Selden, S. (1999). *Inheriting shame: The story of eugenics and racism in America.* New York: Teachers College Press.
Sherman, S. Lu, DeFries, J. C., Gottesman, I. I., Loehlin, J. C., Meyer, J. M., Pelias, M. Z., Rice, J., and Waldman, I. (1997). Recent developments in human behavioral genetics: Past accomplishments and future directions. *American Journal of Human Genetics* 60: 1265–1275.
Wingerson, L. (1998). *Unnatural selection: The promise and the power of human gene research.* New York: Bantam Books.
Wright, W. (1998). *Born that way: Genes, behavior, personality.* New York: Knopf.

FROM: GENETICS AND CRIMINAL BEHAVIOR
Edited by David Wasserman and Robert Wachbroit,
Cambridge University Press, 2001,
Pages 47 to 78.

Chapter 3

Separating Nature and Nurture

ELLIOTT SOBER

Plant and animal breeders routinely attempt to disentangle the contributions of nature and nurture when they think about what makes corn grow tall or cows produce more milk. To apply the same concepts to human characteristics such as intelligence and violence, however, is politically explosive.

The discipline of quantitative genetics separates the relative contributions of genes and environment by deploying a set of technical concepts. The main one is called "*variance*," which measures how much a trait varies in a population. Nature and nurture are analyzed by discussing variance in its different forms; there is phenotypic variance, genetic variance, environmental variance, and variance due to gene-environment interaction. Is it solely political considerations that make some people resist applying these humdrum scientific concepts to human beings? Or are there purely scientific considerations that block the easy transfer of these concepts from one domain to the other? I do not pose this question to disparage the significance of political questions; science is a human activity, and whether a scientific question should be pursued depends on what the consequences for human welfare would be of pursuing it. However, my goal in this chapter is to explain why the issue is not purely political. I am not going to argue that human beings and their traits are somehow outside the scope of biology, whatever that might mean. A human being develops a level of intelligence and attitudes toward violence because of the genes he or she possesses and the environments he or she inhabits. The very same thing is true of height in corn plants and milk yield in dairy cattle.

I am grateful to Andre Ariew, Ned Block, Richard Lewontin, and Steve Orzack for useful discussion.

No, the problem is not that we are outside the biological realm. The concepts of phenotypic, environmental, and genetic variance apply to human beings just as much as they do to cows and corn. For any trait we care to name, there is a fact of the matter concerning what the heritability of that trait is within this or that human population, whether we want to know about that fact or not. The problem is that our current level of knowledge frequently prevents us from ascertaining what the relevant facts are.

In what follows, I explain the relevant mathematical concepts in a way that is both simple and general. The math involved never goes beyond the arithmetic you learned in elementary school. The explanation, however, is quite general, applying as it does to any trait in any population of organisms. I am hoping that this chapter will make the meaning of heritability and related concepts completely transparent even to those with serious cases of math phobia. After defining these quantities, I explain two procedures that are frequently used for estimating the heritability of traits in populations of organisms. Once again, the discussion of estimation procedures involves nothing more than simple arithmetic. At this point in the exposition I explain why it is hard to find out how heritable many traits are.

The upshot of this discussion is not that there is a vitally important property of a trait – its degree of heritability – that we unfortunately are cut off from apprehending. Quite the contrary. In addition to explaining what heritability is, I discuss the question of why it matters and why the everyday concept of a trait's being "inherited" differs in several respects from the technical concept of its having a high degree of "heritability." Once we see heritability for what it is, the question arises of why the heritability of a trait is interesting; another question I ask is whether the heritability of a trait will become less interesting as science learns more about its underlying genetics.

THE ANALYSIS OF VARIANCE

Nuts and Bolts

Consider a farmer who grows two fields of corn. In the first, the corn plants are genetically identical – all have genotype *G*1 – and they receive one unit of fertilizer (*E*1). In the second field, the corn plants also are genetically identical, but they have genotype *G*2; in this second field, the plants receive two units of fertilizer (*E*2). At the end of the growing season, the farmer sees that the plants in the first field are one unit tall (on average), whereas those in the second field average four units of height. These observations can be recorded as two entries in a two-by-two table:

		Genes	
		G1	G2
Environment	E1	1	-
	E2	-	4

The farmer wants to answer a question about nature and nurture: Do the corn plants in the two fields differ in height because they are genetically different, because the plants grew in different environments, or for both these reasons? And if both genes and environment are responsible for the difference between the two fields, which mattered more? With the data described so far, the farmer has no way to answer these questions. The reason is that the genetic and the environmental factors are perfectly *correlated;* *G*1 individuals always inhabit *E*1 environments and *G*2 individuals always live in *E*2 environments.

The way for the farmer to make headway on this problem is to break the correlation. Let him plant a third field in which corn plants have *G*1 genotypes and receive two units of fertilizer; let him plant a fourth field in which *G*2 plants receive one unit of fertilizer. This will allow him to enter data in the other two cells in the two-by-two table just displayed. From these data, the farmer can make an inference concerning how genetic differences and differences in fertilizer treatment contributed to variation in plant height.

The experiment just described might generate different observational outcomes. Here are four possibilities to consider:

	G1	G2
E1	1	1
E2	4	4

(i)

	G1	G2
E1	1	2
E2	3	4

(ii)

	G1	G2
E1	1	3
E2	2	4

(iii)

	G1	G2
E1	1	4
E2	1	4

(iv)

In outcome (i), the genetic factor makes no difference; whether the plants have genotype *G*1 or *G*2 does not affect their height; it is the environmental factor – the amount of fertilizer the plants receive – that explains all the observed variation.[1] Outcome (iv) is the mirror image of (i). In (iv), the fertilizer treatment makes no difference; the genetic variation explains all the variation in height. Outcomes (i) and (iv) support

monistic explanations of the variation in plant height; each suggests that only one of the factors considered made a difference in the observed outcome.

Outcomes (ii) and (iii), on the other hand, support *pluralistic* conclusions. Both suggest that genetic and environmental factors made a difference. They disagree, however, as to which factor mattered more. In outcome (ii), changing the fertilizer treatment yields two units of change in height, whereas changing from one genotype to the other produces only a single unit of change. In this case, the environmental factor makes more of a difference than the genetic factor. By the same reasoning, we can see that outcome (iii) suggests that genetic variation was more important than the environmental factor considered.

Although the four possible outcomes described so far differ in various respects, they have something in common. In each of these data sets, the change effected by moving from $G1$ to $G2$ does not depend on which environmental condition one considers. Similarly, changing from $E1$ to $E2$ has the same impact on plant height, regardless of which genotype the plant possesses. Results (i)–(iv) are thus said to be "additive" (or to show no "gene-environment interaction"). This is not the case for the following possible results:

	G1	G2
E1	1	7
E2	7	4

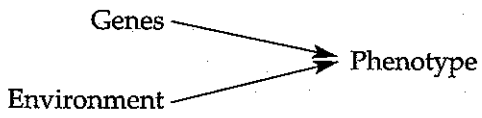
(v)

	G1	G2
E1	1	1
E2	1	4

(vi)

In outcome (v), going from one unit of fertilizer to two increases plant height for plants with genotype $G1$, but reduces height for plants that have genotype $G2$. In (vi), changing genotype has an effect on plant height within one fertilizer treatment but not within the other.

This simple two-by-two experiment illustrates the method that statisticians call the analysis of variance (ANOVA). By generalizing the scheme just described, we can introduce some terminology to label the relevant concepts. In the example just discussed, genotype and fertilizer treatment combine to make a plant grow to a certain height. Height is a phenotypic trait; amount of fertilizer is an environmental condition. There are two possible causes and one effect:



What, in general, do the words "gene," "environment," and "phenotype" mean in ANOVA investigations?

It is hard to answer this question with precision and without circularity. Phenotypes are often described as any feature of an organism's morphology, physiology, or behavior. However, psychological and cultural characteristics also are parts of the phenotypes of some organisms; knowing Korean and liking rock and roll are phenotypic traits, just as much as an individual's height and blood type. What is excluded from the phenotype is the sequence of nucleotides found in the strands of DNA in each cell of an organism's body. In a sense, we can view the organism's phenotype as all the traits it has that are caused by its genes and/or its environment. This way of defining "phenotype" entails that having a certain set of genes is not part of the organism's phenotype, and living in a certain environment isn't either. It will be convenient in what follows to think of phenotypes as quantitative features; they may come in integer values (like number of fingers), or they may be continuous (like height and weight).

