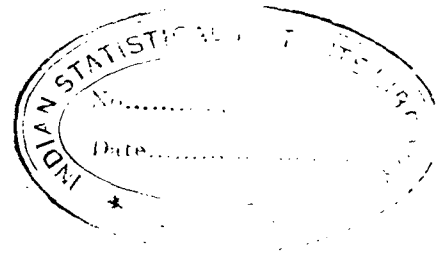
*Reprinted without change of pagination from the*
Proceedings of the Royal Society, B, *volume* 135, 1948




CROONIAN LECTURE

# The formal genetics of man

BY J. B. S. HALDANE, F.R.S.

(*Delivered* 7 *November* 1946—*Received* 18 *December* 1947)

Man has obvious disadvantages as an object of genetical study. The advantages are that very large populations are available, and that many serological differences and congenital abnormalities have been intensively investigated.

Some characters are found to obey Mendel's laws with great exactitude. In others the deviations are such as to suggest the existence of a considerable selective mortality, perhaps prenatal. In yet other cases the observations are biased because we only know that we are investigating the progeny of two heterozygotes when the family includes at least one recessive. Statistical methods which eliminate this bias are described.

Still more complex methods are needed for the detection and estimation of linkage. Several such cases have been detected with greater or less certainty, and the frequency of recombination between the loci of the genes for colour-blindness and haemophilia is now estimated at $10 \pm 4$ %. If the theory of partial sex-linkage be accepted, it is possible to make a provisional map of a segment of the human sex chromosome.

When a gene is sublethal, as are those for haemophilia and achondroplasic dwarfism, its elimination by natural selection is in approximate equilibrium with its appearance by mutation, and the frequency of the latter process can be estimated. The mutation rates at five human gene loci lie between $4 \times 10^{-5}$ and $4 \times 10^{-6}$ per locus per generation. These are the only estimates available for vertebrates. The rates per generation are rather higher than those in *Drosophila*, but those per day are so small that much, or even all, human mutation may be due to natural radiations and particles of high energy.

In 1906 Bateson delivered a lecture to the Neurological Society on 'Mendelian heredity and its application to man' in which he described the genetics of brachydactyly and congenital cataract, which are dominant to the normal (the word is used loosely, since the abnormal homozygote is unknown). He suggested that albinism and alcaptonuria were recessive, and he described the laws of inheritance of haemophilia and colour-blindness, though he did not, of course, give the explanation of these laws which is now accepted.

In the ensuing 40 years a very large number of pedigrees have been collected, unfortunately with very variable standards of accuracy. These show that more than a hundred different human abnormalities are certainly due to single gene substitutions, and that several hundred more are probably so. For example, Cockayne (1933) listed eighty abnormalities of the skin, hair, nails, and teeth which are probably due to dominant autosomal genes, eighteen to autosomal recessives, and thirteen to sex-linked genes. In over half the cases the evidence is adequate.

On the other hand the genetical analysis of the normal polymorphism of a race such as our own for colour, size, and shape has not gone far. The genetics of eye colour, for example, are far more complex than was originally thought, and stature is undoubtedly determined by a large number of genes, as well as by environmental influences. Still less progress has been made in the analysis of the genetics of those differences in skin colour and hair shape which exist between the major human races. However, immunology has revealed a polymorphism existing in all races which was wholly unexpected when Bateson wrote. Its genetical basis is exceedingly simple; perhaps because antigens are direct products of gene action, while pigments are the end products of complex chains of metabolic processes in which many, if not all, of the steps are controlled by different genes; and the processes of morphogenesis are even more complicated. Meanwhile, genetics have developed along many lines, of which three are especially important. It has been shown that genes are material structures located at definite points in the chromosomes. If we can homologize the genes of organisms which conjugate and the 'transforming principles' of bacteria, which can apparently transfer them to one another without conjugation, just as they carry out a communal metabolism, the work of Avery, MacLeod & MacCarty (1944) suggests that genes, at least in some phases of their life cycles, may consist wholly of desoxy-ribose nucleic acid.

We have learned a good deal about the causal chain between a gene and its manifestation. Goldschmidt was a pioneer in this work. You, Mr President, played an important part in the analysis of the genetical control of anthocyanin production in flowers. Beadle, Tatum and others were able to specify the stages in the production of arginine and other essential metabolites controlled by different genes in *Neurospora*. In this country, Grüneberg, in the mouse, and Waddington, in *Drosophila*, investigated the genic control of morphological development.

Finally, a number of workers, notably Dobzhansky, Dubinin, Fisher, Haldane, Teissier & l'Heritier, Tsetverikov, and Wright have investigated the genetics of

populations both practically and theoretically, and they and others have discussed the bearing of their results on the problem of evolution.

All these methods are applicable to our own species. There is, however, a widespread belief that what I may call formal genetics, that is to say the study of heredity and variation, based on the description and counting of individuals, has ceased to be important, and that in future genetics will consist mainly of the study of biochemical and morphogenetic processes controlled by genes, and of evolutionary changes in populations; while the mere enumeration of the results obtained from various matings, and deductions drawn from such enumeration, are no longer of great interest. As I propose to devote this lecture to the pure or formal genetics of man, I may perhaps be pardoned if I state what seem to me to be the legitimate aims of human genetics, and so to justify what some will regard as a reactionary standpoint.

The final aim, perhaps asymptotic, should be the enumeration and location of all the genes found in normal human beings, the function of each being deduced from the variations occurring when the said gene is altered by mutation, or when several allelomorphs of it exist in normal men and women. In addition, information would be gathered on the effect of changes involving sections of chromosomes, such as inversions, translocations, deficiencies, and duplications.

The number of genes in a human nucleus almost certainly runs into thousands, possibly tens of thousands. Each has, so far as one can judge, a highly specific biochemical function. The end result of such a genetical study as I have adumbrated would be an anatomy and physiology of the human nucleus, which would be incomparably more detailed than the anatomy and physiology of the whole body as known at present. This end will perhaps be achieved in part by non-genetical methods, such as ultramicroscopic operations on the nuclei of human cells in tissue culture.

No doubt one result of such a study will be the possibility of a scientific eugenics, which may bear the same relation to the practices now or recently in vogue in certain countries as chemotherapy bears to the bleedings and purgations of early medicine. But other results may be more important. A knowledge of the human nucleus may give us the same powers for good or evil over ourselves as the knowledge of the atomic nucleus has given us over parts of the external world.

In this lecture I shall be largely concerned with the localization of genes in human chromosomes. A simple example will show why this is important. One of the common causes of blindness is retinitis pigmentosa. Ten years ago it could be said that in some pedigrees this disease was transmitted as a dominant, in others as a recessive of the ordinary type, occasionally as a sex-linked recessive. In 1936 I argued that some pedigrees showed partial sex-linkage, a phenomenon which I shall describe later. We can now say tentatively that one of the genes, the abnormality of which causes this condition, is carried in that segment of the sex chromosomes of which women possess two, and men only one; another, which may give dominant or recessive mutants, in that segment of the same chromosomes of which both sexes

possess two; while other such genes, how many we do not know, are carried in the other chromosomes. It is reasonably sure that they control different processes. And this is borne out by the fact that the partially sex-linked recessive type is never associated with deafness, while one of the autosomal recessives is so associated.

Pathologists will have to work out the aetiology of the different genetical types. They can hardly hope to do so until they are distinguished, if, as seems probable, each gene controls a different process. And just as the methods for the cure of bacillary and amoebic dysentery are very different, so it is unlikely that the same therapeutic measures will succeed against diseases, however similar in their symptoms, which are due to different genes. They certainly do not do so in *Drosophila melanogaster*. In that species at least four different recessives give eye colours which are scarlet because they lack a yellow pigment found in the normal eye, which is a derivative of tryptophane. The eye colour of the mutant *vermilion* can be made normal by injecting the larvae with kynurenine, for the gene present in normal flies but inactive in *vermilion* flies is concerned in the oxidation of tryptophane to kynurenine. But *cinnabar* flies are not cured, because they cannot catalyze a further stage in the pigment formation; and *cardinal* and *scarlet* flies are not cured because their eye rudiments cannot take up the pigment precursor (Ephrussi 1942). The four genes in question are carried at different loci in three chromosomes.

The first step in formal genetics is to establish that certain characters are inherited in accordance with Mendel's laws, and in particular that segregation occurs in Mendelian ratios.

