CHAPTER 15

J. B. S. Haldane & the Evolution of the Hardy-Weinberg Model

FRED SINGER

☐ INTRODUCTION

Many of you have enjoyed taking a long automobile trip. Perhaps your destination was a large city, but in other cases, large cities were obstacles that you needed to navigate as quickly and painlessly as possible. In either case, you may have pulled out a piece of paper with gray, blue, and red lines drawn on it and engaged in the time-honored sport of trying to drive a car and read a map at the same time.

A map is a type of **model**. Like all models, it allows the user to make hypotheses about reality. By using the model, you should be able to arrive at or near your destination. Or, if your goal was to bypass the city, you should be able to navigate around it. Like all models, a road map is a simplified version of reality. If all the information about the city were on the map, you would be hopelessly confused by the excessive detail.

PROBLEM

Think about three different types of maps of your own choosing. How do these maps differ regarding (a) the details they include and exclude, (b) the regions they model, (c) their ease of portability, and (d) their ease of interpretation? Maps are designed for different purposes. How do your answers reflect the different purposes for which each type of map was designed?

All models change. In some cases, as knowledge about a subject increases, the models are made more complex in order to give a more accurate description of reality. In some cases, the process or structure that is being modeled changes, and the model must be modified to reflect the new reality. For example, when a trout stream that ran through New York City was diverted to build Wall Street, the blue line on the city map was replaced with a red line. Finally, the original purpose of the model may become irrelevant, and the model may be either abandoned or changed to reflect some new purpose.

Models of evolution are like maps; they represent the world in a simplified form. Maps allow people to get from one geographical point to another and to predict the

efficiency of different potential routes. Likewise, evolutionary models allow biologists to see how populations go from one point in evolutionary history to another and to predict the route that a population may follow as it changes.

The Hardy-Weinberg model was created to solve two specific problems in genetics. Later, J. B. S. Haldane and other researchers modified the model to answer a variety of questions about evolution. Today this model is being applied by forensic scientists to assess the uniqueness of an individual's DNA fingerprint.

A POPULATION GENETICS PUZZLE

R. C. Punnett, one of the founders of mammalian genetics and originator of the Punnett Square, prompted the development of the Hardy-Weinberg model. Following one of his lectures, Punnett was asked: "If brown eyes are dominant over blue eyes, why doesn't a population eventually become all brown-eyed?" Punnett was baffled by this question, as were most biologists during the early days of genetics. The term *dominant* brings to mind a powerful trait that should overwhelm a weaker, "recessive" trait. Shouldn't the dominant trait always increase within the population?

Punnett mentioned this problem to his good friend Godfrey H. Hardy, one of the world's greatest mathematicians. Hardy and Punnett both worked at Cambridge University and had become acquainted by playing and watching cricket matches together. The next day Hardy rewarded Punnett with an answer that was published in the widely read journal *Science* in 1908.

Unknown to both Punnett and Hardy, the German physician Wilhelm Weinberg had developed a similar model six months earlier to help him investigate the question of whether there is a genetic basis to giving birth to twins. Weinberg published his findings in an obscure journal, and therefore scientists ignored his version of the model for another 35 years.

HARDY'S SOLUTION TO PUNNETT'S PUZZLE

Let's return to the map and use it as a model to guide you through a large city that is on the way to a summer vacation paradise (Figure 15.1). You pull out your map and have the choice of taking either I-5 or Route 99 through the city. Which route do you take? What factors do you use to make this decision? What information is on the map that helps you make this choice? You might notice that some features of interest for making this decision are missing from this model. For example, there is no information about traffic lights, mountains, and good places to eat. This particular model makes an assumption that these missing features are not important to the map reader.

Like the city map example, all models contain **assumptions**. To understand the Hardy-Weinberg model, we need to first understand its assumptions. For sexually reproducing organisms, Hardy assumed that all individuals of a population have an equal probability of combining their gametes (random mating). Though gametes have many thousands of genes, Hardy's model looked at only one gene, which was present in the population in two alternate forms, or **alleles**.

