

7

Resemblance Between Relatives

Version 8 August 2022

In the previous two chapters, we found that the phenotypic variance of a trait can theoretically be partitioned into a number of genetic and environmental components. However, the significant practical issue of how these components can be estimated remains. The key to this matter, first pointed out by Fisher (1918) and Wright (1921b), is the fact that various genetic and environmental sources of variance contribute differentially to the resemblance between different types of relatives.

Assuming for the time being an absence of genotype \times environment interaction, and recalling Equation 6.1, let $z_x = G_x + E_x + e_x$ and $z_y = G_y + E_y + e_y$ be the phenotypic values of two members of a particular relationship, such as parent and offspring. As in previous chapters, G , E , and e denote genotypic values, general environmental effects, and special environmental effects, respectively. The phenotypic covariance between relatives x and y thus becomes

$$\begin{aligned}\sigma_z(x, y) &= \sigma[(G_x + E_x + e_x), (G_y + E_y + e_y)] \\ &= \sigma_G(x, y) + \sigma_{G,E}(x, y) + \sigma_{G,E}(y, x) + \sigma_E(x, y)\end{aligned}\quad (7.1a)$$

Because the special environmental effects are random residual deviations, they are uncorrelated among individuals and do not contribute to the resemblance between relatives, i.e., $\sigma_e(x, y) = 0$. The middle two terms, $\sigma_{G,E}(x, y) + \sigma_{G,E}(y, x)$, refer to the covariance of the genotypic value of one member of the pair and the general environmental effect of the other, while the final term, $\sigma_E(x, y)$, is the covariance between general environmental effects. Experiments can often be designed so that all three covariance terms involving E have expected values equal to zero. For now, we ignore the issue of genotype-environment covariance, i.e., we assume $\sigma_{G,E}(x, y) = \sigma_{G,E}(y, x) = 0$. This assumption reduces Equation 7.1a to

$$\sigma_z(x, y) = \sigma_G(x, y) + \sigma_E(x, y)\quad (7.1b)$$

The genetic covariance between relatives, $\sigma_G(x, y)$, merits special attention. Such covariance is a natural consequence of relatives inheriting copies of the same genes. As in the case of the genetic variance, the genetic covariance between relatives can be partitioned into components attributable to additive, dominance, and various epistatic effects. Each term consists of one of the familiar components of genetic variance (Chapter 5) weighted by a coefficient that describes the number of shared genes in pairs of relatives. These coefficients are the first focus of our attention. Once they are understood, it is a relatively simple step to use the results in Chapter 5 to derive a general expression for the genetic covariance between relatives.

We will first consider the ideal situation in which mating is random and loci are unlinked and in gametic-phase equilibrium. The complications that are introduced with linkage, gametic-phase disequilibrium, and assortative mating will then be evaluated. Some of these complications make for difficult reading, but they are realities that cannot not be ignored. Next, the environmental causes of the resemblance between relatives will be discussed. This does not really complete the picture, as still other complications such as sex-linkage, maternal genetic effects, and inbreeding may be of substantial importance in particular cases, but we leave most of our discussion of these to later chapters. Finally, we provide a broad overview of the concept of heritability, a central parameter in many quantitative-genetic formulations.

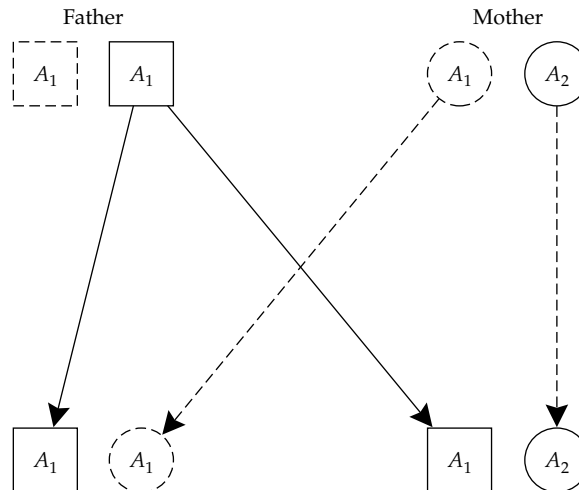


Figure 7.1 The transmittance of genes of two parents to two offspring. All A_1 alleles are identical *in state*. However, of the two A_1 genes carried by the offspring on the left, only the one in the square is identical by *descent* with the A_1 allele carried by the offspring on the right.

MEASURES OF RELATEDNESS

Many relatedness measures have found their way into the population-genetic and sociobiological literature (Wright 1922; Cotterman 1940; Malécot 1948; Denniston 1974; Jacquard 1974; Orlove and Wood 1978; Michod and Hamilton 1980; Grafen 1985). Not all of these play a central role in quantitative-genetic formulations, but they all share two essential features.

First, relatedness can only be defined with respect to a *specified frame of reference*. Technically speaking, all members of a species or population are related to each other to some degree for the simple reason that they contain copies of genes that were present in some remote ancestor in the phylogeny. We avoid this problem by letting the reference population be the base of an observed pedigree. If, for example, no individuals further back than the parental generation have been observed, the usual procedure is to treat that generation as the base and to assume that its members are unrelated. As we detail in Chapter 8, marker information allows this restriction to be significantly relaxed.

Second, all measures of relatedness are based upon the concept of **identity by descent (IBD)**. Genes that are identical by descent are *direct* descendants of a specific gene carried in some ancestral individual (specifically, the two copies trace back to a single ancestral DNA molecule). The distinction between identity by descent and **identity in state** (also **alike in state**) is critical. Two genes that have identical nucleotide sequences but have descended from different copies in the reference population are identical in state but not by descent. On the other hand, genes that are identical by descent are necessarily identical in state, barring mutation. The distinction between these two types of identity is clarified in Figure 7.1, where the parental generation is treated as the base population. Although the first offspring contains two A_1 genes, only one of them is identical by descent with the A_1 gene in the second offspring.

Our treatment of relationships is initially based on classical methods for obtaining the *predicted* degree of relationship given a known pedigree. The recent ability to score an essentially endless number of molecular markers in just about any species allows for two important advances, which are examined in Chapter 8. The first is that a known pedigree is often not needed, as close, to moderately distant, relationships can be reasonably estimated using sufficiently dense marker data. Given that “known” pedigrees can still contain errors (e.g., incorrectly assigned paternities), marker data also provides a check of pedigree data quality. The second is the important distinction between the predicted (or **expected**) value of a

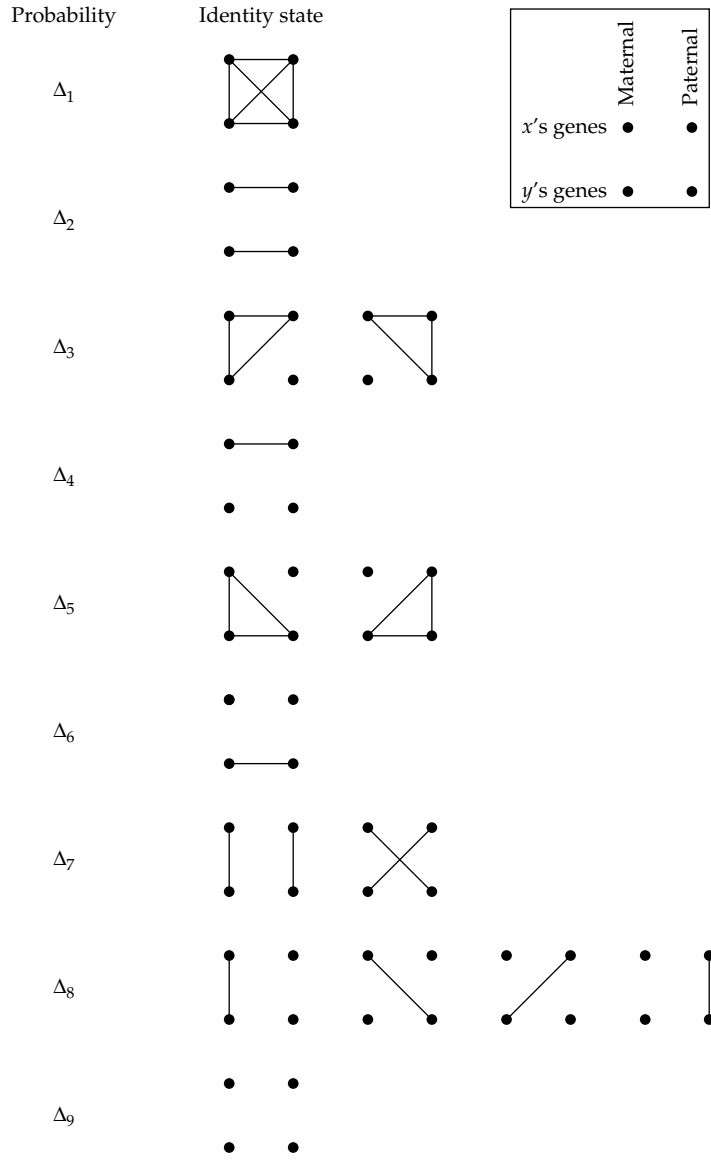


Figure 7.2 The 15 possible states of identity by descent for a locus in individuals x and y , condensed into nine classes, denoted as **identity states**. Alleles that are identical by descent are connected by lines, with horizontal lines implying that the individual is inbred (states 1–6).

relationship for specific types of relatives, and the actual, or **realized**, relationship that occurs for a particular pair of such relatives. A pedigree estimate provides the former, a marker-based estimate the latter. Hence, even with a known and fully accurate pedigree, high quality marker data provide additional information on the true values of relationships.

Coefficients of Identity

Consider a single locus in two diploid individuals. For the four genes involved, there are 15 possible configurations of identity by descent due to the fact that identity may exist *within* as well as *between* individuals (Gillois 1964; Figure 7.2). Individuals that contain pairs of alleles that are identical by descent are said to be **inbred**. If we ignore the distinction between maternally and paternally derived genes, the 15 possible configurations reduce to nine **identity states**. These range from state 1 in which the two individuals are inbred and share

Table 7.1 Summary of the properties of each of the nine identity states given in Figure 7.1. Here, Θ_i is the coefficient of coancestry associated with that identity class.

Class	Θ_i	x inbred	y inbred
1	1	Yes	Yes
2	0	Yes	Yes
3	1/2	Yes	No
4	0	Yes	No
5	1/2	No	Yes
6	0	No	Yes
7	1/2	No	No
8	1/4	No	No
9	0	No	No

a gene that is identical by descent (so that all four genes are identical by descent) to state 9 in which none of the four genes are identical by descent. Many situations exist in which further simplification can be justified. For example, in large random-mating populations, the probability of the first six identity states is essentially zero.

Associated with each of the nine identity states are probabilities, Δ_1 to Δ_9 , which Jacquard (1974) called **condensed coefficients of identity** (Figure 7.2). These coefficients provide a complete description of the probability distribution of identity by descent between single loci of two individuals. The values that they take on depend on the relationship. For example, suppose that x is a parent and y its offspring, and that neither is inbred. Then, because an individual inherits one and only one gene from its parent, $\Delta_8 = 1$ and all other $\Delta_i = 0$. For noninbred full sibs, there is a 0.5 probability of inheriting the same paternal gene and, independently, a 0.5 probability of inheriting the same maternal gene. Thus, $\Delta_7 = 0.25$ (both pairs of genes are identical by descent), $\Delta_8 = 0.50$ (one pair of genes is identical by descent), and $\Delta_9 = 0.25$ (there is no identity by descent at the locus), and all other $\Delta_i = 0$.

Coefficients of Coancestry (Θ) and Inbreeding (f)

Suppose now that single alleles are drawn randomly from individuals x and y . The probability that these two alleles are identical by descent, Θ_{xy} , is called the **coefficient of coancestry**. (In the literature, Θ_{xy} is sometimes referred to as the **coefficient of consanguinity**, **coefficient of kinship**, or **coefficient de parente**). Table 7.1 gives the Θ values associated with each of the nine identity states. For example, either allele from x in a state 3 relationship is IBD one of the alleles in y , but not the other. Hence, $\Theta = 1/2$ for this class. The same logic holds for state 5, with the roles of x and y reversed. Likewise, a similar argument shows that Θ also equals $1/2$ for state 7. Recalling that Δ_i is the probability that a randomly chosen gene pair from relatives x and y corresponds to the i th identity state, Table 7.1 implies that

$$\Theta_{xy} = \Delta_1 + \frac{1}{2}(\Delta_3 + \Delta_5 + \Delta_7) + \frac{1}{4}\Delta_8 \quad (7.2a)$$

Here, we have simply weighted each identity state by its associated Θ value.

An important feature of Θ_{xy} follows by considering a hypothetical offspring (z) of x and y . By the above definition, Θ_{xy} is the probability that the two alleles at a locus in individual z are identical by descent, as the two alleles in z are the result of drawing a single random allele from each of its parents, x and y . The probability that two alleles at a locus in z are IBD is usually called Wright's (1922) **inbreeding coefficient**, f_z . Thus, *an individual's inbreeding coefficient is equivalent to its parents' coefficient of coancestry*,

$$f_z = \Theta_{xy} \quad (7.2b)$$

Note that Equation 7.2 dispels a common misunderstanding about inbreeding. If we cross two fully inbred parents (x and y), the resulting offspring are inbred *only* when $\Theta_{xy} > 0$.

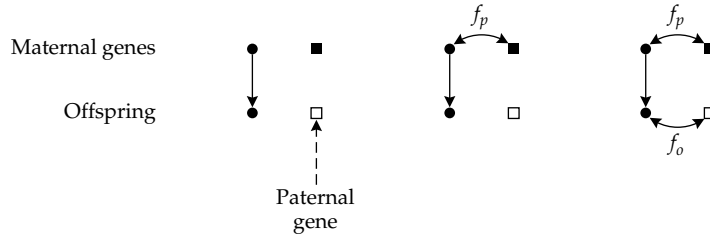


Figure 7.3 The identity of genes by descent for a parent (mother) and her offspring. Circles and squares represent, respectively, maternally and paternally derived genes. **Left:** The mother (denoted by p in the text) is not inbred and her mate (denoted by q) is not a relative (so the offspring is not inbred). **Center:** The mother is inbred ($f_p > 0$) but unrelated to her mate ($\Theta_{pq} = 0$). **Right:** In addition to the mother being inbred, she is related to her mate, so that her offspring is also inbred, with $f_o = \Theta_{pq}$.

If the two parents are unrelated, then the offspring are outbred even when the parents are fully inbred.

We now proceed by example to demonstrate how estimates of Θ_{xy} are derived. The first problem to be tackled is the coefficient of coancestry of an individual with itself, Θ_{xx} . This may seem like a nonsensical task. However, we will soon see that Θ_{xx} is an essential element of all coancestry estimates. Denote the two genes carried by individual x as A_1 and A_2 , and then randomly draw a gene from the locus, replace it, and randomly draw another. Θ_{xx} is the probability that the two genes drawn are identical by descent. There are four ways, each with probability $1/4$, in which the genes can be drawn: A_1 both times, A_1 first and A_2 second, A_2 first and A_1 second, and A_2 both times. If two A_1 genes are drawn, they must be identical by descent because they are copies of the same gene. The same applies to a draw of two A_2 genes. Thus, provided that genes A_1 and A_2 are not identical to each other by descent, then Θ_{xx} is simply $(1/4)(1) + (1/4)(1) = 1/2$. We should, however, recognize the possibility that individual x is inbred, in which case the probability that the gene A_1 is identical by descent with the gene A_2 is f_x . Thus, a general expression for the coefficient of coancestry of an individual with itself is

$$\Theta_{xx} = \frac{1}{4}(1 + f_x + f_x + 1) = \frac{1}{2}(1 + f_x) \quad (7.3a)$$

The impact of inbreeding is thus to increase the coefficient of coancestry, which will be a common theme throughout. We can rearrange Equation 7.3a to yield the useful identity

$$f_x = 2\Theta_{xx} - 1 \quad (7.3b)$$

Note that if we *self* x , the level of inbreeding in its offspring o is $f_o = \Theta_{xx}$. Hence, from Equation 7.3a we obtain the recursion equation for the effect of selfing on the level of inbreeding,

$$f(t+1) = \frac{1}{2}[1 + f(t)] \quad (7.3c)$$

A slightly more complicated situation arises in calculating the coefficient of coancestry between a parent and its offspring. In order to simplify the discussion, we will call the parent (p) of interest the mother, but the same results apply to fathers, provided the locus is autosomal. We first consider the situation in which neither the mother nor her offspring (o) are inbred, i.e., the mother's parents are unrelated, and she is unrelated to her mate (who we denote by q). In that case, of the four ways in which single genes can be drawn from the mother and the child, only one involves a pair that is identical by descent (Figure 7.3, left). Therefore, $\Theta_{po} = 1/4$. Suppose, however, that the mother is inbred (Figure 7.3, center), so that the probability that both of her alleles are identical by descent is f_p . This is the same

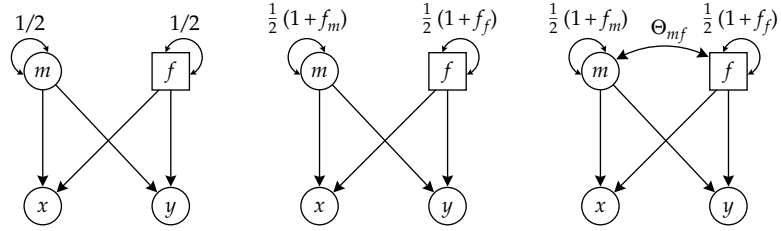


Figure 7.4 Path diagrams (Appendix 2) for analyzing the probability that random genes from two full sibs are identical by descent. The path coefficients along single-headed arrows are always equal to $1/2$. **Left:** The parents, m and s , are neither related nor inbred. **Center:** The parents are unrelated, but inbred. **Right:** In addition to being inbred, the parents are related with coefficient of coancestry Θ_{mf} , so that the offspring are inbred to level $f_0 = \Theta_{mf}$.

as the probability that the maternal gene inherited by the offspring is identical by descent with the maternal gene not inherited. The probability of drawing such a gene combination is $1/4$. Therefore, inbreeding in the parent inflates Θ_{po} to $(1 + f_p)/4$. With complete inbreeding ($f_p = 1$), both parental alleles are identical by descent, increasing Θ_{po} to $1/2$. Finally, we allow for the possibility that the parents of o are related, so that the offspring is inbred with coefficient $f_o = \Theta_{pq}$ (Figure 7.3, right). It is now necessary to consider the implications of drawing a paternally derived gene from the offspring, the probability of which is $1/2$. Because f_o is equivalent to the probability that maternally and paternally derived genes are identical by descent, the additional parent-offspring identity induced by inbreeding is $f_o/2$. In summary, the most general expression for the coefficient of coancestry for a parent and offspring is

$$\Theta_{po} = \frac{1}{4}(1 + f_p + 2f_o) \tag{7.4a}$$

Again note that the impact of inbreeding is to inflate Θ . Recalling Equations 7.2b and 7.3b, we can alternatively express Equation 7.4a as

$$\Theta_{po} = \frac{1}{4} [1 + (2\Theta_{pp} - 1) + 2\Theta_{pq}] = \frac{1}{2} (\Theta_{pp} + \Theta_{pq}) \tag{7.4b}$$

This result foreshadows a simple recursion method for obtaining the Θ values over a very complex pedigree, which will be examined shortly. Often in the literature, Θ_{po} is simply considered to be $1/4$. It should now be clear that this implicitly assumes the absence of inbreeding, which inflates the value of Θ .