Genes are possible causes of phenotypes. But there is also the contribution of the environment. What does "environment" mean? Once again, we must realize that the concept of environment is used in ANOVA as a garbage can category; an environmental factor is anything that is not genetic. The most important point to recognize here is that "genetic" is not synonymous with "biological." This is a common confusion in discussions of nature and nurture. Fertilizer is a "biological" cause of corn plant height, but it counts as an environmental, not a genetic, factor.

If phenotype is defined by contrast with what is genetic, and if the environment is defined by contrast with what is genetic, how are phenotype and environment distinguished? Well, an organism's environment causally contributes to the phenotype it has, but that is not a sufficient answer. After all, it also is true that the organism lives in a particular environment in part because of the phenotype it has. For example, lizards, being cold-blooded creatures, seek out warm environments and shun cold ones. Where the organism lives is a consequence of its physiological makeup.

It suffices for our purposes to distinguish phenotype and environment in the following rough-and-ready way: the organism's phenotype includes only the traits it has in virtue of what is going on inside its own skin (Sober 1984, sec. 1.5).[2] The organism's environment includes only the traits the organism has in virtue of what is going on outside its own

skin. Living in a warm place is an environmental trait; being cold-blooded is a feature of the organism's phenotype (Dawkins 1982).[3]

In the corn plant example, there were two genetic conditions and two environmental conditions. Let's generalize. Each organism studied is in one of *m* possible genetic states and experiences one of *n* possible environmental conditions. An experiment that examines all combinations of these genetic and environmental conditions will have *n*-times-*m* treatment combinations. Organisms within each treatment cell are measured for some phenotypic trait. As in the simpler two-by-two example, cell entries record the average phenotype of individuals in each treatment. To simplify exposition, I'll assume that each cell contains the same number of individuals:

	G1	G2	G3	...	Gm	
E1	x_{11}	x_{12}	x_{13}	...	x_{1m}	$x_{1.}$
E2	x_{21}	x_{22}	x_{23}	...	x_{2m}	$x_{2.}$
	.	.	.	.	.	
	.	.	.	.	.	
	.	.	.	.	.	
En	x_{n1}	x_{n2}	x_{n3}	...	x_{nm}	$x_{n.}$
	$x_{.1}$	$x_{.2}$	$x_{.3}$	...	$x_{.m}$	M

In addition to the x_{ij} entries in the *m*-by-*n* table itself, the table also provides some numbers that are written down along the margins. These are called, appropriately enough, the *marginal averages*. They describe the average phenotype for individuals who experienced the same environment (but different genes), or the same genes (but different environments). The table also states, as a final entry in the lower right-hand corner, the grand mean *M* – the average phenotype in the entire population of individuals.

The phenotypes represented in the *m*-by-*n* table – the different x_{ij} entries – display a certain amount of variation; the numerical values may be tightly clustered or they may be spread out. The standard mathematical measure of the amount of "spread" a phenotype has in a population is given by the phenotypic *variance* (V_p). To compute this quantity, one finds the difference between each x_{ij} and the grand mean, squares this difference, and then computes the average of these squared differences:

$$V_p = \Sigma_{ij} (x_{ij} - M)^2/nm.$$

Intuitively, the amount of variation present in this population can have two sources. First, there is the fact that individuals live in different environments; second, there is the fact that individuals have different genes. This idea is captured by the fact that we can decompose the total (phenotypic) variance in the population into two parts – the genetic variance and the environmental variance:[4]

$$V_g = \Sigma_j (x_{.j} - M)^2/m.$$

$$V_e = \Sigma_i (x_{i.} - M)^2/n.$$

Each of these variances is computed by seeing how much the marginal averages (the $x_{.j}$'s and the $x_{i.}$'s) vary from the grand mean.

Just to demystify this way of representing and decomposing variation, let's analyze a very simple data set, which the farmer might obtain in his two-by-two experiment:

	G1	G2	
E1	1	3	2
E2	5	7	6
	3	5	4

The marginal averages and the grand mean are duly recorded, from which we can compute the three variances:

$$V_p = [(1-4)^2 + (3-4)^2 + (5-4)^2 + (7-4)^2]/4 = 5$$

$$V_g = [(3-4)^2 + (5-4)^2]/2 = 1$$

$$V_e = [(2-4)^2 + (6-4)^2]/2 = 4.$$

Note that in this example $V_p = V_g + V_e$. This defines what it means for a data set to be additive.

However, the sums do not come out this way in the following data set, which involves an interaction:

	G1	G2	
E1	1	3	2
E2	5	11	8
	3	7	5

Here are the values for the three variances:

$$V_p = [(1-5)^2 + (3-5)^2 + (5-5)^2 + (11-5)^2]/4 = 14$$

$$V_g = [(3-5)^2 + (7-5)^2]/2 = 4$$

$$V_e = [(2-5)^2 + (8-5)^2]/2 = 9.$$

In the present case, we introduce a quantity I to represent the difference between V_p and V_g+V_e. In this example, the gene-environment interaction term has a value of unity.

As the various hypothetical data sets I have described make clear, we should not assume in advance that the data produced in an experiment will turn out to be additive. Rather, we should describe the total phenotypic variance as decomposing into three parts:

$$V_p = V_g + V_e + I. \quad (1)$$

The data we obtain may show us that $I = 0$. Indeed, it may turn out that $V_g = 0$ or that $V_e=0$, as was true in data sets (i) and (iv). These are empirical matters, which will vary with the population studied and the trait considered.

Because V_g and V_e are quantities that describe how much variance is associated with the different genes and the different environments considered in the experiment, we may compare these two numbers to say which of them induced the larger amount of variation. It is customary to do this by taking ratios of each of these quantities, relative to the total phenotypic variance. If we divide both sides of (1) by V_p, we obtain:

$$1 = V_g/V_p + V_e/V_p + I/V_p. \quad (2)$$

The three right-hand terms describe, respectively, the proportion of the phenotypic variance that is due to genes, to environment, and to gene-environment interaction; they total 100 percent.

The first of the ratios in proposition (2) defines the concept of "heritability"[5] $h^2 = V_g/V_p$. A phenotype's heritability is the proportion of its variance that is caused by genetic variance. Notice that heritability is a property of phenotypes, not genes. In ordinary parlance, we talk of genes as well as phenotypes (like eye color) being "inherited"; however, "heritable" and "inherited" are not synonymous. Genes are not heritable. We examine other differences between the concepts of "heritable" and "inherited" in the next section.

Notice that proposition (1) cites *three* possible causes of phenotypic variation. This means that the genetic contribution to variation cannot

be defined as the contribution that is not environmental. Thus, the analysis of variance provides a finer-grained representation of causal contribution than the intuitive one with which we started. We began with the idea that phenotypes are caused by genes and by environment, where "gene" and "environment" are defined so that there is no other type of factor that can cause an organism's phenotype. Although this intuitive picture is good enough when it comes to talking about the traits of an individual *organism*, it is not adequate as a description of what can produce phenotypic variance in a *population*. For the latter task, the mathematics of ANOVA requires that we recognize genetic variance, environmental variance, and variance due to gene-environment interaction. We have moved from *two* causes to *three*.[6]

Some Philosophical Comments

The type of experiment just described can provide information about the relative causal contributions of genes and environment to the phenotypic variation found in a population. However, that information must not be misinterpreted.

The analysis of variance can provide information about causality only if the population exhibits variation in the effect term studied. Consider a population of human beings in which everyone has exactly two hands. The grand mean is 2, and each genotype/environment treatment has 2 as its average number of hands. There is no variation to explain; and there is no phenotypic variation due to genes nor any due to environment; $V_p = V_g = V_e = 0$. This example shows another respect in which "heritability" and "inherited" differ in meaning. Hand number may seem like an obvious example of an "inherited" characteristic; however, its heritability in the population just described is not defined, since $V_g/V_p = 0/0$.