This is certainly true in many cases where large numbers have been studied. Thus according to theory a member of blood group $AB$ produces equal numbers of $A$ and $B$ gametes. Table 1* shows that this is the case, the deviation from theory being less than the standard error of sampling. In the mating $A \times AB$ the $A$ children are derived from $A$ gametes of the $AB$ parent, the $B$ and $AB$ children from $B$ gametes; and so on. Such an agreement implies not only that the two types of gamete are

TABLE 1

| | children in group | | | | |
|---|---|---|---|---|---|
| parents | $O$ | $A$ | $B$ | $AB$ | total |
| $O \times AB$ | 8 | 633 | 646 | 3 | 1290 |
| $A \times AB$ | 0 | 533 | 247 | 312 | 1092 |
| $B \times AB$ | 2 | 183 | 406 | 232 | 823 |
| $AB \times AB$ | 0 | 28 | 36 | 65 | 129 |

total $A$ gametes 1609
total $B$ gametes 1647
$(d-1)\, n^{-\frac{1}{2}} = 0{\cdot}648.$
total homozygotes from $AB \times AB$ 64.
total heterozygotes from $AB \times AB$ 65.

* From Wiener (1943), p. 190. This includes all data published since 1931. Before this date only groups of over 250 children are included. This criterion omits three children of $AB$ mother assigned to group $O$ by two workers who had tested seven and nine children of such mothers, and whose findings have perhaps received undue attention.

formed in equal numbers, but that there is no marked selective mortality of either type of zygote. Thirteen children occur in unexpected groups. These represent the combined effects of illegitimacy, technical errors, and conceivably mutation or abnormal segregation of chromosomes. Clearly these causes combined will produce results smaller than sampling error.

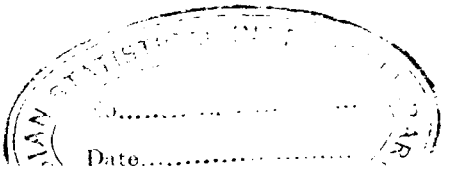
In the case of the pair of allelomorphic genes which respectively produce the *M* and *N* agglutinogens, there is at least prima facie evidence for very abnormal segregation. Where such cases have been investigated in animals, they have so far almost always been found to be due to selective mortality, or in fact natural selection, one genotype having a higher deathrate in the early stages of life than another. Taylor & Prior (1939*a*) found that the progeny from the marriages of heterozygotes which they had examined included a considerable excess of heterozygotes. They were convinced that this was not due to technical errors, and found a similar excess in the pooled data of other workers, giving 54·85 % of heterozygous children, the excess being 3·08 times its standard error. Wiener (1943) believes the excess to be an artefact due to the use of incompletely absorbed testing fluids, and that no such excess occurs where this error is avoided. Now if the sera used give too many *MN* children they should also give too many *MN* parents. Taylor & Prior (1939*b*) showed that in most series of data the square of the number of *MN* individuals approximates to four times the product of the numbers of *M* and *N*, while a few show a great excess. I have therefore applied their test to all the series in Wiener's table 46, and eliminated cases where $\chi^2$ exceeds four for the parents.† I have also omitted two small series containing 40 children between them, and two Japanese and one German series which were not available in English libraries, and have inserted one German series. The results are shown in table 2. The values of $\chi$ are calculated by Taylor & Prior's method, a positive value denoting an excess of *MN* over expectation among the parents. It will be seen that there are 54·85 % of heterozygotes. The excess is 3·16 times its standard error. The probability of so large an excess or defect by sampling error is 0·0016. If it were due to the systematic use of incompletely absorbed testing fluids, we should expect to find a lower percentage of heterozygotes in those series where $\chi$ is negative. The value found is 56·77 %, which is slightly, but not significantly higher. The excess of *M* over *N* children is just below twice its standard sampling error, and not statistically significant.

If the numbers of heterozygotes (*MN*) and homozygotes ($M+N$) in the different groups are compared, we find $\chi^2 = 17{\cdot}21$ for 11 degrees of freedom. Thus $P = 0{\cdot}10$, and the data are not significantly heterogeneous, as they would probably be if some workers had used faulty methods.

Table 3 shows the totals found for all types of marriage by the twelve authors cited in table 2. Illegitimacy, technical errors, and mutation clearly account for

† Wiener gives the reasons for excluding Lattes & Garrasi's (1932) data. The other relevant values of $\chi$ are: *Dahr* (1940) +4·51, Hirzfeld & Kostuch (1938) +2·16, Landsteiner & Wiener (1941) +2·002, Wiener & Sonn (1943) −2·20. The last three values may well all be due to sampling error.





very few unexpected classifications. In the case of the $MN \times M$ and $MN \times N$ marriages, the differences between the classes where equality is expected are less than twice their standard errors. It will be seen that the mean number of children examined per $MN \times MN$ marriage was decidedly less than that in the other groups. The figure 2·73 is not of course the mean fertility per marriage, as sterile marriages were excluded, it was not always possible to examine all the living children, and

TABLE 2

| authors | number of families | $\chi$ | children $M$ | children $MN$ | children $N$ |
|---|---|---|---|---|---|
| Landsteiner & Levine | 11 | +0·017 | 17 | 31 | 7 |
| Wiener & Vaisberg | 25 | −0·006 | 29 | 58 | 29 |
| Schiff | 33 | +1·808 | 18 | 48 | 22 |
| Crome | 9 | +0·083 | 4 | 10 | 4 |
| Clausen | 70 | −0·777 | 38 | 74 | 28 |
| Blaurock | 23 | +1·002 | 25 | 40 | 25 |
| Moureau | 53 | +0·444 | 45 | 102 | 41 |
| Hyman | 32 | +0·906 | 10 | 41 | 16 |
| Matta* | 20 | −0·781 | ·9 | 45 | 10 |
| | 20 | −1·741 | | | |
| Dàhr & Bussmann | 30 | −0·514 | 38 | 70 | 18 |
| Taylor & Prior (*a*) | 56 | −0·349 | 10 | 38 | 8 |
| Holford | 34 | +1·185 | 24 | 37 | 14 |
| total | 416 | — | 267 | 594 | 222 |

* One group in Egypt, one in Glasgow.

TABLE 3. PROGENY OF DIFFERENT MARRIAGES INVOLVING $M$ AND $N$

| parents | number of families | $M$ | $MN$ | $N$ | total | mean children per family |
|---|---|---|---|---|---|---|
| $M \times M$ | 147 | 425 | 3! | 0 | 428 | 2·91 |
| $M \times N$ | 151 | 1! | 477 | 2! | 480 | 3·18 |
| $N \times N$ | 74 | 0 | 0 | 232 | 232 | 3·13 |
| $MN \times M$ | 397 | 597 | 662 | 4! | 1263 | 3·18 |
| $MN \times N$ | 292 | 2! | 428 | 483 | 913 | 3·13 |
| $MN \times MN$ | 396 | 267 | 594 | 222 | 1083 | 2·73 |
| total | 1457 | — | — | — | 4399 | 3·02 |

investigators tend to choose large families. Nevertheless, it suggests that such marriages are less fertile than the average. The shortage of total children and of homozygotes can both be explained if homozygotes have a higher deathrate (probably prenatal) than heterozygotes. The prenatal and infantile fitness of the homozygotes is about 82 % of that of the heterozygotes, so at least 18 % of them must die at an early stage. If there had been 594 homozygotes instead of 489, the mean family size would have been 3·00. The hypothesis of selective death implies that 105, or 2·3 % of a group of 4504, human zygotes were eliminated, probably before birth. If the $MN \times MN$ marriages had been as fertile as the rest, we should have

expected 155 more children from them, making a total of 4659, of whom 3·4 % were eliminated.

This is a substantial fraction of all conceptions, and it would seem that if a scientific study of the problem of human population is to be undertaken, it would be desirable to investigate a group of say 5000 married couples (including sterile couples) serologically, in order to discover whether certain types are less fertile than others, and whether certain human genotypes are eliminated prenatally. It would be essential, in such a study, to tabulate the results of reciprocal unions such as $MN♀ \times M♂$, and $M♀ \times MN♂$, separately. Unfortunately, many of the authors cited did not do so. It is of course possible that Wiener's hypothesis is correct. Nevertheless, the matter seems sufficiently important to warrant further study.

In the case of the *Rh* group of genes it is known that certain classes of offspring are killed off because they immunize their mother, and their blood corpuscles are destroyed by her antibodies. Such a mechanism will not explain the results found with *M* and *N*. Moreover, the elimination of homozygous offspring of two heterozygous parents would make the equilibrium between the two genes unstable, whereas in fact their frequencies in different peoples are much less variable than those of other genes responsible for serological differences.

Whatever may be the final answer to these questions, I hope I have shown that the exact investigation of the segregation of common genes is not a matter of merely academic interest.