We now move on to the coefficient of coancestry of two individuals that share the same father and mother (full sibs). We assume a species with separate sexes so that the mother and father are different individuals, and we again start with the simplest situation, progressively allowing the parents to be inbred and/or related (Figure 7.4). For the analysis of full sibs, as well as more complicated relationships, the method of path analysis (Appendix 2) provides a useful tool. The elements in Figure 7.4 no longer represent gametes (as in Figure 7.3), but rather individuals.

Let m represent the mother, s the sire (father; we use s in place of f to avoid confusion with use of the latter for inbreeding), and x and y their two offspring. When the parents are neither inbred nor related, there are two paths by which the same gene can be passed to both x and y : $x \leftarrow m \rightarrow y$ and $x \leftarrow s \rightarrow y$. Because both paths have identical consequences, we will simply consider the first of them. First, we note that the probability that both x and y receive the same maternal gene is $1/2$. This is the coefficient of coancestry of the (noninbred) mother with herself, Θ_{mm} , and is represented by the double-headed arrow in the figure. Second, we note that the probability of randomly drawing a maternal gene from individual x is $1/2$, and that the same is true for individual y . Thus, the probability of drawing two maternal genes, identical by descent, one from x and the other from y , is $\Theta_{mm}(1/2)^2 = 1/8$. Adding the same contribution from the paternal path, $x \leftarrow s \rightarrow y$, we obtain the coefficient

of coancestry $\Theta_{xy} = 1/4$. Path analysis (Appendix 2) provides a simple way to obtain this result. First, set the path coefficients on all of the single-headed arrows in Figure 7.4 equal to $1/2$. Then, note that the contribution of a path to a correlation between two variables is equal to the product of the path coefficients and the correlation coefficient associated with the common factor (in this case, Θ_{mm} or $\Theta_{ss} = 1/2$).

We now allow for the possibility that the parents are inbred with inbreeding coefficients f_m and f_s , a condition that inflates the coefficient of coancestry of an individual with itself. This is the only necessary change for the path diagram in Figure 7.4 (center). There are still only two paths that lead to genes identical by descent in x and y , and their sum is

$$\Theta_{xy} = \frac{1}{4} (\Theta_{mm} + \Theta_{ss}) = \frac{1}{4} \left(\frac{1+f_m}{2} + \frac{1+f_s}{2} \right) = \frac{1}{8} (2 + f_m + f_s) \quad (7.5a)$$

Finally, we allow for the possibility that m and s are related, such that the probability of drawing two genes (one from each of them) that are identical by descent is Θ_{ms} . It is then necessary to consider two additional paths between x and y : $x \leftarrow m \leftrightarrow s \rightarrow y$ and $x \leftarrow s \leftrightarrow m \rightarrow y$ (Figure 7.4, right). Again taking the coefficients on the single-headed arrows to be $1/2$, it can be seen that each of these two new paths makes a contribution $\Theta_{ms}/4$ to Θ_{xy} , where (Equation 7.2b) Θ_{ms} is the level of inbreeding in the offspring, f_o . Adding these to our previous result, we obtain a general expression for the coefficient of coancestry of full sibs,

$$\Theta_{xy} = \frac{1}{8} (2 + f_m + f_s + 4f_o) \quad (7.5b)$$

which reduces to $\Theta_{xy} = 1/4$ under random mating. Again, we see that inbreeding inflates Θ . Using Equations 7.2b and 7.3b, we can express Equation 7.5b entirely in terms of coefficients of coancestry,

$$\Theta_{xy} = \frac{1}{8} [2 + (2\Theta_{mm} - 1) + (2\Theta_{ss} - 1) + 4\Theta_{ms}] = \frac{1}{4} (\Theta_{mm} + \Theta_{ss} + \Theta_{ms}) \quad (7.5c)$$

Up to now, we have assumed an autosomal locus. The rules change when the locus of interest is sex-linked. Assuming the male is the heterogametic sex, a male cannot receive an X-linked gene from his father, whereas fathers pass on their X-linked genes to daughters with probability one. Females have two X chromosomes and pass each one on to sons or daughters with the usual probability of one-half. Thus, the protocol for obtaining the coefficient of coancestry for an X-linked locus is similar to that used for autosomal loci except that path coefficients leading from fathers to daughters are replaced by a 1 and those leading from fathers to sons are replaced by a 0. For all paths containing two consecutive males, the probability that X-linked genes are identical by descent is zero.

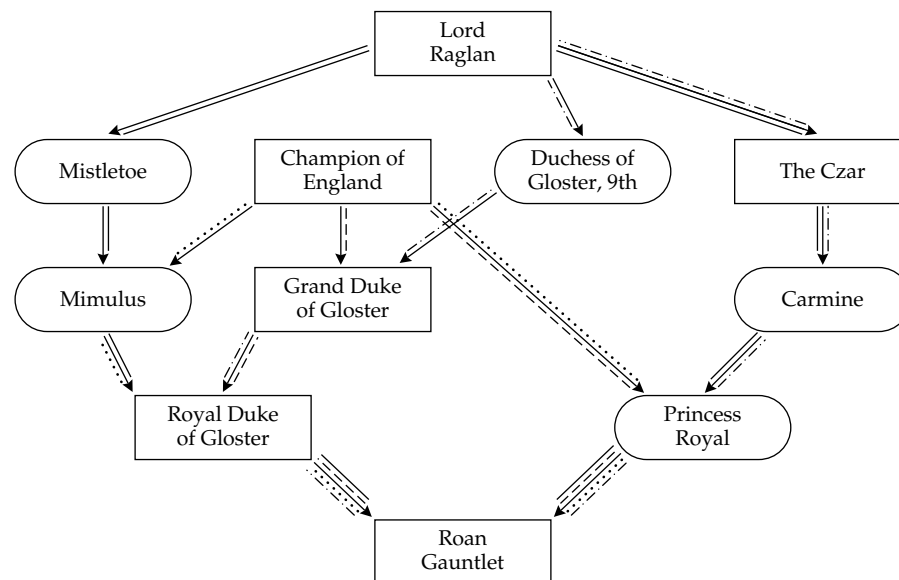
Pedigree-based Estimates of Θ : Chain-counting

The preceding path technique is readily extended to more distant relationships and more complicated schemes of relatedness (Wright 1921b). The coefficient of coancestry is always the sum of a series of two types of paths between x and y . The first type of path leads from a *single common ancestor* to the two individuals of interest, while the second type passes through *two remote ancestors that are related to each other*. Neither type of path is allowed to pass through the same ancestor more than once. This procedure is summarized by the following equation

$$\Theta_{xy} = \sum_i \Theta_{ii} \left(\frac{1}{2} \right)^{n_i-1} + \sum_j \sum_{j \neq k} \Theta_{jk} \left(\frac{1}{2} \right)^{n_{jk}-2} \quad (7.6)$$

where n_i is the number of individuals (including x and y) in the path leading from common ancestor i , and n_{jk} is the number of individuals (including x and y) on the path leading from two different but related ancestors, j and k . Formal proof of this equation can be found in Boucher (1988). Equation 7.6 is often called the method of **chain-counting**.

Example 7.1. One of the first pedigrees to which Wright (1922) applied his theory of inbreeding is that of Roan Gauntlet, an English shorthorn bull. In the following figure, rectangles and ovals refer to bulls and cows, respectively.



We wish to compute the coefficient of coancestry of the Royal Duke of Gloster and Princess Royal. This is the same as the inbreeding coefficient of their son, Roan Gauntlet. The four possible paths by which alleles identical by descent can be inherited by the Royal Duke and Princess Royal are indicated by the coded lines adjacent to the arrows in the pedigree. Two of these paths (the dotted and the dashed lines) contain four individuals and two (the solid and the dot-and-dash lines) contain seven. Thus, assuming that the remote ancestors, Lord Raglan and Champion of England, are not inbred (so that for both, $\Theta_{ii} = 1/2$) and unrelated, the coefficient of coancestry of the Royal Duke and Princess Royal is $[2(1/2)^4 + 2(1/2)^7] = 0.141$. This is a slightly closer relationship than that for half sibs (for which $\Theta = 0.125$). Relative to the base population, the alleles at 14% of the autosomal loci in the offspring, Roan Gauntlet, are expected to be identical by descent.

What is the coefficient of coancestry of the Royal Duke of Gloster and Princess Royal with respect to X-linked loci? Of the four paths between these two individuals, only the one traced by the dotted line does not contain consecutive males. Champion of England passes on his X chromosome to each of his daughters, Mimulus and Princess Royal, with probability 1. Mimulus passes that chromosome on to the Royal Duke of Gloster with probability 1/2. The probability of drawing a specific X chromosome from the Royal Duke is 1 (because he is a male) and from Princess Royal is 1/2. The coefficient of coancestry for X-linked loci is therefore $1 \cdot 1/2 \cdot 1 \cdot 1/2 = 1/4$. This is substantially greater than the coefficient for autosomal loci over the same path, which is $(1/2)^4 = 1/16$.

Pedigree-based Estimates of Θ : Tabular Method

Chain-counting (Equation 7.6) is a reasonable approach for computing the coefficient of coancestry between a specific pair of relatives in a known pedigree. One caveat is that it is easy to miss paths in complex pedigrees, which results in underestimating Θ . Further, using chain-counting to compute *all* $n(n-1)/2$ pairwise Θ values in a pedigree of n individuals is very inefficient. An alternative approach that easily deals with both of these issues is a simple recursive scheme called the **tabular** or **recursive** method (Crudent 1949; Emik and Terrill 1949). Here one starts with the **founding members** of a pedigree (individuals with both parents unknown) and then proceeds down the pedigree (i.e., considering all their

descendants) in a recursive fashion (Example 7.2). The algorithm is as follows:

- (i) *Founding individuals of the pedigree are assumed to be unrelated and not inbred* (although this can also be accommodated). These founders are assigned values of $\Theta_{ii} = 1/2$.
- (ii) *Any unknown parent of any future individual is assumed to be unrelated to all others in the pedigree and not inbred*, and they are also assigned a value of $\Theta_{ii} = 1/2$.
- (iii) To compute the self coancestry for individual i , let S_i and D_i denote their sire and dam (father and mother). Applying Equations 7.2b and 7.3a yields

$$\Theta_{ii} = (1 + \Theta_{S_i D_i})/2$$

which reduces to $1/2$ when $\Theta_{S_i, D_i} = 0$, namely when the parents are unrelated.

- (iv) The coancestry between individuals i and j (who are descendants of the founders) is computed as

$$\Theta_{ij} = (\Theta_{i S_j} + \Theta_{i D_j})/2 = (\Theta_{j S_i} + \Theta_{j D_i})/2$$

The logic here follows directly from the definition of the coefficient of coancestry. With probability $1/2$, a random allele from j will have come from its sire, and the probability that this allele is IBD with the random allele from i is just $\Theta_{i S_j}$, while with probability $1/2$ the allele from j is from its mother, giving the IBD probability as $\Theta_{i D_j}$. By symmetry, we can reverse the roles and consider a random alleles from j and the sire or dam of i . Note that Equation 7.4b (the coefficient of coancestry between a parent p and its offspring o) is just a special case of this expression. Take $i = p$ (the parent) and j as its offspring, with $S_j = p$ and $D_j = q$. Substituting yields $\Theta_{po} = (\Theta_{pp} + \Theta_{pq})/2$.

This recursive approach has been extended by Karigl (1981) to compute all nine condensed coefficients of identity (Figure 7.2) for arbitrary pedigrees.

Example 7.2. Again consider Wright’s shorthorn pedigree (Example 7.1). Note that Lord Raglan and Champion of England are founders of the pedigree (both parents are unknown, as indicated by the lack of any paths to them from ancestors). Likewise, four other individuals (Mistletoe, Duchess of Gloster, Czar, and Carmine) have an unknown parent (dams in all cases), indicated by a path to them from only one ancestor. We can compactly express all of the pedigree relationships in tabular form, starting with the founders, as follows:

Individual	Number	Sire (S_i)	Dam (D_i)
Lord Raglan	1	unknown	unknown
Champion of England	2	unknown	unknown
Mistletoe	3	1	unknown
Duchess of Gloster, 9th	4	1	unknown
The Czar	5	1	unknown
Mimulus	6	2	3
Grand Duke of Gloster	7	2	4
Carmine	8	5	unknown
Royal Duke of Gloster	9	7	6
Princess Royal	10	2	8
Roan Gauntlet	11	9	10

Applying the recursive rules, for the founders we have $\Theta_{11} = \Theta_{22} = 1/2$, and $\Theta_{12} = 0$. For Mistletoe ($i = 3$), $S_3 = 1$ and $D_3 = ?$, giving $\Theta_{33} = (1 + \Theta_{1?})/2 = 1/2$, where $\Theta_{1?}$ between her sire (1) and (unknown) dam is zero by rule (ii). Likewise,

$$\Theta_{31} = [\Theta_{1S_3} + \Theta_{1D_3}]/2 = (\Theta_{11} + \Theta_{1?})/2 = (1/2)/2 = 1/4$$

$$\Theta_{32} = [\Theta_{2S_3} + \Theta_{2D_3}]/2 = (\Theta_{21} + \Theta_{2?})/2 = 0$$

For Duchess of Gloster ($i = 4$), $S_4 = 1$ and $D_4 = ?$, giving $\Theta_{44} = (1 + \Theta_{1?})/2 = 1/2$, while

$$\Theta_{34} = [\Theta_{3S_4} + \Theta_{3D_4}]/2 = [\Theta_{31} + \Theta_{3?}]/2 = (1/4 + 0)/2 = 1/8$$

Proceeding in this fashion down the pedigree, and recalling that $\Theta_{ij} = \Theta_{ji}$, computes all of the other Θ values. Note that this algorithm (with the pedigree coded by associating each pedigree member with the two index numbers for its parents) is easily programmable. Typically, all of the pairwise coancestries for a pedigree are compactly presented as a matrix (Chapter 8). In this case, the result is a (symmetric) matrix with 11 rows and columns, with the element in row i and column j corresponding to Θ_{ij} . The upper left part of this matrix for the first four relatives here is

$$\begin{array}{cccccc} & 1 & 2 & 3 & 4 & \cdots \\ 1 & 1/2 & 0 & 1/4 & 1/4 & \cdots \\ 2 & 0 & 1/2 & 0 & 0 & \cdots \\ 3 & 1/4 & 0 & 1/2 & 1/8 & \cdots \\ 4 & 1/4 & 0 & 1/8 & 1/2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

Predicted versus Realized Values of Θ

While Equation 7.2a gives the *expected* value of Θ_{xy} for the pair of relatives x and y , there is a *distribution* of potential Θ values around this mean. Consider two full sibs from noninbred, and unrelated, parents. While the expected value of Θ_{xy} is $1/4$, for any particular locus, the actual value is either $1/2$, $1/4$, or 0 , corresponding, respectively, to sharing 2, 1, or 0 IBD alleles (identity states 7, 8, and 9) with associated probabilities of $1/4$, $1/2$, and $1/4$. Hence, random sampling of Θ values over all loci in the sib pair results in a *distribution of values* around the expected value of $1/4$ (or, equivalently, a distribution around the expected sharing of 50% IBD alleles). As a result, some full-sib pairs are more similar ($\Theta > 1/4$) and others are less similar ($\Theta < 1/4$) than the expected pedigree-based value. While this concept was fully appreciated in the days of pedigree-based estimates (Fisher 1949; Franklin 1977; Stam 1980), researchers at the time had no ability to actually score the realized amount of IBD sharing. With dense SNP data, we now have this ability (Figure 7.5). As well will see in several later chapters, exploitation of this variance in relationships is an important tool in modern quantitative genetics.

From Table 7.1, the *variance* (at a single locus) in Θ for a specific pair of relatives is given by

$$\sigma^2(\Theta_{xy}) = \sum_{i=1}^9 \Delta_i(xy) (\Theta_i - \bar{\Theta}_{xy})^2 \quad (7.7a)$$

where $\bar{\Theta}_{xy}$ denotes the value given by Equation 7.2a, Θ_i is the value for identity state i , and $\Delta_i(xy)$ is the probability of the relatives x and y being in identity state i . From Table 7.1, there are only four possible Θ_i values: 0 for states 2, 4, 6, and 9; $1/4$ for state 8; $1/2$ for states 3, 5, and 7; and 1 for state 1. Hence,

$$\begin{aligned} \sigma^2(\Theta_{xy}) = & \bar{\Theta}_{xy}^2 (\Delta_2 + \Delta_4 + \Delta_6 + \Delta_9) + (1/4 - \bar{\Theta}_{xy})^2 \Delta_8 \\ & + (1/2 - \bar{\Theta}_{xy})^2 (\Delta_3 + \Delta_5 + \Delta_7) + (1 - \bar{\Theta}_{xy})^2 \Delta_1 \end{aligned} \quad (7.7b)$$

For only outbred relatives (identity states 7–9), this reduces to

$$\sigma^2(\Theta_{xy}) = \bar{\Theta}_{xy}^2 \Delta_9 + (1/4 - \bar{\Theta}_{xy})^2 \Delta_8 + (1/2 - \bar{\Theta}_{xy})^2 \Delta_7 \quad (7.7c)$$

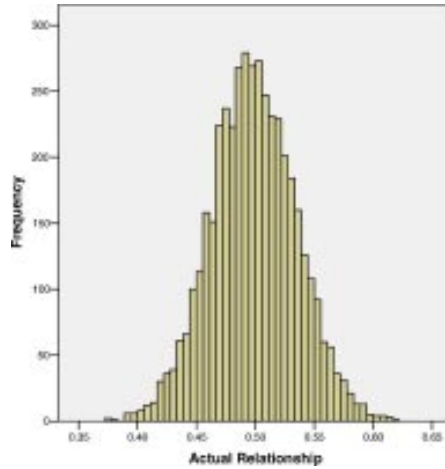


Figure 7.5 Empirical distribution of the actual fraction of shared IBD alleles for 4400 pairs of human full sibs. The realized value of Θ is simply half this value. (After Visscher et al. 2006.)

