Perspectives

Anecdotal, Historical and Critical Commentaries on Genetics
Edited by James F. Crow and William F. Dove

Hardy, Weinberg and Language Impediments

James F. Crow

Genetics Department, University of Wisconsin, Madison, Wisconsin 53706

THE Hardy-Weinberg law is the cornerstone of diploid population genetics. Yet it seems trivially obvious, a routine application of the binomial theorem. And indeed it was so regarded by Hardy when he wrote his famous paper, a masterpiece of clarity:

> To the Editor of Science: 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. However, some remarks of Mr. Udny Yule, to which Mr. R. C. Punnett has called my attention, suggest that it may still be worth making.. . .
>
> Suppose that *Aa* is a pair of Mendelian characters, *A* being dominant, and that in any given generation the number of pure dominants (*AA*), heterozygotes (*Aa*), and pure recessives (*aa*) are as p:2q:r. Finally, suppose that the numbers are fairly large, so that mating may be regarded as random, that the sexes are evenly distributed among the three varieties, and that all are equally fertile. A little mathematics of the multiplication-table type is enough to show that in the next generation the numbers will be as $(p+q)^2:2(p+q)(q+r):(q+r)^2$, or as $p_1:2q_1:r_1$, say.
>
> The interesting question is—in what circumstances will this distribution be the same as that in the generation before? It is easy to see that the condition for this is $q^2 = pr$. And since $q_1^2 = p_1r_1$, whatever the values of *p*, *q*, and *r* may be, the distribution will in any case continue unchanged after the second generation (Hardy 1908).

Britain's leading mathematician must have had a poor impression of the quantitative skills of geneticists. The statement to which he took exception concerned the dominant trait, brachydactyly. In discussing a paper by Punnett, Yule said that eventually one would expect three brachydactylous persons to one normal.

I have always found Yule's statement surprising. It was Yule who pointed out that Karl Pearson's parent-offspring correlation of 1/3 applied only to a single locus with complete dominance and that without dominance it became 1/2, closer to the observed value. He also emphasized that environmental effects should be taken into account. Most important, as Provine has said: "Yule was ahead of his time. In 1906 he was probably the only biometrician in England who recognized not only that Mendelism and biometry were compatible but

also, even more crucial, that Mendelism and Darwin's idea of continuous evolution were compatible" (Provine 1971, p. 85). Yule's statement was a curious slip for a man who had introduced so much clarity into the rancorous debates between the mendelists and the biometricians. I suppose that even the greatest are entitled to one mental lapse.

When I began teaching genetics, this principle was called Hardy's law. Later, Stern (1943) called attention to an article of Weinberg (1908), who showed the same principle at the same time (for an English translation of Weinberg's paper, see Boyer 1963, pp. 4–15). Weinberg went farther. He showed that the principle would work for multiple alleles, which he postulated, not knowing that they had actually been discovered. He also pointed out that the approach to a multilocus equilibrium was asymptotic rather than immediate. Not knowing of linkage, he assumed Mendelian independence.

Since Stern's article this has been called the Hardy-Weinberg (HW) law. It was soon pointed out that both Pearson and Castle had still earlier used the HW principle for special cases, but the cumbersome designation "Castle-Pearson-Hardy-Weinberg law" soon fell under its own weight. Of course, a principle as simple as this must have occurred to many geneticists in the early days of the century. Sewall Wright once said that he had used the idea in his own early calculations long before he had heard of either Hardy or Weinberg.

Why was Weinberg's paper, published the same year as Hardy's, neglected for 35 years? The reason, I am sure, is that he wrote in German. At the time, genetics was largely dominated by English speakers and, sadly, work in other languages was often ignored. We saw in last month's Perspectives (Epperson 1999) that the great accomplishments of Gustav Malécot were unknown to such as Fisher and Wright, mainly because he wrote in French. Even those fluent in French were not likely to read the often obscure journals in which he mainly published. There was, of course, another reason: both Weinberg and Malécot wrote papers that were difficult, even for native speakers. Malécot used a great

deal of higher mathematics. Weinberg used only elementary mathematics—he avoided calculus—but even elementary mathematics can be difficult to follow when the subject is complicated. Although both of these men suffered the kind of neglect that Mendel had, they at least had some appreciation during their lifetimes—although not nearly enough.

