# Critical assumptions of classical quantitative genetics and twin studies that warrant more attention

Peter J. Taylor
Programs in Science, Technology & Values and Critical & Creative Thinking
University of Massachusetts, Boston, MA 02125, USA. peter.taylor@umb.edu

31 March 2009
Comments welcome

Abstract:
Quantitative genetics (QG) analyses variation in traits of humans, other animals, or plants in ways that take account of the genealogical relatedness of the individuals whose traits are observed. This article focuses on "classical" QG where the analysis of variation does not involve data on measurable genetic or environmental entities or factors. Classical QG analysis employs models of genes with simple Mendelian inheritance and direct contributions to the trait (assumption 1). However, the data analysed in QG are of traits, not genes, so it must be possible to analyze the variation without making reference to hypothetical genes. Five other assumptions then warrant new or renewed attention: 2. All other things being equal, similarity in traits for relatives is proportional to the fraction of all the genes that vary in the population which the relatives share, e.g., fraternal or dizygotic twins are half as similar as identical or monozygotic twins. 3. In human twin studies, genotype-environment-interaction variance can either be discounted or be incorporated into the heritability estimates. 4. High heritability of a trait can guide decisions about whether to investigate genetic factors underlying the development of that trait (and analogously for other components of variance). 5. When similarity among a set of close relatives (such as twin pairs) is associated with similarity of (yet-to-be-identified and measured) genes or genetic factors, those factors are the same from one set of relatives to the next. 6. In human twin studies, residual variance is a within-family environmental effect. Allowing for alternatives to these assumptions can change markedly the results or interpretation of classical QG. Lines of further investigation are noted for classical QG, for current QG analyses that incorporate measurable genetic or environmental factors, and for debates in social

science about how to accommodate transmission in families of both environmental and genetic factors. However, examination of those topics lies beyond the scope of this article.

---

Quantitative genetics (QG) is the analysis of variation in traits of humans, other animals, or plants in ways that take account of the genealogical relatedness of the individuals whose traits are observed. The variation in traits, such as IQ test scores, height, litter size, yield, etc., is analyzed for three primary purposes: to interpret outcomes of artificial and natural selection and guide further selective breeding; to assess the relative contributions of yet-to-be identified genetic and environmental entities or "factors" (to be defined shortly) underlying the development of the trait in question; and to decide whether to investigate what the specific factors are. Using the term "classical QG" to denote analysis that involves no data on measurable genetic or environmental factors, this article examines six assumptions of classical QG and plausible alternatives for each. The first assumption and its alternative are key to the subsequent discussion. Although classical QG analysis requires models of genes with simple Mendelian inheritance and direct contributions to the trait—assumption 1—the data analysed in QG are of traits, not genes, so it must be possible to analyze the variation without making reference to hypothetical genes. Once a gene-free QG alternative is considered, the other assumptions, most of which have been noted in some form before, warrant fresh attention. The alternatives can change markedly the results or interpretation of classical QG in ways that are difficult to pick up from standard accounts of QG (e.g., Falconer and Mackay 1996, Lynch and Walsh 1998) and human QG (Plomin et al. 1997, Rijsdijk and Sham 2002) and in accounts critical, to various degrees, of human QG (e.g., Layzer 1974, Jacquard 1983, Otto et al. 1995, Turkheimer 1998, Kamin and Goldberger 2002). Before laying out the six assumptions and the alternatives, some preliminary remarks are needed.

Scope and Exposition: The scope of this article does not extend to QG analyses that incorporate data on measurable genetic or environmental entities or factors (e.g., Plomin et al. 2003) or to debates in social science about how to accommodate transmission in families of both environmental and genetic factors (Turkheimer 2004). Although the assumptions and alternatives discussed here have potentially important implications for those topics (noted in the final section of the article), the focus here is on classical QG. This focus, which keeps us close

to first principles, seemed necessary in order to show that fresh perspectives are possible on issues that many researchers and critical commentators treat as settled or satisfactorily covered. By taking little for granted and by not attempting to encompass all of QG, the exposition should also be more accessible to readers who are not specialists.

In several places the discussion centers on the case of human twin studies, but, except where explicitly indicated (e.g., assumptions 3 and 6), the points are intended to apply to classical QG in general. As noted below, several other assumptions specific to twin studies have been extensively debated and reviewed elsewhere, and so are not examined here.

Terminology: Conventional terminology obscures some distinctions that are important to discussion of assumptions and alternatives, so some non-standard terms are employed. "Factor" is used in this article in a non-technical sense simply to refer to something whose presence or absence can, at least in principle, be observed or whose level can be measured. Measurable genetic factors include the presence of pairs of alleles (variant forms of a single gene) at a specific locus on paired (diploid) chromosomes, repeated DNA sequences, reversed sections of chromosomes, etc. Measurable environmental factors include income level of the family of upbringing, maltreatment when a child, etc.

In genetics a genotype is the set of genetic factors an individual possesses (or at least the subset held to be related to some given trait). In QG, however, the label “genotype” is applied to groups of individuals that are genetically identical ("pure lines") or whose mix of genetic factors can be replicated (such as an open pollinated plant variety), or to groups whose relatedness by genealogy can be characterized (such as human twins). No knowledge of actual genotypes is entailed in the QG use of the term. Similarly, the label “environment” is applied in QG to the situations or places in which the genotypes are raised without knowledge of the relevant environmental factors. In this article, the agricultural terms “variety” and “location” are used instead. A human variety consists of the offspring of a pair of parents; being raised within a family becomes an instance of a location.

In adopting the terms factor, variety, and location, the intention is to counter any conceptual slippage from analysis of observations of a given trait to claims about “genetic” and “environmental” differences, given that such claims suggest misleadingly that classical QG analyses of variation in traits address the measurable genetic and environmental factors involved in the development of the trait. (For a similar reason, phenotype is not used here to refer to the

traits.) This distinction between traits and underlying measurable factors is not clearly made or is not consistently maintained in most accounts of classical QG analyses, including accounts critical of human QG (see references cited earlier). Most notably, heritability, which is calculated in QG without reference to measurements on genetic (or environmental) factors, is commonly described in terms like the "contribution of genetic differences to observed differences among individuals" (Plomin et al. 1997, 83) or "fraction of the variance of a phenotypic trait in a given population caused by (or attributable to) genetic differences" (Layzer 1974, 1259). The terminology in this article is chosen to help keep clear that classical QG looks at differences among variety (or "genotypic") averages (for the trait averaged across locations and replicates), not at differences in genes or genetic factors. Translation from descriptive QG analyses to hypotheses about causal factors is far from direct and depends on assumptions of interpretation discussed in this article.

Analysis of variance versus path analysis: In QG the statistical analysis of variation in traits employs techniques of path analysis (or its generalization as structural equation modeling) or Analysis of Variance (ANOVA). Path analysis is more common in QG, but this article uses an ANOVA formulation of data analysis for two reasons: The ANOVA formulation helps keep the distinction clear between traits and measurable factors underlying the development of the trait; and it makes explicit that analyses of variation are based on additive ("linear") models that connect the values of the trait for an individual to the summation of several contributions. (Indeed, the exposition in this article involves no techniques of algebra, calculus, and statistics beyond simple sums of variables and use of subscripts to index some of those variables.) In any case, any ANOVA can be translated into a path analysis. (This translation is illustrated in the Appendix, which also makes the restrictive assumptions of path analysis explicit.)

Twins raised together versus twins raised apart: Although not the case for QG in general, the most important data for human QG have been observations of identical and fraternal twins. The similarity of pairs of identical or monozygotic (MZ) twins (which share all their genes) is compared with the similarity of pairs of fraternal or dizygotic (DZ) twins (which do not share all their genes). One variant of this comparison considers only twins raised together (in the same family); the other includes twins raised apart (at least for some of their lives). In both kinds of study, it makes intuitive sense that the more that the similarity of MZ twins exceeds the

similarity of DZ twins, the more that genetic factors are influencing the trait. (This can be said without identitying which factors are having that influence.)