What would happen if we considered a larger population in which the number of hands does vary? We can construct an example of this sort by augmenting a population of two-handed individuals with individuals who are born with a smaller number of hands. These individuals may have one hand or none solely because they have some genetic defect; alternatively, they may have been born without hands solely because of a feature of their fetal environment (e.g., perhaps their mothers took a certain drug while pregnant). If the enlarged population includes handless individuals solely of the first type, then hand number will turn out to be highly heritable; if the enlarged population includes handless

individuals solely of the second type, then hand number will have zero heritability. And if both types of individuals are included, the resulting phenotypic variation will have both a genetic and an environmental explanation.

This example provides an interesting lesson. It is possible to assess the heritability of a trait without having any understanding of the developmental processes that lead individuals to exhibit the trait. Even if we know nothing about how an individual's genes and environment conspire to insure that he or she develops hands, we nonetheless can tell whether *variation* in hand number is genetic or environmental or both. This point was already visible in the farmer's experiments discussed in the previous section. He need not understand *why* a particular combination of genes and environment yields plants that average four units of height. It suffices for him to observe that this has happened. The analysis of variance permits one to infer *how much* a cause contributes to an outcome without understanding *how* the cause manages to have its effects. In part, this is because ANOVA aims to explain the variation of traits in a population, not to explain why individual members of the population have the traits they do.[7] It is a fact about development that the genes in an organism's body help explain why that individual ends up with two hands; it is a quite separate matter whether genetic differences in a population help explain differences in hand number.

Perhaps the most important point about interpreting ANOVA data as evidence about causal contribution is that the inferences are specific to the phenotypic trait considered, the range of environments and genotypes studied, and the population studied. In the two-by-two experiments contemplated before, we considered the results of a specific pair of genetic traits ($G1$ and $G2$) and a specific pair of environmental variables ($E1$ and $E2$). The results obtained in that restricted domain say nothing about what would happen if some new genotype $G3$ were taken into account, or if some new environment $E3$ were brought into the problem. For example, it could turn out that the difference between $G1$ and $G2$ makes no difference in plant height, but that $G3$ makes all the difference in the world. The same could happen for the environmental effect; the difference between one unit of fertilizer and two might not matter, even though three units cross a threshold that matters a lot. In short, an ANOVA experiment does not ascertain how much genes *in general* matter, or how much the environment *in general* matters. What the experiment investigates is a *specific* set of genetic factors and a *specific* environmental treatment.

Another respect in which ANOVA yields specific results, not general ones, is that the relative importance of genes and environment to a phenotype can change as a population ages. Suppose the farmer conducts his two-by-two experiment, computes the environmental and genetic effects after the corn plants are three weeks old, and finds that most of the variation is due to genetic differences. If he then follows the plants for an additional three months, it may turn out that variation in height at that later date is mainly due to environmental variation. Here we have another difference between the technical concept of heritability and the commonsense idea of a trait's being inherited. According to the commonsense concept, if you inherit a trait, it remains an inherited trait as long as you have it; however, the heritability of a trait, because it is a property of a population, can change as the population changes.

Two more details are worth mentioning in connection with the fact that ANOVA is specific to the range of environments and genotypes considered and the specific population under study. Different subgroups in the same overall population may show very different patterns of genetic and environmental variation. Imagine a four-by-four experiment; environmental treatments $E1$, $E2$, $E3$, $E4$ are paired with genetic conditions $G1$, $G2$, $G3$, $G4$. Let us consider two subsets of this entire experiment. In the upper left-hand corner of the four-by-four data table we are imagining, we find a description of what happens when $E1$ and $E2$ are paired with $G1$ and $G2$. In the lower right-hand corner, we find $E3$ and $E4$ paired with $G3$ and $G4$. It is entirely possible that genes make a great deal of difference to the resulting phenotype in the first case, but little or no difference in the second. Nor is this possibility merely hypothetical. Suppose we were studying variation in skin color among various populations in North America. Variation in skin color will be mainly genetic, if we focus on the people who live in New York City. But if we compare people who live in North Dakota with people who live in Utah, the variation will be mainly environmental (Block and Dworkin 1976).

The last wrinkle of this sort that I want to describe is this: even if genes matter a lot and environment matters only a little within each of two populations, this does not mean that the difference between the two populations is mainly due to their being genetically different. Lewontin (1970) illustrated this point by describing an experiment in which a genetically heterogeneous collection of corn seeds is used in two experiments. In the first, seeds are drawn from this collection and are planted in standard potting soil; in the second, seeds are drawn from the same collection and are planted in potting soil from which various trace ele-

ments have been removed. Within each experiment, all the variation in height will be due to genetic differences; however, the difference between the two experimental populations will be entirely environmental. This point was central to the controversy in the 1960s and 1970s concerning how the observed IQ difference between American whites and American blacks should be explained. Jensen (1970) argued that the between-group difference is partly due to genetic differences. Lewontin (1970) replied that Jensen's reasoning was fallacious – that one can't conclude that the between-group difference has a genetic component just from the fact that there is a genetic explanation of within-group variation. Jensen (1972) replied that his reasoning committed no such fallacy.

I now want to make a point that is specifically about the concept of heritability. High heritability does not imply that a trait is difficult or impossible to manipulate by changing the environment – an especially important point, because many individuals who have argued that IQ is highly heritable have inferred from this that it is futile to look for environmental interventions that boost IQ. One reason this is a fallacy connects with a point I made before. Because ANOVA is specific to a given range of environments, nothing follows about how the trait would respond to a *new* environmental variable, one not covered in the initial analysis. For example, before the invention of eyeglasses, poor vision was highly heritable. However, this did not mean that environmental interventions were bound to fail. As it turned out, eyeglasses have done wonders. In effect, this invention created a new environment, one that had a profound effect on the ability to see (Goldberger 1979).

Another example of this sort is provided by PKU disease – phenylketonuria. Individuals with two copies of the relevant recessive gene are unable to digest phenylalanine. The result of accumulating this substance is a severe retardation. However, if individuals with the genetic condition are at birth placed on special diets that don't include phenylalanine, they develop normally. Before the disease was diagnosed and all individuals had diets that contained phenylalanine, the retardation was completely heritable; all the variation in phenotype was explained by variation in genes. Once the disease was understood, a new environment was created that had a profound effect on the phenotype. Notice that in this example understanding the *genetic* causes of a condition opened the door for constructing a new *environmental* manipulation.

There is an additional reason why high heritability does not imply that environmental change will make little difference in the resulting phenotype. Even when we consider just the environments analyzed in

an ANOVA study, high heritability does not mean that shifting a particular person from one environment studied to another will have little effect. Consider the following hypothetical data set:

	G1	G2	G3	G4	
E1	1	2	3	4	1.75
E2	1	2	3	2	2
E3	1	2	3	3	2.25
	1	2	3	2	2

The phenotype described here is highly heritable; $V_p = 2/3$ and $V_g = 1/2$, so $h^2 = V_g/V_p = 0.75$. Notice that individuals with genotypes *G*1, *G*2, and *G*3 do not exhibit different phenotypes when their environments are changed. However, matters are quite different for genotype *G*4. It would be a mistake to infer that environmental manipulation has little effect on individuals with genotype *G*4 on the ground that the trait is highly heritable.

The reason that heritability provides little guidance about the effects that environmental change will have on particular individuals is that heritability is a summary statistic about the whole population; it distills the *n*-times-*m* pieces of data in an ANOVA table into a single number. A much better guide to the issue of malleability is provided by the data in the table itself. If you know an individual's genotype, you can look down the relevant column in the ANOVA table and obtain an estimate of how changing the environment will produce changes in phenotype, within the range of environments considered. This information is sometimes presented graphically, as in Figure 3.1, by plotting, genotype by genotype, how an environmental circumstance induces a particular phenotypic condition. These graphs represent what is called the genotype's *norm of reaction*.