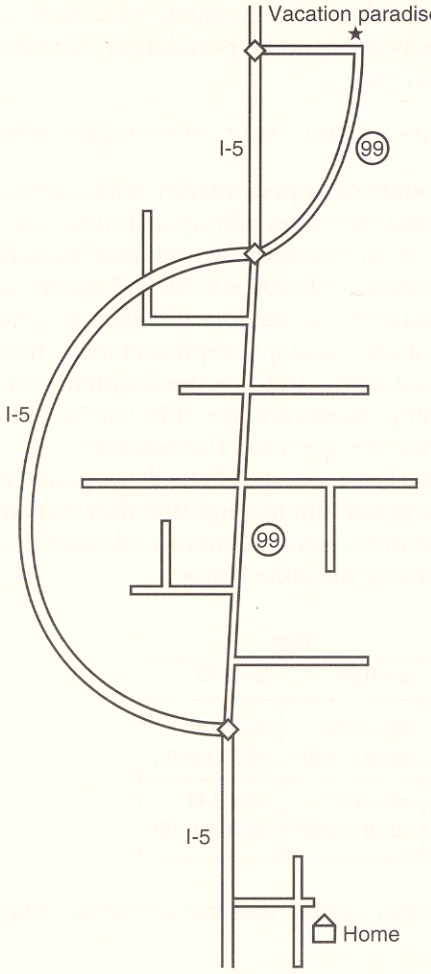

FIGURE 15.1 Road map from home to vacation paradise.

If the brown-eye trait is dominant over the blue-eye trait, *EE* and *Ee* people will have brown eyes, while *ee* people will have blue eyes. Let's begin with a population made up of 9 homozygous dominant individuals (genotype *EE*), 42 heterozygous individuals (genotype *Ee*), and 49 homozygous recessive individuals (genotype *ee*). The total number of *E* and *e* alleles in the population is represented in the table below.

Genotype	Number of Individuals	Number of Copies of *E* Allele	Number of Copies of *e* Allele
EE	9	18	
Ee	42	42	42
ee	49		98
Total	100	60	140

The genotype frequency is the number of individuals of a particular genotype, divided by the total number of individuals in the population. Thus the initial genotype frequencies of this population are:

$$EE = 9/100 = 0.09,\ Ee = 42/100 = 0.42,\ ee = 49/100 = 0.49.$$

These frequencies can also be thought of as percentages, with 9 percent *EE*, 42 percent *Ee*, and 49 percent *ee*. Note that they then add up to 100 percent.

The gene pool is made up of 200 alleles because each individual has two alleles. Therefore the frequency of the *E* allele = 60/200 = 0.30, and the frequency of the *e* allele = 140/200 = 0.70. Just as mapmakers use dots and stars as symbols to represent cities, geneticists use the symbols p and q to represent allele frequencies, with p = the frequency of the dominant allele and q = the frequency of the recessive allele. In this case, $p = 0.30$ (or 30 percent) and $q = 0.70$ (or 70 percent). Because this population only has two alleles, $p + q = 1$ (or 100 percent).

Using these allele frequencies, Hardy asked: What will the genotype frequencies be in the next generation, assuming random mating? Will they remain the same, or will the brown-eyed allele increase in frequency? You can calculate the genotype frequencies for the next generation using the table below.

Sperm	**Eggs**	
	E = 0.30	*e* = 0.70
E = 0.30	*EE* = 0.09 (0.30 X 0.30)	*Ee* = 0.21 (0.30 X 0.70)
e = 0.70	*eE* = 0.21 (0.70 X 0.30)	*ee* = 0.49 (0.70 X 0.70)

According to Hardy's model, we can predict that the genotype frequencies in the next generation will be:

$$EE = 0.09$$
$$Ee = 0.21 + 0.21 = 0.42$$
$$ee = 0.49$$

These predicted frequencies are identical to the starting frequencies, and, as Hardy demonstrated, will always remain the same so long as the model's assumptions are not violated.

