Perspectives

Anecdotal, Historical and Critical Commentaries on Genetics

Edited by James F. Crow and William F. Dove

Eighty Years Ago: The Beginnings of Population Genetics

I suppose that every teacher of elementary genetics has at one time or another encountered the belief that in the absence of counteracting factors there should be three times as many dominant as recessive phenotypes in the population. It was this statement that induced G. H. Hardy in 1908 to write his famous paper. He started out somewhat apologetically, saying: "I am reluctant to intrude in a discussion concerning matters of which I have no expert knowledge, and I should have expected the very simple point which I wish to make to have been familiar to biologists."

The principle was independently published a few months earlier by Weinberg (1908), but this paper remained unknown to most English-speaking geneticists, so for many years the principle was called Hardy's law. Since the 1940s, thanks to Curt Stern's (1943) setting the record straight, it is referred to as the Hardy-Weinberg law. It is so self-evident that it hardly needed to be "discovered"; Sewall Wright, among others, used it before he had heard of either Hardy or Weinberg. Yet, trivial as it appears, the H-W principle is the foundation for diploid population genetics.

The law can be stated in two ways. First, it says that if mating is at random in a large population and mutation, migration, and selection are absent, the genotype proportions do not change from generation to generation. This is almost a tautology: if nothing changes the frequencies, they won't change. But the principle does make clear that, with inbreeding or assortative mating, the genotype frequencies can change while the allele frequencies do not. Second, and much more usefully, the law permits the prediction of genotype frequencies from knowledge of gene frequencies. If alleles A and a are in the proportions p and q, the three zygotic types AA, Aa, and aa are in the proportions p^2, $2pq$, and q^2. Thus, equations can be written in terms of the more basic units of allele frequencies, and hypotheses about how phenotypes are inherited can be tested from population data.

Most natural populations show approximate agreement with Hardy-Weinberg expectations. The main reason for this is that, unlike most equilibria, this one is attained within a single generation rather than asymptotically. Thus, there is no cumulative departure from H-W proportions; although allele frequencies may change cumulatively by random drift, each generation is close to H-W expectations for *its* allele frequencies. Furthermore, a single generation of random mating undoes all the effects of nonrandom mating that may have gone before. All this was clearly pointed out by Hardy in his two-page paper.

In his 1908 paper Weinberg generalized the law to multiple alleles and in his 1909 paper he extended it to multiple loci. He realized that with more than one locus the equilibrium is not attained in a single generation as it is with one locus but is approached asymptotically at a rate determined by the amount of recombination. By this date the basic foundations of diploid, randomly mating populations had been established. The only problem was that Weinberg's work was hardly noticed, mainly I suspect because most British and American geneticists were not fluent in German. Weinberg suffered from a neglect similar to that of Mendel.

The H-W principle has been most useful for studying inheritance in nonexperimental populations. It has been especially helpful in working out the mode of inheritance of common Mendelian traits, which do not lend themselves to pedigree studies. As far as I know, the earliest application of this law to elucidate the inheritance of a trait was Wright's (1917) demonstration, from herdbook records, of the single-locus inheritance of red, roan, and white colors in Shorthorn cattle, now a standard textbook example. Another example is the human ABO blood groups. Although it was known from the turn of the century that these are inherited, the mode of inheritance was not known until Bernstein (1924, 1925) applied gene frequency methods. Still another is Fisher's analysis of the Rh factors, summarized by

him in 1947. Now, of course, these techniques are part of the standard equipment of human geneticists.

HARDY was a British mathematician, one of the greatest, and WEINBERG was a German physician. As far as I know, they never met, although WEINBERG did review HARDY's paper; he didn't think much of the derivation.