Graphing norms of reaction provides a handy way of summarizing some of the concepts we have already described in ANOVA. Imagine an ANOVA experiment in which genotypes *G*1 and *G*2 are tested in a range of environmental settings. We can predict the outcome of this experiment from knowledge of the genotypes' underlying norms of reaction. Graph (a) in Figure 3.1 depicts a situation in which there is no genetic variance; all the phenotypic variation will be due to environmental variation. In (b) we have the opposite situation. Environmental variation will make no difference; all the variation will come from the difference

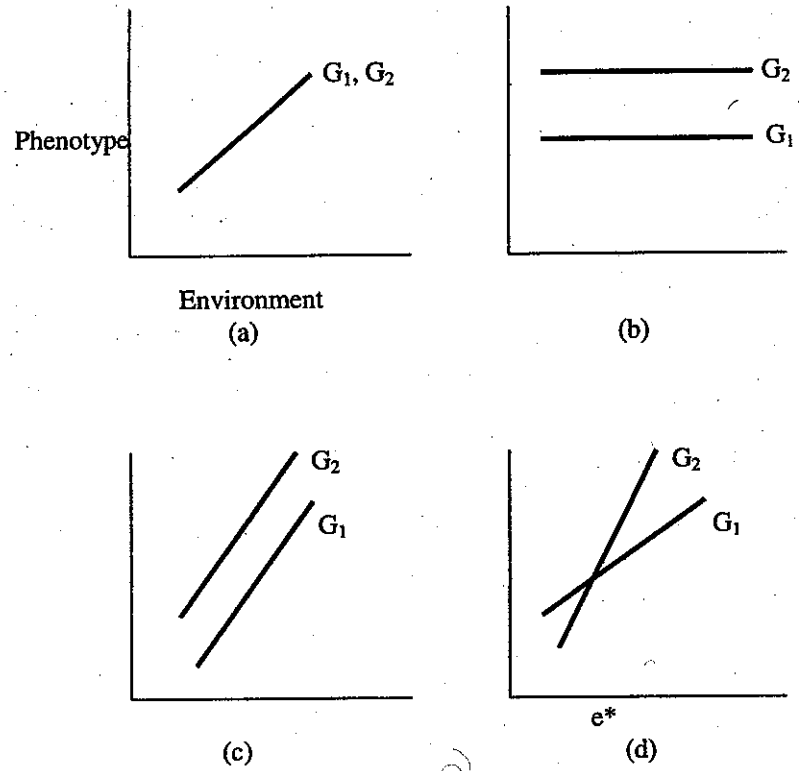


Figure 3.1. Hypothetical norms of reaction for two genotypes, *G*1 and *G*2.

in genotype. In (c), both genetic and environmental variations make a difference; because the norms of reaction are straight lines running in parallel, the situation is additive. In (d), the phenotypic variation will be due to genetic variation, environmental variation, and gene-environmental interaction. Whether *G*1 usually has a higher phenotypic score than *G*2, or the reverse is true, or they come out with the same average performance, will depend on how the environments investigated in the experiment are selected. If they come entirely from points to the left of e^*, *G*1 will do better; if they come entirely from points to the right, *G*2 will do better; and if they are equally spaced around e^*, then there will be a tie.

In an ANOVA experiment, it need not be true that all the individuals in the same environmental treatment experience *identical* environments. There is no way to ensure such uniformity. The plants in the corn experiment described earlier may be unequal distances from the barn; air currents in the experimental plots may differ slightly in direction and speed. No ANOVA study could be performed if perfect uniformity were required within environmental treatments. It might be thought that experimenters are able to attain the more limited goal of having plants within the same experimental treatment differ only in ways that have no effect on the phenotype under study. But here again, the demand is too strong. If we are studying the impact of fertilizer on height, we might set up the experiment so that each experimental plot receives the same amount of water. If there are 100 plants in a plot, however, it is practically certain that some plants will get a bit more water than others. If we are studying the effect of fertilizer, then what is required is that variation in *other* environmental factors that affect the phenotype be the same across plots.

It is at this point that theoretical understanding impinges on an analytic technique that in many ways seeks to get by without the backing of theory. I mentioned before that an ANOVA investigation can proceed without an understanding of the developmental processes that lead individual organisms to end up with their phenotypes. However, there is another type of causal information that is not so dispensable. You must know whether differences in environmental treatments besides the one under study are causally inconsequential for the phenotype under study.

There is no precise level of theoretical understanding that the experiment must satisfy – no cutoff point that marks the boundary between adequate and inadequate studies. Rather, the relevant consideration needs to be stated as a matter of degree. The less you know about which environmental factors influence the phenotype besides the environmental manipulation under study, the less certain you can be that your study has the proper controls in place. Far more is known about the environmental factors that influence the height attained by corn plants than is known about the environmental factors that influence the IQs attained by human beings. Because of this difference, we must be more circumspect in the human case when we say that two individuals grew up in environments that were "the same."

Symmetrical points apply to the concept of a genetic treatment in ANOVA. In an ANOVA experiment individuals said to have "the same" genetic characteristic need not be genetically *identical*. Although desirable, and characteristic of studies that focus exclusively on identical twins, the logic of ANOVA experiments does not demand this.

Individuals have many genes. Suppose *G*1, *G*2, . . . , *G*m are alterna-

tive states of a single gene, and we are doing an ANOVA experiment in which these *G* conditions constitute the genetic variable under study. Individuals alike in the *G* trait they have may differ with respect to other genetic characteristics. Individuals who are all *G*2 may differ in whether they have *H*1 or *H*2, whether they have *J*1 or *J*2, and so on. Some of these other genes may affect the phenotype under study. What is required is that the individuals in one genetic treatment have the same *distribution* of other genetic traits as the individuals in the other genetic treatments, for all the *other* genetic traits that may affect the phenotype.

In this section, I have made several points about how one goes about interpreting the components of variance inferred from an ANOVA experiment. These points do not show that ANOVA is useless as a device for apportioning causal responsibility; rather, the message is that ANOVA must be understood for what it is. ANOVA describes how much of the observed phenotypic variation is due to environmental variation and how much to genetic variation, for the phenotype considered, the range of environmental and genetic variation considered, and the population at hand.

ESTIMATING COMPONENTS OF VARIANCE

In the ANOVA procedure explained so far, investigators specify the various genetic and environmental traits they want to study. Once these are identified, the procedure is to find one or more individuals in each of the *m*-by-*n* treatment cells. To perform this type of study, we must know, not just the phenotype of each individual, but its genetic and environmental characteristics.

If we are interested in a phenotype as prosaic as height or weight, which genetic traits should we examine? If we already know that variation in *G*1, *G*2, . . . *G*m influences height, or that variation in *H*1, *H*2, . . . , *H*m does not, there is no point in doing a study that will tell us what we already know. On the other hand, if we don't know which genes matter, how should we proceed? Do we simply move arbitrarily from one array of genes to another, testing whether variation in a randomly selected gene helps explain variation in height? The same question arises with respect to environmental influences on height; if we know that an environmental factor matters, or that it does not, we will learn little[8] from an ANOVA study. On the other hand, if we focus only on environmental factors whose role is unknown, how do we decide which traits in that

infinity of possibilities are worth studying? Naive empiricism is not a recipe for efficient inquiry. What alternative strategy will do better?

There are two standard alternatives to this strategy of hit-or-miss. The first is to examine identical twins reared apart. The second is to compare identical twins and fraternal twins, both reared by their biological parents. In these studies, one of course needs to decide which phenotype one wishes to study. But having decided this, one does not need to further specify which environmental variables and which genetic traits one wishes to consider. The reason is that these studies are designed to estimate the heritability of the trait with respect to the full range of genetic and environmental variation found in the population as a whole. This convenience, though enormous, also has its price. One need not know which genes affect the phenotype in question to do a heritability study, but the upshot is that the study does not tell you which genes make a difference. Although heritability is defined in terms of concepts drawn from ANOVA, there is a big difference between the way causal variables are treated in ANOVA studies and the way they are treated in heritability studies.

Let's now consider how heritability in the whole population is estimated by examining pairs of identical (i.e., monozygotic) twins who were "reared apart," meaning that the twins were separated at birth or shortly thereafter and raised in different environments. This procedure attempts to ascertain the genetic variance and the environmental variance of a trait in a population by looking at a very special subpopulation. Here is an approximate statement of the idea behind such studies: monozygotic twins have exactly the same genes. If they are reared apart, and if they differ from each other phenotypically, then this difference can be attributed entirely to environmental causes. If it turns out that these twins usually are more similar to each other than are two randomly selected individuals from the population, we can conclude that there is a genetic cause of the phenotypic variation in the population as a whole.