I must pass on to the methods which are used in the investigation of the segregation of rare genes. When the rare gene is a dominant there are no statistical difficulties provided the gene manifests itself in all heterozygotes, and early in life. We cannot possibly expect to find Mendelian ratios for such a character as Huntington's chorea, whose mean age of appearance is about 35 years. We should expect to find good results in the case of hereditary skin diseases, which are easily and accurately diagnosed, and mostly manifested at an early age. Table 4 shows the children from unions of affected and normal persons in the sixteen diseases inherited as dominants of which Cockayne (1933) in his classical treatise was able to collect records of over 100 such children. A few of my numbers differ slightly from his totals through the exclusion of doubtful pedigrees. As a result of sampling error we should expect a normal distribution about zero with unit variance of the values of $(d-1)n^{-\frac{1}{2}}$, when $d$ is the number of affected minus that of normal, and $n$ is their sum. There are three aberrant values. The low incidence of neurofibromatosis may possibly be accounted for by its variable age of onset and sublethal character. Some individuals carrying the gene may have died prenatally, others may not yet have developed it when observed. Hypoplasia of the enamel is due to genes in at least two different chromosomes (Haldane 1937) and therefore presents complications. Tylosis, which is an abnormal thickening of the skin of the palms and soles, generally develops in the first year of life. It seems to present a definite exception to the usual rules, and demands further investigation. The similar anomalous cases which occur in the literature of dominant abnormalities of other organs are easier to explain by faulty diagnosis. There seems

no reason to doubt that the segregation of most human dominant abnormalities follows Mendel's laws.

The ratios in which a gene pair segregates cannot be obtained so simply when one allelomorph is fully recessive. This is due to the fact that the compilation of a pedigree introduced a certain bias. The bias may be of a very simple kind. Birch (1937) in Chicago and Andreassen (1943) in Copenhagen collected 146 pedigrees of haemophilia, a sex-linked recessive condition, transmitted to and from males

TABLE 4. PROGENY OF INDIVIDUALS AFFECTED WITH DOMINANT ABNORMALITIES OF THE SKIN, HAIR, NAILS, AND TEETH

| abnormality | affected | normal | $(d-1)\,n^{-\frac{1}{2}}$ |
|---|---|---|---|
| piebaldness | 133 | 118 | +0·88 |
| cutaneous xanthomatosis | 98 | 111 | −0·83 |
| telangiectasis | 320 | 302 | +0·43 |
| epidermolysis bullosa simplex | 193 | 163 | +1·54 |
| epidermolysis bullosa dystrophica | 147 | 181 | −1·82 |
| monilethrix | 92 | 89 | +0·15 |
| porokeratosis | 70 | 91 | −1·58 |
| tylosis plantaris et palmaris | 594 | 483 | +3·35 |
| ichthyosis vulgaris | 86 | 98 | −0·81 |
| alopecia congenita | 130 | 118 | +0·70 |
| onychogryphosis | 242 | 253 | −0·45 |
| hypoplasia of enamel | 84 | 50 | +2·81 |
| neurofibromatosis | 115 | 160 | −2·61 |
| naevus aplasticus | 53 | 61 | −0·66 |
| fistula auris | 63 | 60 | +0·18 |
| angioneurotic oedema | 182 | 206 | −1·17 |
| total | 2602 | 2544 | +0·79 |

TABLE 5. SONS OF HAEMOPHILICS' DAUGHTERS

| | families | normal sons | haemophilic sons |
|---|---|---|---|
| mothers of patients | 17 | 11 | 26 (−17) |
| other mothers | 26 | 25 | 23 +1? |
| total | 43 | 36 | 49 +1?(−17) =32+1? |

through females. Each pedigree began with a patient whose relatives were then traced. In order to verify that the condition is due to a single gene we must show, among other things, that the daughters of haemophilics bear equal numbers of normal and haemophilic sons. If we study the daughters of haemophilics in the pedigrees we find a considerable excess of haemophilic sons. However, a further analysis (table 5) shows that this excess is confined to the mothers of the patients from whom the compilation of the pedigree started.

The reason is simple. The mothers of patients were investigated because the patient was discovered to be haemophilic. Hence at least one of their sons must have been haemophilic. The other daughters of haemophilics may have borne no

haemophilic sons, indeed one was fortunate enough to bear three normal sons and no haemophilic. We can allow for this bias by subtracting one haemophilic from each family including a patient. The total then becomes 36 normal sons, with 32 haemophilics and one doubtful, a very good approximation to equality. A similar but more complicated analysis shows that about half the sisters of haemophilics transmit the disease. A neglect of this elementary point has led to the most remarkable conclusions as to the fertility of human stocks afflicted with hereditary disease. For if a character is passed on to half the children of an afflicted person, it will not be recognized as hereditary unless at least one child possesses it. We shall thus exclude all families of no children, half the families with only one child, a quarter of those with two, and so on, thus giving a wholly false impression of the fertility of such stocks. Where, on an average, the character only appears in one-quarter of the children, the exaggeration is still greater.

Unfortunately, the very simple type of correction which was applicable to the pedigrees of haemophilia cannot always be applied.

Consider a recessive character such as albinism or amaurotic idiocy which, by analogy with animals, is to be expected in one-quarter of the progeny of unions between two heterozygotes of normal appearance. We have in general no evidence that a pair of parents is heterozygous, except that they have produced at least one recessive. We cannot therefore study the progeny of a number of pairs of known heterozygotes, as we can in animal experiments. We can only study the progeny of those pairs which have produced at least one recessive.

Clearly the frequency of recessives in such sibships* is greater that the expected quarter. For it is 100 % in sibship of one, and over 50 % in sibships of two. The method for assessing the frequency $p$ which would be found in a very large sibship from data on small ones depends on how the data are collected.

Let $a_{rs}$ be the number of sibships of $s$ members, of which $r$ are abnormal.

Let $t_s = \sum_r a_{rs}$, i.e. the total number of sibships of $s$ members.

Let $N = \sum_{r,s} a_{rs}$, i.e. the total number of sibships.

Let $R = \sum_{r,s} (ra_{rs})$, i.e. the total number of abnormals.

Let $S = \sum_{r,s} (sa_{rs})$, i.e. the total number of sibs.

Let $q = 1-p$.

Now consider two ideal cases. In the first case a whole population is surveyed, and all sibships containing at least one abnormal are tabulated. This is possible in a small European country. Thus Sjögren (1931) probably tabulated over 90 % of the Swedish families in which a case of juvenile amaurotic idiocy had occurred in the twentieth century. In this case the estimate of $p$ is given by:

$$\frac{R}{p} = \sum_s \left[\frac{st_s}{1-q^s}\right],$$

* The word sibship means a set of siblings, that is to say brothers and/or sisters.

and its standard error is given by:

$$\sigma_p^{-2} = \frac{R}{p^2 q} - \frac{1}{q^2} \sum_s \left[ \frac{s q^s t_s}{(1-q^s)^2} \right].$$

Unfortunately, this cumbrous equation, due to Haldane (1938), can be shown to yield a result with a smaller standard error than a simpler one due to Weinberg. Perhaps a quicker but equally efficient method may be devised.

In the second ideal case all children leaving school in a certain year, or better all children born in a certain 6 months, are examined. The sibs of all abnormal children are tabulated. Clearly if a sibship contains three abnormals, it is three times as likely to be tabulated as if it contains only one, and so on, apart from an obvious correction for twins. In this case the estimate of $p$ is

$$p = \frac{R-N}{T-N}, \quad \text{and} \quad \sigma_p^2 = \frac{(T-R)(R-N)}{(T-N)^3},$$

an elegant result due to Weinberg (1927).

Applying these methods to Pearson, Nettleship & Usher's (1913) collection of 411 sibships from normal parents including at least one albino, 864 out of 2435, or 35·48 %, were albinos.

Applying the first correction

$$p_1 = 0{\cdot}3082 \pm 0{\cdot}0107,$$

applying the second

$$p_2 = 0{\cdot}2238 \pm 0{\cdot}0092.$$

The Mendelian value of $\frac{1}{4}$ lies between these two estimates, and there is reason to think that if an exact correction were possible, the Mendelian ratio would be found. A simple example will show the need for other corrections. According to Andreassen (1943), in the year 1941 there were 1,820,000 males in Denmark, of whom 81 were haemophilic. Almost all their families were investigated, and valuable results were obtained. However, haemophilics have a much shorter average life than ordinary males (about 18 years in Denmark). So a large fraction, probably the majority, of the haemophilics born between 1910 and 1930 were dead by 1942. A family into which three haemophilics were born in that time was more likely to contain a living haemophilic in 1942 than one into which only one was born. But it was not exactly three times as likely. If half the haemophilics had died, and there was no correlation in the age of death between haemophilic sibs, it was 1·75 times as likely. Hence the true value of $p$ would lie somewhere between the two extreme estimates. Special methods could and should be developed for such cases.

They are important because they offer a possibility of investigating selective prenatal death, of verifying the general applicability of genetical principles to man, and of developing, in comparatively simple cases, the quite peculiar statistical methods which are required when the genotypes of parents must be deduced from the phenotype of the children, with an accuracy which increases with the size of the

sibship. These methods may be said to play the same part in human genetics that standard culture methods play in animal and plant genetics. Without them qualitative conclusions may be drawn, but quantitative work is impossible in the case of many genes.