HARDY spent most of his life at Cambridge and, along with his friend J. E. LITTLEWOOD, formed the most productive mathematical partnership of all time. Some of their greatest work was in number theory and complex analysis, very deep and very creative work. HARDY was also the first to recognize the greatness of the self-taught Indian genius RAMANUJAN; other mathematicians had failed to understand and appreciate him. RAMANUJAN's hundredth birthday anniversary was recently celebrated. HARDY, a confirmed bachelor who lived for mathematics, cricket, and conversation, said that the discovery of RAMANUJAN was the one romantic event in his life. When he first saw the letter from RAMANUJAN containing several hand-scrawled theorems, he was astounded. "I had never seen anything in the least like them before. A single look at them is enough to show that they could only be written down by a mathematician of the highest class. They must be true because, if they were not true, no one would have had the imagination to invent them." RAMANUJAN's rapidly converging expression for π, which included a five-digit numerical constant, was recently used to compute the value to 17 million decimal places; only after this was a rigorous proof worked out. What concatenation of genes can produce such special genius?

RAMANUJAN lived only a short time after being brought to England by HARDY, who tells a story about visiting him while he was hospitalized. In an effort to make conversation, HARDY told RAMANUJAN that the taxi on which he had ridden had the license number 1729, a rather dull number. RAMANUJAN replied that, on the contrary, it was a very interesting number, the smallest that can be expressed as the sum of two cubes in two different ways. In summarizing his life, HARDY said: "I still say to myself when I am depressed, and find myself forced to listen to pompous and tiresome people, 'Well, I have done one thing *you* could never have done, and that is to have collaborated with both LITTLEWOOD and RAMANUJAN on something like equal terms.'"

HARDY was a pure mathematician's pure mathematician. He abhored any "practical" mathematics. For him, pure mathematics was beautiful and useless, while useful mathematics was dull and ugly. In his acerbic, opinionated, idiosyncratic, yet charming *A Mathematician's Apology* (1940) he writes: "I have never done anything 'useful.' No discovery of mine has made, or is likely to make, directly or indirectly, for good or ill, the least difference to the amenity of the world." He took the same view of physics; practical physics was ugly. Curiously, the branches of physics that he regarded as most beautiful, and hence most useless, were relativity and quantum mechanics.

It must have embarrassed him that his mathematically most trivial paper is not only far and away his most widely known, but has been of such distastefully practical value. He published this paper not in the obvious place, *Nature,* but across the Atlantic in *Science.* Why? It has been said that he didn't want to get embroiled in the bitter argument between the Mendelists and biometricians. I would like to think that he didn't want it to be seen by his mathematician colleagues.

WEINBERG was a physician, general practitioner and obstetrician, in Stuttgart. He attended at more than 3500 births. Despite this busy life, he somehow found the time and energy to make fundamental discoveries. He published a method for calculating the proportion of monozygotic and dizygotic twins from the proportion of like-sexed twins as early as 1901, at a time when the biological origin of the two types of twins was still only an assumption. He also concluded, correctly, that a propensity to dizygotic twinning, but not to monozygotic, is inherited.

WEINBERG was the first to recognize and correct ascertainment bias. In his 1901 paper on the inheritance of twinning he astutely used the sibs of twins to determine the proportion of twinning in families that had been identified through a twin pair. Later (1912), on reading BATESON's comment that there are more than ¼ albinos among children of heterozygous parents, he realized that this was also an ascertainment problem. Families with no affected children were not discovered, and thus the proportion of albinos is inflated in those families that are included. He invented the sib and proband methods to correct for it; both utilize the ratio in the sibs of affected, with the probands removed. WEINBERG was the founder of segregation analysis. It was developed further by HALDANE and FISHER, and greatly extended by MORTON (1982), who worked out methods, now in wide use, for separating out sporadic and polygenic components.