In this article, when the six assumptions are discussed in relation to human QG, the context is the comparison of raised-together MZ and DZ twins. Five of the assumptions also pertain to the twins-raised-apart comparison (i.e., all except #3), but examination of the demanding additional assumptions or conditions required in those studies (and of studies of relatives of other degrees) is beyond the scope of this article. So is the assumption that the experience of growing up in the same family is no more similar for MZ twins than for DZ twins. Those assumptions and other contentious issues (which include correct ascertainment and representative sampling of zygosity of twins; definition of traits such as IQ and schizophrenia; and the size of "genotype-environment" [here: variety-location] correlations) have been debated extensively elsewhere (for accessible reviews, see Nuffield Council on Bioethics 2002, Parens 2004).

## Six assumptions and alternatives

The full meaning of the assumptions and their alternatives has to emerge through a systematic sequence of steps, but a brief overview may help orient readers to what lies ahead. Classical QG analysis requires models of genes with simple Mendelian inheritance and direct contributions to the trait (assumption 1). However, the data analysed in QG are of traits, not genes, so it must be possible to analyze the variation without making reference to hypothetical genes. Of course, genealogical relatedness has to be taken into account even under a gene-free QG. Classical QG assumes that, all other things being equal, similarity in traits for relatives is proportional to the fraction of all the genes that vary in the population which the relatives share, e.g., fraternal or dizygotic twins are assumed to be half as similar as identical or monozygotic twins (assumption 2). However, this proportionality is an unreliable heuristic; instead, similarity in traits associated with different degrees of genealogical relatedness can be left for empirical determination. In human studies, the conditions required for empirical determination also allow estimation of the variety-location-interaction variance ("genotype-environment-interaction" variance) component. That component is, however, often discounted or incorporated into the heritability estimates (assumption 3).

These first three assumptions are made in QG analysis. The other three are implicated in almost all interpretations of the results of the analysis: High heritability of a trait can guide decisions about whether to investigate genetic factors underlying the development of that trait (and analogously for other components of variance) (assumption 4); when similarity among a set of close relatives (such as twin pairs) is associated with similarity of (yet-to-be-identified and measured) genes or genetic factors, those factors are the same from one set of relatives to the next (assumption 5); and, in human studies, residual variance is a within-family environmental effect (assumption 6). Meaningful alternatives, which will be identified in due course, exist for each of these interpretive assumptions provided the distinction between traits and underlying measurable factors is kept clear.

**Assumption 1. QG analysis requires models of genes with simple Mendelian inheritance and direct contributions to the trait.** An alternative analysis takes genealogical relatedness into account without making reference to hypothetical genes or genetic factors.

Let us review the classical QG approach, then consider the alternative, "gene-free" QG analysis. The standard models of QG (Falconer and Mackay 1996; Lynch and Walsh 1998) can be constructed through a sequence of steps. The first step is to consider the case of a trait governed by a pair of alleles of a single gene (i.e., at a single locus) where all the individuals are raised in a single location. In that location, the genes contribute directly to the trait in the sense that the presence or level of such a trait depends only on whether the individual has two copies of one allele (i.e., is "homozygote" for that allele), two of the other, or one of each ("heterozygote"). For example, phenylketonuria (PKU) in humans is associated with having two copies of a non-functioning allele for the enzyme phenylalanine hydroxylase (PAH). The development of such individuals is extremely impaired by the level of phenylalanine present in normal diets. In this "normal-diet location" relatives will resemble each other more than unrelated individuals because if, say, a twin has PKU, both parents have at least one copy of the non-functioning PAH allele so the other twin is more likely to have two non-functioning PAH alleles than is an unrelated individual (i.e., one chosen at random from the population).

Resemblance of relatives for any given trait can be quantified as an "intraclass correlation." If the variation among all the individuals is divided into two parts, namely, the average variance for the trait within the class (e.g., twin pairs) and the variance among the

averages of the different classes, the intraclass correlation is the ratio of the among-class averages part to the sum of both parts. (This can be shown to be equivalent for classes of size two to the usual linear correlation of the two values when the order in each pair is arbitrary, as would be the case if we wanted to know the correlation of heights in same sex couples; Howell 2002.) For a trait governed by alleles at a single locus, the intraclass correlation for MZ twins is 1—there is no variation within the pairs—and for DZ twins is .5. (Recent research shows discordance between MZ twins at the genetic level, e.g., Bruder et al. 2008, but the simplication of an intraclass correlation of 1 for MZ twins is preserved in this article. Note also that the DZ intraclass correlation value is independent of the frequency in the population of the two alleles and of the degree of "dominance," which refers to degree that heterozygote individuals depart from the intermediate between the other two homozygote forms. The algebra is available from the author on request.)

Few traits are dictated only by alleles at a single locus, so the standard models of QG envisage the influence of alleles at many loci adding up to shape the traits to be analyzed. If each pair of alleles is assumed to add a small direct contribution to the trait in the sense that each contribution is independent of the others, then the ratio of the variation among twin-pair averages to the total variance is unchanged and the intraclass correlations for MZ and DZ pairs remain the same. (This assumption might be labeled a "sub-assumption" to distinguish it from the six main assumptions discussed in this article, but the context makes clear which level of assumption is being referred to so the prefix will not be added in this and later instances. Note also that, because these intraclass correlations incorporate no environmental or unsystematic influences, they have been termed "genetic similarity" or "genetic correlation." However, to avoid any risk of implying that similarity in traits as analyzed in classical QG has a direct relation with similarity in genetic factors, the ambiguous adjective "genetic" is not used in this article's discussion of similarity.)

Next, the models allow for some noise (from measurement error or unsystematic variation among the replicates of the variety), which, in the case of twins, is assumed to be equal for both kinds of twins, that is:

$$Y = V^{+} + E \quad (1)$$

where

Y is the total variance for the trait across all replicates,

E is the noise contribution, and

$V^+$ refers both to the variance among the MZ twin-pair averages and to the variance among the DZ twin-pair averages combined with the average variance within the DZ twin pairs. (The superscript will be explained in due course.)

The intraclass correlations in the single location are no longer 1 and .5, but are given by:

$$I_{MZ} = V^+ / Y \quad (2)$$

$$I_{DZ} = .5\ V^+ / Y \quad (3)$$

where I is the intraclass correlation and the subscript refers to MZ or DZ twins.

Finally, to allow for the variety to be raised in a number of locations, the standard models of QG incorporate into the equations a term for variance across locations of the average value of the trait in each location (or, in short, "among-location-average variance"). Again, the models assume that this term—denoted here as L—is equal for both kinds of twins, that is:

$$Y = V^+ + L + E \quad (4)$$

and

$$I_{MZ} = (V^+ + L) / Y \quad (5)$$

$$I_{DZ} = (.5\ V^+ + L) / Y \quad (6)$$

Now, if the intraclass correlations for the trait (in conventional terms: the "phenotypic correlations") are calculated from observations for a number of MZ and DZ twin pairs, then the fractions of the total variance made up by $V^+$ and L can be estimated using the following algebraic rearrangements of equations 5 and 6:

$$V^+/ Y = 2 (I_{MZ} - I_{DZ}) \quad (7)$$

$$L/ Y = 2 I_{DZ} - I_{MZ} \quad (8)$$

Equations 7 and 8 are the basis for the estimation of heritability and "shared environmental effect" (i.e., among-location-average fraction of the overall variance), respectively, using twin studies (e.g., Rijsdijk and Sham 2002).