WILHELM WEINBERG, 1862–1937

Weinberg's physical life was uneventful, being that of a busy physician, but his intellectual life was something else. He produced one new idea after another. In those days when phenotypic observations of breeding experiments were almost the sole basis for genetic inferences, the human species was particularly refractory. More than anyone else of his time, Weinberg showed that clever mathematical trickery could provide answers to difficult questions that would be trivially easy in an experimental species with large numbers of progeny. With the techniques now available for the study of human genetics, it is hard to imagine how difficult and limited the subject was at a time when only superficial phenotypes were observed (there were no CEPH families or traits adaptable to such data).

Weinberg was born in Stuttgart and was an outstanding student at the Gymnasium. He studied medicine in Tübingen and Munich, receiving his M.D. in 1886. He returned to Stuttgart in 1889 and remained there until his retirement. In his later years he was in poor health and had a hard time making ends meet. He retired to Tübingen a few years before his death in 1937.

According to Curt Stern's deeply sympathetic short biography, he spent 42 years as a busy private physician (Stern 1962). In addition, he was a physician to the poor. Among other things in his busy life, he delivered 3500 babies. Somehow, he managed to fit into this crowded schedule time to write papers, many of them long and full of carefully analyzed data. Some were path-breaking in their originality. He wrote more than 160 papers, plus reviews and comments. Yet, he received almost no recognition outside Germany.

He worked alone and had neither students nor colleagues. Indeed, he appears to have had few friends. He remained outside the circle of geneticists. In his writings he was often argumentative and abusive. His criticisms were pointed and often personal. He clearly felt that he was not being properly recognized. He must qualify as a "difficult" personality, yet he was benevolent and clearly had a strong social conscience and sense of justice. In an obituary Luxenberg wrote that Weinberg "succeeded to his own harm—to keep carefully secret the high measure of benevolence, good will to men, and sense of justice which had been his" (Stern 1962).

Weinberg's early work, done at the turn of the century, grew out of his obstetrical practice. He interpreted the excess of like-sexed twins as a clue to there being
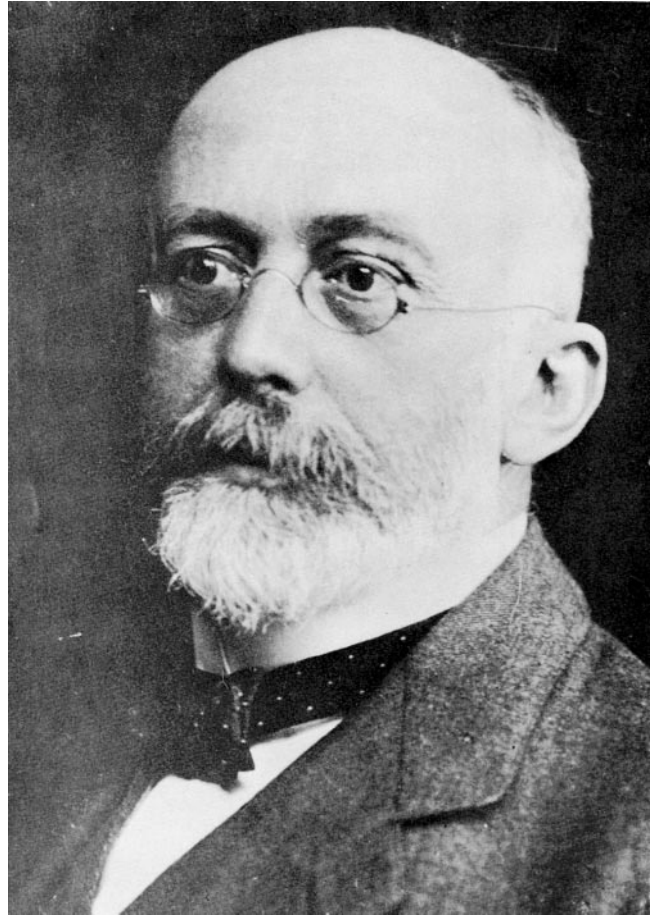

Figure 1.—Wilhelm Weinberg (from Stern 1962).

two kinds of twins and correctly inferred that these were of one-egg and two-egg origin. He used this excess of like-sexed pairs as a way of determining the relative frequency of the two types. Among many findings, he concluded that dizygotic twinning was inherited, although this could not be proven for monozygotics.