Let us now examine this line of reasoning with more care. No new mathematical concept will be introduced, and the only mathematics that will occur in what follows is some subtracting and replacing of equals with equals. I'll discuss the different types of variance that were explained before – the quantities V_p, V_g, V_e, and I, which describe the population as a whole. In addition, I'll talk about the quantities V_p(mono-twin), V_g(mono-twin), V_e(mono-twin), and I(mono-twin),

which characterize the subpopulation of monozygotic twins reared apart.

Our goal is to infer the genetic and environmental variances represented in proposition (1). Let's begin by reminding ourselves of what we actually *observe* in such twin studies and what we must *infer* from our observations. Suppose the phenotype of interest is height. What we observe is how tall various people are in the whole population and how tall pairs of monozygotic twins are. From these numbers we deduce values for V_p and of V_p(mono-twin). V_p(mono-twin) describes how much difference there is in height in the average pair of twins. These two phenotypic variances are quantities that we know by observation. In addition, we know that monozygotic twins are genetically identical. We want to use this information to infer what the relative contributions are of genes, environment, and gene-environment interaction. As stated, the structure of this problem should be puzzling. Proposition (1) tells us that an observational quantity, V_p, is the sum of three quantities whose values we do not know by direct observation. How are we to infer three theoretical quantities from the observed phenotypic variance? It looks as if there are too few equations and too many unknowns.

We examine twins reared apart as a device for solving this problem about the full population. Although proposition (1) describes what is true in the population as a whole, the same set of relationships obtains within the subpopulation composed of the monozygotic twins in our study:

$$V_p(\text{mono-twin}) = V_g(\text{mono-twin}) + V_e(\text{mono-twin}) + I(\text{mono-twin}). \quad (3)$$

Notice that (3) by itself does not allow you to assign a value to V_e(mono-twin), based on the observed value of V_p(mono-twin). However, we should take note of the fact that

$$V_g(\text{mono-twin}) = 0. \quad (4)$$

There is no genetic variation within pairs of monozygotic twins. Propositions (3) and (4) allow us to deduce that

$$V_p(\text{mono-twin}) = V_e(\text{mono-twin}) + I(\text{mono-twin}). \quad (5)$$

Let us now assume that the relationship between gene and environment in this subpopulation of twins is additive – that is, that the interaction term is 0:

$$I(\text{mono-twin}) = 0. \quad (6)^*$$

Propositions (5) and (6) entail that

$$V_p(\text{mono-twin}) = V_e(\text{mono-twin}) \quad (7)$$

We now have solved for one of the theoretical "unknowns" in proposition (3). We began by knowing only what the twins' phenotypic variance is; we now are able to assign a value to the environmental variance that exists within the twin population.

How do we use this result to estimate genetic and environmental variances in the whole population? We begin by making another assumption – that the variances are purely additive in the whole population:

$$I = 0. \quad (8)^*$$

Proposition (8) combines with proposition (1) to yield:

$$V_p = V_g + V_e. \quad (9)$$

We now further assume that twins reared apart live in environments that tend to be just as varied as the environments occupied by two randomly selected individuals in the whole population:

$$V_e(\text{mono-twin}) = V_e. \quad (10)^*$$

Proposition (9) combines with (10) to yield:

$$V_p = V_g + V_e(\text{mono-twin}). \quad (11)$$

Propositions (11) and (7) entail that

$$V_p = V_g + V_p(\text{mono-twin}). \quad (12)$$

Proposition (12) rearranges to yield

$$V_g = V_p - V_p(\text{mono-twin}) \quad (13)$$

and (13) and (9) imply that

$$V_e = V_p(\text{mono-twin}). \quad (14)$$

We are done. The last two equations tell us that the values for the genetic and the environmental variances in the whole population are identical with phenotypic quantities that we can measure by observation. All we have to do is find out how height varies in the population as a whole and how it varies in the subpopulation of monozygotic twins reared apart, and we can calculate what the underlying theoretical quantities are.

As the reader may have guessed, I have placed asterisks besides two crucial assumptions that are used in this derivation. Propositions (6) and

(8) assert that the system is additive; proposition (10) says that twins reared apart tend to live in environments that are just as similar as the environments of two randomly selected people in the population. Both these assumptions are open to question.

Recall what the assumption of additivity means. It means that shifting from one environment to another "adds" the same change in phenotypic value for all individuals, regardless of their genotype. It is not hard to see why this condition can fail. Again, let us consider the example of height. Nutrition in early childhood affects height. But is it plausible to think that shifting from 1,500 calories to 1,600 calories per day will have the same effect on everyone's height, regardless of what their genotype is? This could easily fail to be true. Individuals of different genotype differ in their metabolism; genotypes therefore may differ in how efficiently they convert additional calories into additional height. Notice that the example of height is by no means unusual; just as additivity cannot be assumed a priori for a trait such as height, neither can it be assumed for psychological traits such as intelligence or propensity to violence. The claim of additivity is an empirical claim, and must be supported by evidence.[9]

I am not trying to propose an a priori argument for the claim that additivity always fails to obtain. There can be no a priori argument on this question, one way or the other. Maybe some traits in some populations are additive. Figure 3.2 illustrates the fact that height fails to be additive in a population of the plant *Achillea millefolium*. My main point is to warn against the idea that it is somehow a "safe" general assumption that the trait one is studying is additive.

The assumption that $V_e = V_e$(mono-twin) is also problematic, but for a reason that derives from the specifics of how human adoption agencies work, not from general biological considerations. When monozygotic twins are separated from each other, how are their new environments selected? Sometimes they are adopted into the homes of relatives. At other times, an adoption agency places one or both into a new home. It is well known that adoption agencies give strong preference to adults seeking to adopt who have high socioeconomic status. Both these considerations suggest that V_e(mono-twin) will be smaller than V_e, but by an amount that is difficult to estimate.

The assumptions I have singled out for criticism are very common ones in twin studies; however, the derivation I have described does not depend absolutely on their being true. By this, I mean not that the assumptions could simply be *removed* and the derivation would still go

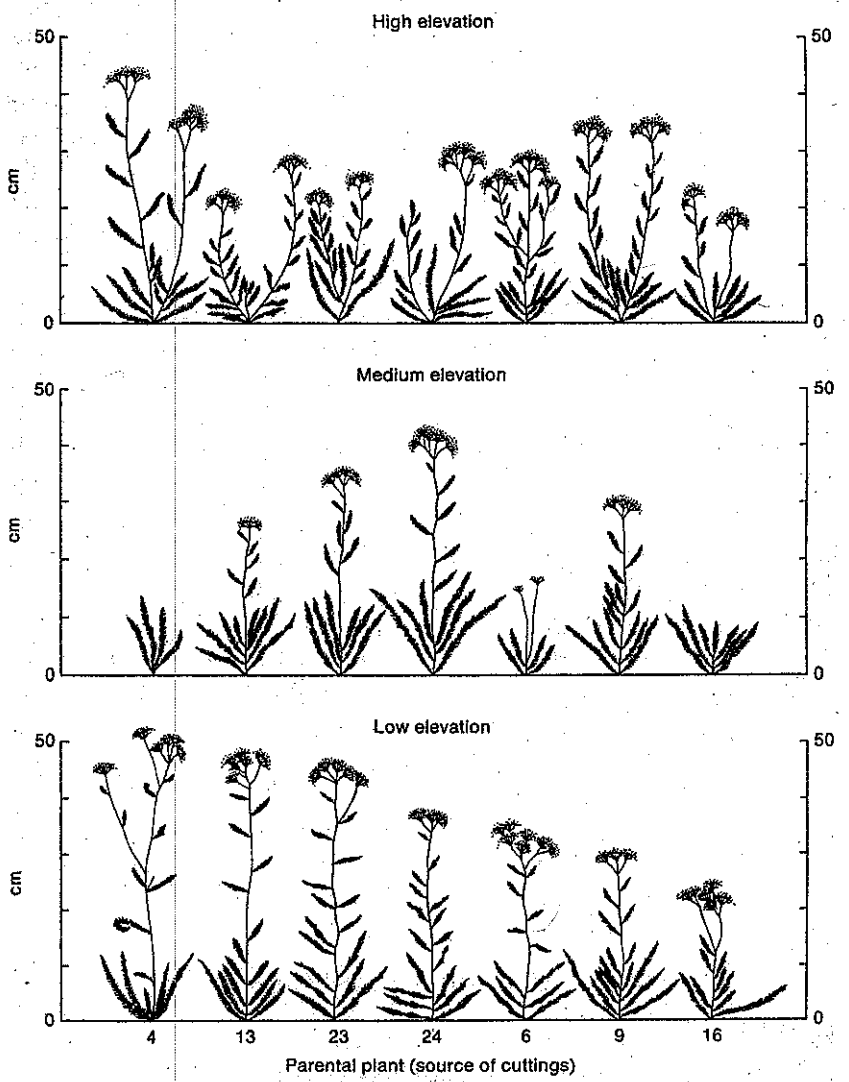


Figure 3.2. Empirically determined norms of reaction for seven genotypes of the plant *Achillea millefolium* grown at three different elevations (reprinted from Griffiths et al. 1993, 14). From *An introduction to genetic analysis* by Griffiths et al. © 1996, 1993, 1989, 1986, 1981, 1976 by W. H. Freeman and Company. Used with permission.