Alternative sub-assumptions, such as MZ and DZ twins not having equal location and noise variances and the others mentioned at the end of the preliminary remarks, add complexities to the construction of this basic model. However, let us consider an alternative analysis that does not even begin from assumption 1. In this analysis the genealogical relatedness is taken into account without reference to the hypothetical multiple genes and the additional sub-assumptions

that went into the preceding construction of equations 7 and 8. Consider first the general case of an agricultural evaluation trial, where it is possible to observe a set of animal or plant varieties in each of a set of locations, and to raise replicates for each variety-location combination. Later, we will bring back in restrictions that match the situation in human twin studies.

If the trait is recorded in all replicates, the data can be fitted to an additive ("linear") model that connects the values of the trait for an individual to the summation of several contributions:

$$y_{ijk} = m + v_i + l_j + vl_{ij} + e_{ijk} \quad (9)$$

where $y_{ijk}$ denotes the measured trait y for the $i^{th}$ variety in the $j^{th}$ location and $k^{th}$ replication;

m is a base level for the trait;

$v_i$ is the additional contribution of the $i^{th}$ variety;

$l_j$ is the additional contribution of the $j^{th}$ location;

$vl_{ij}$ is the additional contribution from the $i,j^{th}$ variety-location combination—in statistical terms, the "variety-location-interaction" contribution; and

$e_{ijk}$ is an unsystematic or "noise" contribution adding to the trait measurement (see assumption 6 for further discussion of this contribution).

Such an additive model can be converted to a model that adds up variances related to these contributions, or reciprocally, to partition the variance of the trait into these component variances, i.e., the "analysis of variance" (ANOVA). The conversion to variances is made as follows: If each kind of contribution is uncorrelated with any of the others, and m is set at the average of the trait over all varieties, locations, and replicates, then the average of each of the other contributions is zero. If m is subtracted from both sides, which are then squared and divided by the total number of individuals to arrive at the average of these squared contributions, equation 9 translates into the following partitioning of <u>variance</u>:

$$Y = V + L + VL + E \quad (10)$$

where

Y denotes the variance of the $y_{ijk}$ observations as a whole,

V denotes the variance of the $v_i$ terms, etc.

When the observations are fitted to this model, dividing the right hand side of equation 10 by Y gives the fractions of the overall variance for, respectively, among-variety-averages (which is another name for heritability), among-location-average fraction ("shared environmental effect"), among-variety-location-interaction contributions, and noise (or unsystematic or residual effects).

As in any such partitioning, the results of fitting observations to an additive model are conditional on the specific set of varieties and set of locations observed—the results are not an indication of causes or properties of the varieties that apply more generally (Taylor 2006). One way to keep this conditionality in mind is to consider how $v_i$ is fit to the data. To use the simplest case, when the number of replicates observed for every variety-location combination is the same, the value of $v_i$ will be the average of $y_{ijk}$'s for variety i across all the observed locations and replicates minus the average over all varieties, locations, and replicates. In other words, the contribution $v_i$ is not a property of the variety i alone. Similarly, for any location contribution, the values that fit the data involve an average across all varieties. (Although partitioning of variance is often conducted without making explicit the original additive model, such as equation 9, or estimating the values of the various contributions that fit the data, the partitioning can always be related to such a model.)

Now consider a special case of the agricultural situation in which the varieties are replicated as MZ or DZ twins. Equation 9 becomes:

$$y_{ijk} = m + \bar{v}_i + t_{ik} + l_j + \bar{vl}_{ij} + tl_{ijk} + e_{ijk} \qquad (11)$$

where $y_{ijk}$, m, $l_j$, $e_{ijk}$ are as before;

$t_{ik}$ denotes an additional contribution from the $k^{th}$ twin (replicate) in the $i^{th}$ variety;

$tl_{ijk}$ is an additional contribution from the $k^{th}$ twin in the $i,j^{th}$ variety-location combination;

$\bar{v}_i$ and $\bar{vl}_{ij}$ replace the $v_i$ and $vl_{ij}$ contributions in equations 9. (The superscript indicates that the new contributions would tend to be smaller given the contributions of $t_{ik}$ and $tl_{ijk}$ to differences among twins [replicates].)

Note that for MZ twins $t_{i1} = t_{i2}$ and $tl_{ij1} = tl_{ij2}$. Let us assume for simplicity that the variance of the trait over all individuals is the same as in the general case, namely, Y. The partitioning of variance in equation 10 then becomes:

$$Y = \bar{V} + T + L + \bar{VL} + TL + E \qquad (12)$$

where $\bar{V}$ denotes the overall variance of the $\bar{v}_i$ terms, etc.

The formulas for intraclass correlations in this situation can be derived by identifying the contributions that do not vary within the twin pairs (and are thus included in the twin-pair averages), summing the variance of these contributions, and dividing by the overall variance:

$$I_{MZ} = (V^- + T + L + VL^- + TL) / Y \quad (13)$$

$$I_{DZ} = (V^- + L + VL^-) / Y \quad (14)$$

The difference between the formulas for MZ twins and DZ twins can be understood by recognizing that the equal $t_{ik}$ and equal $tl_{ijk}$ values for MZ twins means that these components do not vary within MZ twin pairs. Equations 12-14 can be rearranged to yield the following:

$$2 (T + TL) / Y = 2 (I_{MZ} - I_{DZ}) \quad (15)$$

$$(V^- - T + L + VL^- - TL) / Y = 2 I_{DZ} - I_{MZ} \quad (16)$$

$$E / Y = 1 - I_{MZ} \quad (17)$$

Equation 17 estimates E, the interpretation of which will be discussed later (under assumption 6). Equations 15 and 16, however, do not resemble the formulas given in equations 7 and 8 for the fractions of the total variance made up by $V^+$ and L (i.e., heritability and "shared environmental effect"). The connection emerges if the data meet the following empirical condition:

$$V^- + VL^- = T + TL = .5 (V + VL) \quad (18)$$

Under this condition, equations 15 and 16 simplify to:

$$(V + VL) / Y = 2 (I_{MZ} - I_{DZ}) \quad (19)$$

$$L / Y = 2 I_{DZ} - I_{MZ} \quad (20)$$

Equations 19 and 20 are now identical to 7 and 8 except that $V^+$ has been replaced by V + VL. (In the Appendix, the same result is derived using path analysis.)

What would be required to show empirical support for the equality condition in equation 18? Are there theoretical reasons for expecting the equality condition to be met? In other words, for the move from the basic agricultural model (equation 9) to the situation in which replicates are twins (equation 11), which splits the overall variety variance (V) and variety-location-interaction variance (VL) into two parts, should we expect to find that the two parts are equal? These questions, as well as the equivalence of $V^+$ to V + VL and whether heritability should be equated to $V^+/Y$ or V/Y are examined under assumptions 2 and 3 below.

What is important to note here, in relation to assumption 1, is that equations 19 and 20 allow observations to be analyzed in a way that takes into account the genealogical relatedness but makes no reference to or assumptions about hypothetical genes. No models of the contribution of such genes to the trait are needed in this derivation of the formulas that use intraclass correlations to estimate heritability and the fraction of the total variance made up by the among-location-average variance, L, (i.e., "shared environmental effect"). Moreover, gene-free derivations of analyses for comparing relatives of other degrees (e.g., full sibs versus half-sibs) can be derived by the same approach, each derivation having its own equivalent of the equality condition.

Let us now move from the agricultural case to the restricted circumstances of human twin studies, where each variety is observed only in a single, randomly chosen location (one location for each variety) and the twins are the two replicates for each of those variety-location combinations. (Non-random assignment of varieties to locations ["genotype-environment correlation"] is possible, but will not be considered here; Jacquard 1983.) Instead of partitioning the variance based on equation 11, we need to employ a variant of equation 11 that replaces the subscript j with the subscript i. Provided the locations are randomly chosen, the expected estimates of the variances in equation 12 remain the same, and, given this, the expected results of the formulas based on intraclass correlations (equations 19 and 20) are the same. In short, the alternative derivation in the agricultural case still holds in human twin studies. (The rationale and empirical basis of the equality condition, equation 18, still remains to be discussed; see assumptions 2 and 3.)

