J. Genet. Vol. 71, No. 3, December 1992, pp. 73–79. © Printed in India.

Haldane and the emergence of theoretical population genetics, 1924–1932*

SAHOTRA SARKAR
Theoretical Biology Group, Boston University, 745 Commonwealth Ave., Boston, MA 02215, USA

MS received 4 December 1992

1. Introduction

It is well known that R. A. Fisher, J. B. S. Haldane and S. Wright were primarily responsible for laying down the foundations of theoretical population genetics in the 1920s (Sarkar 1992a). There were others such as S. Chetverikov (see Adams 1968), who were no doubt important for the conceptual innovations that they introduced but, nevertheless, were not remotely as influential as the canonical triumvirate in the period 1918–1932 when theoretical population genetics came of age. What is far less clear is the exact nature of the individual contributions of Fisher, Haldane and Wright, that is, the differences between, and the relative importance of, these contributions. For Fisher and Wright, at least part of such an analysis has been previously attempted (see, e.g., Provine 1985, 1986) but Haldane's role has largely been ignored probably because of his comparatively lesser influence on theoretical population genetics since 1935.

The purpose of this paper is to use the occasion of Haldane's birth centenary to begin to remedy this lacuna. It will first be shown (section 2) that what was most important about Haldane's contributions to population genetics was a clear exposition of the general structure of genetic models of natural selection that made it possible for such models to be constructed in any context. Some of the more important models that Haldane constructed in the 1920s will then be briefly discussed (section 3). Finally (section 4), it will be observed that Haldane's contributions to population genetics after the mid-1930s was much more episodic than the systematic work of the 1920s and, although individual insights continue to be relevant today, from the historical point of view, this later work is less important than what he achieved in the 1920s.

2. The structure of models of selection

Between 1924 and 1934, Haldane published a set of ten papers all (except one) of which were entitled "A Mathematical Theory of Natural and Artificial Selection". Ultimately, these were his most important contributions to theoretical population genetics. Haldane was neither the first to attempt to apply mathematics to Mendelian genetics, nor even among the earliest to construct mathematical models of selection.

* Based on a talk given at the Haldane Centenary Symposium held on 6 November 1992 at Ahmedabad as part of the 58th Annual Meeting of the Indian Academy of Sciences.

Mendelian genetics had been rediscovered around 1900 and mathematical exploration of its consequences had begun almost immediately, though in a rather critical vein, with the work of Pearson (1904). The first truly significant result, the Hardy–Weinberg law, was obtained in 1908: for a one-locus system with two alleles, **A** and **a**, if p and q are the frequencies of these alleles in any generation in an infinite population with random mating, the **AA**, **Aa**, and **aa** genotypes have frequencies p^2, $2pq$ and q^2 in the next and all subsequent generations. What selection (or the violation of any of the assumptions of this model, e.g. random mating) does is distort these ratios. During the following decade, H. C. Warren (1917) and H. J. T. Norton (among others) began to develop models of selection acting on Mendelian populations.

Norton wrote a thesis at Trinity College, Cambridge, in 1911 exploring the consequences of selection on Mendelian populations. However, his only published results were in a table in Punnett's (1915) *Mimicry in Butterflies*. Haldane arrived at Cambridge in 1923, ostensibly as a biochemist, and took up residence at Trinity (see Sarkar 1992b). From Norton he learnt how little was known about the kinetics of selection. He set about to rectify the situation. The result was the set of ten papers, "A Mathematical Theory of Natural and Artificial Selection". Unlike Warren and Norton, Haldane was not only interested in simple cases but in developing a general theory that could be applied to any of the wide variety of biologically plausible structures of populations. Right from the beginning, Hadane was interested in the consequences of Mendelism *in general*, not just in special cases. What emerged was a cluster of models of impressive variety but, more than that, a general picture of how to construct these models *ab initio*.