Hence, there is variation in Θ_{xy} for any pair of relatives, *unless* one of the $\Delta_i = 1$. This only occurs for clones ($\Delta_1 = 1$, $\bar{\Theta}_{xy} = 1$) and, in outbreds, for parent-offspring pairs ($\Delta_8 = 1$, $\bar{\Theta}_{xy} = 1/4$). The latter occurs because a parent and its offspring (in the absence of inbreeding) always share *exactly* one IBD allele.

An important point is that Equation 7.7b is the variance in Θ associated *with a particular locus*. We then average the locus-specific Θ values over the entire genome. If one has n independently assorting loci, then the genome-wide variance would be $\sigma^2(\Theta_{xy})/n$. In reality, the situation is much more complex, as while chromosomes assort independently, sites within a chromosome are correlated due to linkage. We examine this more fully in Chapter 8.

Example 7.3. Consider the (single-locus) variance in Θ for noninbred half- and full-sibs. For noninbred full sibs, $\Delta_7 = 1/4$, $\Delta_8 = 1/2$, $\Delta_9 = 1/4$, and $\bar{\Theta} = 1/4$. Equation 7.7c yields

$$\begin{aligned}\sigma^2(\Theta_{xy}) &= (1/4)^2(1/4) + (1/4 - 1/4)^2(1/2) + (1/2 - 1/4)^2(1/4) \\ &= 1/32 = 0.03125\end{aligned}$$

for a standard deviation of 0.177 and a coefficient of variation of $CV = 0.177/0.25 = 0.707$. Turning to half sibs, $\Delta_8 = 1/2$, $\Delta_9 = 1/2$, $\bar{\Theta} = 1/8$, and Equation 7.7c yields

$$\sigma^2(\Theta_{xy}) = (1/8)^2(1/2) + (1/4 - 1/8)^2(1/2) = 1/64 = 0.015625$$

for a standard deviation of $1/8$, and a CV of 1. This increase in CV as relatives become more distant is a general phenomenon (Hill and Weir 2011).

The Coefficient of Fraternity, Δ_{xy}

Up to now we have been considering the identity of *single alleles* by descent. Another useful measure is the probability that *single-locus genotypes* of two individuals are identical by descent. The formulation of such a measure, which we denote as Δ_{xy} , is attributable to Cotterman (1954) and was called the **coefficient of fraternity** by Trustring (1961). The problem is set

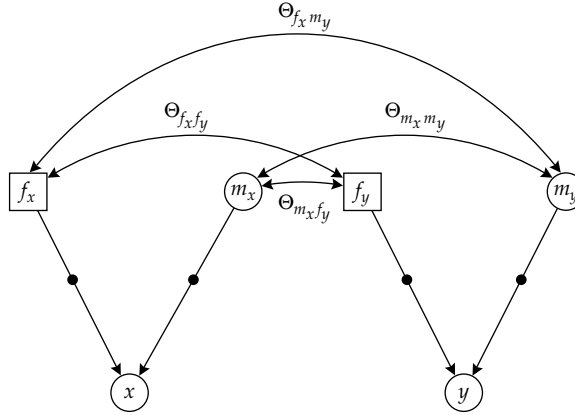


Figure 7.6 The analysis of the identity by descent of genotypes of individuals x and y . s_x and s_y represent sires (fathers) (which may be the same individual) of x and y , respectively, whereas m_x and m_y represent their mothers. Double-headed arrows between two parents represent coefficients of coancestry.

out in Figure 7.6. Here we denote the mothers of individuals x and y as m_x and m_y , and the fathers (sires) as s_x and s_y . The coefficients of coancestry $\Theta_{m_x m_y}$, $\Theta_{m_x s_y}$, $\Theta_{s_x m_y}$, and $\Theta_{s_x s_y}$ provide measures of the probability of drawing genes identical by descent from all four combinations of parents.

There are two ways by which the genotype of x can be identical by descent with that of y : (1) the gene descending from m_x is identical by descent with that descending from m_y and that from s_x is identical by descent with that from s_y , or (2) the gene from m_x is identical by descent with that from s_y , and that from s_x is identical by descent with that from m_y . Thus, the coefficient of fraternity is obtained as

$$\Delta_{xy} = \Theta_{m_x m_y} \Theta_{s_x s_y} + \Theta_{m_x s_y} \Theta_{s_x m_y} \tag{7.8}$$

In terms of the condensed coefficients of identity, $\Delta_{xy} = \Delta_1 + \Delta_7$, which reduces to Δ_7 in the absence of inbreeding. Ovaskainen et al. (2008) noted that Equation 7.8 is an approximation, requiring that the four probabilities (i.e., the Θ_{ij}) determining Δ_{xy} be independent. This is usually not a serious problem unless the pedigree is highly inbred.

Two examples will suffice to illustrate the use of this equation. First, consider the situation when x and y are full sibs, in which case the mothers are the same individual ($m_x = m_y = m$), as are the fathers ($s_x = s_y = s$). Equation 7.8 then reduces to

$$\Delta_{xy} = \Theta_{mm} \Theta_{ss} + \Theta_{ms}^2 \tag{7.9a}$$

If the parents are unrelated, then $\Theta_{ms} = 0$; and if the parents are not inbred, then $\Theta_{mm} = \Theta_{ss} = 1/2$. Substituting these values into the above expression, we obtain $\Delta_{xy} = 1/4$. Expressed in terms of parent (f_s, f_m) and offspring (f_o) levels of inbreeding, Equation 7.9a becomes

$$\Delta_{xy} = \frac{1}{4} [(1 + f_s)(1 + f_m)] + f_o^2 \tag{7.9b}$$

showing that any amount of inbreeding inflates the value above 1/4.

Now consider the case of paternal half sibs, in which case the fathers are the same individual ($s_x = s_y = s$), but the mothers are different. Now,

$$\Delta_{xy} = \Theta_{m_x m_y} \Theta_{ss} + \Theta_{s m_x} \Theta_{s m_y} \tag{7.9c}$$

Provided that the parents are unrelated, then $\Theta_{ss} = 1/2$ and $\Theta_{m_x m_y} = \Theta_{m_x s} = \Theta_{s m_y} = 0$, which yields $\Delta_{xy} = 0$. The genotypes of two individuals cannot be identical by descent if their maternally (or paternally) derived genes come from unrelated individuals. The identity coefficients for several common relationships are summarized in Table 7.2.

Table 7.2 Identity coefficients for common relationships under the assumption of no inbreeding, in which case Δ_1 to $\Delta_6 = 0$.

Relationship	Δ_7	Δ_8	Δ_9	Θ_{xy}	Δ_{xy}
Parent–offspring	0	1	0	$\frac{1}{4}$	0
Grandparent–grandchild	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{8}$	0
Great grandparent–great grandchild	0	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{1}{16}$	0
Half sibs	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{8}$	0
Full sibs, dizygotic twins	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
Uncle (aunt)–nephew (niece)	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{8}$	0
First cousins	0	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{1}{16}$	0
Double first cousins	$\frac{1}{16}$	$\frac{6}{16}$	$\frac{9}{16}$	$\frac{1}{8}$	$\frac{1}{16}$
Second cousins	0	$\frac{1}{16}$	$\frac{15}{16}$	$\frac{1}{64}$	0
Monozygotic twins (clonemates)	1	0	0	$\frac{1}{2}$	1

Example 7.4. Returning to the figure in Example 7.1, what is Δ_{xy} for $x =$ Royal Duke of Gloster and $y =$ Princess Royal? Designate the parents as $s_x =$ Grand Duke of Gloster, $m_x =$ Mimulus, $s_y =$ Champion of England, and $m_y =$ Carmine. Noting that Champion of England is the father of Grand Duke of Gloster and Mimulus, $\Theta_{s_x s_y} = \Theta_{m_x s_y} = (1/4)$. Counting the number of individuals in the paths of descent between the remaining two pairs of parents, $\Theta_{m_x m_y} = \Theta_{s_x m_y} = (1/2)^5$. Substituting into Equation 7.8, the probability that x and y have identical genotypes by descent at an arbitrary autosomal locus is

$$\Delta_{xy} = (1/4)(1/2)^5 + (1/2)^5(1/4) = (1/2)^6 \simeq 0.016$$

or about 6% of the (outbred) full-sib value of 0.25.

THE GENETIC COVARIANCE BETWEEN RELATIVES

The bulk of the credit goes to Fisher (1918) and Wright (1921b) for elucidating the connection between the phenotypic resemblance of relatives and the types of genetic variance in populations. However, despite the tremendous advances in these papers, a number of problems remained. Wright did not pursue nonadditive gene action beyond dominance and even then ran into difficulties. Fisher incorporated epistasis but only of the additive \times additive type. In 1954, Cockerham and Kempthorne (Cockerham's major professor at the time) independently published papers outlining how the earlier results could be generalized to include any type of gene action. Both authors arrived at the same result by very different routes, but Kempthorne's approach is much simpler and is the one that we will pursue. Cockerham's name will appear many times throughout the book in other contexts.

At the outset, it must be emphasized that the simple results that will emerge below are not obtained without making a number of assumptions: (1) all of the genetic variation is attributable to diploid, autosomal loci; (2) mating is random; (3) all loci are unlinked and in gametic-phase equilibrium; (4) there is no genetic variation for maternal effects; (5) genotype-environment covariance and interaction are unimportant; (6) there is no sexual dimorphism; and (7) selection is not operating on the population. In due course, we will relax all of these conditions.

Our first task is to decompose the total genetic covariance between relatives into fundamental components that describe the various types of gene action. We accomplish this by following the logic used in Chapter 5 to partition the genetic variance. Consider a collection of pairs of individuals all of the same type of relationship, and let x and y represent the members of a random pair. From Equation 5.5, taking things out only to the two-locus effects, the genotypic values of the two individuals may be written as

$$\begin{aligned}
 G_{ijkl..}(x) &= \mu_G + [\alpha_i^x + \alpha_j^x + \alpha_k^x + \alpha_l^x + \dots] + [\delta_{ij}^x + \delta_{kl}^x + \dots] \\
 &\quad + [(\alpha\alpha)_{ik}^x + (\alpha\alpha)_{il}^x + (\alpha\alpha)_{jk}^x + (\alpha\alpha)_{jl}^x + \dots] \\
 &\quad + [(\alpha\delta)_{ikl}^x + (\alpha\delta)_{jkl}^x + (\alpha\delta)_{kij}^x + (\alpha\delta)_{lij}^x + \dots] + (\delta\delta)_{ijkl}^x + \dots \\
 G_{ijkl..}(y) &= \mu_G + [\alpha_i^y + \alpha_j^y + \alpha_k^y + \alpha_l^y + \dots] + [\delta_{ij}^y + \delta_{kl}^y + \dots] \\
 &\quad + [(\alpha\alpha)_{ik}^y + (\alpha\alpha)_{il}^y + (\alpha\alpha)_{jk}^y + (\alpha\alpha)_{jl}^y + \dots] \\
 &\quad + [(\alpha\delta)_{ikl}^y + (\alpha\delta)_{jkl}^y + (\alpha\delta)_{kij}^y + (\alpha\delta)_{lij}^y + \dots] + (\delta\delta)_{ijkl}^y + \dots
 \end{aligned} \tag{7.10a}$$

where i, j and k, l represent alleles at the first and second loci. Fisher (1918) showed that just as the different types of effects are uncorrelated within individuals, they are also uncorrelated between individuals, provided the preceding assumptions are met. Consequently, the genetic covariance between relatives can be expanded into a series of terms, each describing the covariance between the same kinds of effects in two individuals:

$$\sigma_G(x, y) = \sigma_A(x, y) + \sigma_D(x, y) + \sigma_{AA}(x, y) + \sigma_{AD}(x, y) + \sigma_{DD}(x, y) + \dots \tag{7.10b}$$

Note that if $x = y$, Equation 7.10b reduces to Equation 5.8, the usual expression for the genetic variance.

The remaining task is to express the terms in Equation 7.10b in terms of variance components and coefficients of relationship. We will do this only for the first three terms and then give the general result. First, we evaluate the additive genetic covariance at locus 1. Because the mean value of the effects is zero by definition (Chapter 5), the covariance between x and y caused by the additive effects is equal to the expectation of the cross-product, $E[(\alpha_i^x + \alpha_j^x)(\alpha_i^y + \alpha_j^y)]$. Consider one of the four terms in the expansion, $E[\alpha_i^x \alpha_i^y]$. The two genes of interest may be identical by descent, with probability Θ_{xy} , in which case $E[\alpha_i^x \alpha_i^y] = E[\alpha_i^2]$, which is half the additive genetic variance attributable to locus 1. If the two genes are not identical by descent, then they must be distributed independently so that $E[\alpha_i^x \alpha_i^y] = [E(\alpha_i)] \cdot [E(\alpha_i)] = 0$. These same arguments can be applied to the remaining three terms in $E[(\alpha_i^x + \alpha_j^x)(\alpha_i^y + \alpha_j^y)]$. Thus, the additive genetic covariance is $4\Theta_{xy}E[\alpha_i^2]$, which is twice the additive genetic variance at the locus times the probability that randomly drawn genes from x and y are identical by descent. Noting that this result applies to all loci and that the distributions of effects at different loci are independent under the assumptions of the model, the additive genetic covariance, obtained by summing over loci, reduces to

$$\sigma_A(x, y) = 2\Theta_{xy}\sigma_A^2 = r_{xy}\sigma_A^2 \tag{7.11a}$$

where $r = 2\Theta$ is **Wright's coefficient of relatedness**.

We now move on to the dominance genetic covariance, which for locus 1 is $E[\delta_{ij}^x \delta_{ij}^y]$. If x and y are identical by descent for both genes at this locus, then $E[\delta_{ij}^x \delta_{ij}^y] = E[\delta_{ij}^2]$, which is the dominance genetic variance attributable to locus 1. The probability of such identity is Δ_{xy} . On the other hand, if x and y do not have identical genotypes by descent, the dominance effects must be distributed independently, and hence $E[\delta_{ij}^x \delta_{ij}^y] = (E[\delta_{ij}]) \cdot (E[\delta_{ij}]) = 0$. Again, because these arguments apply to every locus, the dominance genetic covariance, obtained by summing over all loci, is

$$\sigma_D(x, y) = \Delta_{xy}\sigma_D^2 \tag{7.11b}$$

Table 7.3 Coefficients for the components of genetic covariance between different types of diploid relatives under the assumptions of random mating, free recombination, and gametic-phase equilibrium. To obtain the covariance expression for a particular type of relationship, multiply each variance component by its coefficient and sum. For example, the genetic covariance between half sibs is $(\sigma_A^2/4) + (\sigma_{AA}^2/16)$. Blanks indicate values of zero.

Relationship	σ_A^2	σ_D^2	σ_{AA}^2	σ_{AD}^2	σ_{DD}^2
Parent–offspring	$\frac{1}{2}$		$\frac{1}{4}$		
Grandparent–grandchild	$\frac{1}{4}$		$\frac{1}{16}$		
Great grandparent–great grandchild	$\frac{1}{8}$		$\frac{1}{64}$		
Half sibs	$\frac{1}{4}$		$\frac{1}{16}$		
Full sibs, dizygotic twins	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$
Uncle (aunt)–nephew (niece)	$\frac{1}{4}$		$\frac{1}{16}$		
First cousins	$\frac{1}{8}$		$\frac{1}{64}$		
Double first cousins	$\frac{1}{4}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{64}$	$\frac{1}{256}$
Second cousins	$\frac{1}{32}$		$\frac{1}{1024}$		
Monozygotic twins (clonemates)	1	1	1	1	1

Finally, we consider the epistatic genetic covariance caused by additive \times additive effects. For any pair of loci, this involves the 16 cross-product terms in the expectation $E[\{(\alpha\alpha)_{ik}^x + (\alpha\alpha)_{il}^x + (\alpha\alpha)_{jk}^x + (\alpha\alpha)_{jl}^x\}\{(\alpha\alpha)_{ik}^y + (\alpha\alpha)_{il}^y + (\alpha\alpha)_{jk}^y + (\alpha\alpha)_{jl}^y\}]$. Because the results are the same for all terms, it is sufficient to evaluate only one of them. The term $E[(\alpha\alpha)_{ik}^x(\alpha\alpha)_{ik}^y]$ is equivalent to $E[(\alpha\alpha)^2]$ provided that two conditions hold: a random gene drawn from the first locus in x must be identical by descent with one drawn from y , and the same condition must hold at the second locus. If identity by descent does not arise simultaneously for gene pairs drawn from both loci, $E[(\alpha\alpha)_{ik}^x(\alpha\alpha)_{ik}^y] = [E(\alpha\alpha)]^2 = 0$. Now, under the assumption of gametic-phase equilibrium, the probability of identity by descent at locus 1 is independent of that at locus 2. Both probabilities are Θ_{xy} . Thus, the probability of joint identity by descent is Θ_{xy}^2 , and the covariance caused by additive \times additive epistasis between loci 1 and 2 is $16 \Theta_{xy}^2 E[(\alpha\alpha)^2]$. Noting that $E[(\alpha\alpha)^2]$ is one-fourth the additive \times additive genetic variance for a single pair of loci, and summing over all pairs,

$$\sigma_{AA}(x, y) = (2\Theta_{xy})^2 \sigma_{AA}^2 = r_{xy}^2 \sigma_{AA}^2 \quad (7.11c)$$

All three of the evaluated terms from Equation 7.11b reduce to simple functions of identity coefficients (Θ and Δ) and a component of genetic variance, and the same is true for all higher-order terms. Due to the independent distributions of genes at different loci under the assumptions of the model, any covariance due to higher-order epistatic effects is equal to the product of the identity for each component additive effect, the probability of identity for each component dominance effect, and the corresponding variance component. Thus, the covariance attributable to additive \times dominance epistasis is

$$\sigma_{AD}(x, y) = 2\Theta_{xy}\Delta_{xy}\sigma_{AD}^2 \quad (7.11d)$$

because there is one additive and one dominance effect involved, while that caused by dominance \times dominance epistasis is

$$\sigma_{SD}(x, y) = \Delta_{xy}^2 \sigma_{DD}^2 \quad (7.11e)$$

because there are two dominance but no additive effects involved. Letting n be the number of additive effects and m be the number of dominance effects in a type of gene action, the expression for the covariance between relatives becomes

$$\begin{aligned}\sigma_G(x, y) &= \sum (2\Theta_{xy})^n \Delta_{xy}^m \sigma_A^2 \sigma_D^m \\ &= 2\Theta_{xy}\sigma_A^2 + \Delta_{xy}\sigma_D^2 + (2\Theta_{xy})^2\sigma_{AA}^2 \\ &\quad + 2\Theta_{xy}\Delta_{xy}\sigma_{AD}^2 + \Delta_{xy}^2\sigma_{DD}^2 + (2\Theta_{xy})^3\sigma_{AAA}^2 + \dots\end{aligned}\quad (7.12)$$

Drawing from the coefficients Θ_{xy} and Δ_{xy} given in Table 7.2, explicit expressions for the genetic covariances of common types of relatives are given in Table 7.3. Although these expressions are only expanded to include two-locus epistasis, several things are immediately apparent. First, gene action involving dominance only rarely contributes to the covariance between diploid relatives. It requires that each parent of x be related to a different parent of y (i.e., a nonzero value of Δ_{xy}). Such relationships (full sibs, double first cousins [the offspring from a pair of brothers married to a pair of sisters], and monozygotic twins) are said to be **collateral**. Second, the coefficient for σ_{AA}^2 declines more rapidly with the distance of the relationship than does that for σ_A^2 . As noted in Chapters 4 and 5, because the additive genetic variance is a function of all higher-order types of gene action, these results should not be misconstrued to mean that the resemblance between relatives is influenced only slightly by dominance and epistatic gene action.