In summary, the formulas in classical QG for data analysis are based on models constructed through a series of steps that build on the case of traits governed by alleles at a single locus in a single location. Each step involves additional assumptions, e.g., in twin studies, unsystematic variation within twin pairs for MZ twin pairs is equal to that within DZ pairs. The same formulas, except for one modification, can be derived without referring to or making assumptions about hypothetical genes provided the partitioning of the variance meets a certain condition. In the example of twins, the condition is that, when averaged over all locations, the variance among DZ twin-pair-averages in the single location is equal to average variance within DZ twin pairs. (See assumption 2 for discussion of this condition; equivalent conditions are needed for gene-free analyses of data from relatives of other degrees.) The one modification

(which will be discussed under assumption 3) is that the term for variance among variety contributions (e.g., among the MZ twin pairs for the twin-pair averages) in a single location is replaced by the sum of the variances of variety and variety-location-interaction contributions as derived by fitting an additive model to the observations.

**Assumption 2. All other things being equal, similarity in traits for relatives is proportional the fraction of all the genes that vary in the population which the relatives share, e.g., fraternal or dizygotic twins are assumed to be half as similar as identical or monozygotic twins.** Alternatively, similarity does not have to follow this proportionality.

The discussion of this assumption and the alternative centers on MZ and DZ twins, but the issue of similarity in the trait in relation to the fraction of shared genes applies to relatives of any degree.

As noted earlier, for a trait observed at a single location and governed only by alleles at a single locus, the intraclass correlation for MZ twins is 1 and for DZ twins is .5. In this case, DZ twins are half as similar as MZ twins. Can we extrapolate from this fact and conclude that the same ratio holds for other kinds of traits? The answer might be yes if there were evidence for the assumptions built into the subsequent steps in the derivation of the standard QG formulas (i.e., assumptions such as, each pair of alleles adds a small contribution independently of the others). Such evidence is lacking, which is not surprising given the problems inherent in trying to discriminate among the contributions of many different loci (Lewontin 1974). In the absence of such evidence, QG researchers invoke a heuristic that connects the set of genes as a whole to similarities in the trait, namely, all other things being equal, DZ twins are half as similar as MZ twins <u>because</u> DZ twins share half the genes that vary in the population, while MZ twins share them all (e.g., Kendler and Prescott 2006, 42). (The common statement that DZ twins share half of their genes is not correct; it fails to acknowledge the large proportion of genes shared by all humans.)

The heuristic is not reliable. It is straightforward to invent plausible models of the contributions of multiple genes to a trait that result in ratios of DZ: MZ similarity that are not .5 and that vary considerably around their average. Consider, for example, a disease trait modeled in the following way: The trait occurs when the combined "dosage" from many loci exceeds a threshold, where each pair of alleles contributes a full, zero, or half dose according to whether

the alleles are, respectively, both the same for one variant, same for the other, or one of each. In this case, the intraclass correlation varies according to the frequency of alleles, level of dominance, and so on, and the varying values are predominantly well above .5. (The calculations are available from the author on request.) Of course, more complicated models of the interaction of genes and the timing of their contributions during development are possible. However, the point here does not depend on the validity of the model just mentioned or of any particular hypothetical model of multiple genes contributing to the trait. The reason that the heuristic is unreliable is that the relevant correlations need to be based on observed traits and, as such, cannot be directly given by the proportion of shared genes involved in the development of those traits. (For the same reason, heuristic values of the similarity of relatives of other degrees, which are ubiquitous in QG path analyses, are also unreliable.)

The alternative analysis without assumption 2 allows for similarities in a single location other than .5. This alternative can be achieved with the basic model under assumption 1 by allowing for variety-location ("gene-environment") correlation and "epistasis"—the small direct contributions to the trait from each locus not to be independent of each other (Otto et al. 1995). It can also be achieved by considering a general case of the move discussed under assumption 1 from the basic agricultural model (equation 9) to the situation in which replicates are twins (equation 11). Recall that this splits the overall variety variance (V) and variety-location-interaction variance (VL) into two parts. Suppose that the data meet the following empirical condition:

$$(V^{-} + VL^{-}) / f = T + TL = (V + VL) / (f + 1) \quad (21)$$

where f is a positive number (not necessarily equal to 1; i.e., the two parts are not necessarily equal).

The standard twin studies formulas for heritability and among-location-average variance fraction ("shared environmental effect") can then be estimated using observations across a number of locations:

$$(V + VL)/Y * 2 / (f+1) = 2 (I_{MZ} - I_{DZ}) \quad (22)$$

$$L/Y + (V + VL)/Y * (f-1)/(f+1) = 2 I_{DZ} - I_{MZ} \quad (23)$$

If f > 1, the left hand side of Eq. 22 is smaller than the left hand side of Eq. 19 (which assumes f =1); similarly, the left hand side of Eq. 23 is larger than the left hand side of Eq. 20. Conversely,

if f < 1. These discrepancies disappear, however, if equations 19 and 20 are replaced by the following generalized forms:

$$(V + VL)/ Y = (f + 1) (I_{MZ} – I_{DZ}) \quad (24)$$

$$L / Y = I_{DZ} - f (I_{MZ} – I_{DZ}) \quad (25)$$

If equations 24 and 25 are used, the empirical question of whether the equality condition (equation 18) is met is superseded by the empirical question of what value f takes (to be discussed further under assumption 3 below). (Note that equations 24 and 25 would result from the standard Mendelian construction of twin studies described under assumption 1 if that construction began with intraclass correlation of $\gamma$ for DZ twins for a trait governed by a single locus observed in a single location, set f to $\gamma / (1- \gamma)$ and, as before, took $V^+$ as equivalent to V + VL. However, the justification for a DZ intraclass correlation other than .5 in the single-locus, single-location case is not obvious.)

In summary, similarity in traits for relatives need not, all other things being equal, be proportional to the fraction of all the genes that vary in the population which the relatives share, e.g., fraternal or dizygotic twins need not be half as similar as identical or monozygotic twins. Such proportionality is an unreliable heuristic. The derivation of the standard classical QG formulas can be adjusted, by incorporating an empirically determined parameter, to accommodate similarity for relatives that does not necessarily match that heuristic.

**Assumption 3. In human twin studies, variety-location-interaction variance ("genotype-environment interaction" variance) can either be discounted or be incorporated into the among-variety-averages fraction of the total variance, i.e., into the heritability.** The alternative analysis allows for variety-location-interaction variance when deriving formulas that link intraclass correlations to the partitioning of the total variance (even if the interaction variance turns out to be a small fraction of the total variance observed in the trait).

First a note of clarification: The term "genotype-environment interaction" is used in range of ways, including cases in which "genotype" denotes a value of a measured genetic factor and/or "environment" denotes a value of a measured environmental factor (e.g., Plomin et al. 1977, Moffitt et al. 2005). The substitute term "variety-location interaction" is used here to refer only to the analysis of trait variation made without reference to measured factors (see equation

9). In everyday terms, a high degree of variety-location interaction means that the responses of the observed varieties across the range of the observed locations do not parallel one another, that is, one variety may be highest for the trait in one location, but another variety may be highest in another location—or, at least, the difference between any two varieties may change location to location.