through but that the assumptions could be *replaced* by other assumptions that are equally substantive. The assumption of additivity could be replaced by assigning a value other than zero to the interaction term. The same holds for the assumption that $V_e = V_e$(mono-twin); this could be excised from the argument and replaced by some other characterization of how the two environmental variances are related. However, the problem is not solved by *stipulating* a couple of new assumptions; rather, it must be demonstrated empirically that the assumptions are plausible for the case under analysis (Layzer 1972).

I now want to consider a second popular methodology for inferring genetic and environmental variances. Instead of looking at monozygotic twins reared apart, one compares fraternal (dizygotic) twins reared together with monozygotic twins reared together. The rough idea in this inference procedure is as follows: monozygotic twins have all their genes in common, whereas fraternal twins on average share half their genes. It is taken to follow that if monozygotic twins are more similar to each other than dizygotic twins, then this difference furnishes a valid estimate of the genetic contribution to the phenotypic variance in the population as a whole.

Let us now consider the details. As before, we assume that the interaction term for monozygotic twins reared together is zero; this, and the fact that V_g(mono-twin) = 0, allow us to write

$$V_p(\text{mono-twin}) = V_e(\text{mono-twin}). \qquad (15)$$

Because V_p(mono-twin) is a quantity we can observe, proposition (15) tells us what value we should assign to the underlying parameter V_e(mono-twin). If we assume that the interaction term is zero for dizygotic twins as well, then

$$V_p(\text{di-twin}) = V_g(\text{di-twin}) + V_e(\text{di-twin}) + I(\text{di-twin}).$$

reduces to

$$V_p(\text{di-twin}) = V_g(\text{di-twin}) + V_e(\text{di-twin}). \qquad (16)$$

The question is how we can use (15) and (16), which concern subpopulations of twins, to draw conclusions about the components of variance in the population as a whole.

We now introduce a further assumption – that the environments experienced by monozygotic twins reared together are just as similar as the environments experienced by dizygotic twins reared together:

$$V_e(\text{mono-twin}) = V_e(\text{di-twin}). \qquad (17)^*$$

Propositions (16) and (17) entail that

$$V_p(\text{di-twin}) = V_g(\text{di-twin}) + V_e(\text{mono-twin}). \qquad (18)$$

If we combine (15) and (18), we obtain

$$V_p(\text{di-twin}) = V_g(\text{di-twin}) + V_p(\text{mono-twin}), \qquad (19)$$

which rearranges to yield

$$V_g(\text{di-twin}) = V_p(\text{di-twin}) - V_p(\text{mono-twin}). \qquad (20)$$

Because the two phenotypic variances in (20) are observable, (20) tells us what value to assign to the genetic variance for dizygotic twins. We now need to appeal to a genetic claim – that dizygotic twins differ genetically from each other on average half as much as do randomly chosen individuals from the population at large:

$$V_g(\text{di-twin}) = (1/2)V_g. \qquad (21)^*$$

Propositions (20) and (21) combine to yield a formula that shows how the genetic variance in the whole population can be computed from the difference between the observable phenotypic variances for the two classes of twins:

$$V_g = 2[V_p(\text{di-twin}) - V_p(\text{mono-twin})]. \qquad (22)$$

Finally, (22) combines with (9) to yield a formula for estimating the environmental variance from the observed phenotypic variances:

$$V_e = V_p - 2[V_p(\text{di-twin}) - V_p(\text{mono-twin})]. \qquad (23)$$

In addition to the familiar assumption of additivity, propositions (17) and (21) are noteworthy. Proposition (21) depends on the assumption of random mating in the parental generation; this will often be false. For example, for many phenotypic traits, similar individuals tend to pair up to have children; this is true for height, socioeconomic status, IQ, and so on. The effect of assortative mating is to make dizygotic twins more than twice as genetically similar as pairs of individuals drawn at random; how much more depends on the intensity of the assortative process.

Another complication that affects the assessment of proposition (21) is that dizygotic twinning occurs more frequently in some genotypes than it does in others (Falconer 1981, 160). This is a problem, because the basic idea of twin studies is that the population of twins provides a representative sample of the genetic composition of the population as a whole.

Another problem arises in connection with proposition (17). For many psychological traits, it is questionable whether the environments

of identical twins reared together are just as similar, on average, as the environments of fraternal twins reared together. Perhaps parents treat identical twins more similarly than they treat fraternal twins. For example, identical twins have the same sex, but fraternal twins often do not, and this may affect the way the twins are reared.[10]

There is an observation that throws light on this question. We know by observation that fraternal twins reared together are more similar in IQ than are nontwin siblings reared together (Plomin and Fries 1980). Because fraternal twins and nontwin sibs have the same degree of genetic similarity, it follows (if the system is additive) that fraternal twins experience more similar environments than do nontwin sibs. Of course, we cannot deduce from this that proposition (17) is incorrect; comparing fraternal and nontwin sibs is not the same as comparing fraternal and identical twins. Still, it would be naive simply to assume that (17) is correct.

Just as was true in the discussion of identical twins reared apart, the point about the present procedure is not that (17) and (21) are essential; other specific assumptions could be substituted for them and would allow the derivation to go through. Rather, the point is that assumptions of these types are needed, and must be defended empirically, before estimates of heritability can be obtained by comparing identical and fraternal twins.

In this section, I have emphasized the types of assumptions that need to be made in using data from twin studies to estimate heritability. However, it is well to remember, in addition, how little one would know, even if these assumptions were entirely correct. In a controlled ANOVA study in which each individual is measured for its environment, its genotype, and its phenotype, one can say which genes make a difference and how much difference they make; one also can say which environmental treatments raise phenotypic scores and which lower them. To be sure, the developmental processes that link causes to effects remain opaque. But at least one knows, from such a study, something about the identity of the causes.

None of this information is provided by a twin study that estimates heritability, even when that study is methodologically sound. If we infer that $h^2 = 0.6$, we have no idea which genes make a difference to the phenotype in question. Nor do we know which environmental variables are responsible for the fact that V_e/V_p has the value it does. What one knows is that certain *existence claims* are correct. There exist genetic differences that help explain phenotypic differences and there exist environmental differences that do the same thing. The fact that one can assign numbers to these causal contributions should not obscure the fact that these estimates say very little.

GENE-ENVIRONMENT CORRELATION

In my explanation of ANOVA, I assumed that each of the m-by-n treatment cells contains the same number of individuals. This setup is ideal in a controlled experiment, but the real world of natural populations rarely conforms to this tidy arrangement. In reality, there often are correlations between the genes that individuals have and the environments they tend to occupy. In this section, I want to explain how this fact further complicates the task of estimating heritability.

Let's begin with a simple example that illustrates how gene-environment correlation can affect the total phenotypic variance. Consider the following data that are drawn from a study of four hundred individuals.

	G1	G2
E1	1	3
E2	3	5

If there are 100 individuals in each of the four treatment cells, then the phenotypic variance is 2. But now suppose that 199 individuals are in the upper left cell and 199 are in the lower right cell; the phenotypic variance now has a value of approximately 4. On the other hand, if 199 individuals are in the upper right cell and 199 are in the lower left, the phenotypic variance will be close to 0.