We now pass to the methods for the location of genes on the human chromosomes. A serious beginning has been made with the mapping of two sections. One is the segment of the *X* chromosome which is responsible for sex determination, in which the ordinary sex-linked genes are located. The other is the segment common to the *X* and *Y* chromosomes. With regard to the remaining twenty-three there is fairly good, but never conclusive, evidence, for the compresence in one chromosome of two genes. For each type of location appropriate statistical methods have been developed.

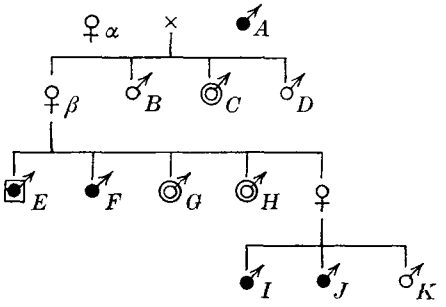



FIGURE 1. ♂, normal male; ◎♂, deuteranopic male, not haemophilic; ●♂, haemophilic male, vision untested; ◧♂, haemophilic trichromatic male.

Let us begin with the differential segment of the *X*. A large number of sex-linked recessive genes have been located there. The mode of inheritance of the characters determined by them is highly characteristic. The abnormality determined by any one of them is far commoner in men than women. It is not transmitted from a father to his son. But it occurs in about half the sons of heterozygous women, who include the daughters of affected men. The *X* chromosomes of *Drosophila* species have been mapped by studying the segregation of genes in the progeny of mothers who are heterozygous for two or more pairs of sex-linked allelomorphs.

Cytological studies have shown that the maps so obtained depict real material structures, as X-ray diffraction and reflexion have shown that the structural formulae of the organic chemist depict real objects.

At one locus in the human *X* chromosome abnormalities are very common. About 8 % of all men are colour-blind or anomalous colour-matchers. Hence if we wish to estimate the percentage of recombination between the loci of haemophilia and colour-blindness we must search for colour-blindness among haemophilics and their brothers. Seventeen pedigrees are known in which both abnormalities are found. Of the total information available from them, about a third was collected by Bell & Haldane (1936), another third by the Dutch physician Hoogvliet (1942) and the remainder by five others.

The method employed can be illustrated by two simple examples. Figure 1 shows a pedigree in which $A$ was a haemophilic, while $\alpha$ carried the gene for colour-blindness in one of her two $X$ chromosomes. She also gave it to $\beta$, who had two colour-blind sons. So $\beta$ had the genes for these two defects in different chromosomes (in the *trans* position, if we like a metaphor from organic chemistry, or in repulsion in Bateson & Punnett's terminology). Of $\beta$'s three surviving sons, one was a haemophilic trichromat; two were colour-blind, but had normal blood. Now if $x$ is the frequency of recombination between the two loci concerned, the probability of $\beta$ producing just these three sons is $(1-x)^3$, given that she had one haemophilic and two non-haemophilics. If one of the non-haemophilics had had normal vision it would have been $x(1-x)^2$, and so on. On this pedigree taken alone the best estimate of $x$ is clearly zero.

Figure 2, which is part of Bell & Haldane's (1937) pedigree A, of 98 members, raises a rather more subtle problem.

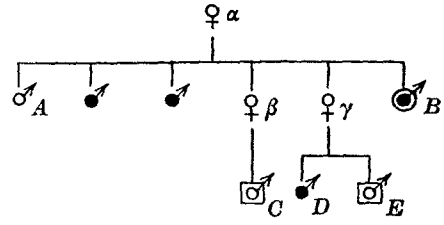



FIGURE 2. ♂, normal male; ♂, haemophilic male, vision untested; ♂, trichromatic male, not haemophilic; ◉♂, haemophilic deuteranopic male.

This was solved approximately by Bell & Haldane (1936), but my colleague Dr Smith has produced a more accurate method (Haldane & Smith 1947). We ask what is the probability that seven persons related in the manner shown should have just these phenotypes. If $p_c$ and $p_h$ are the frequencies of the genes **c** and **h** for colour-blindness and haemophilia, we ask what is the probability that $\alpha$ should have been heterozygous for both of them. Clearly it is $4p_c(1-p_c)\,p_h(1-p_h)$. If $\alpha$ was doubly heterozygous and her husband normal, there are eight possible sets of events in the formation of ova by $\alpha$, $\beta$, and $\gamma$ which could have given the observed results. They are shown in figure 3 and the probability of each is given, putting $y = 1-x$. It is the product of five factors representing the probabilities of the formation of five different eggs. These are shown in each case. Since $\gamma$ had a haemophilic son who died in infancy we know that she received the gene **h** from her mother. This excludes sixteen other possibilities whose probability is zero. It does not exhaust the possibilities. For though we can be sure that neither $\alpha$ nor her husband was haemophilic, we cannot be sure that one or both of them was not colour-blind. So the total probability has two more terms, each containing $p_c^2$. It is

$$\tfrac{1}{8}p_c p_h[3(1-p_c)\,(1-2x+4x^2-4x^3+x^4)+p_c(x^2-x^4)].$$

Since $p_c$ is of the order of 0·01 it can in practice be neglected in this case, but not in all cases.

The corresponding expression for the pedigree of figure 1 is

$$2^{-11}p_c p_h(1-x)^3.$$

These probabilities are of course small, partly because the genes in question are rare, partly because the particular pattern of segregation found is one of a vast number which are possible, like the 635,013,559,600 equiprobable bridge hands. However, each is maximal for some value of $x$ between 0 and 1 inclusive; in the cases considered, for $x = 0$. In other pedigrees a cross-over has occurred, i.e. the genes **c** and **h** have entered a woman in one gamete and left her in different ones or vice versa. In these the polynomial is maximal for some other value of $x$. The product of the seventeen polynomials derived from the different pedigrees is maximal when $x = 0{\cdot}098$, and we may estimate the frequency of recombination between the two loci as $9{\cdot}8 \pm 4{\cdot}2$ %.

White (1940) found $64{\cdot}8 \pm 12{\cdot}7$ % of recombination between the loci of colour-blindness and of myopia with nystagmus, so the genetical map of the human $X$ chromosome is likely to be as long as that of *Drosophila melanogaster*, though probably shorter than that of *Gallus domesticus*.

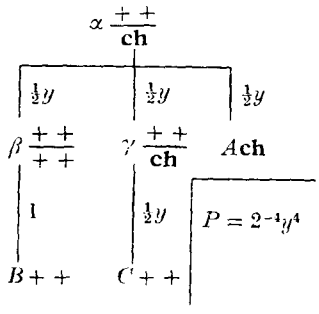
Ten years ago I accounted for the peculiar inheritance of certain characters on the hypothesis that the genes concerned in their determination were located in that segment of the sex chromosomes which is common to the $X$ and $Y$, and may be exchanged between them (Haldane 1936). Such genes are said to be partially sex-linked.

At that time I put forward the hypothesis with considerable misgiving, but it has been generally accepted, and I shall therefore state it with comparative confidence. Penrose (1946*b*) has recently given an alternative explanation for some cases which appear to conform to it, but he does not think that this will explain all the cases. First, consider a dominant gene in this segment. If a woman has it, necessarily in an $X$ chromosome, she will transmit it, on an average, to half her children, regardless of sex. If a man has it in his $Y$, he has inherited it from his father, and will probably transmit it to most of his sons but a few of his daughters. If he has it in his $X$ he has inherited it from his mother, and will transmit it to most of his daughters but a few of his sons. Thus the sex of the affected children of affected males will generally be the same as that of the affected paternal grandparent. The fewer the exceptions, the nearer the locus to the differential segments containing the sex determiners, and (in the $X$) the loci of such genes as haemophilia.

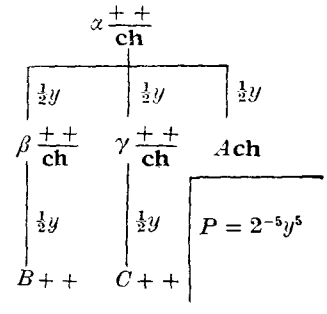
Only one such partially sex-linked dominant is known, namely retinitis pigmentosa in some pedigrees. Penrose (1946*b*) has, I think disproved Pipkin & Pipkin's (1945) claim to have found a second such, zygodactyly, or webbing of the toes. It is possible that my own claim in the case of retinitis pigmentosa will equally be disproved.