There are four aspects of Haldane's systematic development in these papers that deserve special emphasis. The *first* was his assumption that what models of selection must try to do is provide a theory by which natural selection accounts for the facts of evolution. The facts of evolution were taken for granted. The *second* was his assumption that any such theory had to be quantitative. The *third* was his realization that such a quantitative theory, to be successful, had to explain the rate of change of the relevant characteristics of a population. This emphasis on rates was characteristic of Haldane throughout his career. In retrospect it was particularly insightful on two counts: (i) ultimately, any quantitative theory of evolution must be a dynamical theory, where rates of change are the most fundamental quantities of interest; and (ii) as Darwin had already realized in the 1860s, for a theory of evolution to be successful in explaining the history of life on earth, it must show that the amount of time available sufficed to account for the hypothesized evolutionary changes. The *fourth*, and most important, aspect of Haldane's thinking in this period was a fully general and clear conception of the structure of genetic models of selection. Haldane gave an explicit characterization of that structure. At the beginning of the first paper (Haldane 1924a, p. 19) of the series, he notes:

> A satisfactory theory of natural selection must be quantitative. In order to establish the view that natural selection is capable of accounting for the known facts of evolution we must show not only that it can cause a species to change, but that it can cause it to change at a rate that will account for present and past transmutations. In any given case we must specify:
>
> (1) The mode of inheritance of the character considered,
> (2) The system of breeding in the group of organisms studied,
> (3) The intensity of selection,
> (4) Its incidence (e.g. on both sexes or only one), and
> (5) The rate at which the proportion of organisms showing the character increases or diminishes.

It should then be possible to obtain an equation connecting (3) and (5).

What is most important about this characterization is that it provided a clear method for the construction of genetical models of selection, beyond those that Haldane constructed himself. This structure continues to serve as the basic template for constructing models of selection to this day (see, e.g., Crow and Kimura 1970). In particular, Haldane was clear on the point that the most important parameters in these models were the items specified by (3) and (5), though he failed to note that the former was an independent and the latter a dependent parameter. Unlike (3), the three other assumptions of a model were reflected by the structure of the equation obtained, rather than a parameter to be measured experimentally to test a model. In 1924 neither Fisher, Wright, nor anybody else had given so detailed a characterization of these models. As a theorist, the characterization of this structure was Haldane's most significant achievement.

3. The models

Equally impressive in "A Mathematical Theory of Natural and Artificial Selection" is the variety of the explicit models that Haldane constructed. In Part I, he looked at what appeared to be all the biologically plausible one-locus selection models with complete dominance, non-overlapping generations, an asexual or a Mendelian mode of inheritance, and either random mating or complete self-fertilization (Haldane 1924a). Complications introduced by sex and linkage were routinely incorporated. The most famous of the models was one in which the allele A was dominant over **a**, and was also selected. Assuming that A referred to the melanic form of the peppered moth, Haldane showed that in the sooty Manchester environment, where its frequency had increased from less than 1% to at least 99% between 1848 and 1901, the dominants were selected over the recessives at least in a ratio of 3:2. At that time so intense a selection was considered highly improbable. However, the fieldwork of Kettlewell (1956), three decades later, showed it to be plausible and Haldane's application remains a textbook example of natural selection at work in a wild population. More than that, it established theoretical population genetics as an experimental discipline: its quantitative predictions were potentially subject to tests in the field. This is just the best known of Haldane's explicit attempts to connect theory with data. A subsidiary theme that runs throughout this series of papers was the importance of performing laboratory or field studies to test the genetical models. It is arguable that Haldane was at least implicitly aware of the problem posed by the radical underdetermination of theory by data in population genetics. In fact, it will be argued below that this was why he shied away from the controversies surrounding the field since the mid-1930s, as disciples of Fisher and Wright began to engage in disputes about the relative merits of their respective theories of evolution.

Part II of the series (Haldane 1924b) relaxed some of the earlier assumptions to consider partial self-fertilization, inbreeding, assortative mating and partial fertilization. Incomplete dominance was discussed in Part III but, more importantly, that paper introduced the consideration of multiple loci and tetraploidy (Haldane 1926). Meandering from what is now considered as population genetics, Part IV (Haldane 1927a) is demographical and analyses the absolute growth of a population rather than just frequency changes. The most important result was an independent

proof of Lotka's stable age distribution theorem. Returning to conventional population genetics, and drawing on Fisher (1922), Haldane (1927b) calculated the properties of the equilibrium between mutation and selection in Part V. This calculation was particularly important because they provided the basis for the later discussions of genetic loads (see below).