We saw under assumption 1 that the standard and alternative derivations of the formulas for analyzing observations from twin studies give the same results if $V^+$ is equivalent to V + VL. By definition heritability is V/Y. Equation 7 provides, therefore, an estimate of heritability inflated by VL/Y. To understand where VL becomes incorporated into $V^+$ during the standard derivation, consider equation 9 in some fixed location, say, J:

$$y_{iJk} = m + v_i + l_J + vl_{iJ} + e_{iJk} \quad (26)$$

Collecting the terms that are constant and the terms that vary from one variety to the next,

$$y_{iJk} = (m + l_J) + (v_i + vl_{iJ}) + e_{iJk} \quad (27)$$

and

$$Y = (V + VL) + E \quad (28)$$

Equation 28 reminds us that, in equation 1, which refers to a single location, the variance for the trait among the MZ twin pairs, $V^+$, combines variance that is specific to that location with variance that would, if the varieties were raised across all locations, show up as differences among the averages for the varieties. The combination of location-specific (i.e., the variety-location-interaction variance) and across-location (i.e., the variance among variety averages)—in short, the summation of VL and V—is preserved through the subsequent steps of the derivation of the standard models of twin studies (and into equations 7 and 8). The combination of VL and V in human QG should not be surprising. In the case of agricultural evaluation trials, in which a set of varieties is observed in each of a set of locations, it is possible to estimate V and VL separately. In studies of human twin raised together, in contrast, when each variety is observed in one location with two replicates (twins) for each of those variety-location combinations, there is no way to ascertain how much the variation among varieties would change if the varieties were observed in locations (families) other than the ones where they were actually observed.

When the standard twin studies methods label $V^+/Y$ as heritability, what is being assumed is either that the variety-location-interaction variance can be discounted or that it can be subsumed in the variance among variety averages across locations. With respect to the first

assumption: While it is possible that in some cases the fraction of the total variation due to interaction variance is small, any method that assumes that cannot demonstrate it. Because the standard method takes $V^+$ as an estimate of variance among variety averages, it cannot demonstrate that the interaction is negligible. With respect to the second assumption: When agricultural and laboratory breeders raise each generation of plants or animals in the same location or conditions as in the previous generation, it is reasonable to incorporate the interaction variance into the heritability estimate. What they are interested in that case is what is technically called within-location heritability (Lynch and Walsh 1998, 669). For humans, however, control of locations so they are the same from one generation to the next is not possible.

The size of the interaction variance has to be estimated by anyone wanting to claim that variation among locations (families) is of small importance (or of smaller importance than had been believed). To support such a claim entails showing not only that the among-location-averages variance ("shared environmental effect") is a small fraction of the total variance, but so also is the variety-location-interaction variance. There is good and bad news on this count. The good news is that there are human QG methods that can separate out the variety-location-interaction variance fraction. To do so it is necessary to combine studies of twins raised together with two other kinds of study: twins raised apart in randomly chosen locations; and groups of individuals from different varieties raised together, that is, each group in one randomly chosen location. Making use of the twins-raised-apart studies requires that the choice of locations of twins raised apart is truly random, that is, they are no more similar to each other on average than any two of the possible locations. Making use of the three kinds of studies together also requires the "equal environment assumption," that is, the treatment or experience of the twins or unrelated individuals within a family is unaffected by whether they are MZ or DZ twins, non-twinned siblings, or unrelated. Under these conditions, it becomes possible to use intraclass correlations to estimate values for $V^-$, T, L, $VL^-$, and TL (each as a fraction of Y) without assumption 2 (or related assumptions about similarity of relatives of other degrees) (Taylor 2007). From the the fractions for $V^-$, T, L, $VL^-$, and TL we can determine the size of the fraction for V (and thus heritability), and for variety-location-interaction variance (i.e., $VL^- + TL$), as well as estimate the value of f and thus resolve the empirical question raised under assumption 2. Bouchard and McGue's (1981) summary of available studies of IQ test scores yields, if the conditions are

assumed to hold, estimates, as a fraction of Y, of V, .72; L, .24; VL, -.10; and E, .14; which translates to f, 1.4 and γ, .58 (calculations available from the author on request).

The bad news is that suitable data sets are rare and may not meet the conditions. In particular, data on twins raised apart are not always accompanied by data on unrelated individuals raised together. This prevents the separation of L from $VL^-$ and only allows bounds to be put on the value of f. Even when data are available, the estimates of some of the values turn out to be quite negative, which is hard to interpret and could indicate that the two required conditions above do not hold. Indeed, whether the two conditions hold in any actual study remains under debate (Richardson and Norgate 2005), not the least because twins share at the very least environmental conditions before birth.

Suppose that we decide <u>not</u> to employ observations where those conditions are disputed, but to rely only on the raised-together MZ:DZ comparison. Given the absence of any better estimate for the variety-location-interaction variance fraction, the systematic fraction (i.e., 1 - the noise fraction) could be divided by three (because there are three systematic components, V, L, and VL). This comes out as $I_{MZ}$ /3 (see equation 17), leading us to the following adjusted formulas for heritability:

$$V / Y = 5I_{MZ}/3 - 2I_{DZ} \quad (29)$$

(in place of equations 7 and 19), and

$$V / Y = (f + 2/3)\, I_{MZ} - (f+1)\, I_{DZ} \quad (30)$$

(in place of equation 24). The first formula (equation 29) reduces most human heritability estimates to values that in almost all cases are below the fractions for among-location-average variance ("shared environment effect") and variety-location-interaction variance (Table 1, using data cited in Falconer 1960, 185 and Falconer and Mackay 1996, 173, and data given in Nichols 1978). Results from the second formula (equation 30) depend on the unknown value of f. If a plausible range for f could be given, say, from observations of other traits for twins raised apart or from some model of the influence of alleles at many loci, a range could be given for heritability estimates derived from the second formula. Note, however, that f would have to be approximately 2 for such adjusted estimates to be as high the ones derived using the commonly used formulas (calculations not shown here).

Table 1. Heritability estimates subject to a simple adjustment that excludes variety-location-interaction variance

| Trait | Intraclass correlations | | Fractions of total variance | | | | |
|---|---|---|---|---|---|---|---|
| | $I_{MZ}$ | $I_{DZ}$ | heritability* | adjusted heritability | interaction fraction | location fraction | noise fraction |
| From Falconer 1960, 185 | | | | | | | |
| height | 0.93 | 0.64 | 0.58 | 0.27 | 0.31 | 0.35 | 0.07 |
| weight | 0.92 | 0.63 | 0.58 | 0.27 | 0.31 | 0.34 | 0.08 |
| IQ | 0.88 | 0.63 | 0.50 | 0.21 | 0.29 | 0.38 | 0.12 |
| birth weight | 0.67 | 0.58 | 0.18 | -0.04 | 0.22 | 0.49 | 0.33 |
| From Falconer and Mackay 1996, 173** | | | | | | | |
| Finger-ridge count | 0.96 | 0.47 | 0.98 | 0.66 | 0.32 | -0.02 | 0.04 |
| Height | 0.90 | 0.57 | 0.66 | 0.36 | 0.30 | 0.24 | 0.10 |
| IQ score | 0.83 | 0.66 | 0.34 | 0.06 | 0.28 | 0.49 | 0.17 |
| Social maturity score | 0.97 | 0.89 | 0.16 | -0.16 | 0.32 | 0.81 | 0.03 |
| From Nichols 1978 | | | | | | | |
| Verbal comprehension | 0.78 | 0.59 | 0.38 | 0.12 | 0.26 | 0.40 | 0.22 |
| Verbal fluency | 0.67 | 0.52 | 0.30 | 0.08 | 0.22 | 0.37 | 0.33 |
| Reasoning | 0.74 | 0.50 | 0.48 | 0.23 | 0.25 | 0.26 | 0.26 |
| Spatial visualization | 0.64 | 0.41 | 0.46 | 0.25 | 0.21 | 0.18 | 0.36 |
| Perceptual speed | 0.70 | 0.47 | 0.46 | 0.23 | 0.23 | 0.24 | 0.30 |
| Memory | 0.52 | 0.36 | 0.32 | 0.15 | 0.17 | 0.20 | 0.48 |

* Heritability calculated as 2 ($I_{MZ}$ - $I_{DZ}$). ** $I_{DZ}$ values are for same-sex twins. Estimates of error in the original are not reported by Falconer and Mackay.

