology* 26:415-19
57. Wright, S. 1946. Isolation by distance under diverse systems of mating. *Genetics* 31:39-59
58. Wright, S. 1948. On the roles of directed and random changes in gene frequency in the genetics of populations. *Evolution* 2:279-94
59. Wright, S. 1949. Adaptation and selection. In *Genetics Paleontology and Evolution*, ed. G. L. Jepsen, G. G. Simpson, E. Mayr. Princeton, N.J.: Princeton Univ. Press
60. Wright, S. 1949. Population structures in evolution. *Proc. Am. Philos. Soc.* 93:471-78
61. Wright, S. 1951. The genetical structure of populations. *Ann. Eugenetics* 15:323-54
62. Wright, S. 1951. Fisher and Ford on "The Sewall Wright Effect." *Am. Sci.* 39:452-59
63. Wright, S. 1968. *Evolution and the Genetics of Populations*, Vol. 1. *Genetic and Biometric Foundation*. Chicago: Univ. Chicago Press
64. Wright, S. 1969. See Ref. 63, Vol. 2. *The Theory of Gene Frequencies*.
65. Wright, S. 1977. See Ref. 63, Vol. 3. *Experimental Results and Theoretical Deductions*
66. Wright, S. 1978. See Ref. 63, Vol. 4. *Variability Within and Among Natural Populations*.
67. Wright, S. 1980. Genic and organismic selection. *Evolution* 34:825-43
68. Yule, G. U. 1902. Mendel's laws and their probable relation to interracial heredity. *New Phytol.* 2:192-207, 222-38