Let's see if the allele frequency has changed.

p = frequency of *EE* + 1/2 (frequency of *Ee*) = 0.09 + 1/2 (0.42) = 0.30
q = frequency of *ee* + 1/2 (frequency of *Ee*) = 0.49 + 1/2 (0.42) = 0.70

In Hardy's population, if mating is random, the allele frequency also remains constant from generation to generation. Note that there is no tendency for a dominant trait to increase in frequency from one generation to the next. Thus Hardy pro-

vided his friend Punnett with an answer to his question of why the dominant allele doesn't increase from generation to generation.

The Role of Assumptions in Modeling

In order to be useful, a model must make simplifying assumptions. For example, mapmakers assume that users are not interested in a three-dimensional scale model of all the buildings in a town. Such a model, though wonderfully accurate, would have difficulty fitting into the typical glove compartment. The Hardy-Weinberg model simplifies reality by assuming random mating. The other major assumption is that the four evolutionary factors of natural selection, gene flow, genetic drift, and mutation are not operating. This is an imposing list of assumptions, and on the surface you might wonder why we even bother with this model, given that it applies to no real population. There are three important reasons why this particular model is important, despite its unrealistic assumptions.

First, the general prediction of the Hardy-Weinberg model is true, even if we violate its assumptions to some degree. This is an important characteristic, because mating is never completely random, and evolutionary factors will be affecting allele frequencies to some degree in most populations. In the case of the Hardy-Weinberg model, even if the assumptions are not met, the genotype and allele frequencies will be similar (though not exactly identical) from one generation to the next.

Second, this model can be used as a diagnostic tool. Just as a physician bangs your knee with a hammer to test your reflexes, population geneticists use the Hardy-Weinberg model to determine if its assumptions hold for a particular population. If the genotype frequencies are changing, they know that the population has violated at least one assumption of the model in a major way. Then they can look for evidence of nonrandom mating or for evolutionary factors such as natural selection or genetic drift that may be operating in the population.

Finally, the history of this model shows that ways of thinking that develop for one purpose may often be applied for many other purposes. After the Hardy-Weinberg model was developed, it sat idle for over a decade, having successfully resolved the confusion over why dominant traits don't increase in frequency within a population. After World War I, however, the theoretical biologists R. A. Fisher, J. B. S. Haldane, and Sewell Wright built upon the model, ultimately providing the foundation for the new field of population genetics.

JOHN BURDON SANDERSON HALDANE

J. B. S. Haldane (Figure 15.2) had a unique scientific background. His father, a world authority on respiratory physiology, devoted a significant portion of his professional activities to investigating atmospheric conditions in mines. Haldane routinely accompanied his father on these investigations, and was once required by his father to stand up in an alcove of a mine shaft and recite Mark Antony's famous "Friends, Romans, countrymen" speech from *Julius Caesar*. After a few sentences, Haldane passed out

FIGURE 15.2 J. B. S. Haldane (background) with colleague, conducting research in a pressure chamber. *Source:* Ronald W. Clark, *JBS: The Life and Work of J.B.S. Haldane,* New York: Coward-McCann, Inc., 1969.

from inhaling methane that had collected at the ceiling of the alcove, effectively learning that methane is lighter than air and generally does not kill when inhaled. Such experiential learning formed a significant component of his informal education.

A second important episode in Haldane's informal education was World War I, during which he designed, tested, and lobbed experimental bombs at enemy artillery. Haldane relished these activities, earning for himself the nickname "Bombo" and a reputation as the bravest and dirtiest officer in the army. He used his respiratory physiology background to design a gas mask (using himself as the "guinea pig") to provide protection against poisonous gas attacks launched by the Germans.

This man of action who designed sensitive mechanisms for grenades and bombs and personally hurled his creations at the enemy contrasts with Haldane the laboratory scientist. With the exception of his specialized respiratory equipment, which he could use quite effectively, Haldane was very clumsy when it came to operating delicate equipment or handling small specimens for his genetics research. We don't know whether this ineptitude guided his research agenda at all, but most of Haldane's work was done with his brain, a pencil and paper, and the language of mathematics.