Part VI discussed an island model of isolation (Haldane 1930). The analysis of the interaction between migration and selection in this model predated the better-known analysis of Wright (1931) even though Haldane was loath to attribute too much importance to these models. Part VII was even more interesting: Haldane (1931a) analysed the relation between competition and selection. Competition was modelled by what has come to be known as truncation selection and Haldane showed that the intensity of selection may, under certain circumstances, decrease with an increase in the intensity of competition. This paper demonstrated the power of quantitative analysis. As Haldane put it, it showed that the assumption "often made that when competition is extremely intense at any stage in a life cycle, natural selection is bound to be intense also" was false. Quantitative analysis thus exposed the limitations of qualitative argument. Whether or not Haldane's model of competition is correct, it nevertheless was the first systematic treatment of the peculiar properties of truncation selection.

Part VIII looked at metastable populations (Haldane 1931b); Part IX looked at some special properties of rapid selection (Haldane 1932b); Part X explored the properties of artifical selection (Haldane 1934). Part VIII is perhaps the most interesting of the papers in this series. Haldane considered a two-locus two-allele model (with complete dominance and infinite population size) where the double homozygotes **aabb** and **AABB**, and the heterozygotes **AaBb, AaBB** and **AABb** (which have the same phenotype as **AABB**) have higher fitnesses than any of the intermediate phenotypes. Assuming some symmetry properties between the fitnesses of the other heterozygotes (**aaBB** and **aaBb** have the same fitness as do **Aabb** and **AAbb**), he showed that there were no stable polymorphic equilibria that included heterozygotes. He noted that heterosis with respect to either locus could result in a stable polymorphism. If there were n loci (with two alleles each), the space of genotypes form an n-dimensional hypercube. If in this hypercube, fitnesses were randomly distributed, there would at most be 2^{n-1} stable points, that is, genotypes that were fitter than any neighbours separated from them by an allelic difference only at one locus. Populations at such points were "metastable", that is, stable with respect to their nearest neighbours.

This remains an interesting multiple locus model which has been explored by Aggarwal and Sarkar (1992) and others. But Haldane went no further, and only observed (Haldane 1930, pp. 141–142): "It is suggested that in many cases related species represent stable types ... and that the process of species formation may be a rupture of the metastable equilibrium. Clearly such a rupture will be specially likely where small communities are isolated." It was a clear anticipation of Wright's (1931) shifting-balance theory of evolution, but no more than an anticipation that Haldane never even bothered to mention after that theory became central to population genetics after the mid-1930s.

It is impossible not to wonder why Haldane did not pursue this insight further. The reason is probably the same as that for his refusal to become involved in the disputes between Fisher and Wright since the 1930s or, for that matter, his "retirement" from population genetics after the mid-1930s (see below). There were no experimental

data that could resolve these disputes and Haldane was temperamentally disinclined to theorize in the complete absence of potential evidence. His preferred mode of theorizing was the construction of models, not the derivation of general theorems (cf. Fisher (1930) for a contrast). As far as Haldane was concerned, there was no reason yet to assume any single mechanism as the preferred mode of evolutionary change: in the absence of any potentially corroborating evidence, he was not willing to espouse any single *theory of evolution*. Should either Fisher or Wright (or, from the next generation, Kimura) turn out to be correct about their univocal theories of evolution, Haldane's hesitation will justify his customary placement on a lesser pedestal than them in the pantheon of theoretical population genetics. If not, he will have shown more wisdom, though not have had more influence on the development of the field since the mid-1930s. The future will tell.

4. After 1932

Meanwhile, in 1932, Haldane published *The Causes of Evolution*, based on lectures given in 1931. In a single text, he managed to incorporate a discussion of population genetics, palaeontology and the new chromosomal genetics of the Morgan school. In the Appendix he summarized the state of mathematical population genetics and provided a comprehensive review of his own work along with that of Fisher and Wright. He even introduced some new models: in particular, he produced the first quantitative attempt to explain altruism on the basis of kin selection. It was a remarkable achievement but it also marked Haldane's departure from the mainstream of mathematical, if not also theoretical, population genetics. Parts IX and X of "A Mathematical Theory of Natural and Artificial Selection", which came later, were at most of incident importance.