Weinberg's outstanding work, I believe, was his analysis of the correlation of relatives. In these articles (Weinberg 1909a,b, 1910) he anticipated much of the later work of Fisher and Wright. In particular, he partitioned the total phenotypic variance into genetic and environmental components, which Fisher did not, and got the effect of dominance correct, which Wright did not. Weinberg must be included with Wright and Fisher as pioneers in quantitative genetics. His articles were extremely difficult for British and American geneticists to read, partly because they were in German and partly because of Weinberg's notation, which is quite different from the Fisher-Wright usage that is now conventional. Hill (1984, p. 13) has done a great service by providing a table comparing Weinberg's expressions with the usual ones of Falconer. Also, in the same volume (pp. 42–57), Karin Meyer has provided a most welcome translation of the key article (Weinberg 1910).

Weinberg was the first to recognize the problem of

ascertainment bias. When, in his early twin studies, he wanted to determine the frequency of twin births in families in which a pair of twins had occurred, he realized that this should be based on the twinning frequency among the sibs of the twins, omitting the index twins. In another problem, others had noted that the proportion of albino children from normal parents considerably exceeded the expected 1/4. Weinberg realized that families in which no albino child occurred were not included in the data and worked out several ways for correcting for the bias. He proposed the "sib" and "proband" methods, by which the sibs of affected individuals are counted and each family is appropriately weighted. In the proband method the weight is the number of independent ascertainments of the sibship. These methods were all refined and further developed by other workers much later, especially Fisher and Morton (for a review, see Crow 1965). Although part of the human geneticist's tool kit, these methods are now much less frequently used, thanks to the more direct approaches made possible by molecular and computer methods.

Weinberg was the first to deal with ascertainment issues in other problems. He explained the greater fertility of parents compared with their children as a simple consequence of the fact that children necessarily come from fertile parents. He proposed using the fertility of sibs of the parents to compare with that of the children. He also explained *anticipation*, the earlier onset of a disease in later generations, as the consequence of lesser severity and later onset in those individuals who reproduced. Galtonian regression would account for the greater expression in the children. As a specific mechanism, Penrose noted that unlinked modifiers could be involved (Laxova 1998). Weinberg did not live long enough to discover that some of the most striking cases of anticipation are not the statistical artifacts that he predicted, but rather have a mechanistic basis in the tendency of trinucleotide repeats to increase in length and in the severity of their consequences.

Weinberg pioneered in the use of identical versus fraternal twins for separating genetic from environmental causes. His method was the now-standard one—find a twin affected with whatever trait is being studied and then ask how often the co-twin is affected. He realized that what he really wanted was not the proportion of cases in which the co-twin had the trait at the time, but the probability of the co-twin developing the trait during a lifetime. So, he worked out a correction, which had the usual Weinberg touch of cleverness and elegance.

MUTATION AND PATERNAL AGE

Most of Weinberg's methods are now standard or have become obsolete because of later developments. But one idea introduced by Weinberg is a subject of active contemporary research, made much more precise by molecular techniques. This is one of his most remarkable observations and deserves to be brought up to date. He made a detailed study of the dwarfism trait, achondroplasia, which he knew to be inherited as a Mendelian dominant. Specifically, he noted that an affected child born from normal parents tended to be among the last-born children in the sibship. From this he suggested that these were new mutations. In his words: "If a more exact analysis of birth order indeed confirmed a high incidence in last-born children, this would speak for the formation of the initial predisposition for dwarfism by mutation" (Weinberg 1912, p. 717).

This is a remarkable statement for its time. Mutation was an extremely vague concept in those days and, of the little that was known, it is not clear how much Weinberg knew. The clarifying Drosophila work was just getting started. Weinberg did not try to distinguish between maternal age, paternal age, or birth order. That was to come later, in fact not until some 40 years later. Penrose (1955) was able to separate these causes and show that paternal age is the main, if not the only, one. Of course there is a birth order and maternal age effect, but these are accounted for by the correlation of ages between husbands and wives and of paternal age with later births. For an account of Penrose's work, life, and character by one who knew him well, see an earlier Perspectives (Laxova 1998).