Haldane was fascinated by the fruit fly research of T. H. Morgan and H. J. Muller, experiencing "all the satisfaction of reading a first-rate detective story, much enhanced by the fact that the story was true" (see Chapter 5). He realized that he could help tie together the fields of genetics and evolution if he could derive a quan-

titative theory to show how natural selection changes allele frequencies. Haldane had two problems. First, he needed to develop a model that related the intensity of natural selection to the rate of evolutionary change. Second, he needed to apply his model to a species that had changed genetically in some clear and significant manner in the recent past.

Haldane knew that Hardy's model predicted no change in genotype frequency. He also knew that Hardy's model assumed that natural selection was not operating in the population. Haldane reasoned that one approach to developing his model would be to modify Hardy's model by including the effects of natural selection. But he needed a measure of natural selection's intensity.

Haldane defined fitness (w) as the relative ability of an organism to pass on its genes to the next generation. Fitness is affected by two factors: an individual's ability to survive to sexual maturity and an individual's reproductive success. Both of these depend on how well adapted the individual is to a particular environment. According to Haldane's definition, the most successful genotype has a fitness (w) = 1.0. A genotype that produces, on average, only 20 percent as many offspring as the most successful genotype has $w = 0.20$.

Haldane defined the selection coefficient (s) as a reduction of fitness suffered by the less successful genotype; it is equal to $1 - w$. In this example, the more successful genotype has no reduction in fitness, so $s = 0$, and the less successful genotype has an 80 percent reduction in fitness, so $s = 1.0 - 0.20 = 0.80$.

One of the best-known populations that had changed genetically in Haldane's recent past was the peppered moth (see Chapter 1). Between 1848 and 1901, the population near Manchester, England, went from almost 100 percent light-winged moths to almost 100 percent dark-winged moths. Haldane suspected that natural selection was responsible for this change, and he used this example to calculate how strong natural selection must be to bring about such a dramatic change in genotype frequency in only 53 generations.

PROBLEM

Propose at least two different factors that could give a selective advantage to the dark-winged moths in Manchester.

By incorporating the selection coefficient into the basic Hardy-Weinberg equation, Haldane was able to show how a population would change in cases where selection was operating against either the recessive allele or the dominant allele. Haldane knew that light wing color was due to a recessive allele in the peppered moths. According to his new formula, the change in the frequency of a recessive allele (Δq) in one generation is $\Delta q = -spq^2/(1 - sq^2)$.

Haldane assumed there were a few dark-colored moths in Manchester in 1848, and he began with the frequency of the dark-winged moths = 0.01 and the frequency of the light-winged moths = 0.99. He then plugged in different possible values for the selection coefficient, repeating this process 53 times (without a calculator) to represent the 53 generations between 1848 and 1901. For each generation, the allele frequencies changed in response to the intensity of natural selection. Haldane estimated that an average dark-winged moth must produce three surviving

offspring for each two offspring produced by the average light-winged moth in the Manchester population.

PROBLEM

Based on Haldane's estimates of rates for the two forms of the peppered moth, what are the values of *w* and *s* for the dark- and light-winged forms of the peppered moth?

Approximately 30 years later, H. B. D. Kettlewell conducted field experiments that measured the selective advantage to the dark-winged moths in Manchester. While there were some differences between Haldane's estimation and Kettlewell's field measurement, the two values were reasonably close. This demonstrated that Haldane's model of evolution could make testable predictions about evolution. During the 1920s and 1930s, Haldane, Fisher, and Wright also modified the basic model to show how genetic drift, gene flow, and mutation could bring about evolutionary change. Ultimately, these modeling efforts tied together the fields of genetics and evolutionary biology.

☐ EPILOGUE

One of the reasons for focusing discussion on the Hardy-Weinberg model is that it provides a good example of how a model may be modified and applied to solve different problems. The Hardy-Weinberg model is now being applied by forensic scientists to estimate the uniqueness of DNA fingerprints.