In this example, individuals in $E2$ have a higher phenotypic value than individuals in $E1$; and individuals with $G2$ have a higher phenotypic value than individuals with $G1$. If higher-valued genotypes tend to occur in higher-valued environments, we have a *positive* association of genes and environment; this association tends to boost the total phenotypic variance. On the other hand if higher-valued genotypes tend to occur in lower-valued environments, there is a *negative* association of genes and environment, which tends to reduce the total phenotypic variance.

What this means is that proposition (1) provides an incomplete list of the possible sources of phenotypic variance. Proposition (1) said that

$$V_p = V_g + V_e + I.$$

It should be replaced with the following:

$$V_P = V_g + V_e + I + 2Cov(g,e). \qquad (1)^*$$

Cov(g,e) is the gene-environment *covariance.* The covariance, which measures the strength of association of genotypes and environment, ranges from -1 to +1.

How does this complication affect the procedures for estimating heritability reviewed in the previous section? Just as twin studies often assume that the interaction term $I = 0$, they also often assume that there is no correlation between genes and environment. The point made earlier about interaction applies here as well. One cannot simply assume that the covariance is zero; one must estimate it empirically.

With respect to traits such as IQ, the gene-environment covariance is known to be positive (Falconer 1981, 121). Individuals with favorable genes tend to grow up in favorable environments, due to the fact that parents not only pass their genes along to their children but do a great deal to structure the environments in which the children are reared. However, if our goal is to estimate heritability, knowing that the covariance term is positive is not enough. We must be able to estimate its value. This requires a type of theoretical understanding that the simple data drawn from twin studies do not provide.

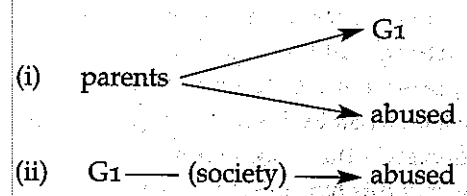
The so-called question of "nature" versus "nurture" or of "genes" versus "environment" suggests that the inferential problem involves saying how important *two* possible causes are. The idea that there are just two causes allows one to think that the contribution of genes can be viewed as a *remainder;* it is what is left unexplained by environmental factors. Proposition (1)* shows that this suggestion is doubly misleading and has the effect of inflating one's estimate of the importance of genes. Only when the interaction and the covariance terms are both zero will genetic variation explain everything that environmental variation fails to explain.

THE CONCEPT OF ENVIRONMENT

Earlier in this chapter, I adopted the rough-and-ready idea that an environmental factor is any property of the world that depends just on what is going on outside the organism's skin. The fertilizer treatment in the farmer's corn plant experiment was judged an environmental factor for just this reason. I now want to explain how quantitative geneticists assign a narrower meaning to the idea of an "environmental factor."

I begin with a simple example, due to Jencks et al. (1972). Suppose it is true in a given population of human beings that red-haired individuals have lower IQ scores than people with other hair colors and that this is true solely because they are treated badly when they are young. It may seem to follow from this description that the lower IQ is due to environment, not genes. For quantitative geneticists, this conclusion is not so straightforward.

I said that redheads have lower IQs because of how they are treated, but I did not say *who* treats them badly. Let's consider two scenarios. In the first, redheads tend to be born to parents who treat their children in ways that reduce their children's IQs. These parents, let us suppose, abuse *all* their children, regardless of what hair color the children happen to have. The point is that red-headed children are born disproportionately into such families. In the second scenario, society as a whole treats redheads in ways that reduce their IQ scores. The parents of redheads are not especially abusive toward their offspring; rather, society treats redheads badly. These two scenarios may be schematized as follows:



The difference between these scenarios is that genotype causes an individual to be abused in (ii), but not in (i). In (ii), society abuses redheads *because* they are red-headed, and redheads have red hair *because* they have genotype *G*1; in (i), it is not true that parents of redheads abuse their children *because* the children have red hair.

In quantitative genetics, abuse counts as an "environmental" factor if it is provided by parents but not if it is provided by society generally. Notice that an adoption study will have quite different results in these two cases. If redheads are removed from their parents and placed in new homes, then they will suffer no deficit in IQ, if (i) is true. However, if scenario (ii) is in place, adoption will make no difference; redheads will continue to show lower IQs.

Quantitative geneticists do not regard abuse as an environmental factor, if the situation is of type (ii). The lower IQ of redheads in scenario (ii) is said to be genetic, rather than environmental, on the grounds that individuals experience abuse because of their genes (Falconer 1981). An environmental factor is not just something that occurs outside the or-

ganism's own skin; in addition, an environmental factor is defined as a factor that is not caused by genes.

Let's see how this mode of description applies to other examples. Suppose we observe that the women in a given population like to knit more than the men do. The question may then be posed of whether this pattern is due to the genetic difference between the sexes or to differences in how boys and girls are reared (or to both). One might have thought that the only way to separate nature and nurture in this case is to see what happens to *XX* individuals who are raised as boys and to *XY* individuals who are raised as girls. Only by breaking the correlation between genotype and rearing environment can nature and nurture be disentangled. This assessment is not correct, according to the practice of quantitative geneticists, if society tends to treat *XX* individuals one way and *XY* individuals another. If the causal relationships conform to pattern (ii), the correlation need not be broken. The conclusion will then be drawn that there is a genetic explanation for why women like to knit more than men.[11]

I hope the reader can infer from this discussion how societal racism will be classified in the quantitative geneticist's separation of genes from environment. Suppose that the difference in IQ scores between blacks and whites in the United States is entirely due to the fact that the United States is a racist society – that people treat blacks worse than whites. It might be thought that this hypothesis constitutes a purely environmentalist explanation of the IQ difference. Actually, this is not correct, according to the standard framework of quantitative genetics. If blacks are badly treated because of their skin color, and their skin color is genetic, then the lower IQ will be assigned to genes, not to environment.

It may be a bit surprising that quantitative geneticists use the terms "gene" and "environment" in this way. However, as odd as this usage may seem, it is not hard to understand why quantitative geneticists feel driven to adopt it. Every gene has the effect it does because of the ambient environment in which it acts. The so-called gene for eye color produces blue or brown eyes only as a result of the somatic environment in which it acts. To say that eye color is genetic means that genetic variation at a particular locus, *given the environment in which those genes exist*, causes variation in eye color.

The same point can be illustrated by returning to the farmer's two-by-two experiment in which fertilizer is the environmental variable. Suppose that *G*1 has a positive effect on plant height because *G*1 produces more of a particular gene product, one that repels an insect pest

that happens to be present in the fields in which the experiment is performed. *G*1 plants grow taller than *G*2 plants because *G*1 plants are treated differently by the insect pests. It is not an objection to the claim that genes make a difference in the determination of plant height that *G*1 and *G*2 make the difference they do only because the insect pest is present. Admittedly, it is quite possible that if a different pest had been present, the very opposite result would have transpired; maybe a different pest would have been *attracted* by larger quantities of the gene product.

In the plant example, we view the insect pest as a background condition and tend to focus on the difference in genotype as the genuine cause. In the hair color example, we view an individual's hair color as given, and tend to focus on societal abuse of redheads as the genuine cause. It is an interesting psychological question why we find some causal factors more salient than others. But from the point of view of quantitative genetics, an environmental factor is one that is not caused by genes. Abuse of redheads can arise by either pathway (i) or (ii). When (i) obtains, abuse counts as an environmental condition; when (ii) is in place, abuse of redheads is not part of the environment but is an effect of genes. Without this convention, it is hard to see how quantitative geneticists would be able to say that genes *ever* have any effect.

Quantitative geneticists differ from the rest of us in the way they tend to use the term "environment," and this difference in usage will probably persist. This means that when quantitative geneticists say that the variation in some phenotype has a genetic component, the rest of us must be very careful. The reason is that a genetic cause, in the quantitative genetics sense, may be what the rest of us would regard as an environmental cause. As I explained earlier, a so-called genetic cause may be changed just by changing the environment. If societal abuse of redheads, women, or blacks is changed, the "genetic" causes of the resulting phenotypes may entirely disappear.