In agricultural plant evaluation trials, VL is typically as large as V. We do not know whether this is also the case for traits in animal or human populations observed in a typical range of locations. The adjusted human heritability estimates in Table 1 should be viewed, therefore, not as the correct estimates, but as exclamation points that emphasize the dependency of the standard human heritability estimates on a fundamental assumption about variety-location-interaction variance. The alternative introduced in this section is to allow for non-zero variety-location-interaction variance when deriving formulas that link intraclass correlations to the partitioning of the total variance.

**Assumption 4. High heritability of a trait can guide decisions about whether to investigate genetic factors underlying the development of that trait (and analogously for other components of variance).** Alternatively, when researchers are making hypotheses to account for any given fraction of variance, they could allow for the integration of underlying genetic and environmental factors.

The conventional wisdom is that, although high heritability provides no clues about the identity of the genetic factors that underlie differences among varieties, it can indicate that a trait is a good candidate for molecular research to expose those factors (Nuffield Council on Bioethics 2002). By the same thinking, a high fraction of variance among locations would indicate that the trait is a good candidate for exposing the environmental factors, and, if variety-location-interaction were estimated, a high value would invite us to investigate the combination of genetic and environmental factors operating (perhaps along the lines of Moffitt et al. 2005). The assumption, in other words, is that partitioning of variation provides insight into the relative strength of the different kinds of factors underlying the development of the trait.

Three observations call the assumption into question (and a fourth pertinent observation will be added under assumption 5):

a. Variation or differences among variety contributions to the values of the observed traits are conceptually distinct from differences in the genetic factors underlying those traits, even though the former quantity is often labeled "genetic variation" and ambiguously described as a measure of "genetic differences" (see preliminary remarks on terminology).

b. An ANOVA based on a model that adds up variety, location, and interaction contributions (e.g., equations 9 and 11) can be undertaken without assuming that any gradient of a measurable genetic factor (or composite of factors) runs through the differences among variety averages. Similarly, such an ANOVA does not require that any gradient of a measurable environmental factor runs through the differences among location averages. (This observation is obvious in the situation where the varieties are drawn from different species. As we move towards the situation in which the varieties are drawn from a single species—perhaps even made up of inbred lines—there is no logical point at which the observation ceases to hold and we can assume that an underlying gradient is indeed present.)

c. As many have noted, the results of fitting observations to an additive model are conditional on the specific set of varieties and set of locations observed. Similarly for

coefficients calculated through "path analyses" based on additive models related to those in ANOVA (Lynch and Walsh 1998, 827ff) and for heritability, which derives from the variety contributions. (For example, as Turkheimer et al. 2003 show for heritability of IQ test scores, estimates decrease when families of low socioeconomic status are included.) Conditionality also means that the contribution of any given variety is not a property of the variety alone (see discussion of assumption 1).

The alternative to assumption 4 is that, when researchers are accounting for any specific fraction of variance, they could consider hypotheses that integrate genetic and environmental factors in ways that may vary among subsets of varieties and locations. Knowledge from sources other than the data analysis is always needed to help researchers generate hypotheses. Other than these points, nothing general can be said about such hypothesis generation.

**Assumption 5. When similarity among a set of close relatives is associated with similarity of genetic factors, those (yet-to-be-identified and measured) factors are the same from one set of relatives to the next (in humans: from one family to the next).** Alternatively, interpretations of heritability should allow for the possibility that genetic and environmental factors underlying the development of observed traits are not the same, i.e., are heterogeneous.

The conventional wisdom that high heritability traits are good candidates for molecular research to expose the underlying genetic factors is also based on the assumption that those (yet-to-be-identified and measured) factors are the same from one set of relatives to the next (in humans: from one family to the next). Similarly, for the other fractions of the variance.

The alternative interpretation, which does not make assumption 5, is that the genetic and environmental factors underlying the development of the trait are not necessarily the same from one set of relatives to the next. Consider the approach that is central to human QG, namely, the comparison of the similarity of MZ twins with the similarity of DZ twins. The more that the former quantity exceeds the latter, the higher the trait's heritability. Suppose, indeed, that the similarity among a set of close relatives is associated with similarity of specific genetic factors. There is nothing in the methods of human QG that requires those factors to be the same from one set of relatives to the next, or from one family (location) to the next. In other words, the factors underlying the observed traits may be heterogeneous. It could be that pairs of alleles at a number of loci, say, AAbbccDDee, subject to a sequence of environmental factors, say, FghiJ, are

associated, all other things being equal, with the same outcome for the trait as are alleles aabbCCDDEE subject to a sequence of environmental factors FgHiJ (Figure 1). (Again, this possibility is obvious in the case where the varieties are drawn from different species.) The possibility of underlying heterogeneity of factors means that heritability is an unreliable basis for judging a trait to be a good candidate for molecular research. Similarly for the other fractions of the variance.

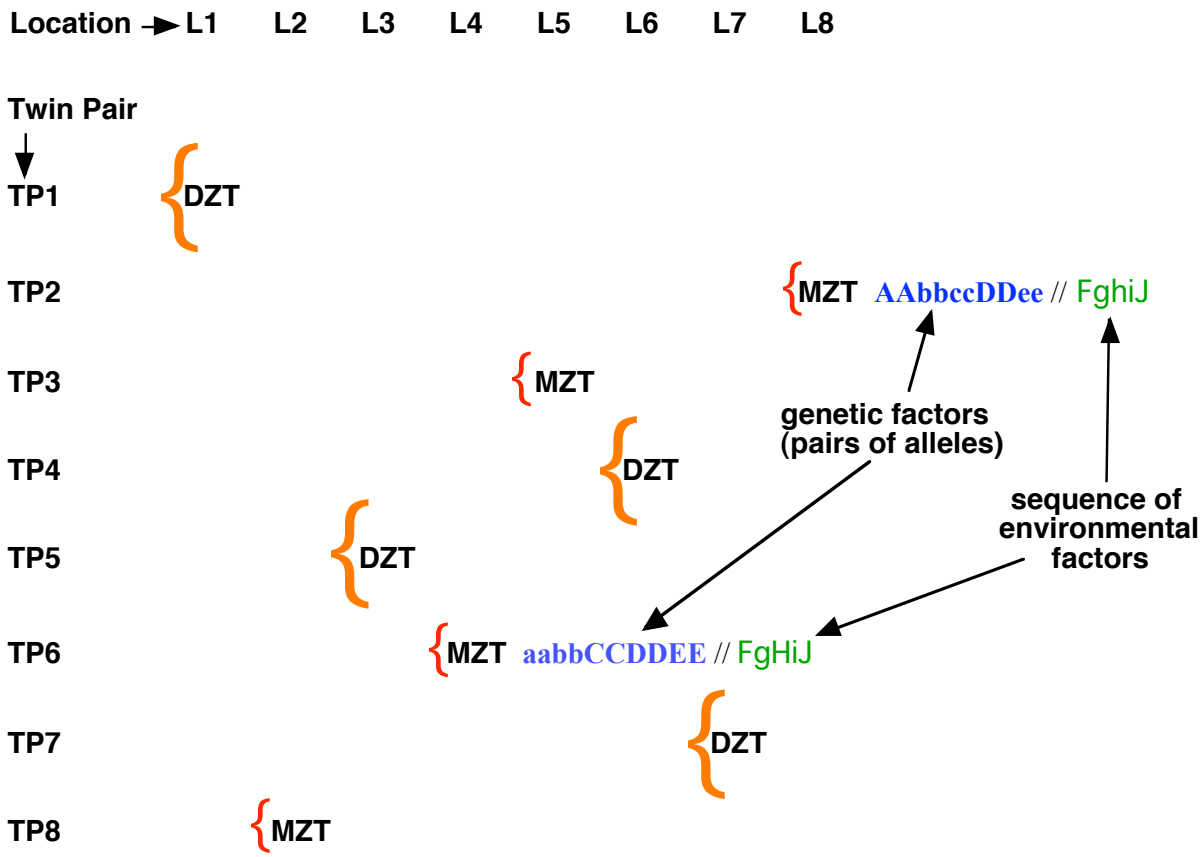


Figure 1. Factors underlying a trait may be heterogeneous even when identical (or monozygotic) twins raised together (MZT) are more similar than fraternal (dizygotic) twins raised together (DZT). The greater similarity is indicated by the smaller size of the curly brackets. The underlying factors for two MZ pairs are indicated by upper and lower case letters for pairs of alleles (A-E) and environmental factors to which they are subject (F-J).

