

The Hardy Criterion and Why the Hardy-Weinberg equilibrium principle holds

John T. Baldwin

April 26, 2020

In this short note we first sketch four proofs of the Hardy-Weinberg principle. The first two are in the modern context of working from the allele frequency. The second two, due to Hardy and Weinberg work more directly with genotype frequency. In particular, Hardy's criterion only require knowing the genotype frequency.

My special thanks to Brendan Boyle of St. John's college who pointed out a mistake in previous version. For more detail and examples see [1].

1 Why the Hardy-Weinberg principle holds

Consider the simplest situation of a monogenic Mendelian trait: a pair of alleles, one dominant A and the other recessive a , and suppose the frequency of the allele A is p and that of a is q .

Hardy-Weinberg Equilibrium Principle: Consider a population satisfying biological conditions 1-6 that are listed below. If in a certain generation the allele frequencies are p and q and the genotype frequencies are $p^2, 2pq, q^2$ then BOTH the genotype and allele frequencies remain the same for as many generations as conditions 1-6 continue to hold.

Hardy implicitly and Weinberg explicitly used the reasoning of a Punnett square of frequencies to verify:

Crucial H-W Insight: Under fully random mating, the frequency of AA homozygotes in the next generation is p^2 , of heterozygotes is $2pq$, and of aa homozygotes is q^2 .

The contribution of Hardy and Weinberg is to see that this principle holds regardless of the allele frequency in the first generation. It was known by a

number of people (e.g. Yule, Pearson) for $p = 1/2$.

We first give a proof that came later as biologists began looking at the evolution of a population rather than the purely genetic question of what happens between parents and their children. So we must introduce some conditions (1 -4) that are absent from the Hardy and Weinberg papers.

Biological Conditions

Suppose the following conditions hold (with respect to a particular gene):

1. There is no migration ("gene flow") in or out of the population.
2. Natural selection is not occurring.
3. Mutation is not occurring.
4. Each member of the population is equally likely to breed.
5. The population is infinitely large.
6. (Full random mating) Each pair from the population is equally likely to breed

It is easy to see using only reasoning about biology that:

Observation 1. As long as a population satisfies biological conditions 1-5 the allele frequencies (p and q) are the same in each generation.

We give two arguments here, which differ in certain details, for the Hardy Weinberg principle. In later sections we review the arguments of Hardy and Weinberg, giving four slight different proofs.

The 'least math' proof When a population satisfies conditions 1-5, Observation 1 ensures that allele frequencies will remain unchanged in every succeeding generation satisfying those conditions. Applying condition 6 and the crucial H-W insight, in each generation, therefore, the genotype frequencies are $p^2, 2pq, q^2$.

A proof separating the 'biological' from the 'random' conditions This version¹ isolates where the biological hypotheses are used.

¹Herron-Freeman[2] has the same general idea –but as in the 'least math' version– building in the evolutionary hypotheses.

Fact 1.1 (From genotype frequency to allele frequency) *In any population basic Mendelian genetics allows us to determine the allele frequencies from the genotype frequencies (in cases where all three genotypes can be identified).*

This is easily seen: In ANY population, the frequency of the dominant allele of a Mendelian pair is the sum of two times the number of dominant homozygotes (AA) + the number of heterozygotes (Aa) divided by the total number of these alleles (i.e twice the number of individuals).

We can view this statement in terms of genotype frequencies. Noting that the number of alleles is 2 times the number I of individuals, this means frequency of the dominant allele is

$$\frac{2 \times (\text{number of AA individuals})}{2I} + \frac{\text{number of Aa individuals}}{2I}.$$

This is:

$$\frac{\text{number of AA individuals}}{I} + \frac{\text{number of Aa individuals}}{2I}.$$

So (*) the frequency of the dominant allele in the second generation is the frequency of AA plus $\frac{1}{2}$ the frequency of Aa.

Now we get the result for one generation which does not involve biological hypotheses 1-4. We assume there are no mutations taking place during the mating process.

One generation HW principle: If an infinite population with allele frequencies p and q for the parental generation has full random mating then at birth the second generation has the allele frequencies p and q .

Proof. By the crucial insight noted in the introduction the genotype frequencies in the second generation are $p^2, 2pq, q^2$.

So, by (*) the frequency p_1 of the dominant allele in the next generation is $p_1 = p^2 + pq$. By basic algebra, $p_1 = p^2 + pq = p(p + q) = p$ and this equation is exactly true because the population is infinite. That is, the allele frequency is the same in both generations.