The location of partially sex-linked recessives is not so simple, but I think some of its results are more certain. Where the parents are first cousins, the sex of the affected offspring is usually the same as that of the paternal grandparent through whom the parents are related. For if this grandparent was a male the father carries

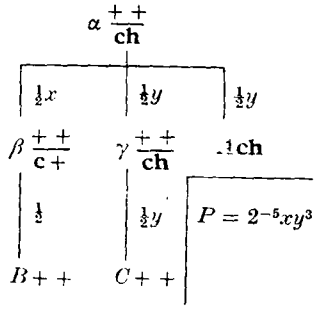
the gene in his $Y$ chromosome, and will transmit it predominantly to his sons, as in figure 4. If she was a female they will mostly be daughters. If the parents are not known to be related, we can only say that in any particular sibship the affected members will be predominantly of one sex, though in all sibships together no such predominance is to be expected, except in the case of very rare conditions where the parents are mostly consanguineous.



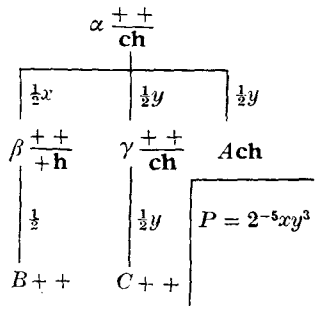

1.



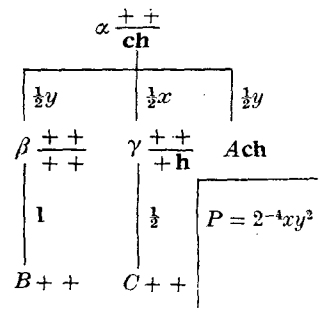

2.



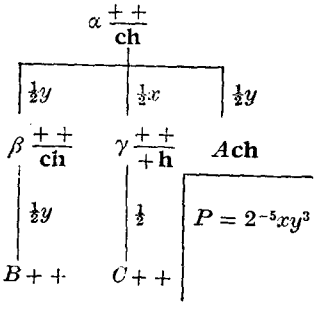

3.



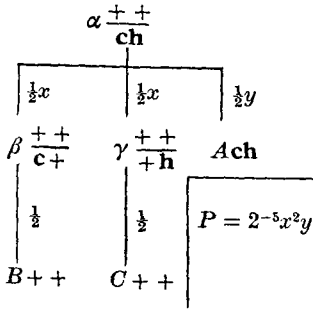

4.



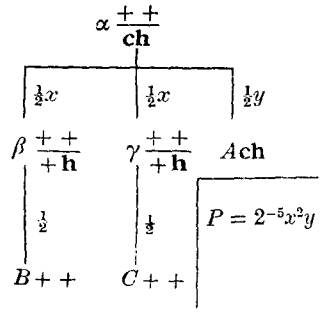

5.



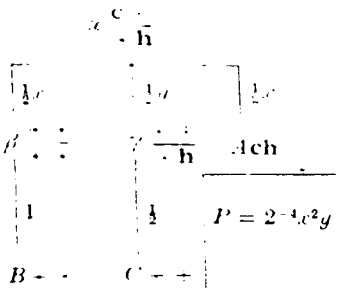

6.



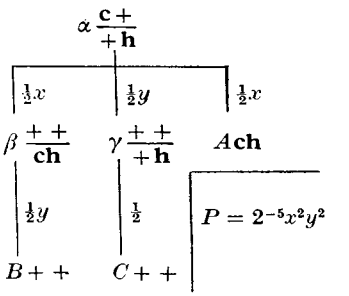

7.



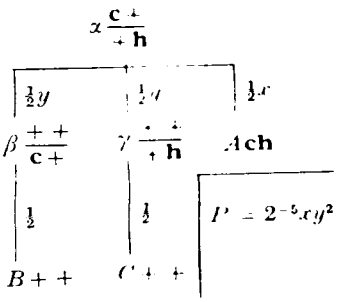

8.



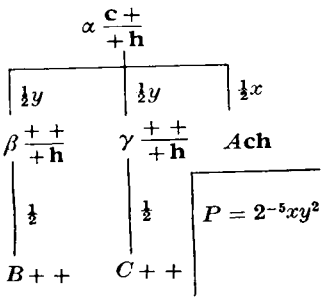

9.



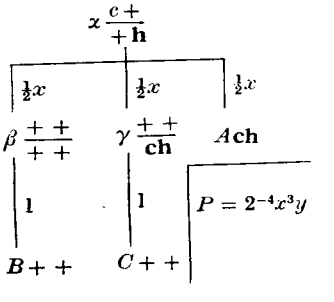

10.



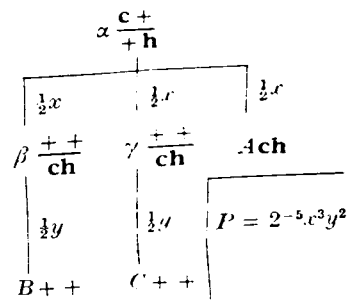

11.



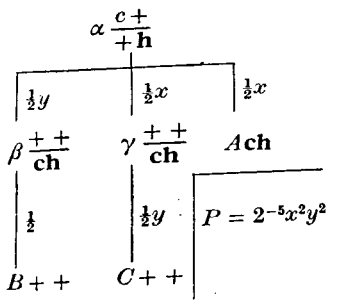

12.



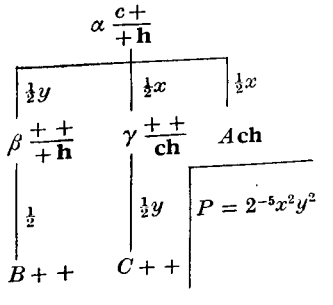

13.




14.




15.




16.

FIGURE 3. Sixteen possible explanations of the pedigree of figure 2. In each case the probability of the five different steps is given, together with their product. The overall probability is the sum of these products, multiplied by the probability that **c** and **h** should be compresent in $\alpha$.

Fisher (1936) developed a most elegant method for the detection of partial sex-linkage in such cases. Suppose a sibship consists of $N$ normal females, $n$ normal males, $A$ affected females, $a$ affected males, and that

$$u = (N - n - 3A + 3a)^2 - (N + n + 9A + 9a),$$

then in the absence of partial sex-linkage the expected value of $u$ is zero, in its presence it is $\frac{1}{9}k(1-2x)^2$, where $k = (N+n+9A+9a)^2 - (N+n+81A+81a)$, and $x$ is the frequency of recombination. Thus if the sum of a large number of $u$ values is significantly positive partial sex-linkage can be inferred, and its intensity estimated.

I have since shown (Haldane 1948) that if we calculate a polynomial for each sibship on the lines developed for the investigation of linkage between sex-linked genes, its logarithm can be expressed as a series in ascending powers of $(1-2x)^2$. The coefficient of $(1-2x)^2$ is Fisher's $u$. Thus the sum of Fisher's $u$ scores gives a perfect test for the presence of linkage, though not a quite unbiased estimate of its intensity. There is, however, a further complication. In the absence of linkage the sampling distribution of $\Sigma u$ is not normal, but positively skew. So a high positive value gives a rather exaggerated estimate of the significance of the evidence for linkage. When allowance is made for this (Haldane 1946) most, but not quite all, of the data formerly regarded as significant remain so.

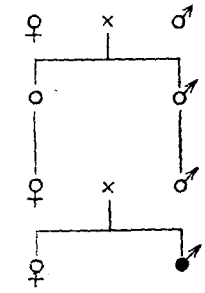


FIGURE 4. ♂, male homozygous for a partially sex-linked recessive. The sex of his maternal grandparent is irrelevant.

The Croonian Lecture was originally intended to be on 'local motion', and I shall therefore illustrate recessive partial sex-linkage by discussing spastic paraplegia, a disease in which the tonus of the limb muscles gradually increases until walking becomes impossible. Bell (1939) collected forty-four pedigrees in which one or more children of normal parents were affected with this disease.

Applying Fisher's method we have $\Sigma u = 1256 \pm 231$, and the estimated frequency of recombination is 17·5 %. The significance is not as high as it appears, since the sampling distribution in the absence of linkage is very skew positively. But it is not in doubt. It is wholly possible that while most of the families are segregating for a partially sex-linked gene, others are segregating for an autosomal one. To determine whether this is so, about five times the present number of families would be required, and it would be necessary to devise new statistical methods. There is also a suggestion, both from the results of the direct method applied to the progeny of cousin marriages, and the indirect method based on Fisher's '$u$' scores, that a few cases diagnosed as spastic ataxia, and perhaps even as Friedreich's ataxia, may be due to partially sex-linked genes. The large majority are not (Haldane 1941*a*).

On the basis of such statistical work I located seven genes on this segment. The standard errors of their distances are so large that I do not think that a map is worth publishing. However, the loci of dominant and recessive genes for retinitis pigmentosa, probably allelomorphs, lie about 30 units from the sex-determining or

differential segment, while the loci of five other genes, namely those for achromatopsia, epidermolysis bullosa, xeroderma pigmentosum, spastic paraplegia, and Oguchi's disease, seem to lie between 21 and 44 units from it. In addition a lethal gene for convulsive seizures with mental deterioration is probably located in this segment (Snyder & Palmer 1943) while a gene concerned in some cases of hare-lip and cleft-palate may be so (Philip & Mather 1940).