The most useful feature of the expressions in Table 7.3 involves their different coefficients, which permit the estimation of the different variance components from linear combinations of different observed genetic covariances between relatives. For example, ignoring higher-order epistasis and environmental sources of covariance, $8 \times$ [parent-offspring covariance $- (2 \times$ half-sib covariance)] has an expected value of $8[(\sigma_A^2/2 + \sigma_{AA}^2/4) - 2(\sigma_A^2/4 + \sigma_{AA}^2/16)] = \sigma_{AA}^2$. Similarly, $2 \times [(4 \times$ half-sib covariance) $-$ (parent-offspring covariance)] has an expected value of σ_A^2 . Subsequent examples in this chapter will illustrate the utility of these kinds of manipulations. With marker-based estimates of Θ and Δ , an even more refined analysis is possible. For example, Figure 7.5 showed that pairs of full sibs vary in their realized Θ value, and one can use this variation to decouple the effects of shared genes from shared environments in humans (Visscher et al. 2006; Chapter 32.)

THE EFFECTS OF LINKAGE AND GAMETIC-PHASE DISEQUILIBRIUM

In deriving the Kempthorne-Cockerham equation (7.12), we assumed that the constituent loci are freely recombining and in gametic-phase equilibrium. We now consider the extent to which the interpretation of observed covariances between relatives needs to be modified in the face of violations of these assumptions. In most practical situations, we have little if any information on either linkage or gametic-phase disequilibrium among the (almost always unknown) loci underlying a particular trait (although we can survey general genomic patterns of LD using random SNPs; Chapter 5) so the following theoretical results provide our only guidance as to the potential seriousness of the matter.

The most complete analyses of this problem were developed by Gallais (1974) and Weir and Cockerham (1977), who allowed for inbreeding as well as linkage and gametic-phase disequilibrium. Neither study went beyond two-locus relationships, as even then the algebraic and notational complexities are enormous. We will take a simpler approach than these authors, first considering the consequences of linkage under the assumption of gametic-phase equilibrium and then evaluating some of the consequences of disequilibrium. In both cases, we will assume that mating is random and that all of the remaining assumptions of the Kempthorne-Cockerham model are fulfilled.

Linkage

Under the assumption of gametic-phase equilibrium, linkage influences only the epistatic components of genetic covariance, which depend upon the multilocus gene *combinations*

inherited through gametes. Recall that in Equation 7.12, each component of epistatic variance is weighted by a coefficient of the form $(2\Theta_{xy})^n \Delta_{xy}^m$. The term Θ_{xy}^n is the probability that randomly drawn pairs of genes (one from x and one from y) will be simultaneously identical by descent at n loci, the 2^n coming in because of diploidy. Similarly, Δ_{xy}^m is the probability that the genotypes of x and y are simultaneously identical by descent at m loci. These definitions of the joint probability of events at multiple loci as the products of probabilities at individual loci assume that identity by descent is *distributed independently* among loci. However, for linked loci, *identity by descent is expected to be positively correlated* because the genes at such loci tend to be inherited together, a feature known as **identity disequilibrium**. Thus, for linked loci, we can anticipate that the multilocus coefficients $(2\Theta_{xy})^n \Delta_{xy}^m$ must be too low, to a degree depending on the recombination frequency.

A general solution to this problem requires the use of **digenic descent coefficients**, which define the probability that two nonalleles (genes from different loci) are copies of genes that were originally contained in the same gamete. Such a condition is known as **equivalence by descent**. The formal theory of digenic descent, developed in great detail by Cockerham and Weir (1968, 1973, 1977a) and Weir and Cockerham (1968, 1969, 1973, 1974, 1977), is algebraically and notationally complex. We will rely upon an approach that is less general but more transparent (Schnell 1961, 1963; Van Aarde 1975). There is no easy entry into this field, but for the adventuresome we suggest the reviews of Weir and Cockerham (1977, 1989) as a starting point.

For our purposes, it will suffice to examine an arbitrary pair of loci, A and B . Consider an individual whose two-locus genotype resulted from the fusion of $A_m B_m$ and $A_f B_f$ gametes (m for mother, f for father). If the recombination fraction for the two loci is c , then this individual will produce gametes in frequencies: $p(A_m B_m) = p(A_f B_f) = (1 - c)/2$ and $p(A_m B_f) = p(A_f B_m) = c/2$. Two gametes randomly drawn from this individual can have four possible identity-by-descent relationships. Identity exists at both loci with probability $p(A, B) = p^2(A_m B_m) + p^2(A_f B_f) + p^2(A_m B_f) + p^2(A_f B_m)$, at neither locus with probability $p(0, 0) = 2p(A_m B_m)p(A_f B_f) + 2p(A_m B_f)p(A_f B_m)$, at only the A locus with probability $p(A, 0) = 2p(A_m B_m)p(A_m B_f) + 2p(A_f B_m)p(A_f B_f)$, and at only the B locus with probability $p(0, B) = 2p(A_f B_f)p(A_m B_f) + 2p(A_f B_m)p(A_m B_m)$. Letting $\lambda = (1 - 2c)^2$ —where $\lambda = 1$ for complete linkage—these gametic-identity probabilities simplify to

$$p(A, B) = p(0, 0) = \frac{1 + \lambda}{4} \quad (7.13a)$$

$$p(A, 0) = p(0, B) = \frac{1 - \lambda}{4} \quad (7.13b)$$

all of which reduce to $1/4$ with free recombination ($c = 0.5$).

As a specific application of gametic identity probabilities, we will consider the case of full sibs, first evaluating the covariance caused by additive \times additive epistasis. As noted previously, for each pair of loci, the genotypic value of each sib contains four ($\alpha\alpha$) terms, so there are 16 combinations of these terms between individuals. For each of the 16 combinations, the quantity of interest is the joint probability that randomly drawn A genes (one from each sib) are identical by descent and that randomly drawn B genes (one from each sib) are identical by descent. We first draw a random A gene from the two sibs. As noted earlier, these are identical by descent with probability $1/4$. Given that identity by descent was obtained at the A locus, we now draw the B genes (again, one from each sib). These genes may be identical by descent through two routes—they may both derive from the parent from which the A genes were drawn (with probability $1/4$) or both from the opposite parent (with probability $1/4$). If they come from the same parent as the A gene, the B genes will be identical by descent with conditional probability

$$p(A|B) = \frac{p(A, B)}{p(A)} = \frac{p(A, B)}{p(A, B) + p(A, 0)} = (1 + \lambda)/2$$

If they come from the opposite parent, they are identical by descent with probability $1/2$.

Table 7.4 Coefficients for the components of genetic covariance between (noninbred) relatives modified to account for linkage (first five columns) and gametic-phase disequilibrium (last two columns). Here GP represents grandparent, GC represents grandchild, $\lambda = (1 - 2c)^2$, and t is the generation number for the common ancestors. Blanks denote values of zero.

Relationship	σ_A^2	σ_D^2	σ_{AA}^2	σ_{AD}^2	σ_{DD}^2	$\sigma_{A,A}(0)$	$\sigma_{D,D}(0)$
Parent–offspring	$\frac{1}{2}$		$\frac{1}{4}$			$\frac{(1-c)^t}{2}$	
GP – GC	$\frac{1}{4}$		$\frac{1+\lambda^{1/2}}{16}$			$\frac{(1-c)^t}{4}$	
Great GP – great GC	$\frac{1}{8}$		$\frac{(1+\lambda^{1/2})^2}{64}$			$\frac{(1-c)^t}{8}$	
Half sibs	$\frac{1}{4}$		$\frac{1+\lambda}{16}$			$\frac{(1-c)^t}{4}$	
Full sibs, dizygotic twins	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{2+\lambda}{8}$	$\frac{1+\lambda}{8}$	$\frac{(1+\lambda)^2}{16}$	$\frac{(1-c)^t}{2}$	$\frac{(1-c)^{2t}}{4}$
Uncle (aunt)–nephew (niece)	$\frac{1}{4}$		$\frac{1+\lambda(1+\lambda^{1/2})/2}{16}$			$\frac{(1-c)^t}{4}$	
First cousins	$\frac{1}{8}$		$\frac{1+(1+\lambda)^3}{128}$			$\frac{(1-c)^t}{8}$	
Double first cousins	$\frac{1}{4}$	$\frac{1}{16}$	$\frac{3+(1+\lambda)^3}{64}$	$\frac{1+(1+\lambda)^3}{128}$	$\frac{(1+\lambda)^4}{256}$	$\frac{(1-c)^t}{4}$	$\frac{(1-c)^{2t}}{16}$
Second cousins	$\frac{1}{32}$		$\frac{1+(1+\lambda)^5}{1024}$			$\frac{(1-c)^t}{32}$	
Monozygotic twins	1	1	1	1	1	$(1-c)^t$	$(1-c)^{2t}$

Summing up, the probability of digenic identity by descent for two full sibs is $1/4 \cdot 1/4 \cdot [(1 + \lambda)/2 + 1/2] = (2 + \lambda)/32$. Given such identity, the contribution to the genetic covariance is $E[(\alpha\alpha)^2]$, which is equivalent to one-fourth of the additive \times additive genetic variance. Thus, after taking account of all 16 pairs of additive \times additive effects, the covariance between full sibs resulting from additive \times additive epistasis is $16 \cdot [(2 + \lambda)/32] \cdot [\sigma_{AA}^2/4] = (2 + \lambda)\sigma_{AA}^2/8$. With free recombination ($\lambda = 0$), this reduces to $\sigma_{AA}^2/4$, our previous result (Table 7.3). Linkage ($\lambda > 0$) causes the additive \times additive genetic variance to be inflated.

We next consider the additive \times dominance covariance between full sibs. Again, from Equation 7.10a, for any pair of loci, there are 16 combinations of $(\alpha\delta)$ terms in the two sibs. If the two individuals possess an additive \times dominance effect that is identical by descent, the contribution to the genetic covariance will be $\sigma_{AD}^2/4$. To evaluate the probability of such an event, we need to determine the joint probability of drawing single genes that are identical by descent at one locus and genotypes that are identical by descent at the second locus. Such identity is only possible in the eight comparisons for which the additive effects involve the same locus in x and y . As usual with full sibs, the probability of drawing two genes (one from each sib) that are identical by descent is $1/4$. Also as noted above, given that the genes drawn at one locus are identical by descent, the genes for the second locus that descended from gametes from the same parent are identical by descent with probability $(1 + \lambda)/2$. The other pair of genes at the second locus are identical by descent with probability $1/2$. Summing up, the additive \times dominance covariance becomes $8 \cdot (1/4) \cdot [(1 + \lambda)/2] \cdot (1/2) \cdot (\sigma_{AD}^2/4) = (1 + \lambda)\sigma_{AD}^2/8$, which reduces to $\sigma_{AD}^2/8$ with free recombination.

Finally, we note that covariance between the single $(\delta\delta)$ term for a pair of loci requires that both sibs inherit maternal gametes that are identical by descent at both loci, and similarly for the paternal gametes. Each of these events occurs with probability $p(A, B)$. Therefore, the dominance \times dominance covariance between full sibs is $p^2(A, B)\sigma_{DD}^2 = (1 + \lambda)^2\sigma_{DD}^2/16$.

The coefficients for full sibs and for other common relationships are summarized in Table 7.4. The main conclusion to be drawn from these results is that *except for parent-offspring and monozygotic twin relationships, linkage inflates the covariance between relatives, unless there are no epistatic sources of genetic variance*. Linkage has no influence on the parent-offspring covariance because the two individuals always have exactly one gene identical by descent at each locus, and therefore share exactly one of the four additive \times additive effects. For grandparent-grandchild and half-sib relations, relative to the situation with free

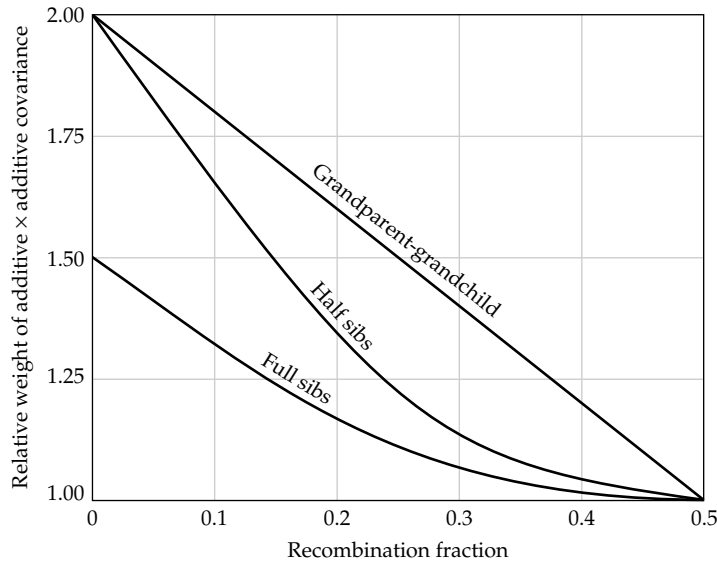


Figure 7.7 Inflation of the additive \times additive covariance between relatives caused by linkage, obtained by use of the coefficients in Table 7.4.

recombination, the additive \times additive covariance can be inflated as much as twofold with completely linked loci ($\lambda = 1$), whereas with full sibs the inflation can be no greater than 50% (Figure 7.7). It is of interest to note that whereas the expressions for the covariance of half sibs and for grandparent-grandchild are identical under free recombination, the latter is inflated to a greater degree by linkage unless linkage is complete ($\lambda = 1$). Thus, at least in principle, a comparison of these two types of covariance may shed some light on the presence of linkage for quantitative-trait loci.

Gametic-Phase Disequilibrium

Further complications arise when loci are in gametic-phase disequilibrium because this can cause a covariance between the effects of genes carried in the same gamete. The problem has been addressed by Weir et al. (1980), who ignored epistasis and assumed random mating, as we do below. In Chapter 5 (Equations 5.23a and 5.23b), we saw that in the presence of gametic-phase disequilibrium, the total genetic variance can be expressed as

$$\sigma_G^2 = \sigma_A^2 + \sigma_{A,A} + \sigma_D^2 + \sigma_{D,D} \quad (7.14)$$

where $\sigma_{A,A}$ is the contribution due to covariance of additive effects of nonalleles within gametes (the **additive disequilibrium covariance**), and $\sigma_{D,D}$ is the contribution of covariance due to dominance effects of different loci within individuals (the **dominance disequilibrium covariance**).

We now consider the genetic covariance between relatives derived from a base population displaying gametic-phase disequilibrium. The covariance between relatives attributable to equilibrium additive and dominance genetic variance (σ_A^2 and σ_D^2) is the same as given above, so it is only necessary to consider the additional contributions resulting from $\sigma_{A,A}$ and $\sigma_{D,D}$. For situations in which the study population is being maintained in a steady state of disequilibrium from generation to generation (by processes such as natural selection, migration, and/or nonrandom mating; WL Chapters 16 and 24), the modifications to the theory are simple—all of the preceding formulations still apply, except that $(\sigma_A^2 + \sigma_{A,A})$ replaces σ_A^2 , and $(\sigma_D^2 + \sigma_{D,D})$ replaces σ_D^2 .

The more interesting situation arises when a study population, initially in gametic-phase disequilibrium, is allowed to mate randomly in an environment in which the forces

maintaining disequilibria are relaxed. In this case, the covariances between relatives change through time as recombination causes the disequilibrium components of covariance to decay towards zero. In the following, we will consider only the simple (and most common) situation in which the ancestors leading to a relationship are all members of the same generation; more general results can be found in Weir et al. (1980).

We start by considering the additive component of disequilibrium covariance. Returning to Equation 7.10a, we make the distinction that allele i at the first locus and allele k at the second locus are inherited in one gamete, whereas genes j and l are inherited in the other gamete. By using the expression for the variance of a sum, Equation 3.11b, the additive equilibrium genetic variance is defined to be

$$\sigma_A^2 = E[\alpha_i^2] + E[\alpha_j^2] + E[\alpha_k^2] + E[\alpha_l^2] \quad (7.15a)$$

On the other hand, the additive disequilibrium covariance in the base population is

$$\sigma_{A,A}(0) = 2E[\alpha_i\alpha_k] + 2E[\alpha_j\alpha_l] \quad (7.15b)$$

Note that terms involving cross-gamete expectations, e.g., $E[\alpha_i\alpha_j]$ and $E[\alpha_i\alpha_l]$, do not appear in this expression because their expected values are zero under the assumption of random mating (as there is no correlation *between* gametes). In Chapter 5, we found that in the absence of restoring forces, gametic-phase disequilibrium (*within* gamete correlation) is reduced by a fraction c after each generation of random mating. Thus, the additive disequilibrium covariance remaining after t generations is

$$\sigma_{A,A}(t) = (1 - c)^t \sigma_{A,A}(0) \quad (7.15c)$$

To obtain the covariance between relatives associated with the additive equilibrium component of variance, we take expectations of the cross-products of additive effects of genes in two individuals. A similar procedure is followed in obtaining the covariance due to $\sigma_{A,A}$, except that we now focus on pairs of genes at different loci, one in each member of the relationship,

$$\begin{aligned} \sigma_{A,A}(x, y, t) = & (E[\alpha_i^x\alpha_k^y] + E[\alpha_i^x\alpha_l^y] + E[\alpha_j^x\alpha_k^y] + E[\alpha_j^x\alpha_l^y] \\ & + E[\alpha_k^x\alpha_i^y] + E[\alpha_k^x\alpha_j^y] + E[\alpha_l^x\alpha_i^y] + E[\alpha_l^x\alpha_j^y]) \end{aligned} \quad (7.15d)$$

For any of the terms in this equation to be nonzero, the genes in them must be equivalent by descent. The key to solving Equation 7.15d is the fact that ***the probability of equivalence by descent for genes at different loci in different gametes is the same as the probability of identity by descent for alleles at the same locus***. The reason for this equivalence is that when parents produce a gamete pool, although the two loci (A and B) may be linked, under random mating the gene at the A locus in one parental gamete is independent of the gene at the B locus in a second gamete. Thus, the probability of any term in Equation 7.15d being nonzero is Θ_{xy} .