In one notorious example, an assailant broke into the house of a computer operator in Orlando, Florida. He covered the victim's face with a sheet and raped her. This same pattern was repeated over the next year, with police suspecting the same man to be responsible for 23 incidents of breaking and entering and attempted assault or rape. On March 1, 1987, police apprehended and arrested a suspected prowler.

The original victim had seen the suspect before her face was covered, so she was called to examine a photo lineup. She immediately identified the suspect from the group of photos presented to her. A problem arose because none of the other victims was able to make a visual identification, and the suspect had two witnesses who swore he never left the house at the time of the crimes.

Fortunately, police had collected sperm samples from the original victim immediately after she reported the crime. The DNA pattern of the rapist's sperm was identical to the DNA from the suspect's white blood cells. The testing laboratory estimated the odds of such a match as approximately 1 in 10 billion. How did the lab make this numerical claim?

Forensic scientists can make such estimates because there are regions of DNA that don't appear to code for any types of protein. Some of these regions are made up of repetitive DNA—regions of DNA that repeat the same base-pair sequences over and over again. Molecular biologists have identified regions that are highly variable in the number of repeating sequences. This type of gene is called a VNTR, for "variable number of tandem repeats." Within a population, there may be hundreds of alleles of a particular VNTR gene, with each allele representing a different number of tandem repeats (Figure 15.3).

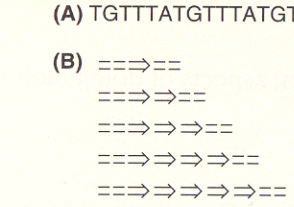

FIGURE 15.3 (A) Repetitive sequence of one strand of DNA, made up of four tandem repeats of the six-base sequence TGTTTA. (B) Five alleles of VNTR genes that differ in number of tandem repeats (= represents nonrepetitive part of the DNA; ⇒ represents tandem repeats). Notice that there are other parts of the genes besides the tandem repeats. Also notice that the genes with more tandem repeats are larger. This size difference allows each allele to be distinguished by forensic scientists.

Because there are so many different VNTR alleles, the frequency of each individual allele is very low. If the frequencies of the two alleles are 0.01 and 0.02, then according to the Hardy-Weinberg model, the genotype frequency of a heterozygote individual will be $2pq$ or 2(0.01)(0.02) or 0.0004. On average, only 4 out of 10,000 individuals in the human population would have that genotype. In this way, the allele frequencies from a large random sample of the population form a baseline from which the expected genotype frequencies are calculated.

Furthermore, forensic scientists can simultaneously look at more than one VNTR gene for a given individual. By assuming that a VNTR genotype at one gene does not affect the probability of having a particular VNTR genotype at a second gene, scientists are able to more effectively discriminate between DNA samples. The assumption that genotype probabilities at two (or more) different genes are unrelated is an example of the assumption of statistical independence and only holds in a population that is in Hardy-Weinberg equilibrium. Under these conditions, the probability of having two matching genotypes is the product of the probability of having each of those genotypes independently (this is called the product rule).

Returning to the previous example, the genotype frequency at the first VNTR gene was 0.0004. Let's consider a second VNTR gene for the same individual, who is heterozygous for two alleles that have frequencies of 0.03 and 0.04. The genotype frequency is $2pq$ = 2(0.03)(0.04) = 0.0024. According to the product rule, the probability of having both of those genotypes is the product of the genotype frequencies within the population = 0.0004 x 0.0024 = 0.00000096. Only 96 individuals would be expected to have that genotype in a population of 100 million.

Most forensic scientists look at four different independent VNTR genes for criminal cases. They believe that this number of genes is sufficient to rule out the possibility of a false conviction. Other geneticists disagree, pointing out that the product rule may not be valid for human populations. If random mating does not occur, then there may be a higher probability that two individuals will, by chance, have the same VNTR genes. Recently, a panel of scientists from the National Academy of Sciences endorsed a more conservative type of analysis that accounts for possible nonrandom mating and gives somewhat lower probabilities of uniqueness of DNA fingerprints. Though this debate continues, the Hardy-Weinberg model lies at the heart of this new technology, bringing criminals to justice and clearing innocent suspects of wrongdoing.