The next step in this investigation will perhaps be the discovery of families in which colour-blindness is segregating along with a partially sex-linked gene such as those for spastic paraplegia or xeroderma pigmentosum. In the latter case in particular we should be able to detect linkage in doubly heterozygous females between genes in the two segments of the $X$, and thus to produce a unified map. Even more valuable would be the discovery of a gene as common as those for colour-blindness and anomalous matching, in either section of the $X$ chromosome. The most hopeful field for such a discovery is among the antigens.

I have deliberately restricted my own work on human linkage to the sex chromosomes, because in every satisfactory pedigree sex as well as abnormality is recorded. Other workers more industrious than myself have looked for linkage between autosomal genes. Since man has twenty-three pairs of autosomes, the probability that a particular pair of loci will lie in the same chromosome is of the order of $\frac{1}{23}$, though rather more, because the chromosomes are of unequal size. On the other hand a recombination value of over 25 % is unlikely to be detected until data accumulate in very considerable quantity.

Two genes are very possibly linked with those for the blood-group antigens. These are the recessive gene for phenylketonuria (Penrose 1946*a*) and the partially dominant gene for allergy (Finney 1940; see also Finney 1941, 1942). A phenylketonuric apparently lacks some enzyme concerned in the metabolism of phenylalanine, and consequently excretes up to 1 g./day of phenylpyruvic acid. As a further consequence there is a shortage of material for melanin formation, and the hair is of a lighter colour than that of other members of the family. Far more important, thought is impossible. Phenylketonurics are usually idiots or imbeciles, at best feeble-minded.

Penrose (1946*a*) has developed a statistical method for dealing with such cases, when the children in a family can be examined, but the parents cannot. Consider a series of sibships in which some members are normal, others are phenylketonurics, some have the $B$ antigen, i.e. belong to group $B$ or $AB$, others belong to other groups. Consider what is to be expected if the loci concerned are on the same chromosome. In some sibships the recessive genes for phenylketonuria and absence of $B$ antigen have been in the same chromosome in both parents. In these, the sibs who are phenylketonurics will probably not possess antigen $B$, or more accurately, will be less likely than normals to possess it. If the genes are in different chromosomes in one or both parents, those who are phenylketonurics will probably possess antigen $B$. Over a group of sibships there will be no association between the two characters, but in any particular sibship they will be positively or negatively associated.

Now if we observe any pair of sibs, they must fall into one of the nine categories of table 6.

TABLE 6. THE NINE POSSIBLE TYPES OF SIB-PAIRS IN A SIBSHIP SEGREGATING FOR THE $B$ ANTIGEN AND PHENYLKETONURIA. EXPLANATION IN TEXT

| | $BB$ | $Bb$ | $bb$ | total |
|---|---|---|---|---|
| $NN$ | $a_{11}+$ | $a_{12}-$ | $a_{13}+$ | $c_1$ |
| $NP$ | $a_{21}-$ | $a_{22}+$ | $a_{23}-$ | $c_2$ |
| $PP$ | $a_{31}+$ | $a_{32}-$ | $a_{33}+$ | $c_3$ |
| total | $b_1$ | $b_2$ | $b_3$ | 1 |

Here $N$ and $P$ denote normality and phenylketonuria $B$ and $b$ the presence and absence of the $B$ antigen. Thus a pair of sibs of whom one is normal and one phenylketonuric, but both belong to group $O$, falls into the category $NP.bb$. If the two gene pairs are located in different chromosomes the expected number $E_{23}$ of sib-pairs in this category will be the product of the numbers of $NP$ pairs and $bb$ pairs, divided by the total, that is to say $b_3c_2/s$. If they are in the same chromosome it will be less. In general, linkage will increase the numbers in the categories labelled + in table 6, and diminish it in the remainder. Penrose (1946$a$) found the figures of table 7.

TABLE 7. NUMBERS OF SIB-PAIR TYPES FOUND BY PENROSE, WITH EXPECTATIONS IN THE ABSENCE OF LINKAGE

| | $BB$ | | $Bb$ | | $bb$ | | total |
|---|---|---|---|---|---|---|---|
| $NN$ | 17 | (9·125) | 11 | (16·947) | 89 | (90·928) | 117 |
| $NP$ | 10 | (15·365) | 34 | (28·535) | 153 | (153·100) | 197 |
| $PP$ | 1 | (3·510) | 7 | (6·518) | 37 | (34·972) | 45 |
| total | | 28 | | 52 | | 279 | 359 |

In this table the bracketed numbers are the expectations in the absence of linkage. Thus $9{\cdot}125 = \frac{117 \times 28}{359}$. In the absence of linkage, Penrose finds that

$$\xi = \left[\frac{a_{11}}{E_{11}} + \frac{a_{13}}{E_{13}} + \frac{a_{31}}{E_{31}} + \frac{a_{33}}{E_{33}} + \frac{4a_{22}}{E_{22}} - \frac{2a_{12}}{E_{12}} - \frac{2a_{21}}{E_{21}} - \frac{2a_{23}}{E_{23}} - \frac{2a_{32}}{E_{32}}\right] \div \left[s\left(\frac{1}{b_1} + \frac{4}{b_2} + \frac{1}{b_3}\right)\left(\frac{1}{c_1} + \frac{4}{c_2} + \frac{1}{c_3}\right)\right]^{\frac{1}{2}}$$

is normally distributed with mean zero and unit variance. A positive value indicates linkage. In this case the value is + 1·51 which is not in itself significant. But Penrose informed me that when the $O$ antigen and the two forms of the $A$ antigen are also taken into consideration, the value rises to 1·78. Since this is in the direction expected on theoretical grounds, the probability of obtaining so large a value by chance is 0·046 or one in twenty-two. Such a value is usually taken as on the borderline of significance. A few more families may well establish this linkage conclusively.

Finney (1940) used modifications of Fisher's $u$ statistics, and concluded from Zieve, Wiener & Fries' (1936) data that the gene for allergy was linked with those for the blood groups. However, it must be emphasized that the genetics of allergy are not so simple as those of phenylketonuria. The probability of obtaining so large a deviation in the expected direction is 0·04, after correcting for skewness. His result and Penrose's are weaker evidence for linkage than appears at first sight because man has twenty-three pairs of autosomes, so the *a priori* probability of linkage is only about 0·04. This means that considerably stronger evidence for linkage is required than in *Drosophila melanogaster* where the *a priori* probability is about 0·5.

Burks (1939) published preliminary results which very strongly suggest linkage between genes for hair colour and defective teeth. Unfortunately only a statistical summary was given, and as I have pointed out elsewhere (Haldane 1941*b*) other explanations besides linkage are possible, though perhaps not very likely. She also obtained evidence of linkage between genes for myopia and eye colour. Penrose (1935) and Rife (1941) obtained suggestions of linkage between blood groups and hair colour, and interdigital pattern and left-handedness respectively.

Finally, Kloepfer (1946) has made a most comprehensive study involving nineteen characters, and obtained evidence suggesting a number of linkages. The most impressive are those between eye colour and flare (or projection) of the ears, and between ability to taste phenyl-thio-urea and ear size. Unfortunately nothing is yet known as to the genetics of ear size and structure.

Finally, there is massive negative evidence showing that various genes, notably those for blood groups, blood types ($M$ and $N$), and ability to taste phenyl-thio-urea, are not linked with one another nor with sex. Such work is inevitably tedious, but it is striking how long a time elapsed before linkage was discovered in poultry or peas, and how rapidly knowledge accumulated once the first linkages were discovered.

Up till now we have considered the behaviour of genes in so far as they reproduce their like (or perhaps better, are copied exactly) at each nuclear division. When this does not happen, a new type of gene arises which generally, but by no means always, reduces the fitness of the organism either (*a*) at once, if it is a dominant; (*b*) when it appears in a male, if it is a sex-linked recessive; or (*c*) when two genes of the new type are contributed by different parents to the same zygote, if it is an autosomal recessive. This process of change is called mutation. Clearly it may be due to a failure of the copying process, or to a change induced in the model between copyings by physical means such as X-rays or chemical means such as $\beta\beta'$-dichlorethyl disulphide. Mutation occurs spontaneously, that is to say under normal conditions, in all organisms so far studied; but as it is a rare process, it can only be measured when vast numbers are available. The rate was first measured in *Drosophila melanogaster*, then in *Zea Mays*, and finally by Gunther & Penrose (1935) and Haldane (1935) in man. More exact estimates, fully confirming these figures, have been made in Denmark in the last 6 years.