The general HW principle: If a population satisfies the Hardy-Weinberg criteria in one generation, the genotype and allele frequencies remain the same for as many generations as conditions 1-6 continue to hold.

Biological conditions 1-4 guarantee that the allele frequency is still $p_1 = p$ when the second generation mates. Repeating the 1-generation argument, the genotype and allele frequencies are the same in the third generation and, indeed, as long as the biological conditions continue to hold.

2 Hardy and Weinberg

In this section we will consider the proofs of Hardy and Weinberg who are studying genetics, not evolution. They completely ignore external influences on the allele pool.

2.1 Hardy's Criterion

Hardy[3] works entirely with genotype frequencies (never actually mentioning the allele frequency) and provides a condition (that we consider below) on genotype frequencies which guarantees that they repeat. We probably don't use this argument now because it concerns a triple proportion: D:H:R, where D, H, R are the frequencies of the homozygous dominant, heterozygote, and homozygous recessive genotype. I have reformulated this proportion in modern mathematical terminology.

In the usual formulations, determining whether H-W eq exists in ONE generation of a population can only be done when both the allele frequencies and the genotype frequencies are known. In practice, of course, for most species the generation time is too long to be able to compare phenotype frequencies over several generations and thus some method is needed to determine whether or not a population is at equilibrium for a given gene at a particular time. Hardy[3] provided an ingenious way to determine H-W eq in a single generation in terms of genotype frequencies alone. By Fact 1.1 we can always compute the allele frequency from the genotype frequency², so this method is only a convenience. But it also provides another proof of the

²The frequency of the dominant ideal is twice the frequency of homozygous individuals plus the frequency of heterozygotes.

general principle; the biological conditions 1-6 imply the genotype frequencies remain constant from generation to generation. According to Hardy, in mathematical terms equilibrium ('stability') exists when $H^2 = DR$, where D,H, and R are defined as follows.

1. D is the number of homozygous dominant individuals,
2. 2H is the number of heterozygous individuals, and
3. R is the number of homozygous recessive individuals.

The frequency of the recessive allele is

$$\frac{R + H}{D + 2H + R} \quad \text{Generation 1.}$$

(If we actually count alleles we double each of R, H, D and then factor out a 2.)

Therefore, if all the criteria are met, the number of homozygous recessive individuals in the next generation is the result of all possible matings among the H Heterozygotes who contribute a recessive and the R homozygous Recessives: $(H + R)^2$. The total number of individuals in the entire next generation is $(D + 2H + R)^2$. The FREQUENCY of homozygous recessive individuals, therefore, is:

$$\frac{(H + R)^2}{(D + 2H + R)^2} \quad \text{Generation 2.}$$

Thus, the frequency of the *homozygous* recessive is the same in both generations if and only if the following genotype equation holds.

$$\frac{(H + R)^2}{(D + 2H + R)^2} = \frac{R}{D + 2H + R}$$

Now we do some pure algebra; to clear fractions, multiply both sides of the equation by $(D + 2H + R)^2$, getting:

$$(H + R)^2 = R(D + 2H + R).$$

So $H^2 + 2HR + R^2 = DR + 2HR + R^2$ and subtracting $2HR + R^2$, we get the required $H^2 = DR$.

2.2 Weinberg's argument

Weinberg [5] gives almost the same argument. He starts with allele frequencies; computes (as proportions) the genotype frequencies in second generation: that is, he writes out the *crucial insight*. Then he uses the same reasoning on the genotype frequencies in the second generation: he squares $D + H + R$ to compute the genotype frequency in the third generation. He then simplifies the algebraic express to see that the second and the third generation have the same genotype frequency. Thus equilibrium has been established.

Interestingly neither Hardy nor Weinberg point out that the allele frequency is also constant.

References

- [1] John T. Baldwin and M.U. Smith. Making sense of hardy-weinberg equilibrium. *American Biology Teacher*, 77:577–582, 2015. <https://doi.org/10.1525/abt.2015.77.8.3>.
- [2] S. Freeman and J. Herron. *Evolutionary Analysis*. Pearson/Prentice Hall, 2004.
- [3] G. W. Hardy. Mendelian proportions in a mixed population. *Science*, XXVIII(706):49–50, 1908.
- [4] Curt Stern. The Hardy-Weinberg law. *Science*, XCVII(2510):137–138, 1943.