Hardy's "Small" Discovery Remembered

Indika Rajapakse, Lindsey Muir, and Paul Martin

> What we do may be small, but it has a certain character of permanence; and to have produced anything of the slightest permanent interest, whether it be a copy of verses or a geometrical theorem, is to have done something utterly beyond the powers of the vast majority of men.
> —*Godfrey Harold Hardy*
> A Mathematician's Apology

The most beautiful mathematics to Godfrey Harold Hardy was that which had no application. For Hardy, mathematics was purely for intellectual challenge. He justified the pursuit of pure mathematics with the argument that its very "uselessness" meant that it could not be used to cause harm. Hardy went so far as to describe applied mathematics as "ugly", "trivial", and "dull" [1].

Despite Hardy's aversion to applied mathematics, he had a profound impact on biology. Mathematicians tend not to realize his contribution, which was downplayed by Hardy himself. At the same time, biologists tend not to appreciate his mathematical brilliance. This note is intended to recognize G. H. Hardy's "little" discovery as a contribution to genetics and to revisit a classic paper [2] that has shaped the field for the past century.

Godfrey Harold Hardy, a graduate of Trinity College, Cambridge, began his mathematics career

Indika Rajapakse is in the Interdisciplinary Research & Training, Program in Biostatistics and Biomathematics, Fred Hutchinson Cancer Research Center, Seattle, WA. His email address is irajapak@fhcrc.org.

Lindsey Muir is in the Molecular and Cellular Biology Program, University of Washington, Seattle. Her email address is lamuir@u.washington.edu.

Paul Martin is in the Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA. His email address is pmartin@fhcrc.org.

in the early 1900s as a fellow at Trinity [3]. He lectured in mathematics for a number of years and published many papers of such significance that he was considered Britain's leading pure mathematician. Hardy was also responsible for bringing the Indian mathematical genius Srinivasa Ramanujan to England, where they published many papers together and developed the field of number theory. While visiting an ill Ramanujan on one occasion, Hardy mentioned that he had traveled in cab number 1729 and "hoped it was not an unfavorable omen" to which Ramanujan replied "...it is a very interesting number; it is the smallest number expressible as the sum of two cubes in two different ways" (it is expressible as $1729 = 1^3 + 12^3$ or $9^3 + 10^3$, now known as the Hardy-Ramanujan number).

In 1908, Hardy published a paper in *Science* that changed the field of population genetics, entitled "Mendelian Proportions in a Mixed Population". The findings were later known as the Hardy-Weinberg Law (Equilibrium) because the same principle was published by Wilhelm Weinberg in the same year [4]. This principle offered a simple solution for the question of how genetic diversity is maintained in a population. For Hardy, the law was trivial and obvious, and he was reluctant to acknowledge its applications. But one hundred years later, the Hardy-Weinberg Law remains a cornerstone of modern computational genetics.

Preservation of genetic diversity in a population requires stability, or equilibrium, of the genotype distribution from one generation to the next. The following is an outline of Hardy's stability condition for Mendelian proportions [2]. This condition holds for a closed system, where the population mates randomly, or for purposes of simplicity, where every individual mates with every other individual once, and where each mating yields a single offspring, with no selection, mutation, migration, or death.

Genotypes	A_1A_1	A_1A_2	A_2A_2	2nd generation frequency
A_1A_1	$\frac{P_{11}}{a} \times \frac{P_{11}}{a}$	$\frac{P_{11}}{a} \times \frac{2P_{12}}{a}$	$\frac{P_{11}}{a} \times \frac{P_{22}}{a}$	$\frac{(P_{11}+P_{12})^2}{a^2}$
A_1A_2	$\frac{2P_{12}}{a} \times \frac{P_{11}}{a}$	$\frac{2P_{12}}{a} \times \frac{2P_{12}}{a}$	$\frac{2P_{12}}{a} \times \frac{P_{22}}{a}$	$\frac{2(P_{11}+P_{12})(P_{12}+P_{22})}{a^2}$
A_2A_2	$\frac{P_{22}}{a} \times \frac{P_{11}}{a}$	$\frac{P_{22}}{a} \times \frac{2P_{12}}{a}$	$\frac{P_{22}}{a} \times \frac{P_{22}}{a}$	$\frac{(P_{12}+P_{22})^2}{a^2}$

Recall that in the mammalian genome, each gene is represented by two of many possible variants, called alleles. For example, the gene, or locus, for eye color can exist in many forms, including a "blue" allele coding for blue eyes and a "brown" allele coding for brown eyes. If we consider only the blue and brown alleles, an individual may have two copies of the blue allele (homozygous for the blue allele), one copy of each allele (heterozygous), or two copies of the brown allele (homozygous for the brown allele). An individual who is homozygous for the blue allele can be designated as having an "A_1A_1" genotype, while an individual who is homozygous for the brown allele is designated as having an "A_2A_2" genotype.

Consider the following case for alleles A_1 and A_2,

P_{11} = the number of individuals with genotype A_1A_1 (the homozygous A_1 case)

$2P_{12}$ = the number of individuals with genotype A_1A_2 or A_2A_1 (the heterozygous case)

P_{22} = the number of individuals with genotype A_2A_2 (the homozygous A_2 case)

Therefore, we can write the first generation proportions of individuals as $P_{11} : 2P_{12} : P_{22}$. Let the total number of individuals in the first generation be represented as a, where $a = P_{11} + 2P_{12} + P_{22}$. The second generation proportions of individuals can be derived from the table seen above.