In 1933 Haldane moved to University College, London. At that point, his interest in theoretical population genetics declined and he only very occasionally returned to it in later life. Instead, he first turned to human genetics. With J. Bell he discovered the first putative case of linkage in humans, that between the loci for haemophilia and colour-blindness (see Bell and Haldane 1937), and made the first attempt to construct a map of a human chromosome (Haldane 1936). In 1937, he extended his earlier calculations (Haldane 1927b) to show that the decline of the average fitness of a population due to mutation of the optimal allele was proportional to the mutation rate but not to the induced decrease in fitness (Haldane 1937). The same point was later (apparently independently) rediscovered by Muller (1950), who chose to look upon the decline in fitness of a population as a "load" due to mutation. Ultimately, Haldane's (1937) discovery, after significant clarification by Crow (1958), came to be called the "mutation load principle" to distinguish it from other kinds of possible loads (see Wallace 1991, pp. 15–17). This principle continues to be important, especially in attempts to gauge the potential genetic ill effects of radiation on natural populations, though the extent of its importance remains controversial (see Wallace 1991). In a similar vein, Haldane (1957) estimated the loss of average fitness of a population due to the substitution of one allele for another. This estimate, which came to be known as the "substitution load", was particularly important as it set the stage for Kimura's development of the neutral theory of evolution (see Kimura 1985).

Finally, towards the end of his life, in the early 1960s, Haldane returned briefly

but systematically to population genetics, in collaboration with S. Jayakar. Mathematically more sophisticated than anything that Haldane had done in the 1920s—this was presumably what Jayakar contributed to the collaboration—several interesting insights emerged. In particular, Haldane and Jayakar (1963) showed that selection in varying directions was sufficient to maintain polymorphism in a population. A more rigorous and careful derivation of this result was given by Cornette (1981). Its importance lies in its novelty as a potential mechanism for the maintenance of allelic polymorphism in a population, as opposed to the more usually invoked hypotheses of heterosis and neutrality. Moreover, even in cases when heterosis is usually assumed as the explanation of polymorphism, such as in the sickle-cell haemoglobin allele in human populations, the Haldane–Jayakar mechanism might well be a superior explanation. If heterozygotes were usually slightly less fit than the normal haemoglobin homozygotes, but more fit only during malaria epidemics, the varying direction of selection would account for the observed polymorphism according to the Haldane–Jayakar mechanism (see Nagylaki 1992, p. 66). The assumption that the heterozygote is more fit than the normal homozygote, even in the absence of a malaria epidemic, which is required in the usual heterosis-based model (Edelstein 1986), is probably much less reasonable.

There are many ways in which this Haldane–Jayakar model, and some of the other results of their collaboration, can be profitably extended. Nevertheless, ultimately, Haldane's most important contribution to theoretical population genetics remains his systematic work in the 1920s. That was when he helped create a new discipline, not only because of the results that were obtained, but because he elucidated a clear method by which others could construct models of selection analogous to his. Compared to this work the later insights, however illuminating, are episodic. Moreover, by always attempting to connect model to data, he ensured that theoretical population genetics remained a part of biology in spite of his often-quoted hope that it would some day become "a new branch of applied mathematics" (Haldane 1932a).

References

Adams M. 1968 The founding of population genetics: Contributions of the Chetverikov school 1924–1934. *J. Hist. Biol.* 1: 23–39

Aggarwal A. and Sarkar S. 1992 On the evolution of complex genomes: Adaptive graph computations with a CM-2 Connection Machine. In *Applied computing: Technological challenges of the 1990's* (eds) H. Berghel, E. Deaton, G. Hedrick, D. Roach and R. Wainwright (New York: ACM Press) Vol. II pp. 798–807

Bell J. and Haldane J. B. S. 1937 The linkage between the genes for colour-blindness and haemophilia in man. *Proc. R. Soc. London* B123: 119–150

Cornette J. L. 1981 Deterministic genetic models in varying environments. *J. Math. Biol.* 12: 173–186

Crow J. F. 1958 Some possibilities for measuring selection intensities in man. *Hum. Biol.* 30: 1–13

Crow J. F. and Kimura M. 1970 *An introduction to population genetics theory* (Minneapolis: Burgess)

Edelstein S. J. 1986 *The sickled cell: From myths to molecules* (Cambridge, USA: Harvard University Press)

Fisher R. A. 1922 On the dominance ratio. *Proc. R. Soc. Edinburgh* 42: 321–341.