The rate can be measured directly, as was done by Mørch (1941) for achondroplasia, or chondrodystrophy. This is the condition found in the familiar short-legged type

of dwarf. In 1938 there were eighty-six such dwarfs in Denmark among 3,800,000 people. They have a very low fertility, but when they breed, about half their offspring are similar dwarfs. The large majority of dwarfs, however, are the offspring of normal parents with no dwarfs in their families. It is clear that the gene for dwarfism arises sporadically by mutation. Out of 132,761 children born of normal parents in hospitals in Copenhagen and Lund over a period of 21 years, eleven were dwarfs of this type. This gives a mutation rate of $4{\cdot}1 \times 10^{-5} \pm 1{\cdot}2 \times 10^{-5}$ per normal gene per generation, or about $1{\cdot}2 \times 10^{-6}$ per year, since the mean age of normal parents is 35 years. The probability that the true value should be as low as $10^{-5}$ per generation is 0·0011, the probability that it should be as high as $10^{-4}$ is 0·0001, so the order of magnitude is certain.

Mørch also estimated the mutation frequency indirectly. Most such dwarfs die at or within 2 days of birth, and a number more in the first year of life, but after this their expectation of life is normal. If 80 % die in the first year, which is his estimate, there would be 415 such dwarfs in Denmark but for this mortality, or a frequency of $1{\cdot}09 \times 10^{-4}$. Now 108 dwarfs had 27 children, and their 457 normal sibs had 582, thus their fitness from a Darwinian point of view is $\frac{27}{108} \times \frac{457}{582}$, or $f = 0{\cdot}204$. That is to say in each generation natural selection effectively eliminates 80 % of the dominant genes, and but for mutation there would be no dwarfs left on earth within seven generations, or say two centuries, if Danish figures are typical. However, the two processes are in approximate equilibrium. So if $x$ is the frequency at birth, the mutation rate $\mu = \frac{1}{2}(1-f)x = 4{\cdot}3 \times 10^{-5}$. The factor $\frac{1}{2}$ arises because we are dealing with a population of chromosomes equal to twice the population of human beings. The two estimates agree very well, but the second is much less accurate, since it depends on the figure for the infant mortality. Mørch, using a rather different argument, gets $4{\cdot}8 \times 10^{-5}$.

Professor Penrose has pointed out to me that Mørch's data are open to three criticisms. In some pedigrees, though not in any of those which he collected, there is evidence that the gene for achondroplasia can fail to manifest itself, as in Richsbieth's (1912) pedigree 608. The cases where two normal parents had more than one achondroplasic child may be due to this cause or to a mutation at an early stage in the development of a gonad. A correction for this possibility makes very little difference to the estimate of the mutation rate, since the gene is detected on its first appearance, even if this be occasionally delayed. Secondly, Mørch did not personally examine all the infants, and it is possible that some may have been wrongly diagnosed. This is plausible, since he himself failed to confirm the diagnosis of achondroplasia made by another worker in a Norwegian family. Finally, the frequency increases with parental age to an extent inexplicable if all the dwarfs born of normal parents are mutants, but explicable if some of them are due to bad prenatal conditions, as with mongoloid imbeciles. It may therefore well be that Mørch's figure is too high. But the true value is almost certainly above $10^{-5}$

For most diseases only the indirect method is available. Andreassen (1943) has applied it to haemophilia in Denmark. However, I believe (Haldane 1947*b*)

that his calculations give rather too low a result. Haemophilia is due to a sex-linked recessive gene. Hence only about one-third of the genes for haemophilia in a population are exposed to natural selection at any moment. More accurately, if $\mu$ and $\nu$ are the mutation rate in the female and male sexes respectively,

$$2\mu + \nu = (1-f)x,$$

where $x$ is the frequency in males at birth. Now there were just eighty-one haemophilics alive in Denmark in 1941, and their mean life is 18 years compared with 55 for Danes in general. So $x = 1{\cdot}33 \times 10^{-4}$. The fitness $f$, that is to say the mean number of progeny, compared with that of the population in general, appears to be 0·28. I have criticized Andreassen's much higher figure, $f = 0{\cdot}59$. It follows that $2\mu + \nu = 9{\cdot}6 \times 10^{-4}$, or a mean mutation rate of $3{\cdot}2 \times 10^{-4}$. Now if the mutation rates were equal in the two sexes, i.e. $\mu = \nu$, nearly a third of all haemophilics would be single cases due to mutation in homozygous mothers. Andreassen has shown that the gene for haemophilia is not completely recessive. Heterozygous women sometimes bleed abnormally, but always have an abnormally long coagulation time, by which they can be detected. Using this technique he has not yet found a case where the mother of a haemophilic was homozygous. Doubtless such a case will be found. But it can be concluded that $\nu$ is much larger than $\mu$, very likely ten times as large. If this is correct we should have, very roughly, $\nu = 8 \times 10^{-5}$, and $\mu = 8 \times 10^{-6}$.

Similar estimates are available for three other conditions. Gunther & Penrose (1935) found $4\text{–}8 \times 10^{-6}$ for epiloia, Philip & Sorsby (unpublished) found $1{\cdot}4 \times 10^{-5}$ for retinoblastoma, and Møllenbach (according to Kemp 1944) finds $5\text{–}10 \times 10^{-6}$ for aniridia (Kemp's figure of double this value appears to be the mutation rate per zygote, not per gene). The median rate is about $10^{-5}$. Unfortunately, this method cannot be applied to autosomal recessive conditions.

Pätau & Nachtsheim (1946) have estimated the mutation rate of the autosomal dominant gene *Pg* which is responsible for the Pelger anomaly, a failure of segmentation of the nuclei of polymorphonuclear leucocytes. *Pg*/+ individuals are thought to be less resistant to disease than +/+, whilst by analogy with rabbits, it is suggested that *Pg*/*Pg* is a lethal genotype. However, too little is known as to the viability of heterozygotes to allow an indirect estimate of the mutation frequency. The authors estimated the frequency of the condition as 0·001, and found that out of twelve persons showing the Pelger anomaly, and both of whose parents could be examined, one parent was affected in ten cases, neither in two cases. This gives a mutation rate of $\frac{1}{2} \times 10^{-3} \times \frac{2}{12}$, or $8 \times 10^{-5}$. Even if the frequency were accurately known, and if 120 cases had been examined instead of twelve, this estimate would be somewhat high, simply because the cases with both parents living are a selected group, and include a higher fraction of cases with normal parents than of those with one affected. However, the order of magnitude agrees well with the figures given above, and is unlikely to be incorrect by a power of ten.

It is certain that these figures are not representative. Consider a well-known and unmistakable dominant such as lobster claw. Five families with this gene are known

in England. They are quite fertile, but presumably their fitness, or net fertility, is a little below the average, or the condition would be commoner. Lewis & Embleton's (1908) pedigree goes back to a son of allegedly normal parents born in 1793. In the case of such a conspicuous abnormality mutation is a far likelier explanation than adultery. But it is extremely doubtful whether the mutation occurs in Britain once in 10 years. Five isolated cases were described in Britain between 1895 and 1918. That is to say its mutation frequency is of the order of $10^{-7}$ per generation. This is probably a much more representative figure than those of $10^{-5}$ or over. Unfortunately, the indirect method becomes quite unreliable when, as in this case, the fitness is near unity. Finally, we have such cases as that of the 'porcupine men' of the Lambert family (literature, see Cockayne 1933), a most striking dominant mutation, perhaps a translocation, carried by the *Y* chromosome. This has only been recorded once, and would have stood a good chance of being recorded in any civilized country, in the last 2000 years. It was twice described in the *Philosophical Transactions* of this Society. The mutation rate is probably below $10^{-10}$.

A man or woman consists of about $2^{48}$ cells, that is to say a representative cell is separated from the fertilized ovum by about fifty mitoses. The primordial ova are all formed at birth, and do not undergo further mitoses. A man may produce $2^{40}$ spermatozoa in a lifetime, so the mean number of mitoses is somewhat greater in the male than the female germ-line, but probably not over 100 in the former.

Thus a mutation rate even of the order of $10^{-4}$ means that the gene-copying process, at worst, goes wrong about once in a million times, whether as the result of a failure of copying, or of a change in between two copying processes. A similar degree of accuracy in crystal growth would give a crystal with under ten flaws per millimetre, and $10^{10}$ successive flawless layers would give a perfect crystal several metres in length. The living substance of our bodies is clearly far more accurately copied than the successive layers of a crystal.