Achondroplasia is only one of a number of conditions under which *de novo* cases show a paternal age effect. A number of other traits show a similar pattern (Risch *et al.* 1987; Vogel and Motulsky 1997). Interestingly, much of this work was done in Germany. The mean age of fathers at the time of conception of an affected child is about 6 years higher than the average age of fathers at conception in the same population. X-linked traits show an increased age of maternal grandfathers, as expected. The hypothesis that is immediately suggested is that the mutation process is replication dependent (or at least correlated with number of cell divisions). According to Vogel and Motulsky (1997), in the male there are 30 cell divisions from zygote to puberty (age 15), 23 per year thereafter, and 6 more from gonial proliferation and meiosis. Thus the number of chromosome replications prior to a sperm produced by a male of age A is then

$$N_A = 30 + 23(A - 15) + 5,$$

because there is only one replication for the two meiotic cell divisions. Thus, in males of age 20, 30, 40, and 50, the number of chromosome replications is 150, 380, 610, and 840. The ratio for a man of age 50 to that at puberty is 840/35 or 24.

Thus, a large paternal age effect is not surprising if mutation is correlated with the number of replications, as seems reasonable. The actual age increase is considerably greater, however. This, I think, is not surprising. We would expect fidelity of transcription, error correction,

and such to deteriorate with age. The pioneering findings of Weinberg and Penrose have been abundantly borne out.

A recent report of congenital heart abnormalities, in which ventricular and atrial septal defects were lumped with patent ductus, showed a small but statistically significant paternal age effect (Olshan *et al.* 1994). The authors concluded that some 5% of the incidence was attributable to an age effect. This suggests that a part of the cases may be due to new dominant mutations. If so, perhaps these could be found by studying an enriched sample of families in which the fathers were unusually old at the time the affected child was conceived. This might be a useful research strategy.

Whatever the age of the parents, there are many more cell divisions in the male than in the female. In the female all the cell divisions take place early, so the number of chromosome replications, 23, is not age dependent. Thus, for a 40-year-old father, the male/female replication ratio is $610/23 \approx 27$, and the mutation ratio should be still higher.

Until recently it was not possible to identify the parental source of a mutation except for X-linked genes. The first person to take advantage of this possibility was Haldane, who estimated, from the excess of carrier mothers of hemophilic sons, that the mutation rate in males was some 10 times higher than that in females (Haldane 1947). Haldane and Penrose were the first to estimate the human mutation rate. Haldane clearly regarded this as one of his greatest accomplishments. He was invited to write his own obituary, which he accepted with alacrity. Not inhibited by false modesty, he wrote: "I am going to begin with a boast. I believe that I am one of the most influential people living today, though I haven't got a scrap of power. Let me explain. In 1932 I was the first person to estimate the rate of mutation of a human gene; and my estimate was not far out." I was told that, in the earliest version, he said *the* most influential, but he thought better of it later.

In more recent data for mutation to X-linked ornithine transcarbamylase (OTC) deficiency, the estimated male/female ratio is 51, although with a large confidence interval (Tuchman *et al.* 1995). Data for mild hemophilia are comparable (Becker *et al.* 1996).

Now that the parental origin of mutations can often be inferred by linkage to molecular markers, we can determine the male/female mutation ratio for autosomal genes. Data are available for the Apert syndrome, multiple endocrine neoplasia (two types), and achondroplasia (Crow 1997; Szabo *et al.* 1997). Altogether more than 150 new mutations have been analyzed and practically all are paternal in origin. The discrepancy is even greater than is expected from the cell division hypothesis, so this may not be the whole story.

A number of other traits, less completely analyzed, show a strong paternal age effect. Additional evidence comes from another source. There is an almost complete absence of affected males for the 13 known dominant X-linked traits that are lethal or sterilizing in females (Thomas 1996). This can be explained easily by a generally high male mutation rate, because such males would come from heterozygous mothers, but such women do not reproduce. And if the female mutation rate is very low, we would expect very few affected males.

There are two striking exceptions to the higher male mutation rate, Duchenne muscular dystrophy and neurofibromatosis (Grimm *et al.* 1994; Lazaro *et al.* 1996). Each of these is an enormous gene with many introns. A substantial share of the mutations in these genes are intragenic deletions or duplications, which do not show an excess of paternal origin. The data actually show a higher female rate, but the numbers are small.