From Equation 7.15b, we see that nonzero terms of the form $E[\alpha_i\alpha_k]$ initially have expected values equal to $\sigma_{A,A}(0)/4$. At time t , however, each of the nonzero terms in Equation 7.15d has expected value $\sigma_{A,A}(t)/4$. Because there are eight terms, the covariance between relatives due to additive disequilibrium covariance is $8\Theta_{xy}\sigma_{A,A}(t)/4$, or from Equation 7.15c,

$$\sigma_{A,A}(x, y, t) = 2\Theta_{xy}(1 - c)^t \sigma_{A,A}(0) \quad (7.16)$$

where t denotes the number of generations that the common ancestors are removed from the base population. Thus, for example, the parent-offspring covariance (with $\Theta_{xy} = 1/4$) resulting from additive disequilibrium covariance is $\sigma_{A,A}(0)/2$ if the parents are members of the base population ($t = 0$), $(1 - c)\sigma_{A,A}(0)/2$ if the parents are second-generation individuals ($t = 1$), and $(1 - c)^2\sigma_{A,A}(0)/2$ if they are third-generation individuals ($t = 3$).

Derivation of the covariance between relatives resulting from dominance disequilibrium covariance follows the same logic just presented. However, in this case, the dominance disequilibrium covariance declines each generation to $(1 - c)^2$ of its previous value. This quadratic decline occurs because dominance disequilibria are only maintained if neither of the gametes fusing to form a zygote have undergone recombination. Thus, the dominance disequilibrium covariance is

$$\sigma_{D,D}(t) = (1 - c)^{2t} \sigma_{D,D}(0) \quad (7.17a)$$

and the covariance between relatives due to this disequilibrium is

$$\sigma_{D,D}(x, y, t) = \Delta_{xy}(1 - c)^{2t} \sigma_{D,D}(0) \quad (7.17b)$$

For example, the covariance between full sibs resulting from dominance disequilibrium covariance is $\sigma_{D,D}(0)/4$, $(1 - c)^2 \sigma_{D,D}(0)/4$, and $(1 - c)^4 \sigma_{D,D}(0)/4$, respectively, when the parents are members of the base population, second, and third generations.

The general coefficients for the disequilibrium covariances for common types of relationships, assuming the common ancestor to be a member of generation t (where $t = 0$ denotes the base population), are given in the last two columns of Table 7.4. Unlike all of the equilibrium components of genetic variance, which always cause positive phenotypic covariance between relatives, the components resulting from disequilibrium covariance may be positive or negative depending upon whether loci are in coupling or repulsion disequilibrium.

Strictly speaking, the modified coefficients in Table 7.4 apply to a single pair of loci with recombination frequency c . These expressions could be refined further to account for the influence of linkage and/or gametic-phase disequilibrium of all loci on the covariance between relatives by summing terms over all pairs of loci, weighting each locus by its specific set of variances and covariances. However, without detailed information on the map structure of the constituent loci for a quantitative trait, such refinements would be of little practical value. An alternative approximation can be obtained by assuming that the loci underlying the trait are distributed randomly throughout the genome and using the average recombination frequency \bar{c} in place of c (WL Chapter 24). Estimates of \bar{c} are given for a number of species in Table 11.2.

It may be argued that linkage is of little consequence for the resemblance between relatives in organisms with high chromosome numbers because most pairs of loci will lie on different chromosomes. In this case, $\lambda \simeq 0$, and there is no inflation of the epistatic components of covariance, regardless of the magnitude of gametic-phase disequilibrium. However, regardless of the degree of linkage, gametic-phase disequilibrium is of special concern, as theoretical arguments and some empirical data suggest that it may reach significant levels for characters under selection (Chapter 5; WL Chapters 16 and 24). In this case, even with free recombination ($c = 0.5$), the contribution of disequilibrium covariance to the resemblance between relatives is nonzero. In the following section, we consider a common situation in which the build-up of positive $\sigma_{A,A}$ is inevitable, even with unlinked loci.

Example 7.5. There is one case where the probability of equivalence by descent is unequal to the coefficient of coancestry, which renders Equations 7.17a and 7.17b inappropriate. In monozygotic twins, both members of the pair are products of the same gametes, so the genes inherited in one twin are not independent of those inherited in the other twin through the same parent. Because monozygotic twins have genetic effects at all loci identical by descent, the covariance between monozygotic twins is equivalent to the expressed genetic variance in the population in the present generation. For twins whose parents are members of the base population, the probability that each ancestral gamete contributing to the twin progeny has not experienced a recombination event (between two loci of interest) is $(1 - c)$. Therefore, the

covariance between monozygotic twins resulting from additive and dominance effects is

$$\begin{aligned}\sigma_A(MZ) &= \sigma_A^2 + (1 - c)\sigma_{A,A}(0) \\ \sigma_D(MZ) &= \sigma_D^2 + (1 - c)^2\sigma_{D,D}(0)\end{aligned}$$

where the disequilibrium covariances refer to the levels in the parental generation. The sum of these two quantities is the total expressed genetic variance in the population in the twin's generation (ignoring epistasis).

ASSORTATIVE MATING

Although we have assumed a randomly mating population throughout this chapter, it is not unusual for mate choice to be based on aspects of the phenotype. Often, individuals will choose mates whose phenotypes resemble their own (**positive assortative mating**, or **homogamy**). For example, Shine et al. (2001) found positive assortative mating for size in a Canadian population of red-sided garter snakes (*Thamnophis sirtalis*). A meta-analysis by Jiang et al. (2013) of 360 species-trait combinations spanning five animal phyla found that positive assortative mating was not uncommon, with a mean correlation between mates of 0.28. They found that negative assortative mating (**diassortative mating**, or **heterogamy**), wherein the phenotype correlation among mates was negative, was rare. In vertebrates, two notable exceptions involve body color: wolves in Yellowstone national park (Hedrick et al. 2016) and white-throated sparrows (Tuttle 2003). In plants, assortative mating based on flowering time is common because individuals often produce viable pollen for only short periods of time. Likewise, self-incompatibility systems could be regarded as a form of examples of heterogamy. Moura et al. (2021) offered the important caveat that patterns of assortative mating can change over time and space, and that the pooling of data over such different settings can easily result in biased estimates of the nature and strength of assortative mating.

In humans there are significant (and almost always positive) correlations between mates with respect to a number of traits, both morphological and behavioral, such as height, skin color, IQ, social status, religion, level of educational attainment, and conditions such as deafness (Spuhler 1968; Vandenberg 1972; Mare 1995; Schwartz and Mare 2005). Because a phenotypic correlation could be caused by either shared genes or shared environments, a subject of some debate among human biologists is the fraction of these associations that are entirely social (i.e., environmental) in origin. If mate choice is entirely social, there should be no genetic consequences. Conversely, if mate choice is based on phenotypes, then to the extent that phenotypes are heritable, there should also be a genetic impact for assortative mating, which we examine below.

Genomic data has recently provided some clarity on this debate (at least for some traits). Several studies (e.g., Domingue et al. 2014; Guo et al. 2014; Zou et al. 2015) showed that couples share an excessive number of anonymous SNP alleles relative to some random standard. However, as noted by Abdellaoui et al. (2014), population stratification (Chapter 20) can generate such an excess, and several of these studies did not fully account for its effect. Robinson et al. (2017) further noted that the expected theoretical correlation generated by assortative mating was sufficiently small that most of these studies were under-powered and likely detected some other signal (such as stratification) rather than one created by assortative mating. A more direct test involves using SNPs that are associated with a trait of interest (GWAS; Chapter 20), such as height. Tenesa et al. (2016) found a correlation between the SNP-predicted height of an individual with the phenotypic value of its mate. This was also seen in a more careful analysis by Robinson et al. (2017), who observed such a correlation for other traits as well, such as BMI (body-mass index) and educational attainment. Hence, at least for some traits in humans, there is a mate-choice component entirely based on phenotype.

Many assortative mating systems are selective, such that some phenotypes of one or both sexes have a greater ability to attract mates than others. Here, however, we confine our attention to nonselective assortative mating, i.e., we assume that all individuals have an equal opportunity to reproduce, the phenotypic distribution of their available mates being the only limitation. This is the type of assortative mating that Fisher (1918) and Wright (1921d) had in mind when they first attacked the problem from a quantitative-genetic perspective. Almost all of the basic results of the theory were produced by these two pioneers. Fisher's (1918) elaborate treatment of the subject is notoriously difficult, and some uncertainty still exists as to exactly what he meant to say (Moran and Smith 1966; Wilson 1973; Vetta and Smith 1974; Yengo and Visscher 2018).

In some ways, positive assortative mating is like inbreeding, but in other ways it is very different. A simple way to discriminate between the two is to note that inbreeding is choice of mates based on similar *genotypes*, while assortative mating is based on *phenotypes*. If there is genetic variance for the characters that are the targets of mate choice, then positive assortative mating must cause identical alleles to come together more often than in the case of random mating. This will cause an increase in the homozygosity of the population, just as inbreeding does. However, while prolonged inbreeding can result in a completely homozygous population, positive assortative mating will not generally induce such an extreme genetic structure. It definitely will not if there are many loci underlying the character upon which mate choice depends, if there is any variation for the trait due to sources other than additive genetic variance, or if the correlation between mates is less than perfect.

It was shown in Chapter 4 (Equation 4.23c) that inbreeding (in the absence of epistasis) can inflate the genetic variance of a population up to twofold. Strong positive assortative mating has the potential to cause an even greater inflation, while disassortative mating results in a reduction of the genetic variance. The change in variance brought about by assortative mating is primarily a consequence of a directional build-up of gametic-phase disequilibria. Positive assortative mating increases the coupling of genes with similar effects, i.e., induces a positive covariance between allelic effects at different loci, while disassortative mating leads to the proliferation of repulsion gametes, in which positive effects at one locus are balanced by negative effects at another. Inbreeding also causes a build-up of gametic-phase disequilibrium, but there is no tendency for extreme gamete types to be formed at the expense of balanced ones, or vice versa.

Example 7.6. Direct observation of positive disequilibrium for several human traits showing assortative mating was seen by Yengo et al. (2018). These authors started with informative SNP markers detected from previous GWAS analyses for the traits of interest (Chapter 20), and then examined the association between the number of **trait-increasing alleles (TIAs)** between even and odd-numbered chromosomes. This was done by computing the correlation, ϕ , between even vs. odd-chromosome based trait predictor scores (weighted sums of TIAs) within an individual. This contrast nicely controls for shared environmental effects, which should equally impact both the even and odd chromosome sets. The potential of population stratification was controlled by using the first 20 principal components of the marker matrix (this approach is examined in Example 8.13). Using a data set of $\sim 400,000$ unrelated individuals of European ancestry, ϕ values of 3.2% and 2.7%, respectively, were found for height and education attainment. These values are consistent with their predicted values under assortative mating (based on trait heritability, fraction of variation accounted for by the predictors, and the phenotypic correlation between mates; see Yengo et al. for details).

A further item of supporting evidence for disequilibrium was seen by comparing the within-individual correlation (ϕ) with the correlations between the SNP-based predictors for a married pair (r_m), the latter computed using an independent data set. Yengo et al. showed that the predicted relationship between these two metrics at assortative mating equilibrium should be $r_m \sim 2\phi$. The slope of a regression for 32 different traits of r_m on ϕ showed a slope of 1.8 (with a standard error of 0.2), consistent with this prediction.

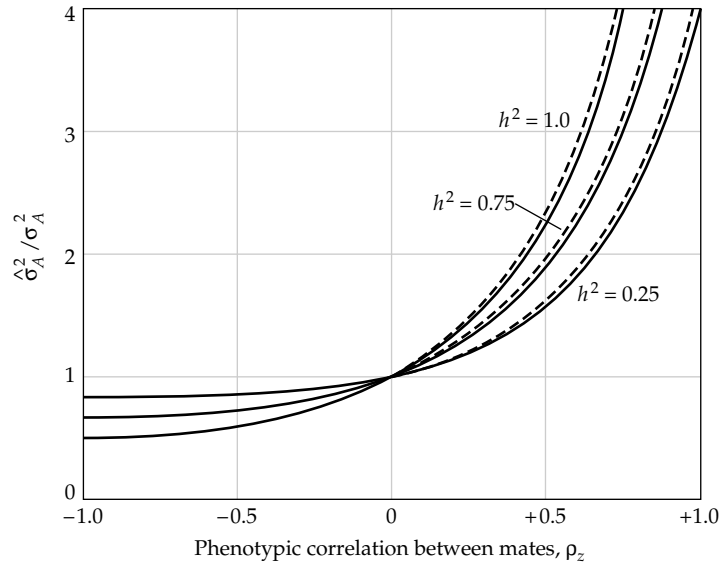


Figure 7.8 Inflation of the additive genetic variance at equilibrium under assortative mating relative to that in an otherwise identical random-mating population. Solid and dashed lines refer to 10 and 20 effective loci, respectively.

Theory

We now take a more quantitative look at these matters, starting with the consequences of assortative mating for the additive genetic variance and heritability (WL Chapters 16 and 24 provide a more detailed development of the theory). We restrict our attention to the situation in which interactions between loci are additive, because epistasis has not yet been incorporated into the theory. We define ρ_z and ρ_g to be, respectively, the phenotypic and additive-genetic correlations between mates and assume that the regression of phenotypes of mates is linear. We also suppose that there are n loci, each contributing equally to the genetic variance of the trait, and define the parameter $\gamma = 1 - [1/(2n)]$.

Starting from a random-mating base population in gametic-phase equilibrium with additive genetic variance σ_A^2 , a single generation of assortative mating will shift the additive genetic variance to

$$\sigma_A^2(1) = \left(1 + \frac{\rho_z h^2}{2}\right) \sigma_A^2 \quad (7.18)$$

where $h^2 = \sigma_A^2 / \sigma_z^2$ is the heritability under random mating (Crow and Kimura 1970). With continued assortative mating, the variance asymptotically approaches the equilibrium

$$\hat{\sigma}_A^2 = \frac{\sigma_A^2}{1 - \gamma \hat{\rho}_g} \quad (7.19a)$$

where $\hat{\rho}_g = \rho_z \hat{\sigma}_A^2 / \hat{\sigma}_z^2$ is the equilibrium genetic correlation between mates. This solution was first obtained by Wright (1921d) for unlinked loci with equivalent effects. Later, Crow and Felsenstein (1968), Bulmer (1980), and Nagylaki (1982) showed that the same result holds with linked loci with arbitrary effects, as long as n is replaced by n_e , the effective number of loci. Thus, the general form of Equation 7.19a holds regardless of the map structure of loci. Gimelfarb (1984) proved that the equilibrium is stable. Equation 7.19a can be rewritten as

$$\frac{\hat{\sigma}_A^2}{\sigma_A^2} = \frac{2 + \left[\sqrt{1 - 4\gamma\rho_z h^2(1 - h^2)} - 1 \right] / h^2}{2(1 - \gamma\rho_z)} \quad (7.19b)$$

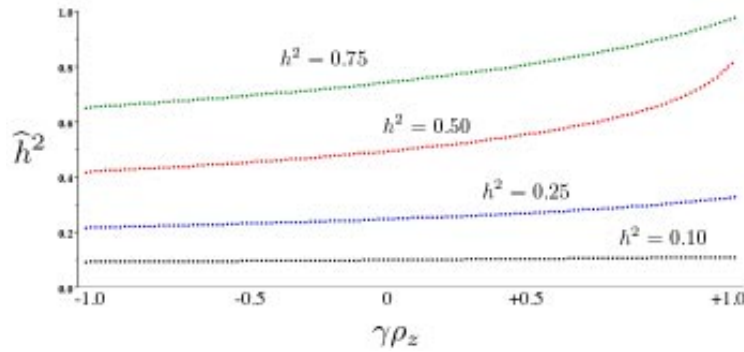


Figure 7.9 The equilibrium heritability, \hat{h}^2 , as a function of the heritability under random mating (h^2) and the strength of the assortative mating ($\gamma\rho_z$).

which gives the inflation of the additive equilibrium genetic variance relative to that in the random-mating base population (Figure 7.8). The difference $\hat{\sigma}_A^2 - \sigma_A^2$ is the additive disequilibrium covariance, $\sigma_{A,A}$, maintained by assortative mating, which is positive under positive assortative mating and negative under disassortative mating. Upon cessation of nonrandom mating, $\sigma_{A,A}$ decays to zero, and the additive variance returns to its random-mating value, σ_A^2 .

Three points can be gleaned from Figure 7.8. First, there is an asymmetry in the response to positive and negative assortative mating. Very strong positive assortative mating can inflate the additive genetic variance to nearly $2n_e\sigma_A^2$ if h^2 is also very high. However, negative assortative mating can depress the variance to no less than $\sigma_A^2/2$. The little empirical work that has been done in this area is qualitatively consistent with this expectation (Breese 1956; McBride and Robertson 1963; Example 7.7). Second, assortative mating must be fairly strong ($\rho_z^2 \geq 0.2$) and combined with high h^2 to induce much change in the variance. Third, the effective number of loci has a negligible effect unless it is very small.