---

**Assumption 6. Residual variation is a within-family environmental effect.** Alternatively, this variation can be viewed simply as what remains after systematic variation (among variety averages, among location averages, and among variety-location-combination averages) is taken into account.

When QG analysis assigns fractions of the variation to heritability and to location variance (or "shared environmental effect"), a residual fraction remains. This fraction is commonly labeled a "non-shared environmental effect." In line with assumption 4, under which partitioning of variation is viewed as providing insight into the relative strength of the different kinds of factors underlying the development of the trait, this labeling has stimulated the search for environmental factors that differentially affect members within the same family. Turkheimer and Waldron (2000) review this research and Turkheimer (2000) concludes that the search has not been very fruitful. Nevertheless, considerable currency is still given to the idea that differences within families are large relative to the effects due to the members of a family growing up in the same location. (In discussions of this idea it is not always clear when effects is being used to refer to fractions of variance or to causal factors.)

The alternative to assumption 6 is that residual variation—E in equations 10, 12, 17 etc.—can be viewed simply as what remains after systematic variation is taken into account (as acknowledged by Turkheimer 2004, 163). To the extent that components of variation can be translated into hypotheses about underlying factors (see discussion of such translation under assumptions 4 and 5), residual variation can be attributed to measurement error and to differences among replicates (twins) within variety-location combinations that are unrelated to variation within other combinations. Such differences provide no basis for expecting the same kind of environmental factor—or the same combination of genetic and environmental factors—to generalize across families.

## Implications

This article has examined six assumptions in classical QG and the analysis of human twins raised together. Alternatives to these assumptions are given little attention in standard QG texts or in critiques of human QG, even though allowing for the alternatives could change markedly the results and interpretations of classical QG. The central contention of this article—

that these assumptions and alternatives warrant more attention—is amplified by additional theoretical and empirical issues that have not been resolved and invite investigation. The following sequence of issues arise from the assumptions or alternatives taken in turn:

1. The models on which classical QG is based, namely, of genes with simple Mendelian inheritance and direct contributions to the trait, have been elaborated by allowing dominance interactions among alleles at separate loci, epistatic interactions among loci, narrow versus broad heritability, and so on. Could such elaborations be recast meaningfully in a gene-free QG (i.e., following the alternative to assumption 1)?

2. If the proportionality used heuristically to link similarity of variety contributions to fraction of variable genes that relatives share is replaced by an empirically determined parameter (f or $\gamma$ for twin studies), what average value and range do these parameters take in agricultural and laboratory populations, where empirical estimation is not difficult? In twin studies, is the empirical ratio of DZ: MZ similarity generally close to .5? Do the same values apply to the similarity of variety-location interaction contributions? If these average values and ranges are extrapolated to human QG, what difference does it make to previously reported results? The same questions could be asked of the equivalent parameters that arise in gene-free derivations of analyses for comparing relatives of other degrees (e.g., full sibs and half-sibs). Is there a linear relation between the empirically determined parameters for relatives of various degrees and the fraction of variable genes that each kind of relative shares?

3. Can existing sets of data for twins raised together and apart and non-related individuals raised together be used to generate estimates for the variety-location-interaction ("genotype-environment interaction") fraction of variance for human QG? This is a fraction that twin studies subsumes in the term for heritability, thus systematically inflating the heritability estimates. Is the variety-location-interaction variance, and thus the amount of inflation, generally negligible? How sensitive are the estimates to the special conditions required for the analyses of twins raised apart (see discussion under assumption 3)?

4. Interpretations of the partitioning of variance could be rewritten to eliminate the unfounded implication that analysis of observations of a given trait translate directly into claims or hypotheses about differences in measurable genetic and environmental factors (see preliminary remarks on terminology). When does the rewriting make a substantive difference, and when does it not? For example, agricultural and laboratory breeders can proceed as if

heritability were related to some underlying gradient in genetic factors, assess whether the results meet their expectations, and, when the results do not, try to compensate by discarding those offspring that do have the desired traits. When does it make a difference to expose the identity of the actual genetic and environmental factors and investigate their dynamic integration?

5. If the method of data analysis does not allow researchers to tell whether or not the genetic and environmental factors underlying the observed trait are heterogeneous, what can researchers do on the basis of knowing a trait's heritability? Taylor (2008a) notes five possible directions that can be pursued. In particular, as with assumption 4, agricultural and laboratory breeders can proceed as if there were no underlying heterogeneity, assess whether the results meet their expectations, and try to compensate when the results do not. When does it make a difference to allow for breeders to allow for the posibility of underlying heterogeneity?

6. Interpretations of the partitioning of variance in human twin studies could be rewritten to eliminate the implication that residual variation is a within-family environmental effect. In what ways does such a revision affect the results and interpretation of the search for environmental factors that differentially affect members within the same family?

Investigation of the preceding issues lies beyond the scope of this article, as does examination of implications beyond classical QG. Briefly however on the latter, suppose that researchers find that in many cases classical QG analyses of empirical data are sensitive to first three assumptions or that the researchers want to avoid interpretations based on the last three assumptions. The obvious next step would be to investigate whether the assumptions have also been worked into more recent QG where the analysis of variation involves data on measurable genetic or environmental factors (Plomin et al. 2003), and whether the results are sensitive to those assumptions. If so, could the assumptions be reformulated given that, in this age of genomics, empirical investigation of measurable genetic factors as well as environmental factors is possible? If reformulation proves difficult, what would QG gain and lose by shifting to analyses of variation that build on different foundations (e.g., Moffitt et al. 2005, Davey-Smith and Ebrahim 2007, Khoury et al. 2007)?

The answers to these various questions might well have implications for long-standing debates between, broadly, those who argue that human QG produces little that is reliable or relevant to social policy and those who argue that social science research on determinants of behaviors is limited by its inattention to hereditary transmission within families (Turkheimer

2004). Moreover, whatever answers are arrived at, historians, philosophers, and sociologists of science might be interested to explore why the questions have not emerged more clearly in previous methodological debates (Taylor 2008b). Researchers who have taken any or all of the six assumptions as given—or, at least, as plausible—might not welcome this questioning. However, understanding and avoiding past oversights could be seen as a constructive step as researchers continue to advance methods for analyzing similarity among genealogically related individuals. At the very least, newcomers to quantitative genetics should be informed that alternatives to the standard assumptions exist so they can take the implications into account as they define their own directions of analysis and interpretation.

**Acknowledgements**

This article is based on research supported by the National Science Foundation under grant SES–0634744. The comments of Marc Feldman, Steve Orzack, Ken Richardson, Hamish Spencer, Erik Turkheimer, and an anonymous reviewer helped in the revision process.

**References**

Bouchard, T., & McGue, M. (1981). Familial studies of intelligence: a review. Science, 212(4498): 1055-1059.

Bruder, C. E. G., A. Piotrowski, et al. (2008). "Phenotypically Concordant and Discordant Monozygotic Twins Display Different DNA Copy-Number-Variation Profiles." American Journal of Human Genetics 82:763-771.

Davey-Smith, G. and S. Ebrahim (2007). Mendelian randomization: Genetic variants as instruments for strengthening causal influences in observational studies. Pp 336-366 in Weinstein, M., Vaupel, J. W., Wachter, K.W. (eds) Biosocial Surveys. Washington, DC, National Academies Press.

Falconer, D. S. (1960). Introduction to Quantitative Genetics. New York, The Ronald Press Company.