QUESTIONS AND ACTIVITIES

1. What does this case show about the following aspects of doing biology?
   — different uses of models
   — the assumptions made by models
   — how models change over time
2. Why is high fitness not the same as dominance? How does understanding this distinction allow you to answer the question posed to Punnett after his lecture?
3. In this chapter, Haldane predicted the selective advantage of the dark-winged form of the peppered moth in industrialized areas, based on how long it took the dark-winged form to spread within a population. As described in Chapter 1, Kettlewell measured the selective advantage in the field. Are the numbers in close agreement? What might account for any differences between Haldane's predicted values and Kettlewell's findings?
4. One of the characteristics of the Hardy-Weinberg model is that its predictions hold even when its assumptions are not completely met. As a prosecuting attorney, tell the jury why this characteristic is important for them to understand when considering DNA fingerprinting evidence.
5. Haldane made a model of evolution in the case where selection operated against a dominant allele as well as in the case of selection against the homozygous recessive allele described above. In which case would you expect evolution to proceed more rapidly? Why?
6. The initial population in the eye-color example in the text was already in Hardy-Weinberg equilibrium. Hardy's original *Science* paper began with a population made up exclusively of homozygous individuals; thus, his population was not in equilibrium. Calculate the allele frequency and the genotype frequencies of the next generation for a population consisting of 30 homozygous brown-eyed individuals and 70 homozygous blue-eyed individuals. How do these frequencies compare with the eye-color example in the text? If a population is disturbed from Hardy-Weinberg equilibrium, how long does it take to return to equilibrium genotype frequencies?
7. Revisit the example from the text where $w = 0.20$ for the less successful form (remember w always equals 1.0 for the most successful form). Realize that this is an example of very intense natural selection, as the most fit genotype is producing five times as many offspring as the less fit genotype. If $w = 0.20$, then $s = 1 - 0.20 = 0.80$. Let's begin with an intermediate frequency of the recessive allele of $q = 0.60$ and see how the allele frequency changes from one generation to the next. The change in the frequency of the recessive allele, $\Delta q = -spq^2/(1-sq^2)$, $= -0.80(0.40)(0.60)^2/(1 - 0.80(0.60))^2 = -0.1152/(1 - 0.288) = -0.1618$. Thus q for the next generation is $0.60 - 0.1618 = 0.4382$. This dramatic decrease in the frequency of the recessive allele occurs because the fitness of the homozygous recessive individual is so low.

Intuitively, do you think evolutionary change will be most rapid at high, intermediate (as in our example), or low frequencies of the recessive allele? Haldane was particularly interested in how fast evolutionary change should occur under different conditions. Help him solve this problem, and test your intuition by using his formula to determine if Δq is greatest at high, intermediate, or low frequencies of the recessive allele.

SUGGESTED READING

Clark, R. W. 1969. *JBS: The Life and Work of J. B. S. Haldane.* New York: Coward-McCann.

Hardy, G. H. 1908. "Mendelian Proportions in a Mixed Population." *Science* 28: 49–50.

Judson, H. F. 1987. "Modeling." In *The Search for Solutions.* Baltimore: Johns Hopkins University Press, pp. 111–129.

Levins, R. 1966. "The Strategy of Model Building in Population Biology." *American Scientist* 54: 421–431.

Lewis, R. 1988. "DNA Fingerprints: Witness for the Prosecution." *Discover* (June): 45–52.

Punnett, R. C. 1950. "Early Days of Genetics." *Heredity* 4: 1–10.

Weinberg, W. 1908. "On the Demonstration of Heredity in Man." In S. H. Boyer, IV, ed., *Papers on Human Genetics.* Englewood Cliffs, NJ: Prentice-Hall, pp. 4–15. (Reprinted from *Jahreshefte des Vereins für Vaterländische Naturkunde in Württemberg, Stüttgart.* 1908. 64: 368–382. Lecture at the scientific evening at Stüttgart, January 13, 1908.)