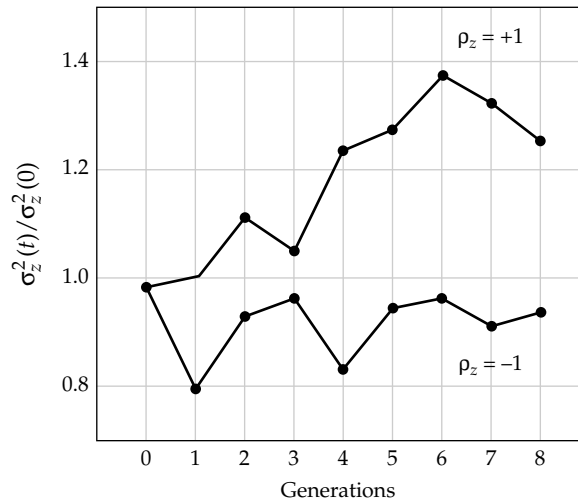
Finally, consider the equilibrium heritability, $\hat{h}^2 = \hat{\sigma}_A^2/\hat{\sigma}_z^2 = \hat{\sigma}_A^2/(\hat{\sigma}_A^2 + \sigma_E^2)$. Following Gomez-Raya and Burnside (1990), Equation 7.19b implies that

$$\hat{h}^2 = \frac{1 - \sqrt{1 - 4\gamma\rho_z h^2(1 - h^2)}}{2\gamma\rho_z(1 - h^2)} \quad (7.19c)$$

Figure 7.9 plots this as a function of the heritability under random mating (h^2) and the strength of the assortative mating ($\gamma\rho_z$). As expected from their impact on the sign of $\sigma_{A,A}$ positive assortative mating increased the heritability, while disassortative mating decreases it. Again, upon return to random mating, the heritability returns to its random-mating value as $\sigma_{A,A}$ decays to zero. Equation 7.19c implies that an observed heritability seen for a trait under assortative-mating equilibrium is a biased estimate of its random-mating value. For example, the observed heritability for human height is around $\hat{h}^2 = 0.8$. Assuming a phenotypic correlation of $\rho_z = 0.28$ between mates (Example 7.8), one can iteratively solve for h^2 using Equation 7.19c (taking $\gamma \simeq 1$), yielding $h^2 = 0.76$. Hence, the positive association between the effects of height alleles at different loci accounts for about 4% of the observed heritability. We show in Example 7.8 that these same parameters give a much larger change in the additive variation (around 30%). The reason for the smaller change in the heritability is that any change in the additive variance inflates (or deflates if $\rho_z < 0$) both the numerator and denominator of the ratio that defines the heritability.

Example 7.7. Gimelfarb (1984) performed an experiment in which lines of *Drosophila melanogaster* were artificially maintained under an absolute regimen of positive and nega-

tive assortative mating ($\rho_z \simeq \pm 1.0$). Each generation, 300 male and 300 female flies from each line were scored for abdominal bristle number, and the line was then propagated by performing 300 rank-ordered matings. The change in the phenotypic variance of abdominal bristle number over an eight generation period is shown in the following figure.



To what extent can these results be reconciled with the theory? Abdominal bristle number in *D. melanogaster* has been the subject of many quantitative-genetic studies. It appears to be nearly completely lacking in nonadditive genetic variance, and random-mating laboratory populations generally exhibit heritabilities of approximately $h^2 = \sigma_A^2/(\sigma_A^2 + \sigma_e^2) = 0.5$. Assuming that the effective number of loci underlying the trait is at least five or so, then $\gamma \simeq 1$, and Equation 7.19b predicts that complete negative assortative mating should ultimately reduce the additive genetic variance to approximately 70% of the level in the base population. Thus, scaling the original phenotypic variance to be one so that $\sigma_A^2 = \sigma_e^2 = 0.50$, we expect complete negative assortative mating to depress the additive genetic variance to approximately $\hat{\sigma}_A^2 = 0.70 \times 0.50 = 0.35$. From the figure, it can be seen that the *phenotypic* variance declined to approximately 90% of the level in the base population. This is close to the theoretical expectation that the phenotypic variance should be 85% of the base-population level, $\hat{\sigma}_z^2/\sigma_z^2 = (\hat{\sigma}_A^2 + \sigma_e^2)/(\sigma_A^2 + \sigma_e^2) = (0.35 + 0.50)/(0.50 + 0.50) = 0.85$.

Such an analysis cannot be performed for the line under positive assortative mating because it is less clear that an equilibrium phenotypic variance had been obtained. Nevertheless, a simple expectation can be pointed out for an experiment of this type. With absolute positive assortative mating, $\rho_z = 1.0$, and a heritability of 0.5, the expected inflation of the additive genetic variance predicted by Equation 7.19b is simply $(2n_e)^{1/2}$. Thus, if this experiment had been carried out to the point at which $\hat{\sigma}_z^2$ had stabilized, an estimate of the effective number of factors for abdominal bristle number would have been possible.

To this point, we have said nothing about the dominance component of variance. When only two loci contribute to the trait, σ_D^2 can change with assortative mating (Reeve 1961). However, Fisher (1918) argued that if the number of loci is even moderate, the effect will be negligible, and this was later confirmed by Vetta (1976). Thus, for all practical purposes, we can assume that the dominance genetic variance is unaltered by assortative mating.

In addition to creating gametic-phase disequilibrium, assortative mating causes the genotype frequencies for the selected trait to deviate from Hardy-Weinberg expectations. Positive assortative mating leads to a *heterozygote deficit*, while negative assortative mating leads to a *heterozygote excess*. Wright (1921d) and Nagylaki (1982) showed that if the effective number of loci is large, Hardy-Weinberg deviations will be minor unless both the

Table 7.5 Coefficients for the additive and dominance components of the covariance between relatives for an equilibrium population undergoing assortative mating. The equilibrium heritability is $\widehat{h}^2 = \widehat{\sigma}_A^2 / \widehat{\sigma}_z^2$. In the absence of assortative mating, $\rho_z = 0$, and all coefficients reduce to those in Table 7.3.

Relationship	$\widehat{\sigma}_A^2$	σ_D^2
Parent–offspring	$\frac{1}{2}(1 + \rho_z)$	
Grandparent–grandchild	$\frac{1}{4}(1 + \rho_z)(1 + \rho_z \widehat{h}^2)$	
Great grandparent–great grandchild	$\frac{1}{8}(1 + \rho_z)(1 + \rho_z \widehat{h}^2)^2$	
Half sibs	$\frac{1}{4}(1 + 2\rho_z \widehat{h}^2 + \rho_z^2 \widehat{h}^2)$	
Full sibs, dizygotic twins	$\frac{1}{2}(1 + \rho_z \widehat{h}^2)$	$\frac{1}{4}$
Uncle(aunt)–nephew(niece)	$\frac{1}{4}(1 + \rho_z \widehat{h}^2)^2$	$\frac{1}{8}\rho_z \widehat{h}^2$
First cousins	$\frac{1}{8}(1 + \rho_z \widehat{h}^2)^3$	$\frac{1}{16}(\rho_z \widehat{h}^2)^2$
Double first cousins	$\frac{1}{4}(1 + 3\rho_z \widehat{h}^2)$	$\frac{1}{16}$
Second cousins	$\frac{1}{32}(1 + \rho_z \widehat{h}^2)^5$	$\frac{1}{64}(\rho_z \widehat{h}^2)^4$
Monozygotic twins	1	1

phenotypic correlation between mates and h^2 are very large. This implies that *the vast majority of the change in genetic variance induced by assortative mating is a consequence of gametic-phase disequilibrium* rather than Hardy-Weinberg disequilibrium, i.e., of allelic associations within, rather than between, gametes.

The covariance between relatives in a population undergoing assortative mating has been considered by Fisher (1918), Crow and Felsenstein (1968), Gimelfarb (1981), and Yengo and Visscher (2018), with the last authors also considering the impact of assortative mating on X-chromosome coded traits. Nagylaki (1978) used path analysis to yield a particularly lucid overview for the case of additive gene action. The coefficients of the additive and dominance components of covariance are given for common types of relationships in Table 7.5. Positive assortative mating inflates the additive genetic covariance between all types of relatives, while disassortative mating depresses it. Moreover, for some types of relationships (uncle-nephew, first and second cousins), assortative mating induces dominance genetic covariance where there otherwise would be none. We emphasize that the additive variance appearing in the resemblance between relatives is $\widehat{\sigma}_A^2$, the variance in the current assortatively mating population, not σ_A^2 , the variance that would exist in a random-mating population. The latter can always be obtained by rearranging Equation 7.19a. Assuming large n_e ($\gamma \simeq 1$),

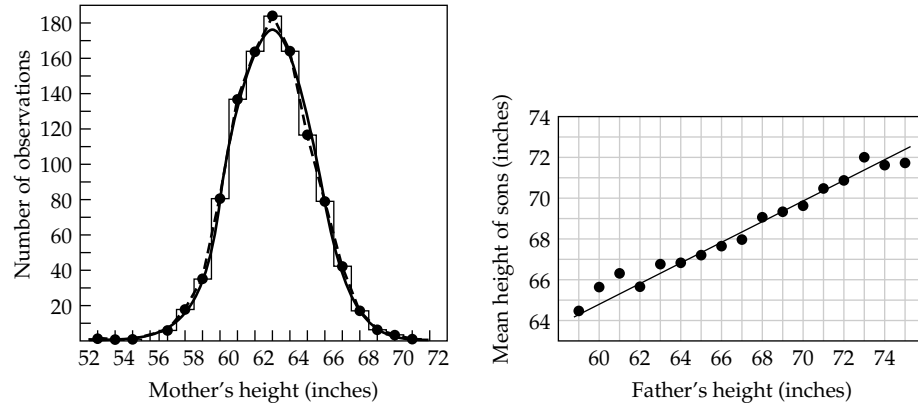
$$\sigma_A^2 = \left(1 - \frac{\rho_z \widehat{\sigma}_A^2}{\widehat{\sigma}_z^2}\right) \widehat{\sigma}_A^2 \quad (7.20)$$

Example 7.8. As an example of the application of the preceding theory, we consider an early data set on height in British families (Pearson and Lee 1903). Pearson recruited college students to obtain data from approximately 1300 families, recording whenever possible the stature of father, mother, and eldest son and daughter (ignoring offspring less than 18 years of age). This was a very large data set for the precomputer era, and it took two years to calculate the statistics by hand. The data are remarkable for their essentially normal distribution and for the linearity of the regressions between relatives.

In the figure below on the left, the histogram gives the observed data for maternal height, while the smooth curve is the fitted normal distribution. On the right, the straight line is the

least-squares regression of sons' height on paternal height, the data having been pooled into one-inch size classes for paternal height.

Pearson and Lee do not report the covariances between relatives, but rather the correlations, but this turns out to be quite convenient for the following analysis. The variance of height tends to be larger for males than for females, and this may be expected to influence the regressions depending upon which sexes are involved. However, because a correlation



coefficient is equivalent to the covariance between two standardized variables (each with unit variance), we can ignore this problem. The data in the following table indicate that the phenotypic correlations for all four sex-specific parent–offspring combinations are quite consistent, yielding a pooled value of 0.506 ± 0.011 . The three full-sib correlations are also consistent with each other and give the pooled value of 0.534 ± 0.019 . Finally, we note that there is highly significant assortative mating for height, the estimated value of ρ_z being 0.280 ± 0.028 . As a point of reference, a recent meta-analysis by Stulp et al. (2015) of 154 studies over a very wide diversity of cultures obtained a value of 0.25 for western populations (and an overall value of 0.23).

Relationship	r	SE
Parent–offspring:		
Father–son	0.514	0.022
Father–daughter	0.510	0.019
Mother–son	0.494	0.024
Mother–daughter	0.507	0.021
Full sibs:		
Brother–brother	0.511	0.042
Brother–sister	0.553	0.019
Sister–sister	0.537	0.033
Husband–wife	0.280	0.028

We first consider the additive genetic variance in the population. Recall that $\hat{\sigma}_z^2 = 1$ on the scale of analysis. From Table 7.5, we see that the expected covariance between parent and offspring is $(1 + \rho_z) \hat{\sigma}_A^2 / 2$. Setting this quantity equal to 0.506, substituting 0.280 for ρ_z , and rearranging, we obtain an estimate for $\hat{\sigma}_A^2$ of 0.791 ± 0.024 (the standard error being obtained by the Taylor expansion delta method; Appendix 1). Thus, unless epistasis is very strong (we have assumed it to be equal to zero), it appears that approximately 80% of the variance in human height is attributable to additive gene action.

Suppose assortative mating were to be eliminated. What is the expected equilibrium value of the additive genetic variance? Assuming a large effective number of loci ($\gamma \simeq 1$), and substituting the estimates of $\hat{\sigma}_A^2$ and ρ_z into Equation 7.20, we obtain $\sigma_A^2 \simeq 0.616$. The additive disequilibrium covariance for height is estimated as $0.791 - 0.616 = 0.175$, showing that through the creation of gametic-phase disequilibrium, assortative mating has induced an

approximately 28% increase in the additive genetic variance.

We now consider whether the full-sib data are consistent with an additive genetic model. Again from Table 7.5, we see that in the absence of dominance, the covariance between full sibs has an expected value of $\hat{\sigma}_A^2(1 + \rho_z \hat{h}^2)/2$. Because we are operating on a scale for which $\hat{\sigma}_z^2 = 1$, this value is also the expected correlation coefficient. Substituting 0.280 for ρ_z and 0.791 for $\hat{\sigma}_A^2$, we obtain the expectation 0.483 ± 0.024 . The difference between the observed and expected value ($0.534 - 0.483 = 0.051$) has a standard error equal to $[(0.024)^2 + (0.019)^2]^{1/2} = 0.031$ and cannot be considered to be significant. Thus, on the basis of the existing data, there are no grounds for rejecting the purely additive model.

In a more recent study, Roberts et al. (1978) performed a census of adult heights in families of a West African population, where assortative mating by height is relatively weak, $\rho_z \simeq 0.10 \pm 0.10$. The correlations between parents and offspring, (0.434 ± 0.015), and between full sibs, (0.378 ± 0.048), are somewhat lower than those observed in the study of Pearson and Lee. Using the procedures outlined above, these data yield estimates of $\hat{\sigma}_A^2 \simeq 0.789$ and $\sigma_A^2 = 0.727$. Thus, the expressed additive genetic variances in the two populations are very similar, although a smaller fraction is associated with disequilibrium covariance in the West African population. Roberts et al. also reported a half-sib correlation equal to 0.198 ± 0.059 , which can be used to provide a further check on the theory. Substituting the estimate of ρ_z and $\hat{h}^2 = 0.789$ into the half-sib expression in Table 7.5, the predicted correlation is 0.229, which is not significantly different from the observed value.

There is an interesting historical note regarding the data of Pearson and Lee. Fisher's (1918) demonstration that the data were quite consistent with a Mendelian hypothesis flew in the face of Pearson's notorious non-Mendelian philosophy. Fisher also extended the analyses to other types of relatives, illustrating their consistency with resemblances between first cousins and between grandparents and their grandchildren. However, the literature that Fisher cites for the latter correlations actually involves characters other than height (in one case, eye color)! Thus, at the time, Fisher apparently believed that there was a universal correlation for all characters within a species. This is an interesting twist because Pearson had similar feelings, as amply documented in Pearson and Lee (1903, p. 379): "Thus for most practical purposes we may assume parental heredity for all species and all characters to be approximately represented by a correlation of 0.5." We now know that this is far from the truth.

POLYPLOIDY

Because of the high incidence of polyploidy in plants, some attention needs to be given to its effects on the covariance between relatives. We will confine our comments to single-locus effects in a randomly mating tetraploid population and assume that the four alleles at each locus assort independently during meiosis. The results for this special case were first worked out by Kempthorne (1953, 1957).

As in the case of diploids, the genotypic values of tetraploids can be partitioned into several independent effects defined (as was the case for diploids) in a least-squares sense. However, because four alleles are present at each locus, in addition to additive (single-gene) and dominance (gene-pair) effects, triallelic and quadra-allelic effects must be considered (Chapter 5). Thus, the single-locus genotypic value of individual x is written as

$$G_{ijkl}(x) = \mu_G + [\alpha_i^x + \alpha_j^x + \alpha_k^x + \alpha_l^x] + [\delta_{ij}^x + \delta_{ik}^x + \delta_{il}^x + \delta_{jk}^x + \delta_{jl}^x + \delta_{kl}^x] \\ + [\gamma_{ijk}^x + \gamma_{ijl}^x + \gamma_{ikl}^x + \gamma_{jkl}^x] + \tau_{ijkl}^x \quad (7.21)$$

As in the case of diploids, the expected value of each of the effects is zero. Thus, the additive and dominance genetic variances for a locus are defined to be $\sigma_A^2 = 4E[\alpha^2]$ and $\sigma_D^2 = 6E[\delta^2]$, while the trigenic and quadragenetic variances are $\sigma_T^2 = 4E[\gamma^2]$ and $\sigma_Q^2 = E[\tau^2]$.

The covariances between tetraploid relatives are obtained in the same manner as in the case of diploids. For example, the additive genetic covariance involves 16 cross-products of additive effects (four in individual x times the four in individual y). Each of these cross-products has expectation $E[\alpha^2]$ if the two genes are identical by descent, the probability of

Table 7.6 Coefficients for Equation 7.22 for describing the covariance between tetraploid relatives. (After Kempthorne 1957.)

Relationship	Θ_{xy}	Δ_{xy}	φ_{xy}	ϕ_{xy}
Parent–offspring	$\frac{1}{8}$	$\frac{1}{36}$	0	0
Grandparent–grandchild	$\frac{1}{16}$	$\frac{1}{216}$	0	0
Full sibs, dizygotic twins	$\frac{1}{8}$	$\frac{1}{27}$	$\frac{1}{48}$	$\frac{1}{36}$
Half sibs	$\frac{1}{16}$	$\frac{1}{216}$	0	0
Uncle (aunt)–nephew (niece)	$\frac{1}{16}$	$\frac{1}{648}$	0	0
Monozygotic twins	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{4}$	1

which is Θ_{xy} , and expectation zero otherwise. Thus, the expected additive genetic covariance between relatives is $16\Theta_{xy}E[\alpha^2] = 4\Theta_{xy}\sigma_A^2$. In the case of digenic interaction, there are 36 cross-products to consider, each of which has expectation $E[\delta^2]$ if both members of the pair of alleles in x are identical by descent with those in the pair from y (probability Δ_{xy}). Therefore, the digenic covariance is $36\Delta_{xy}E[\delta^2] = 6\Delta_{xy}\sigma_D^2$. Letting φ_{xy} be the probability that all members of random three-gene sets are identical by descent, and ϕ_{xy} be the probability that the complete single-locus tetraploid genotypes of x and y are identical by descent, the trigenic and quadrigenic covariances between relatives are $4\varphi_{xy}\sigma_T^2$ and $\phi_{xy}\sigma_Q^2$. Thus, assuming random mating, and no epistasis, linkage, or gametic-phase disequilibrium, the covariance between tetraploid relatives can be summarized as

$$\sigma_G(x, y) = 4\Theta_{xy}\sigma_A^2 + 6\Delta_{xy}\sigma_D^2 + 4\varphi_{xy}\sigma_T^2 + \phi_{xy}\sigma_Q^2 \quad (7.22)$$

The coefficients for a number of common relationships are given in Table 7.6.

An interesting consequence of polyploidy is that the covariance between noncollateral relatives (individuals that do not share both parents) can be influenced by dominance. This effect occurs because parents pass two alleles on to their progeny. Thus, the covariance between parent and offspring is $\sigma_A^2/2 + \sigma_D^2/6$, as compared to $\sigma_A^2/2$ in the case of diploidy. Trigenic and quadrigenic variance components only appear in expressions of resemblance when collateral relatives are involved.