In 1910 WEINBERG published an article[1] on the correlations between relatives in a randomly mating population. At this time, PEARSON and other biometricians thought that the observed correlations between relatives were inconsistent with Mendelian inheritance. WEINBERG showed this to be wrong. He also took environmental effects into account and

[1] Part of this remarkable paper has been translated by KARIN MEYER (see pp. 42–57 of HILL 1984). HILL (*ibid.*, pg. 13) has also provided a useful table of correspondences between WEINBERG's sometimes confusing notation and that now in common use.

utilized the additivity of squared standard deviations, anticipating the analysis of variance.

While WEINBERG was writing in Germany, R. A. FISHER was beginning his studies in England. FISHER was later to become the greatest statistician of his generation if not the greatest ever. While still an undergraduate he had written a paper (1912) that foreshadowed the use of maximum likelihood as an estimating procedure. Yet, for all his promise, he was not able to find a job that suited his talents. He was turned down for military service in World War I because of poor eyesight, and finally found employment teaching physics and mathematics. He taught at Rugby and Haileybury Schools, on a naval training ship, and at Bradfield College. He hated it, and undoubtedly was no good at bringing the subject down to the level of his students.

In these years there was a raging argument in Britain between the "Mendelists" and the "Biometricians." In retrospect the disagreement seems a bit silly, for it now seems obvious that quantitative traits can be explained by postulating a large number of Mendelian factors. In fact the argument never took place in the United States, where this assumption was made from the beginning. The continuing vituperation probably had more to do with the personal differences and sensitive egos of PEARSON and WELDON, representing the biometricians, and BATESON and PUNNETT, representing the Mendelists, than on the scientific evidence.

While still a student at Cambridge, FISHER became convinced that the large number of Mendelian factors was a sufficient explanation of metrical traits. He decided to see if the observed correlations between relatives were consistent with Mendelism. His blockbuster, "The correlation between relatives on the supposition of Mendelian inheritance," was published in 1918. FISHER wrote the paper while teaching school students and completed it in 1916, but it was not accepted for publication by the Royal Society of London. For creativity and depth by someone out of the academic mainstream, this is reminiscent of EINSTEIN's great papers written while he worked in a patent office. FISHER's reviewers were PEARSON and PUNNETT, bitter opponents in the Mendelism-biometrics debate, and neither recommended publication. It has been said that this was the only time that the two ever agreed. The paper was finally published in the *Transactions of the Royal Society of Edinburgh,* and only through the financial help of LEONARD DARWIN, CHARLES DARWIN's son. Ironically, by the time the paper was published, the point had been settled and the paper's main argument was moot.

Yet this paper is remarkable, and more important, in other ways. Not only did FISHER show that biometry and Mendelism were compatible, but he worked out in full detail the theory of correlations between relatives and the apportionment of variance—a term he invented in this paper—between genetic and environmental factors. He further showed that dominance contributed to sib correlations but not to those of parent-offspring, thus accounting for the greater observed value of sib correlations. Nowadays, we would give more emphasis to the greater environmental correlations of sibs, but FISHER's analysis was a remarkable theoretical breakthrough. He also showed how to include epistatic interactions as a component of variance. Finally, he considered in great detail the consequences of assortative mating, in some ways more thoroughly than anyone since.

Although this paper is in many ways the foundation of quantitative genetics, FISHER did little more on this subject. Perhaps he thought he had answered most of the major questions. He did write one more paper, with IMMER and TEDIN (1932), in which he carried the analysis to third moments. It has not been used widely.

In 1919 FISHER finally got a job. This was at the Rothamsted Experiment Station, and it was in connection with this work that he worked out the procedures—analysis of variance and covariance, factorial design, field plot arrangements, design of experiments—that are now everyday practice. He also laid the mathematical foundations for the statistics of small samples. His biological interests turned to evolution, and his book *The Genetical Theory of Natural Selection* (1930) is in many ways the natural successor to *The Origin of Species.* A book-length biography of FISHER is available, written by his daughter (BOX 1978).