CONCLUDING REMARKS

The criticisms I made of the methodology used in twin studies does not mean that such studies will never underwrite reasonable inferences about components of variance and heritability. As we learn more about the issue of gene-environment interaction, and as we learn more about the ways in which environmental factors influence various phenotypic traits, the sophistication of twin studies will improve. The criticisms reg-

istered here apply only to a certain "naive," though pervasive, approach in twin studies; these studies rest on assumptions that are nothing more than assumptions. However, there is no reason in principle why a groundless assumption should not someday be replaced by a statement that is empirically well attested. The new knowledge that will facilitate this improvement in twin studies will not come from surveys, which is what twin studies are, but from theoretical work on how genes and environment work together to produce phenotypes.

The points I made about interpreting ANOVA and heritability have a quite different status. For example, I have emphasized that high heritability of a trait does not mean that it cannot be altered much by the environment. I also explained why we must be careful to separate the task of explaining phenotypic variation *within* groups from the task of explaining variation *between* groups. These points will not disappear once we learn more about genetics. They are permanent features of the conceptual landscape of heritability and ANOVA.

As we learn more about genetics, what will happen to questions about heritability? I have explained how ANOVA studies do not depend on understanding the developmental processes whereby genes and environment conspire to produce phenotypes. The farmer can examine the average height of corn plants in his *m*-by-*n* experiment and obtain a heritability estimate without knowing why one unit of fertilizer and genotype *G*1 combine to produce plants that average one unit in height. Twin studies, which are stopgap measures that scientists use when they cannot manipulate organisms or identify the relevant genotypes that they want to study, also attempt to draw conclusions with little or no information about developmental processes. My suspicion is that as we learn more about these developmental issues, questions about heritability will become increasingly marginalized. For example, why would a population statistic about the heritability of IQ matter to us, if we understood why some interventions in the lives of individual children boost their IQs while others do not? It is already perfectly clear, as a conceptual point, that high heritability does not mean that a trait cannot be modified much by environmental change. I suggest that as we learn more about the norms of reaction of different genotypes, and *why* genotypes differ from each other in their norms of reaction, we will come to care less and less about assigning a number to a trait's heritability. Ironically, the reason that heritability studies will suffer this displacement is not that science will become convinced that radical environmentalism is true. It isn't that we will lose interest in h^2 because we think its value

is lower than we thought before. Rather, heritability will move to the periphery of scientific interest as we learn more about the details of genetic processes. The real challenge to quantitative genetics is not the advocate of nurture over nature, but the developmental geneticist who provides insights into underlying processes.

NOTES

1. Here and in what follows, I am ignoring the way in which statistical inference enters into the interpretation of ANOVA data. One must ask whether the pattern of variation is due to the factors considered or should be attributed to chance. This depends on the number of individuals and on the amount of variation that there is *within* treatment cells. For purposes of getting clear on the bare bones of ANOVA, however, assume that there is very little variation within cells and that the cells each contain a large number of individuals.
2. The philosophical concept of supervenience plays a useful role here. Phenotypic traits supervene just on what is going on inside the skin; environmental traits supervene just on what is going on outside.
3. This proposal goes against the quite reasonable idea that the web a spider weaves is part of its phenotype. However, the distinction of phenotype and environment we are proposing is a convenient one for understanding the basics of ANOVA, and that is all that matters here.
4. There are other components of variance, which will be discussed in due course.
5. This is the so-called broad heritability. In some contexts, it is important to decompose broad heritability into a sum of terms, one of which is the so-called narrow heritability. The narrow concept will not be relevant to our discussion.
6. In the subsequent discussion, we will move from three causes to four, by introducing the concept of gene-environment correlation.
7. There are other contexts in biology in which a population configuration is explained without explaining why individual organisms have the traits they do. I discuss this pattern in connection with the concept of natural selection in Sober (1995).
8. Even if you know that a genetic or an environmental factor positively affects some phenotype, an ANOVA study will add at least some information: it will tell you how much each of them matters, relative to the other. However, the larger question of how important these two factors are, as compared with other traits that were not investigated, remains open.
9. More precisely, the additivity thesis must be *tested* statistically. The question is whether relevant data deviate sufficiently from the predictions of additivity for this to justify rejecting the additivity hypothesis. "Sufficient deviation" is defined to reflect both the size of the difference between predicted and observed values and the sample size.
10. Even when the dizygotic twins considered are of the same sex, there may be

other reasons why monozygotic twins tend to share environments that differ in their degree of similarity from those occupied by dizygotic twins – for example, parents may treat identical twins more similarly (or may encourage differences).

11. This example is due to Dawkins 1972; I discuss it in Sober 1984.

REFERENCES

Block, N., and Dworkin, G. (1974). IQ, heritability, and inequality. *Philosophy and Public Affairs* 3: 331–409, 4: 40–99. Reprinted in Block and Dworkin 1976, 410–540.

(1976). *The IQ controversy.* New York: Pantheon.

Dawkins, R. (1982). *The extended phenotype.* San Francisco: W. H. Freeman.

Falconer, D. (1981). *Introduction to quantitative genetics.* London: Longman.

Goldberger, A. (1979). Heritability. *Econometrica* 46: 327–347.

Griffiths, A., Miller, J. Suzuki, D., Lewontin, R., and Gelbart, W. (1993). *An introduction to genetic analysis.* New York: Freeman.

Jencks, C., Smith, M., Acland, H., Bane, M., Cohen, D., Gintis, H., Heyns, B., and Michelson, S. (1972). *Inequality: A reassessment of the effect of family and schooling in America.* New York: Basic Books.

Jensen, A. (1970). How much can we boost IQ and scholastic achievement? *Harvard Education Review* 39: 1–123.

(1972). Race and genetics of intelligence – a reply to Lewontin. *Bulletin of the Atomic Scientists* (May). Reprinted in Block and Dworkin 1976, 93–106.

Layzer, D. (1972). Science or superstition: A physicist looks at the IQ controversy. *Cognition* 1: 265–300. Reprinted in Block and Dworkin 1976, 194–241.

Lewontin, R. (1970). Race and intelligence. *Bulletin of the Atomic Scientists* (March): 2–8. Reprinted in Block and Dworkin 1976, 78–92.

Plomin, R., and Fries, J. (1980). Genetics and intelligence: Recent data. *Intelligence* 4: 15–24.

Sober, E. (1984). *The nature of selection.* Cambridge, Mass.: MIT Press.

(1994). Apportioning causal responsibility. In *From a biological point of view,* 184–200. Cambridge: Cambridge University Press.

(1995). Natural selection and distinctive explanation. *British Journal for the Philosophy of Science* 46: 384–398.

Chapter 4

Genetic Explanations of Behavior: Of Worms, Flies, and Men

KENNETH F. SCHAFFNER

INTRODUCTION

For several hundred years scientists and philosophers have speculated about the character and scope of explanations of a particular type, where that type might be, for example, "mechanical," "chemical," "electrical," and, more recently, "biochemical," "molecular," "selectional," "developmental," "adaptational," and "genetic."[1] Extensive discussions about "mechanical" explanations are an interesting and useful analogous problem to the one that is the subject of the present chapter, especially in the context of nineteenth-century debates about the reach of mechanics, because it was thought for much of that century that *all* of physics and, ultimately, all natural science were susceptible to a reductive "mechanical explanation."[2] In a sense, in biology, genetics now plays the part that mechanics did for nineteenth-century physics. Philosophy of science in the twentieth century has been less interested in general unificatory or in special-science explanations, though there are important exceptions, including some in recent philosophy of biology. E. Nagel's analysis of "What is a mechanical explanation?" (1979, 153–174) is one exception that discusses a special-science explanation,

This research was partially supported by National Institutes of Health Grant R13 HG00703 to the University of Maryland, and by the National Science Foundation's Studies in Science, Technology, and Society Program. I would also like to express my gratitude to the NIH for inviting me to a special August 1993 workshop convened under the auspices of the NICHD at which work in progress by Drs. Bargmann and Chalfie on *C. elegans,* Drs. Hall and Tully on *Drosophila,* Dr. Hamer's studies on humans, and Dr. Plomin's general methodological approaches were presented and discussed. I am also grateful to Robert Wachbroit and David Wasserman for comments on an earlier draft of this manuscript. This chapter draws partially on my 1998 and 1999 essays, which develop some of the themes in the present chapter in different directions. Those essays will be referenced as appropriate.