ENVIRONMENTAL SOURCES OF COVARIANCE BETWEEN RELATIVES

Up to now, we have focused entirely on the genetic causes of resemblance between relatives. There are, however, many circumstances in which the environmental effects on the phenotypes of relatives are correlated. This is an especially pernicious complication in human genetics. The most obvious situation arises when full sibs are raised in a common familial environment, but the resemblance between other types of relatives can also be modified by shared aspects of the environment. For example, in a spatially heterogeneous environment, the resemblance between parents and offspring may be exaggerated by environmental effects if the latter do not randomly disperse following birth. In humans, and probably in other vertebrates, cultural transmission can create a continuity in the behavioral phenotypes of parents and offspring as well as among more distant relatives. In some circumstances, the environment may actually deflate the resemblance between relatives. Consider, for example, trees with limited seed-dispersal abilities. Parent trees that have grown large due to fortuitous circumstances, such as germination in a light gap, may tend to create relatively poor microhabitats for their offspring as a consequence of shading, attraction of herbivores, and so on.

In an effort to deal with the complications of cultural transmission, human quantitative geneticists have gone to great extremes to incorporate various types of environmental covariance into expressions for the resemblance between relatives (Rao et al. 1974; Eaves 1976; Cloninger et al. 1979a, 1979b; Eaves et al. 1988). The result is a family of models that contain large numbers of parameters. Because the number of observable relationships must equal or exceed the number of parameters for there to be any hope of estimating the latter, these models are of little practical utility for most organisms. However, in humans the difficulties are not insurmountable. Due to the frequency of divorce, adoption, and fostering by relatives, and to the availability of records on parentage, many kinds of familial relationships are observable.

Here we consider some of the possibilities, under the assumption of additive gene action, and negligible linkage, gametic-phase disequilibrium, and inbreeding. To illustrate the major points, we will rely on path analysis (Appendix 2), drawing specifically from examples given in Rao et al. (1974). The basic feature of path analysis that we will exploit is its ability to partition the correlation between any two variables into a series of pathways connecting the two variables through causal components. As noted during our above discussion of coefficients of coancestry, the proportional contribution of each pathway to the correlation between relatives is simply equal to the product of various path coefficients and correlation coefficients along the pathway. We start by providing some general definitions, and then show, for several types of relationships, how path analysis can yield expressions for the expected phenotypic correlations in terms of causal genetic and environmental components. We then complete our discussion of environmental effects by providing a worked example involving a large data set on human birth weight.

Before proceeding, we consider how the phenotypic variance can be partitioned in terms of path coefficients. As usual, we consider an individual's phenotype (z) to be the sum of its genotypic value (G), general (shared) environmental effects (E), and special environmental effects (e). However, we now partition G into three quantities: the population mean (μ_G), the mean genotypic (breeding) value of the parents (\bar{A}), and the deviation of G from \bar{A} caused by segregation of parental genes (S). Thus, an individual's phenotypic value is expressed as

$$z = \mu_G + \bar{A} + S + E + e \quad (7.23)$$

The path coefficients from \bar{A} , S , E , and e to the phenotype are denoted by g , s , c , and d , respectively (for expected genotypic value, deviation due to segregation, common environmental effect, and residual deviation). Because S is a random genetic deviation, it is uncorrelated with the other components of Equation 7.23. However, genotype-environment correlation may exist between \bar{A} and E when offspring are raised by their own parents, and we denote it by ρ_{GE} .

The phenotypic variance among individuals raised by their biological parents is

$$\sigma_z^2 = \sigma_{\bar{A}}^2 + \sigma_S^2 + \sigma_E^2 + \sigma_e^2 + 2\sigma_{\bar{A},E} \quad (7.24a)$$

Dividing all terms by σ_z^2 provides an expression for the partitioning of the phenotypic variance of offspring raised by their own parents,

$$1 = g^2 + s^2 + c^2 + d^2 + 2g\rho_{GEC} \quad (7.24b)$$

In this expression, g^2 , s^2 , c^2 , and d^2 , are, respectively, the fractions of the phenotypic variance attributable to midparent breeding values, segregational, general and special environmental effects, while $2g\rho_{GEC} = 2\sigma(G, E)/\sigma_z^2$ is the fraction resulting from genotype-environment covariance.

Equation 7.24b requires modification in the case of progeny raised in adoptive homes. Under the assumption that there is no correlation between offspring genotype and the environment provided by an adoptive home, then $\rho_{GE} = 0$, and the phenotypic variance of adopted children can differ from that of progeny raised by their own parents. For example, if $\rho_{GE} > 0$ in intact families, the variance among adopted children will be reduced. To

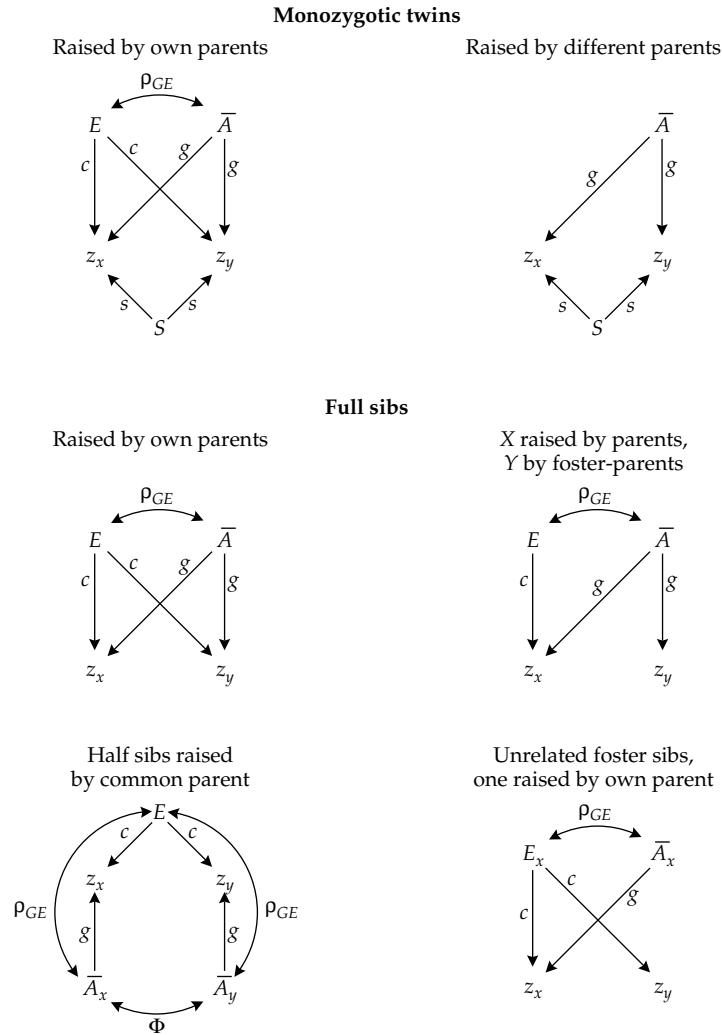


Figure 7.10 Path diagrams for the phenotypic correlation between individuals x and y . All terms are defined in the text.

account for this, Equation 7.24b needs to be modified to ensure that the sum of the various paths is equal to one. For adopted children, the modification is accomplished by letting

$$1 = \theta^2(g^2 + s^2 + c^2 + d^2) \tag{7.24c}$$

where $\theta^2 = 1/(1 - 2g\rho_{GE}c)$ is the ratio of the phenotypic variances for the two types of offspring. Multiplication by θ^2 has the effect of dividing each of the causal path contributions by the phenotypic variance of progeny raised in adoptive homes rather than by the phenotypic variance of progeny raised by their biological parents.

Finally, we note that it is useful to define

$$h^2 = g^2 + s^2 \tag{7.25}$$

as the heritability of the trait, i.e., the fraction of the total phenotypic variance attributable to additive genetic differences among individuals. With assortative mating,

$$g = h\sqrt{(1 + \rho_g)/2} \tag{7.26a}$$

$$s = h\sqrt{(1 - \rho_g)/2} \tag{7.26b}$$

where, as before, ρ_g is the additive-genetic correlation between mates (i.e., the correlation of their breeding values). Under random mating, $g = s = h/\sqrt{2}$.

To illustrate how these general relationships can be extended to the description of the expected phenotypic correlation between individuals, we now focus on the path diagrams for four specific relationships (Figure 7.10). For simplicity, the diagrams only include those factors that contribute jointly to the phenotypes of both members of a pair of individuals. Thus, the residual environmental deviation (e) never appears, while S (as a random genetic deviation) is only relevant in the case of monozygotic twins.

Because they are genetically identical, monozygotic twins (or clonemates) raised by their natural parents share the same general environmental effects, midparent value, and segregation deviation. Allowing for genotype-environment covariance, and summing over all pathways between z_x and z_y , we obtain the phenotypic correlation,

$$\rho(MZ) = s^2 + g^2 + c^2 + 2gc\rho_{GE} = h^2 + c^2 + 2gc\rho_{GE} \quad (7.27)$$

Thus, the correlation between monozygotic twins raised by their biological parents is not particularly informative, because it is a function of additive genetic variance (h^2), variance due to shared environment (c^2), and genotype-environment correlation $2gc\rho_{GE}$.

Now consider the situation in which monozygotic twins are separated at birth, with each being raised in a different adoptive home. The removal of the common-environment effect eliminates the path $z_x \leftarrow E \rightarrow z_y$ (Figure 7.10), and as noted above, the absence of genotype-environment correlation changes the phenotypic variance in the subpopulation of such twins. Thus, the expected phenotypic correlation between twins raised by different foster parents is $h^2\theta^2$. An estimate of the heritability can be acquired after factoring out θ^2 (the ratio of phenotypic variances, defined above). Expressions for twins living in other combinations of home environments are given in Table 7.7. (Note that when one member of a relationship is living with its parents and the other is living in an adoptive home, the correlation is multiplied by θ , rather than θ^2 , because only one member of the pair is from a subpopulation with modified variance.) Chapter 24 treats the issue of twin analysis in considerable detail.

We next consider the correlation between full sibs. The path diagram in this case is identical to that for monozygotic twins, except that the sibs, being products of different gametes, do not share the segregational deviation (S). Thus, the expected correlation between full sibs raised by their biological parents is

$$\rho(FS) = g^2 + c^2 + 2gc\rho_{GE} \quad (7.28)$$

As in the case of monozygotic twins, this expression is simplified in situations where one or both sibs are raised in adoptive environments. For example, suppose one sib is raised by its biological parents, while the other is raised by unrelated foster parents (Figure 7.10). This eliminates the paths $z_x \leftarrow E \rightarrow z_y$ and $z_x \leftarrow \bar{A} \leftrightarrow E \rightarrow z_y$. Moreover, because the second sib comes from a segment of the population without genotype-environment correlation, all of the path coefficients leading to it must be multiplied by θ . Summing over the two paths between x and y , the phenotypic correlation becomes $\theta(g^2 + gc\rho_{GE})$. The issue of sib analysis will be covered in detail in Chapter 23.

A slight complication arises in the case of half sibs. The correlation ρ_g between midparent values (\bar{A}_x and \bar{A}_y) is no longer one, because there is only one common parent. Rao et al. (1974) showed it to be $\phi = (1 + 3\rho_g)/[2(1 + \rho_g)]$, which reduces to $1/2$ under random mating. Summing over all paths between z_x and z_y (Figure 7.10), the correlation between half sibs raised by the parent creating the common environmental effect is

$$\rho(HS) = \phi g^2 + c^2 + 2gc\rho_{GE} \quad (7.29)$$

Again, various simplifications arise when one or both members of the sib pair are raised by foster parents (Table 7.7). For example, when each sib is raised in a different adoptive

Table 7.7 Expected phenotypic correlations between sibs in terms of path coefficients. All coefficients are defined in the text except b , which is the correlation of environments provided by parents of half sibs. Completely additive gene action is assumed. (After Rao et al. 1974.)

Relationship	Phenotypic correlation
Monozygotic twins:	
Reared by own parents	$h^2 + c^2 + 2gc\rho_{GE}$
One reared by own parents, one by foster parents	$(h^2 + gc\rho_{GE})\theta$
Raised by different foster parents	$h^2\theta^2$
Reared together by foster parents	$(h^2 + c^2)\theta^2$
Full sibs:	
Reared by own parents	$g^2 + c^2 + 2gc\rho_{GE}$
One reared by own parents, one by foster parents	$(g^2 + gc\rho_{GE})\theta$
Reared by different foster parents	$g^2\theta^2$
Half sibs:	
Raised by common parent	$\phi g^2 + c^2 + 2gc\rho_{GE}$
One reared by own parents, one by foster parents	$(\phi g^2 + gc\rho_{GE})\theta$
Reared by different foster parents	$\phi g^2\theta^2$
Reared apart by own parents	$\phi g^2 + c^2b + gc\rho_{GE}$
Unrelated foster sibs:	
Reared together by same foster parents	$c^2\theta^2$
Reared together by parents of one of them	$(c^2 + gc\rho_{GE})\theta$

environment, the only path between sib phenotypes is $z_x \leftarrow \bar{A}_x \leftrightarrow \bar{A}_y \rightarrow z_y$, so the correlation is simply $\phi g^2\theta^2$.

Finally, we note that unrelated individuals fostered by the same set of parents can resemble each other as a consequence of the common environment in the adoptive home. The expected phenotypic correlation between such individuals depends on whether the adoptive parents are the biological parents of one of the foster sibs (Figure 7.10, Table 7.7).

Even more complicated scenarios, for additional types of relatives, have been considered by the authors cited above. However, we assume that the basic principles are clear at this point, and will not pursue these any further. One very notable aspect of the models outlined above is their ability to provide estimates of genotype-environment correlation when data are available on sibs raised in various types of home environments. For example, when covariances are observed for monozygotic twins living in the four types of environmental settings outlined in Table 7.7, joint estimates of h^2 , c^2 , and ρ_{GE} can be obtained by setting the observed correlations equal to their expected values and solving. Modifications of all the expressions in Table 7.7 are necessary, however, in the presence of significant sources of nonadditive genetic variance. We close this section with an example of a character whose expression is strongly influenced by shared environmental effects.

Example 7.9. Several large and independent studies have been performed on human birth weight. As can be seen in the following table, the estimated correlations between relatives are quite consistent among studies. For example, the five available full-sib correlations have a narrow range of 0.47 to 0.52. Mi et al. (1986) performed large analyses on several ethnic groups in Hawaii and found only minor differences among them for the correlations between relatives. We will therefore pool the independent estimates where they exist. In keeping with the linear

model just outlined, we will assume that dominance and epistatic sources of variance are of negligible importance. In the absence of conflicting data, we will also assume that assortative mating and genotype-environment correlation are negligible, and that general environmental effects are only transmitted through mothers. Under these assumptions, $\rho_G = 0$, $g^2 = h^2/2$, $\rho_{GE} = 0$, and $\theta = 1$.

Relationship	Estimated Correlations	Prediction
Full sibs ²⁻⁶	0.50, 0.52, 0.47, 0.48, 0.48	$\frac{h^2}{2} + c^2 = 0.50$
Maternal half sibs ³	0.58	$\frac{h^2}{4} + c^2 = 0.42$
Paternal half sibs ³	0.10	$\frac{h^2}{4} = 0.08$
Maternal first cousins ^{2,6}	0.14, 0.13	$\frac{h^2}{8} + \frac{c_G^2}{2} = 0.15$
Paternal first cousins ^{2,6}	0.02, 0.06	$\frac{h^2}{8} = 0.04$
Monozygotic twins ¹	0.67	$h^2 + c^2 = 0.65$
Dizygotic twins ^{2,3}	0.59, 0.66	$\frac{h^2}{2} + c^2 = 0.50$
Half sibs via monozygotic twin parents:		
Maternal ⁵	0.31	$\frac{h^2}{4} + c_G^2 = 0.30$
Paternal ^{5,7}	-0.03, 0.12	$\frac{h^2}{4} = 0.08$

The superscripts here correspond to the references: 1. Penrose (1954a); 2. Robson (1955); 3. Morton (1955a); 4. Billewicz (1972); 5. Nance et al. (1983); 6. Mi et al. (1986); 7. Magnus (1984).

We first consider the additive genetic variance. Inferences about this parameter must be derived from relationships for which shared environmental effects do not influence the covariance. Paternal half sibs and paternal first cousins satisfy these conditions. The expected correlations for these types of relatives are $h^2/4$ and $h^2/8$, respectively, where $h^2 = \sigma_A^2/\sigma_z^2$. Because the observed correlations are 0.10 and 0.04, we obtain independent estimates of h^2 of $4 \times 0.10 = 0.40$ and $8 \times 0.04 = 0.32$. Also available is an average correlation of 0.05 for the offspring of monozygotic twin brothers. Such individuals are genetically equivalent to paternal half sibs (the fathers are different individuals, but identical genetically), so this result yields an additional estimate of $h^2 = 4 \times 0.05 = 0.20$. Averaging over all three types of relationship, $h^2 \simeq 0.30$, i. e., additive genetic variance appears to account for approximately 30% of the phenotypic variance.

The data make it very clear that aspects of the maternal environment have a substantial influence on birth weight. For example, the correlation between maternal half sibs is several times greater than that between paternal half sibs, and the same pattern is seen for maternal versus paternal first cousins. The total variation caused by the maternal environment can be obtained from the maternal half-sib correlation, 0.58, whose expectation is $(h^2/4) + c^2$. Subtracting out the additive genetic contribution, $c^2 = 0.58 - (0.30/4) = 0.50$. Pooling this with an independent estimate of $c^2 = 0.20$ obtained by Magnus (1984), we estimate $c^2 \simeq 0.35$. Thus, aspects of the mother (in excess of the genes that she contributes to her offspring) account for approximately 35% of the variance in birth weight. We examine the genetics of maternal effects in detail in Chapter 28.

There are two ways to partition the maternal effects variance into genetic and environmental components, $c^2 = c_G^2 + c_E^2$. First, monozygotic twin sisters provide the same genetic environment but different home settings for their progeny, which are genetically equivalent to maternal half sibs. The expected correlation between half sibs via this route is therefore $(h^2/4) + c_G^2$. Subtracting $h^2/4$ from the observed correlation, we obtain an estimate of the variance caused by genetic maternal effects, $c_G^2 = 0.31 - (0.30/4) = 0.23$. Second, the covariance between maternal first cousins is unaffected by common maternal environment, but is influenced by half the genetic maternal variance because the mothers are full sibs; their expected correlation is therefore $(h^2/8) + (c_G^2/2)$. Again equating observed and expected correlations, we obtain a second estimate $c_G^2 = 2[0.14 - (0.30/8)] = 0.20$. Thus, of the maternal effects variance, approximately two-thirds $(0.215/0.35)$ appears to be caused by the effects of the

maternal genotype on the uterine environment that an offspring experiences.