Fisher R. A. 1930 *The genetical theory of natural selection* (Oxford: Clarendon Press)

Haldane J. B. S. 1924a A mathematical theory of natural and artificial selection. Part I. *Trans. Cambridge Philos. Soc.* 23: 19–41

Haldane J. B. S. 1924b A mathematical theory of natural and artificial selection. Part II. The influence of partial self-fertilization, inbreeding, assortative mating, and selective fertilization on the composition of Mendelian populations, and on natural selection. *Proc. Cambridge Philos. Soc.* 1: 158–163

Haldane J. B. S. 1926 A mathematical theory of natural and artificial selection. Part III. *Proc. Cambridge Philos. Soc.* 23: 363–372

Haldane J. B. S. 1927a A mathematical theory of natural and artificial selection. Part IV. *Proc. Cambridge Philos. Soc.* 23: 607–615

Haldane J. B. S. 1927b A mathematical theory of natural and artificial selection. Part V. Selection and mutation. *Proc. Cambridge Philos. Soc.* 23: 838–844

Haldane J. B. S. 1930 A mathematical theory of natural and artificial selection. Part VI. Isolation. *Proc. Cambridge Philos. Soc.* 26: 220–230

Haldane J. B. S. 1931a A mathematical theory of natural and artificial selection. Part VII. Selection intensity as a function of mortality rate. *Proc. Cambridge Philos. Soc.* 27: 131–136

Haldane J. B. S. 1931b A mathematical theory of natural and artificial selection. Part VIII. Metastable populations. *Proc. Cambridge Philos. Soc.* 27: 137–142

Haldane J. B. S. 1932a *The causes of evolution* (London: Longmans, Green)

Haldane J. B. S. 1932b A mathematical theory of natural and artificial selection. Part IX. Rapid selection. *Proc. Cambridge Philos. Soc.* 28: 244–248

Haldane J. B. S. 1934 A mathematical theory of natural and artificial selection. Part X. Some theorems on artificial selection. *Genetics* 19: 412–429

Haldane J. B. S. 1936 A provisional map of a human chromosome. *Nature* 137: 398–400

Haldane J. B. S. 1937 The effect of variation on fitness. *Am. Nat.* 71: 337–349

Haldane J. B. S. 1957 The cost of natural selection. *J. Genet.* 55: 511–524

Haldane J. B. S. and Jayakar S. 1963 Polymorphism due to selection of varying direction. *J. Genet.* 58: 237–242

Kettlewell H. B. D. 1956 A résumé of investigations on the evaluation of melanism in the Lepidoptera. *Proc. R. Soc. London* B145: 297–303

Kimura M. 1985 Genes, populations, and molecules: A memoir. In *Population genetics and molecular evolution* (eds) T Ohta and K Aoki (Berlin: Springer-Verlag) pp. 459–481

Muller H. J. 1950 Our load of mutations. *Am. J. Hum. Genet.* 2: 111–176

Nagylaki T. 1992 *Introduction to theoretical population genetics* (Berlin: Springer-Verlag)

Pearson K. 1904 Mathematical contributions to the theory of evolution. XII. On a generalised theory of alternative inheritance, with special reference to Mendel's laws. *Philos. Trans. R. Soc.* A203: 53–86

Provine W. B. 1985 The R. A. Fisher–Sewall Wright controversy. In *Oxford surveys in evolutionary biology* (eds) Paul H. Harvey and Linda Partridge (Oxford: Oxford University Press) vol. 2, pp. 197–219

Provine W. B. 1986 *Sewall Wright and evolutionary biology* (Chicago: University of Chicago Press)

Punnett R. C. 1915 *Mimicry in butterflies* (Cambridge: Cambridge University Press)

Sarkar S. 1992a The founders of evolutionary genetics: The editor's introduction. In *The founders of evolutionary genetics* (ed) S. Sarkar (Dordrecht: Kluwer) pp. 1–22

Sarkar S. 1992b Haldane as biochemist: the Cambridge decade, 1923–1932. In *The founders of evolutionary genetics* (ed) S. Sarkar (Dordrecht: Kluwer) pp. 53–81

Wallace B. 1991 *Fifty years of genetic load: An odyssey* (Ithaca: Cornell University Press)

Warren H. C. 1917 Numerical effects of natural selection acting upon Mendelian characters. *Genetics* 2: 305–312

Wright S. 1931 Evolution in Mendelian populations. *Genetics* 16: 97–159