1918 is significant for another paper, this time in GENETICS. In this year WRIGHT published a paper with the innocent title, "The nature of size factors." This is the forerunner of path analysis, WRIGHT's technique for using partial regression coefficients to assign relative importance to different paths in a complex causal pattern. This method became his greatest contribution to statistical methodology.

Finally, 1988 is the bicentennial of the completion of GIBBON's *Decline and Fall.* Other than as a numerical coincidence, why mention *this* book in *this* context? The reason is that FISHER was greatly influenced by it, and the last chapters of his 1930 book are devoted to the conditions for stability of civilizations. To FISHER's regret his genetically based theory was ignored by both biologists and historians. The last chapters of his otherwise highly influential book are largely unread.

The years 1908 and 1918 are important ones in the history of genetics. FISHER's 1918 paper is still discussed and WRIGHT's method of path analysis is widely applied in the social sciences. And we are still learning things about the HARDY-WEINBERG relationship. In this issue of GENETICS, C. C. LI shows that

random mating is a sufficient, not a necessary condition for H-W ratios.

James F. Crow
Genetics Department
University of Wisconsin
Madison, Wisconsin 53706

LITERATURE CITED

Bernstein, F., 1924 Ergebnisse einer biostatischen zusammenfassenden Betrachtung über die Erblichen Blustrukturen des Menshen. Wien. Klin. Wochenschr. **3:** 1495–1497.

Bernstein, F., 1925 Zusammenfassende Betrachtungen über die erblichen Blutstructuren des Menschen. Z. Indukt. Abstammungs Vererbungsl. **37:** 237–270.

Box, J. F., 1978 *R. A. Fisher: The Life of a Scientist.* John Wiley & Sons. New York.

Fisher, R. A., 1912 On an absolute criterion for fitting frequency curves. Messenger Math. **41:** 155–160.

Fisher, R. A., 1918 The correlation between relatives on the supposition of Mendelian inheritance. Trans. R. Soc. Edinb. **42:** 399–433.

Fisher, R. A., 1930 *The Genetical Theory of Natural Selection.* Oxford University Press, Oxford.

Fisher, R. A., 1947 The Rhesus factor. Am. Sci. **35:** 95–102, 113.

Fisher, R. A., F. R. Immer and O. Tedin, 1932 The genetical interpretation of statistics of the third degree in the study of quantitative inheritance. Genetics **17:** 107–124.

Hardy, G. H., 1908 Mendelian proportions in a mixed population. Science **28:** 49–50.

Hardy, G. H., 1940 *A Mathematician's Apology.* Cambridge University Press, Cambridge.

Hill, W. G. (Editor), 1984 *Quantitative Genetics.* Benchmark Papers in Genetics. Van Nostrand Reinhold, New York.

Morton, N. E., 1982 *Outline of Genetic Epidemiology.* S. Karger, Basel.

Stern, C., 1943 The Hardy-Weinberg law. Science **97:** 137–138.

Weinberg, W., 1901 Beiträgae zur Physiologie und Pathologie der Mehrlingsgeburten beim Menschen. Arch. Ges. Physiol. **88:** 346–430.

Weinberg, W., 1908 Über den Nachweis der Vererbung beim Menshen. Jahresh. Ver. Vaterl. Naturkd. Wuerttemb. **64:** 368–382.

Weinberg, W., 1909 Über Vererbungsgesetze beim Menschen. Z. Indukt. Abstammungs Vererbungsl. **1:** 277–330.

Weinberg, W., 1910 Weitere Beiträge zur Theorie der Vererbung. Arch. Rass. Ges. Biol. **7:** 35–49.

Weinberg, W., 1912 Über Methode und Fehlerquellen der Untersuchung auf Mendelsche Zahlen beim Menschen. Arch. Rass. Ges. Biol. **9:** 165–174.

Wright, S., 1917 Color inheritance in mammals. VI. Cattle. J. Hered. **8:** 521–527.

Wright, S., 1918 On the nature of size factors. Genetics **3:** 367–374.