The causes of approximately 35% of the variance remain to be identified. Offspring sex, birth order, and gestation age account for 2, 3, and 4% of the variance, respectively (Penrose 1954a; Morton 1955a; Billewicz 1972; Magnus 1984). Some of these sources of variance presumably fall in the environmental maternal-effects category. Relatively lower weights for first-born offspring account for another 5% of the variation. Morton (1955a) has argued for the existence of dominance genetic variance, but the following argument suggests that this source of variance is negligible. In principle, the correlation between full sibs is $(\sigma_A^2/2 + \sigma_D^2/4 + \sigma_E^2)/\sigma_z^2$. However, from the above $(\sigma_A^2/2 + \sigma_E^2)/\sigma_z^2 \simeq (0.30/2) + 0.35 = 0.50$, which accounts for the mean observed correlation of 0.49.

The approach that we have taken to analyze these data is not very rigorous from a statistical standpoint, our main objective being to illustrate how correlations derived from several types of relatives can be used to estimate components of variance. Nevertheless, when the estimates of h^2 , c^2 , and c_G^2 are substituted into the expressions for the expected correlations between relatives, the overall fit to the data is quite good (last column in the preceding table). Thus, variation in human birth weight appears to be largely a function of additive gene action, maternal effects, and special environmental effects, each of which accounts for about a third of the total variance.

THE HERITABILITY CONCEPT

We have now seen, in theory and by example, that the analysis of a series of relationships provides the basis for partitioning the phenotypic variance into its elementary components. In practice, however, we are often confronted with difficulties, aside from the problem of finite resources, that prevent us from ever obtaining exact estimates of variance components. Some of the variance, such as that caused by higher-order epistatic interactions, is essentially beyond reach in a statistical sense. Nevertheless, with appropriate experimental designs, most of the fundamental sources of variance (additive and dominance genetic variance, and environmental variance due to common familial environments) can be approximated to a good degree, and levels of confidence attached to them. Most practical applications of quantitative genetics have been concerned with only the additive genetic component of the phenotypic variance, with the remaining components being treated as noise. The ratio σ_A^2/σ_z^2 has come to be known as the **heritability** of a trait (more precisely, the **narrow-sense heritability**).

This brings us to an important conceptual issue that has plagued the field of quantitative genetics almost since its inception (Feldman and Lewontin 1975; Bell 1977; Jacquard 1983). The preoccupation with the additive component of genetic variance stems from the desire for a parameter that describes the genetic resemblance between parents and offspring. At the close of Chapter 3 it was shown that the slope of a regression of offspring phenotype on average parental phenotype has expected value σ_A^2/σ_z^2 , provided that gene action is purely additive and all of the assumptions underlying the Kempthorne-Cockerham model are met. Moreover, we showed that if selection changes the mean phenotype in the parental generation by S units, the expected evolutionary advance in the offspring generation (relative to that of the parents before selection) is $S\sigma_A^2/\sigma_z^2$. Based on this reasoning, many studies have accepted uncritically the slope of a midparent-offspring regression ($b_{o\bar{p}}$) (or equivalently, twice the slope of mother-offspring or father-offspring regression, $2b_{op}$) as an estimate of σ_A^2/σ_z^2 . However, over the last several pages, we have found that the validity of this interpretation requires, among other things, random mating, gametic-phase equilibrium, absence of additive \times additive epistatic genetic variance, and absence of shared environmental effects. We cannot expect all of these conditions to be fulfilled in many natural populations.

Jacquard (1983) provided a useful discussion of the problems of interpretation of $b_{o\bar{p}}$ and $2b_{op}$ and suggested that these statistics simply be labeled **biometric heritability** without prejudice regarding the mechanisms causing similarity. However, because of the fundamental importance of the ratio σ_A^2/σ_z^2 , particularly in selection theory, we will continue to call

Table 7.8 Independent estimates of h^2 , obtained with Equation 7.30, for temperature tolerance in the marine copepod *Eurytemora affinis*. (From Bradley 1978.)

Relationship	Females	Males
Parent–offspring	0.14 ± 0.44	0.72 ± 0.26
Full sibs	0.20 ± 0.05	0.82 ± 0.04
Paternal half sibs	0.40 ± 0.18	0.84 ± 0.35
Maternal half sibs	0	0.73 ± 0.32

the latter quantity the heritability, denoting it as h^2 in keeping with Wright's original usage of h as the path coefficient σ_A/σ_z (Appendix 2). Providing an explicit definition eliminates the ambiguity of the usage of h^2 in theoretical contexts, but highlights the practical problems of estimation.

It should now be clear that heritabilities can often be approximated by reference to sets of relatives other than parents and offspring. The general logic behind this approach is that the first term in any genetic covariance expression is $2\Theta_{xy}\sigma_A^2$. Thus, under the assumption that the additive genetic variance is the dominant source of phenotypic covariance,

$$h^2 \simeq \frac{\text{Cov}(z_x, z_y)}{2\Theta_{xy}\text{Var}(z)} \quad (7.30)$$

should provide a good approximation to the heritability. Violations of the assumptions of the ideal additive model will usually cause $\text{Cov}(z_x, z_y)/2\Theta_{xy}$ to be an upwardly biased estimator of σ_A^2 .

A simple means of evaluating the likelihood of bias in heritability estimates arises when estimates of the phenotypic covariance are available for more than one type of relative. Such a test was performed by Clayton et al. (1957) on abdominal bristle number in a laboratory population of *Drosophila melanogaster*. The estimates of h^2 derived from four types of relatives were consistent with each other: mother–daughter (0.54 ± 0.11), mother–son (0.48 ± 0.11), half sibs (0.48 ± 0.11), and full sibs (0.53 ± 0.07). For this population, the evidence is strong that approximately 50% of the total variance for abdominal bristle number is attributable to additive genetic variance and that the remainder is a function of special environmental effects.

A second example in which heritability estimates are consistent across relationships involves a study of the susceptibility of the marine copepod *Eurytemora affinis* to high temperature shock (Table 7.8). Within each sex, four different relationships give fairly consistent results, but there is a clear sexual dimorphism—approximately four times as much variance in males is accounted for by additive genetic variance as in females. The fact that estimates from full sibs and maternal half sibs are not inflated implies that dominance genetic variance and maternal-effects variance are of minor significance. Thus, the data are consistent with the hypothesis that $h^2 \simeq 0.2$ in females and 0.8 in males with the residual variance being attributable to special environmental effects.

Such results in which additive genetic variance is the only source of resemblance between relatives are by no means universal in quantitative-genetic analyses. We have already encountered a striking exception with human birth weight, and more will appear in the following chapters. While there is a general tendency for heritability estimates based on parent–offspring and full-sib analyses to be consistent with each other (Figure 7.11), in any particular study, it is incumbent upon the investigator to evaluate whether the inconsistencies between different estimates of h^2 are significant. High levels of dominance genetic variance often exist for fitness-related characters (Crnokrak and Roff 1995; WL Chapter 6).

For practical reasons, the components of variance of natural populations are frequently estimated by assaying a segment of the population in a laboratory setting. Although the goal of such studies is generally to infer the genetic properties of the wild population,

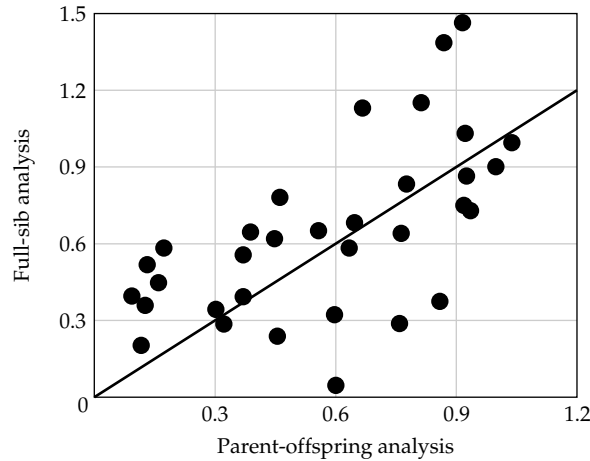


Figure 7.11 The relationship between heritability estimated as $2\text{Cov}(PO)/\text{Var}(z)$ from parent-offspring analysis and as $2\text{Cov}(FS)/\text{Var}(z)$ from full-sib analysis, for studies in which both estimates are available. Individual data points are for physiological and morphological characters for various natural populations of animals. The straight line gives the expected pattern under a perfect correspondence of the two estimates. (From Mousseau and Roff 1987.)

laboratory settings often impose a rather substantial change in the environment. It is tempting to speculate that heritability estimates derived from controlled laboratory experiments will be inflated relative to those expressed in the natural environment, where the environmental component of variance might be expected to be magnified by spatial and temporal heterogeneity. However, other outcomes are possible. For example, homeostatic mechanisms, such as habitat selection, which are operable in the field may be rendered inoperable in the laboratory. It is also conceivable that a shift in the environment may induce a change in the additive genetic variance by altering gene expression. We revisit the estimation of h^2 in natural populations using laboratory data in Chapter 22.

The few attempts that have been made to evaluate the sensitivity of heritability estimates to environmental change have yielded a diversity of results. Contrary to expectations, Mackay (1981) found that parent-offspring regressions for sternopleural bristle number and body weight in *Drosophila melanogaster* were increased by varying the environment temporally and spatially in the laboratory. For the same species, Coyne and Beecham (1987) found that the parent-offspring regression for abdominal bristle number was not affected by raising the parents in the laboratory (as opposed to the field), while that for wing length increased by 150%. The change was a consequence of a reduction in the environmental component of variance in the lab-reared parents. Simons and Roff (1994) also observed a general increase in the heritabilities of life-history traits when crickets were raised in the lab due to a reduction in the environmental component of variance as well as an increase in the genetic component. In a broad survey of the existing data on a diversity of organisms, Weigensberg and Roff (1996) found that there are no *systematic* differences in heritability estimates obtained in the laboratory and in the field; if anything, the latter tend to be slightly higher on average.

In another review, Hoffmann and Parsons (1991) found that heritabilities tend to increase in stressful environments. This observation may be of relevance to the interpretation of some laboratory analyses, in that the laboratory may constitute a form of stress in some cases. However, there are many exceptions to the pattern suggested by Hoffmann and Parsons. In natural populations of birds, for example, heritabilities of bone lengths and body size tend to decline, sometimes to undetectable levels, under poor growth conditions (Gebhardt-Henrich and van Noordwijk 1991; Larsson 1993). Thus, no strong generalizations emerge from existing studies as to how heritabilities are likely to change in laboratory vs. field situations, benign vs. harsh environments, novel vs. usual conditions, and so forth.

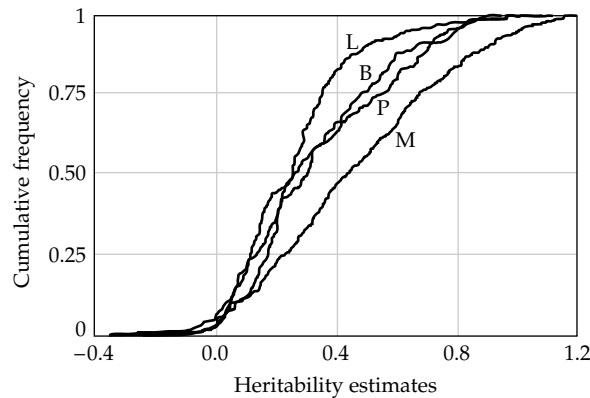


Figure 7.12 Cumulative frequency distributions for heritability estimates derived from numerous wild animal populations. L = life history, B = behavior, P = physiology, M = morphology. These data, from Mousseau and Roff (1987), do not include *Drosophila* studies, which yield a similar pattern (Roff and Mousseau 1987).

The best that can be said is that heritabilities do respond to environmental change, and that substantial care should be taken in extrapolating results beyond the environment in which they are obtained.

If one's sole interest in performing a quantitative-genetic analysis is to demonstrate that the character of interest is heritable, there is probably little point in expending the effort. The outcome is virtually certain. Almost every character in almost every species that has been studied intensely exhibits nonzero heritability. This should come as no surprise, because mutation brings in a small amount of new genetic variation each generation (Chapter 15; WL Chapter 28). The interesting questions remaining are, How does the magnitude of h^2 differ among characters and species, and why?

One weak generalization that has emerged is that morphological characters tend to have higher heritabilities than life-history traits, with behavioral and physiological characters falling at intermediate levels (Figure 7.12). Although there are numerous exceptions, these results are consistent with the intuitive concept that natural selection will most efficiently reduce the genetic variation for characters closely related to fitness by rapidly driving beneficial genes to fixation and eliminating deleterious ones (Robertson 1955; Fisher 1958; Falconer 1989). However, there are other explanations. As emphasized by Price and Schluter (1991), the relatively low heritabilities of life-history traits may be as much a consequence of relatively high levels of environmental variance as of unusually low levels of genetic variance for such traits. One possible reason for this is that the environmental variance of life-history traits is a function of the variance of all of the other morphological, physiological, and behavioral characters that influence their expression. Alternatively, characters closely related to fitness (life-history traits) may be relatively canalized genetically (Stearns and Kawecki 1994; Chapter 14), such that their expression is relatively insensitive to new mutations. This would result in low levels of genetic variation maintained under selection-mutation balance. Finally, while selection may derive the additive variance to lower values, it may not have a major impact on *nonadditive* genetic variances (WL Chapters 5 and 6).

Evolvability

In comparing the evolutionary potential of different traits/species, it is clearly desirable to use a dimensionless parameter, and one such measure is the heritability (which standardizes the additive variance by comparison to the phenotypic variance). Recall the traditional expression for the rate of evolution of a trait,

$$\Delta\mu = h^2 S \quad (7.31a)$$

Table 7.9 Heritabilities and coefficients of additive genetic (CV_A) and residual (CV_R) variation for representative traits in *Drosophila melanogaster*. Both CV_A and CV_R values are multiplied by 100. Here n is the number of studies, and the median estimates are reported. (After Houle 1992.)

Trait	n	h^2	CV_A	CV_R
Sternopleural bristles	21	0.44	8.39	7.97
Wing length	31	0.36	1.56	2.09
Fecundity	12	0.06	11.90	39.02
Longevity	7	0.11	9.89	27.73

This equation neatly separates the forces of selection from the properties of inheritance, such that h^2 can be viewed as the efficiency of the response to selection. The change in the mean in phenotypic standard deviations provides another useful descriptor of evolutionary potential, with

$$\frac{\Delta\mu}{\sigma_z} = h^2 i \quad (7.31b)$$

where $i = S/\sigma_z$ is the standardized selection differential, i.e., the change in the mean caused by selection in units of phenotypic standard deviations.

Houle (1992) has suggested that heritability may not be the best measure of the evolutionary potential of a trait, arguing for the use of the **coefficient of additive genetic variation**, i.e.,

$$CV_A = \frac{\sigma_A}{\mu} \quad (7.31c)$$

where σ_A is the square root of the additive genetic variance and μ is the mean of the trait. In other words, standardizing the additive variance by the *trait mean* (**mean standardization**), rather than the *trait variance* (**variance standardization**). He termed the CV_A metric (originally suggested by Charlesworth 1984) **evolvability**. Because it relies on the mean, this metric is limited to traits that are nonnegative and which have a natural value of zero. Height and weight are examples of traits that fit these restrictions, but date is not, because the zero date on a calendar is arbitrary. Using the definition of CV_A , Equation 7.31a can be rearranged to give the **proportional response** in a trait as

$$\frac{R}{\mu} = \frac{\sigma_A^2 S}{\sigma_z^2 \mu} = \left(\frac{\sigma_A^2}{\mu \sigma_z} \right) i = \left(\frac{\sigma_A}{\mu} \right) \left(\frac{S}{\sigma_z} \right) \left(\frac{\sigma_A}{\sigma_z} \right) = CV_A i h \quad (7.31d)$$

Example 7.10. As an example of the difference between the proportional response and the response expressed in standard deviations, consider two traits, both with $h^2 = 0.5$. Trait one has a mean of 10 and a variance of 4, while trait two has a mean of 50 and also a variance of 4. Suppose that $S = 4$, then the change in mean for both traits is $h^2 S = 2$, resulting a response of one standard deviation. Trait one has a new mean of 12, or 120% of its previous mean, while trait two has a new mean of 52, or 104% of its previous mean. Under the change in standard deviations framework, both traits show comparable responses, while as a proportion of the current mean, the change in trait one is much more dramatic.

Houle (1992) and Hansen et al. (2011) observed that h^2 is essentially uncorrelated with evolvability, so that a trait with a lower h^2 could still have high evolvability (i.e., potential for a significant proportional change in the mean), and vice versa. Strikingly, while life-history traits tend to have lower heritabilities, they often display *higher* evolvability values. Table 7.9 shows Houle's (1992) original analysis. Lower values for h^2 for life-history traits appear

to arise not from smaller additive variances, but rather from larger values of the residual (environmental plus nonadditive) variance. WL Chapter 6 examines this in more detail.

Finally, consider the special case in which the character is fitness (W). Recall from Chapter 3 that the selection differential is equivalent to the phenotypic covariance between the character and relative fitness ($w = W/\bar{W}$), i.e., $S = \sigma(z, w) = \sigma(z, W/\bar{W}) = \sigma(z, W)/\bar{W}$, where \bar{W} is mean fitness on the absolute scale. If z is fitness ($z = W$), then $\sigma(z, W) = \sigma(W, W) = \sigma^2(W)$, hence $S = \sigma^2(W)/\bar{W}$, where $\sigma^2(W)$ is the phenotypic variance of fitness. Equation 7.31d then reduces to

$$\frac{\Delta\bar{W}}{\bar{W}} = \frac{\sigma_A^2(W)}{\bar{W}^2} = \sigma_A^2(w) \quad (7.31d)$$

where $\sigma_A^2(W)$ and $\sigma_A^2(w)$ are the additive genetic variances of absolute and relative fitness, respectively. Thus, the proportional rate of evolution in mean fitness is equal to the squared coefficient of additive genetic variation of absolute fitness, or equivalently, to the additive genetic variance of relative fitness. This is **Fisher's (1958) fundamental theorem of natural selection**, which is examined in detail in WL Chapter 6.

These alternative formulations merely serve to illustrate that there are several ways to measure evolutionary potential, each of which has its own merits in particular contexts. All of the measures are interchangeable provided that information is available on the phenotypic variance, additive genetic variance, and mean phenotype. As emphasized by Houle (1992), however, many quantitative-genetic studies simply report the heritability of a trait, with no mention of the mean or variance components. This greatly limits the scope of investigation that can be performed with published data. The theory of selection response, including all of the above issues, is discussed in considerable detail in the second volume of this series, Walsh and Lynch (2018).