This suggests the hypothesis that base substitution mutations are replication dependent and show large male and paternal age effects. In contrast, deletions and duplications are not replication dependent and are associated with neither the gender of the parent nor paternal age.

But in biology, the situation is rarely simple. Hemophilia provides an example. Most cases, especially mild ones, show a high male rate for point mutations and a higher female rate for deletions (Becker *et al.* 1996). But severe cases are often caused by specific X chromosome deletions, almost all of paternal origin (Antonarakis *et al.* 1995). There is no elevation of paternal age, as if the inversions occur during meiosis. Another complication is that almost all the mutations for achondroplasia occur at one CpG site.

So I do not want to overgeneralize from a small number of diseases. But we should know more soon, because appropriate studies are going on. Furthermore, the molecular techniques now available provide for a deeper, quantitative analysis of these processes and we shall soon see how well this hypothesis holds up.

Although I have compared relative mutation rates in males and females and for different male ages, I have said nothing about the absolute rates. In particular, it is important to measure this, not for isolated genes, but on a genome-wide basis. This will be the subject of a forthcoming Perspectives by Keightley and Eyre-Walker.

GEOFFREY H. HARDY, 1877–1947

Let us return to Hardy. Both he and Weinberg were brilliant and abrasive. Both were strikingly original. And both did far more profound work than is represented by the Hardy-Weinberg law. But here the resemblance ceases. While Weinberg was delivering babies and giving medical care to the poor, Hardy was doing mathematics in the morning, watching cricket in the afternoon, and drinking port at a Cambridge high table in the evening. Weinberg's work was very practical, while Hardy disdained practicality. In his cloistered world, applied mathematics was ugly; he loved the purest of the pure,

and the more impractical the better. He was strange, original, and enigmatic; but he was Britain's leading pure mathematician. And he could certainly use the English language. For all its idiosyncrasies, parts of his "A Mathematician's Apology" (Hardy 1967) are sheer poetry. I have written more about Hardy in an earlier Perspectives (Crow 1988).

According to Hardy, the one romantic episode in his life was his bringing to England the Indian phenomenon, Ramanujan. This untaught genius found an astonishing number of deep mathematical relationships, and how he did it no one knows. Hardy remarks that Ramanujan was remarkably adept with numbers and had a remarkable memory. But that is surely not a sufficient explanation of his genius. Here is one example from the fascinating list that he sent to Hardy from India.

$$1 - 5\left(\frac{1}{2}\right)^3 + 9\left(\frac{1.3}{2.4}\right)^3 - 13\left(\frac{1.3.5}{2.4.6}\right)^3 + \ldots = \frac{2}{\pi}.$$

One might suspect that he found this by calculating a few terms and seeing the convergence, but this can hardly be. You might enjoy checking this on your own computer. You will find that it does, in fact, approach the proper limit, but *very* slowly. In the first dozen terms it is nowhere near the correct value, but after 10 million it is getting close, giving the value 3.14215. However he divined this, Ramanujan surely did not sum millions of terms. To this non-mathematician, it is black magic. It is beautiful and utterly impractical. This is surely the kind of thing that Hardy loved.

The work of Hardy and Weinberg had little in common, save for the famous rule that forever joined their names. I am sure that neither regarded this as a significant contribution.

LITERATURE CITED

Antonarakis, J. P., J. P. Rossiter, M. Young, J. Horst *et al.*, 1995 Factor VIII gene inversions in severe hemophilia: results of an international consortium study. Blood **86:** 2206–2212.

Becker, J., R. Schwaab, A. Möller-Taube, U. Schwaab, W. Schmidt *et al.*, 1996 Characterization of the factor VIII defect in 147 patients with sporadic hemophilia A: family studies indicate a mutation type-dependent sex ratio of mutation frequencies. Am. J. Hum. Genet. **58:** 657–670.

Boyer, S. H., IV, 1963 *Papers on Human Genetics.* Prentice-Hall, Englewood Cliffs, NJ.