In *Drosophila* the natural mutation rate is of the order of $10^{-6}$ to $10^{-5}$ per generation for the more mutable loci, such as that whose mutation produces a white eye, and considerably lower for the stabler genes. Natural mutation is increased about threefold by a rise of 10° C, and is therefore largely due to a chemical reaction. As a generation in *Drosophila melanogaster* takes about 10 days and a fly contains about $2^{23}$ cells, while the mutation rate of the more labile genes is about one-fifth of that of man per generation, it follows that human mutation rates are about twice those of *Drosophila* per nuclear division, and about one two-hundredth of those of *Drosophila* per day, though the body temperature is about 13° higher. It has been calculated that natural radiations and particles of high energy will account for only 0·001 of the mutations in *Drosophila*. It is clear that if so they may account for about a fifth of those in man, and in view of the uncertainty of our knowledge as to the efficiency of particles from $K^{40}$ and cosmic rays in producing mutations, and the different radiosensitivity of different genes, it is quite possible that radiation may account for most human mutation. Mørch found that the rate of mutation to achondroplasia increased with age, but it was not clear whether maternal or paternal age was most important. If this finding is confirmed it suggests a cumulative effect,

either of radiation, or of successive nuclear divisions, during a lifetime. The apparently higher rate in males suggests that the number of nuclear divisions may be an important factor in human mutation. To sum up, there are three possible known causes of mutation, a chemical reaction with a temperature coefficient, radiation, and imperfections of copying, which might have a positive or negative temperature coefficient. The first predominates in *Drosophila*, the second or third probably does so in man. There must be about a thousand achondroplasic dwarfs in Britain. If the ages of their parents at their births were determined, it would be possible to decide between these alternatives, since the egg of a woman of 45 has undergone no more nuclear divisions than that of a girl of 15. It is worth remarking that it is quite practicable to obtain data of this kind on populations of 40 million human beings, and wholly impracticable to do so on 40 million of any other mammal.

The mutation rate is probably more or less adaptive. Too high a mutation rate would flood a species with undesirable mutations, too low a one would probably slow down evolution. Man and *Drosophila melanogaster* have about the same rate per generation, and this could not be increased ten times without a very great loss of fitness (Haldane 1937). Other species such as five species of *Sciara* (Metz 1938) have far lower rates per generation though not necessarily less than the human rate per day. But it is doubtful whether the human rate could be lowered much further, since a substantial fraction of it is due to natural radiation. In fact a very great prolongation of human life, or at any rate of the reproductive period, might be incompatible with the survival of the human species.

I hope that, in this brief survey, I have shown that human genetics has reached the stage when it can claim to be a branch of biology with its own peculiar problems and methods. I have only dealt with a few of them. This lecture could equally well have been devoted to the human antigenic structure, to human prenatal physiology, or to variation in human sensory and intellectual capacity, all of which a human geneticist must study. If I have confined myself to the more quantitative aspects, my excuse must be that in dealing with a branch of science where erroneous views may have important political consequences, in such a lecture as this it is desirable to concentrate on those problems where political or social bias is least likely to be effective, and where we may hope to raise a solid theoretical structure by methods like those which have been fruitful in the other branches of science.

## References


Andreassen, M. 1943 Haemofili i Danmark. *Opera ex domo biologiae hereditariae humanae Universitatis Hafniensis*, **6** (Copenhagen).

Avery, O. T., Macleod, C. M. & MacCarty, M. 1944 *J. Exp. Med.* **79**, 137–157.

Bateson, W. 1906 *Brain*, pt. 2, pp. 157–179.

Bell, J. & Haldane, J. B. S. 1936 *Proc. Roy. Soc.* B, **123**, 119–150.

Bell, J. 1939 *Treas. Hum. Inher.* **4**, pt. 3.

Birch, C. L. 1937 *Illinois Med. Dent. Monogr.* **1**, 4.

Burks, B. 1939 *Proc. Nat. Acad. Sci., Wash.*, **24**, 512–514.

Blaurock, G. 1932 *Münch. Med. Wschr.* **74**, 1552–1556.

Clausen, J. 1932 *Hospitalstidende*, **75**, 196–206.

Cockayne, F. A. 1933 *Inherited abnormalities of the skin and its appendages.* Oxford: University Press.
Crome, W. 1933 *Dtsch. Z. ges. gerichtl. Med.* **21**, 435–450.
Dahr, P. & Bussmann, R. 1938 *Dtsch. Med. Wschr.* **64**, 818–821.
Dahr, P. 1940 *Z. ImmunForsch.* **97**, 168–188.
Ephrussi, B. 1942 *Cold Spring Harbor Symp.* **10**, 40–48.
Finney, D. J. 1940 *Ann. Eugen.* **10**, 171–214.
Finney, D. J. 1941 *Ann. Eugen.* **11**, 10–30, 115–135.
Finney, D. J. 1942 *Ann. Eugen.* **11**, 224–244.
Fisher, R. A. 1936 *Ann. Eugen.* **7**, 87–104.
Gunther, E. R. & Penrose, L. S. 1935 *J. Genet.* **31**, 413–430.
Haldane, J. B. S. 1935 *J. Genet.* **31**, 317–326.
Haldane, J. B. S. 1936 *Ann. Eugen.* **7**, 28–57.
Haldane, J. B. S. 1937*a* *J. Hered.* **28**, 58–60.
Haldane, J. B. S. 1937*b* *Amer. Nat.* **71**, 337–348.
Haldane, J. B. S. 1938 *Ann. Eugen.* **8**, 255–262.
Haldane, J. B. S. 1941*a* *J. Genet.* **41**, 141–144.
Haldane, J. B. S. 1941*b* *New paths in genetics.* London: Allen and Unwin.
Haldane, J. B. S. 1946 *Ann. Eugen.* **13**, 122–134.
Haldane, J. B. S. 1947 *Ann. Eugen.* **13**, 262–271.
Haldane, J. B. S. 1948 Unpublished.
Haldane, J. B. S. & Smith, C. A. B. 1947 *Ann. Eugen.* **14**, 10–31.
Holford, F. F. 1938 *J. Infect. Dis.* **63**, 287–297.
Hirszfeld L. & Kostuch Z. 1938 *Schweiz. Z. Path. u. Bakt.* **1**, 23.
Hoogvliet, B. 1942 *Genetica*, **23**, 94.
Hyman, H. S. 1935 *J. Immunol.* **29**, no. 3.
Kemp, T. 1944 *Acta Path. microbiol. Scand.* suppl. LIV.
Kloepfer, H. W. 1946 *Ann. Eugen.* **13**, 35–71.
Landsteiner, K. & Levine, P. 1928 *J. Exp. Med.* **47**, 757–775.
Landsteiner, K. & Wienin, A. S. 1941 *J. Exp. Med.* **74**, 309–320.
Lattes & Garrasi 1932 *Atti IV Congr. Naz. Microbiol.* p. 146.
Lewis, T. & Embleton 1908 *Biometrika*, **6**, 26.
Mather, K. & Philip, U. 1940 *Ann. Eugen.* **10**, 403–416.
Matta, D. 1937 *Faculty Med. Publ. Egypt. Univ. Cairo*, no. 11.
Metz, C. W. 1938 Cooperation in research, *Carn. Inst. Wash. Pub.* **501**, 275–294.
Mørch, T. 1941 *Chondrodystrophic dwarfs in Denmark.* Copenhagen: Eijnar Munksgaard.
Moureau, P. 1935 *Rev. Belg. Sci. Med.* **7**, 541–588.
Pätau, K. & Nachtsheim, H. 1946 *Z. Naturforsch.* **1**, 345.
Pearson, K., Nettleship & Usher 1913 *Drap. Co. Res. Mem. Biom.*, Series IX.
Penrose, L. S. 1935 *Ann. Eugen.* **6**, 133–138.
Penrose, L. S. 1946*a* *Ann. Eugen.* **13**, 25.
Penrose, L. S. 1946*b* *J. Hered.* **37**, 285.
Philip, U. & Mather, K. 1940 *Ann. Eugen.* **10**, 403–416.
Pipkin, A. C. & Pipkin, S. 1945 *J. Hered.* **36**, 313.
Richsbieth, H. 1912 *Treas. Hum. Inher.* **1**, 355–553.
Rife, D. C. 1941 *Science*, **94**, 187.
Schiff, F. 1933 *Dtsch. Z. ges. gerichtl. Med.* **21**, 404–434.
Sjögren, T. 1931 *Hereditas*, **14**, 197–425.
Snyder, L. H. & Palmer, D. M. 1943 *J. Hered.* **34**, 207–212.
Taylor, G. L. & Prior, A. M. 1939*a* *Ann. Eugen.* **9**, 18–44.
Taylor, G. L. & Prior, A. M. 1939*b* *Ann. Eugen.* **9**, 97–108.
Weinberg, W. 1927 *Z. indukt. Abstamm.- u. VererbLehre*, **48**, 179–228.
White, M. 1940 *J. Genet.* **40**, 403–438.
Wiener, A. S. 1943 *Blood groups and transfusion.* Springfield: Charles C. Thomas.
Wiener, A. S. & Sonn, E. B. 1943 *Genetics*, **28**, 157–161.
Wiener, A. S. & Vaisberg, M. 1931 *J. Immunol.* **20**, 371–388.
Zieve, M. A., Wiener, A. S. & Fries, J. 1936 *Ann. Eugen.* **3**, 163–178.