Therefore, we can write the second generation proportions of individuals as

$$(P_{11} + P_{12})^2 : 2(P_{11} + P_{12})(P_{12} + P_{22}) : (P_{12} + P_{22})^2$$

Let the total number of individuals in the second generation be represented as b, where $b = (P_{11} + P_{12})^2 + 2(P_{11} + P_{12})(P_{12} + P_{22}) + (P_{12} + P_{22})^2$. Hardy's stability condition requires that the proportion of individuals with any given genotype remains constant across generations and is therefore established as follows

$$(1) \quad \frac{P_{11}}{a} = \frac{(P_{11} + P_{12})^2}{b}$$

$$(2) \quad \frac{2P_{12}}{a} = \frac{2(P_{11} + P_{12}(P_{12} + P_{22}))}{b}$$

$$(3) \quad \frac{P_{22}}{a} = \frac{(P_{12} + P_{22})^2}{b}$$

Solving the above system simultaneously yields $P_{12}^2 = P_{11} \times P_{22}$, the stability condition for the two-allele case. This provides a null hypothesis for biologists investigating the distribution of genetic characteristics in a population.

Since $a^2 = (P_{11} + 2P_{12} + P_{22})^2 = P_{11}^2 + 4P_{12}P_{11} + 4P_{12}^2 + 2P_{11}P_{22} + 4P_{12}P_{22} + P_{22}^2 = (P_{11} + P_{12})^2 + 2(P_{11} + P_{12})(P_{12} + P_{22}) + (P_{12} + P_{22})^2 = b$, this demonstrates that $a^2 = b$, as required by the assumption of a closed system.

Let the haplotype frequency of allele $A_1 = p$, and the haplotype frequency of allele $A_2 = q$, recalling that a haplotype is a combination of alleles at multiple genetic loci that are transmitted together from one generation to the next. Accordingly, and $p = \frac{2P_{11}+2P_{12}}{2a}$, and $q = \frac{2P_{22}+2P_{12}}{2a}$, where homozygous individuals are counted twice, heterozygous individuals counted once, and the number of haplotypes is twice the population size. Hence, $p = \frac{P_{11}+P_{12}}{a}$, and $q = \frac{P_{22}+P_{12}}{a}$. Since $p + q = 1$, the modern population genetics interpretation, $p^2 + 2pq + q^2 = 1$ must hold, which states that $\Pr(A_1A_1) = p^2$, $\Pr(A_1A_2) = 2pq$, and $\Pr(A_2A_2) = q^2$.

We can extend the two allele case to a general case as follows. Under the same assumptions as above and additionally that a given gene includes k alleles, say $A_1, A_2, \cdots, A_k$ and P_{ij} is the number of people with the genotype A_iA_j, where i, j can be any real number.

The following generations will approach values in proportion to those suggested by counting all possibilities of mating (the way Hardy did). Thus, if in addition the values satisfy (or nearly satisfy) the stability condition $P_{ij}^2 = P_{ii} \times P_{ij}$, then we can assume that the probabilities for each genotype in each generation will be the same as the

probabilities for each genotype in the preceding generation. That is, the proportion of individuals with each genotype will stay the same, while genetic diversity will be maintained in a predictable way. This formulation is the k allele analog of the Hardy-Weinberg Law.

Certain violations of the closed system assumption lead to departure from Hardy-Weinberg Equilibrium. One such violation is non-random mating, where preferential mating according to genotype may occur in the population. We will not explore the mathematical details, but we can represent this case as follows: $\Pr(A_1 A_1) = p^2 + pqF$, $\Pr(A_1 A_2) = 2pq(1 - F)$, and $\Pr(A_2 A_2) = p^2 + pqF$, where F is defined as the inbreeding coefficient [5]. F can be thought of as the probability that two alleles are identical due to parents passing on the same allele to their progeny. It is therefore also a measure of the degree of parental relatedness. When $0 < F < 1$, homozygosity in the population increases, which may reduce health and reproductive fitness.

The modern interdisciplinary approach has gathered great minds to work on some of the most challenging problems in biology. Hardy's contribution has greatly influenced the growing field of computational genetics. Today, applied mathematics is a critical component of genetics and has the potential to revolutionize the field and profoundly impact modern medicine. If Hardy were alive today, it would be interesting to know whether he would join these minds or remain steadfast in his pursuit of pure mathematics. Despite his disdain for applied mathematics, Hardy was one of the greatest contributors to contemporary mathematical biology, and at this one hundred year anniversary, his "small" discovery will be remembered as such a great contribution.

Acknowledgements

References

[1] G. H. HARDY, *A Mathematician's Apology*, Cambridge University Press, London, 1940.
[2] ———, Mendelian proportions in a mixed population, *Science* **28** (1908), 49-50.
[3] R. KANIGEL, *The Man Who Knew Infinity*, Charles Scribner's Sons, New York, 1991.
[4] J. F. CROW, Hardy, Weinberg and Language Impediments, *Genetics* **152** (1999), 821- 825.
[5] B. S. WEIR, *Genetic Data Analysis II*, Sinauer, Sunderland, 1996.