Crow, J. F., 1965 Problems of ascertainment in the analysis of family data, pp. 23–24 in *Genetics and the Epidemiology of Chronic Diseases*, Pub. No. 1163, edited by J. V. Neel, M. W. Shaw and W. J. Schull. U.S. Public Health Service, Washington, DC.

Crow, J. F., 1988 Eighty years ago: the beginnings of population genetics. Genetics **119:** 473–476.

Crow, J. F., 1997 The high spontaneous mutation rate: Is it a health risk? Proc. Natl. Acad. Sci. USA **94:** 8380–8386.

Epperson, B. K., 1999 Gustave Malécot, 1911–1998, population genetics founding father. Genetics **152:** 477–484.

Grimm, T., G. Meng, S. Leichti-Gallati, T. Ettecken, C. R. Müller *et al.*, 1994 On the origin of deletions and point mutations in Duchenne muscular dystrophy: most deletions arise in oogenesis and most point mutations result from events in spermatogenesis. J. Med. Genet. **31:** 183–186.

Haldane, J. B. S., 1947 The mutation rate of the gene for hemophilia, and its segregation ratios in males and females. Ann. Eugen. **13:** 262–271.

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

Hardy, G. H., 1967 *A Mathematician's Apology*, Foreword by C. P. Snow. Cambridge University Press, Cambridge.

Hill, W. G., (Editor), 1984 *Quantitative Genetics*, Part I. Van Nostrand Reinhold, New York.

Lazaro, C., A. Gaona, P. Ainsworth, R. Tenconi, D. Vidaud *et al.*, 1996 Sex differences in the mutational rate and mutational mechanism in the NF gene in neurofibromatosis type 1 patients. Hum. Genet. **98:** 696–699.

Laxova, R., 1998 Lionel Sharples Penrose, 1898–1972: a personal memoir in celebration of the centenary of his birth. Genetics **150:** 1333–1340.

Olshan, A. F., P. G. Schnitzer and P. A. Baird, 1994 Paternal age and the risk of congenital heart defects. Teratology **50:** 80–84.

Penrose, L. S., 1955 Parental age and mutation. Lancet **2:** 312–313.

Provine, W. B., 1971 *The Origins of Theoretical Population Genetics.* University of Chicago Press, Chicago.

Risch, N., E. W. Reich, M. M. Wishnick and J. G. McCarthy, 1987 Spontaneous mutation and parental age in humans. Am. J. Human Genet. **41:** 218–248.

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

Stern, C., 1962 Wilhelm Weinberg. Genetics **47:** 1–5.

Szabo, J. K, D. J. Wilkin, R. Cameron, S. Henderson, G. Bellus *et al.*, 1997 The achondroplasia mutation occurs exclusively in the paternally derived fibroblast growth receptor 3 (FGFR3) allele. Am. J. Hum. Genet. **61:** A348.

Thomas, G. H., 1996 High male:female ratio of germ-line mutations: an alternative explanation for postulated gestational lethality in males in X-linked dominant disorders. Am. J. Hum. Genet. **58:** 1364–1368.

Tuchman, M., I. Matsuda, A. Munnich, S. Malcolm, S. Strautnieks *et al.*, 1995 Proportions of spontaneous mutations in males and females with ornithine transcarbamylase deficiency. Am. J. Med. Genet. **55:** 67–70.

Vogel, F., and A. G. Motulsky, 1997 *Human Genetics: Problems and Approaches.* Springer-Verlag, Berlin.

Weinberg, W., 1908 Über den Nachweis der Vererbung beim Menschen. Jahresh. Wuertt. Ver. vaterl. Natkd. **64:** 369–382 [for an English translation, see Boyer 1963].

Weinberg, W., 1909a Über Vererbungsgesetze beim Menschen I. Z. indukt. Abstamm. Vererbungsl. **1:** 377–392, 440–460.

Weinberg, W., 1909b Über Vererbungsgesetze beim Menschen II. Z. indukt. Abstamm. Vererbungsl. **2:** 276–330.

Weinberg, W., 1910 Weitere Beiträge zur Theorie der Vererbung. Arch. Rassen. Ges. Biol. **7:** 35–49 [for an English translation, see Hill 1984].

Weinberg, W., 1912 Sur Vererbung des Zwergwuchses. Arch. Rassen. Ges. Biol. **9:** 710–717.