Falconer, D. S., & Mackay, T. F. C. (1996). Introduction to Quantitative Genetics (4th ed.). Harlow: Longman.

Freedman, D. A. (2005). Linear statistical models for causation: A critical review. Encyclopedia of Statistics in the Behavioral Sciences. B. Everitt and D. Howell. Chichester, Wiley.

Howell, D. C. (2002). "Intraclass Correlation: For Unordered Pairs." http://www.uvm.edu/~dhowell/StatPages/More_Stuff/icc/icc.html (viewed 10 Jan. '07).

Jacquard, A. (1983). "Heritability: One word, Three concepts." Biometrics 39: 465-477.

Kamin, Leon J., and Arthur S. Goldberger. (2002). "Twin studies in behavioral research: A skeptical view." Theoretical Population Biology 61:83-95.

Kendler, K. S. and C. A. Prescott (2006). Genes, Environment, and Psychopathology: Understanding the Causes of Psychiatric and Substance Abuse Disorders. New York, The Guilford Press.

Khoury, M. J., J. Little, M. Gwinn and J. P. Ioannidis (2007). "On the synthesis and interpretation of consistent but weak gene-disease associations in the era of genome-wide association studies." International Journal of Epidemiology 36: 439-445.

Layzer, D. (1974). Heritability Analyses of IQ Scores: Science or Numerology? Science 183(4131): 1259 - 1266.

Lewontin, R. C. (1974). The Genetic Basis of Evolutionary Change. New York, Columbia University Press.

Lynch, M. and B. Walsh (1998). Genetics and Analysis of Quantitative Traits. Sunderland, MA, Sinauer.

Moffitt, T. E., A. Caspi, and M. Rutter (2005). "Strategy for investigating interactions between measured genes and measured environments." Archives of General Psychiatry 62(5): 473-481.

Nichols, R. C. (1978). Twin studies of ability, personality, and interests. Homo 29: 158-173.

Nuffield Council on Bioethics. (2002). Genetics and Human Behavior: The Ethical Context. http://www.nuffieldbioethics.org (viewed 22 June 2007)

Otto, S. P., Christiansen, F. B., & Feldman, M. W. (1995). Genetic and cultural inheritance of continuous traits. Stanford University, Morrison Institute for Population and Resource Studies Working Paper Series #64. http://www.stanford.edu/group/morrinst/pdf/64.pdf (viewed 29 March 2009)

Parens, E. (2004). "Genetic differences and human identities: On why talking about behavioral genetics is important and difficult." Hastings Center Report (January-February): S1-S36.

Pearl, J. (2000). Causality: Models, Reasoning, and Inference. Cambridge, Cambridge University Press.

Plomin, R., DeFries, J. C., & Loehlin, J. C. (1977). Genotype-environment interaction correlation in analysis of human behavior. Psychological Bulletin, 84: 309-322.

Plomin, R., J. C. Defries, G. E. McClearn, M. Rutter (1997). Behavioral Genetics. New York, Freeman.

Plomin, R., J.C. Defries, I. W. Craig, and P. McGuffi (eds.) (2003). Behavioral Genetics in the Postgenomic Era. Washington, DC: American Psychological Association.

Richardson, K. and S. Norgate (2005). "The equal environments assumption of classical twin studies may not hold." British Journal Educational Psychology 75(3): 339-350.

Rijsdijk, F. V. and P. C. Sham (2002). "Analytic approaches to twin data using structural equation models." Briefings In Bioinformatics 3(2): 119–133.

Taylor, P. J. (2005). Unruly Complexity: Ecology, Interpretation, Engagement. Chicago, University of Chicago Press.

Taylor, P. J. (2006). "Heritability and heterogeneity: On the limited relevance of heritability in investigating genetic and environmental factors." Biological Theory: Integrating Development, Evolution and Cognition 1(2): 150-164.

Taylor, P. J. (2007). "The Unreliability of High Human Heritability Estimates and Small Shared Effects of Growing Up in the Same Family " Biological Theory: Integrating Development, Evolution and Cognition 2(4): 387-397.

Taylor, P. J. (2008a). "Underlying heterogeneity: A problem for biological, philosophical, and other analyses of heritability?" Biology & Philosophy 23:587–589.

Taylor, P. J. (2008b). "The under-recognized implications of heterogeneity: Opportunities for fresh views on scientific, philosophical, and social debates about heritability," History and Philosophy of the Life Sciences 30: 423-448.

Turkheimer, E. (1998). Heritability and Biological Explanation. Psychological Review 105(4): 782-791.

Turkheimer, E. (2000). "Three laws of behavior genetics and what they mean." Current Directions in Psychological Science 9(5): 160-164.

Turkheimer, E. (2004). Spinach and Ice Cream: Why Social Science Is So Difficult. Pp. 161-189 in DiLalla, L. (ed.), Behavior genetics principles: Perspectives in development, personality, and psychopathology. Washington, DC: American Psychological Association.

Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic Status Modifies Heritability of IQ in Young Children. Psychological Science, 16(6), 623-628.

Turkheimer, E., & Waldron, M. (2000). Nonshared Environment: A Theoretical, Methodological, and Quantitative Review. Psychological Bulletin, 126(1), 78-108.

---

Appendix. Path Analysis used to derive the alternative analysis without assumption 1.

Path analysis is a data analysis technique that quantifies the relative contributions ("path coefficients") of variables to the variation in a focal variable once a certain network of interrelated variables has been accepted (Lynch & Walsh 1998, 823). Although some researchers interpret "contribution" in causal terms (e.g., Pearl 2000, 135 and 344-5), others criticize such an interpretation (e.g., Freedman 2005). Here contribution refers neutrally to the term of an additive model fitted to data.

The usual starting point for path analysis is an additive regression model that associates the focal variable with several other measured variables, but it is possible to employ the technique when there are no measured variables except the observed focal variable, as is the case in classical QG. This can be done by converting the additive model on which any given Analysis of Variance is based into an additive model of constructed variables that take the values of the contributions fitted to the first model. For example, the path model equivalent to equation 9 is

$$y_x = \quad m \quad +z_{1x} \quad +z_{2x} \quad +z_{3x} \quad +e_x \qquad \text{(A.1)}$$

where

y is the measured trait as before and x denotes the replicate

$z_{1x} = v_i$ if x if a replicate of variety i, or 0 otherwise

$z_{2x} = l_j$ if x if a replicate in location j, or 0 otherwise

$z_{3x} = vl_{ij}$ if x if a replicate of variety i in location j, or 0 otherwise

$e_x = e_{ijk}$ where x is replicate k of variety i in location j

The path coefficients are then set to equal the square root of the ratio of the variance of the contribution (V, L, etc.) to the total variance for the trait (Y). The "equation of complete determination" that lies at the heart of path analysis becomes

$$1 = \Sigma \text{ variance } (z_w) / Y \qquad \text{(A.2)}$$

where w denotes the different contributions in the Analysis of Variance model.

Thus far, path analysis is simply an algebraic reformulation of the Analysis of Variance. However, when the same trait is observed in two relatives, their separate path analyses can be linked in one network and the intraclass correlation between the relatives calculated (Lynch & Walsh 1998, 826)—provided it is assumed that the contributions (and path coefficients) apply to both and that the noise contributions are uncorrelated. For the Analysis of Variance based on equation 11 the corresponding path network is given by Figure A.1. Applying the formulas for linking separate path analyses given in Lynch & Walsh (1998, 826), the intraclass correlation between, for example, DZ twins where both members of any pair are raised in the same location, becomes:

$$I_{DZ} = (V^{-} + L + VL^{-}) / Y \qquad \text{(A.3)}$$

Equation A.3 is the same as equation 14 in the text. This same network diagram can be used to calculate the equations for intraclass correlations for MZ twins, and for both kinds of twins raised apart (i.e., the two members of each pair raised in randomly chosen locations). The estimation of the separate variances as a fraction of the total variance (Y) then proceeds as in the text.

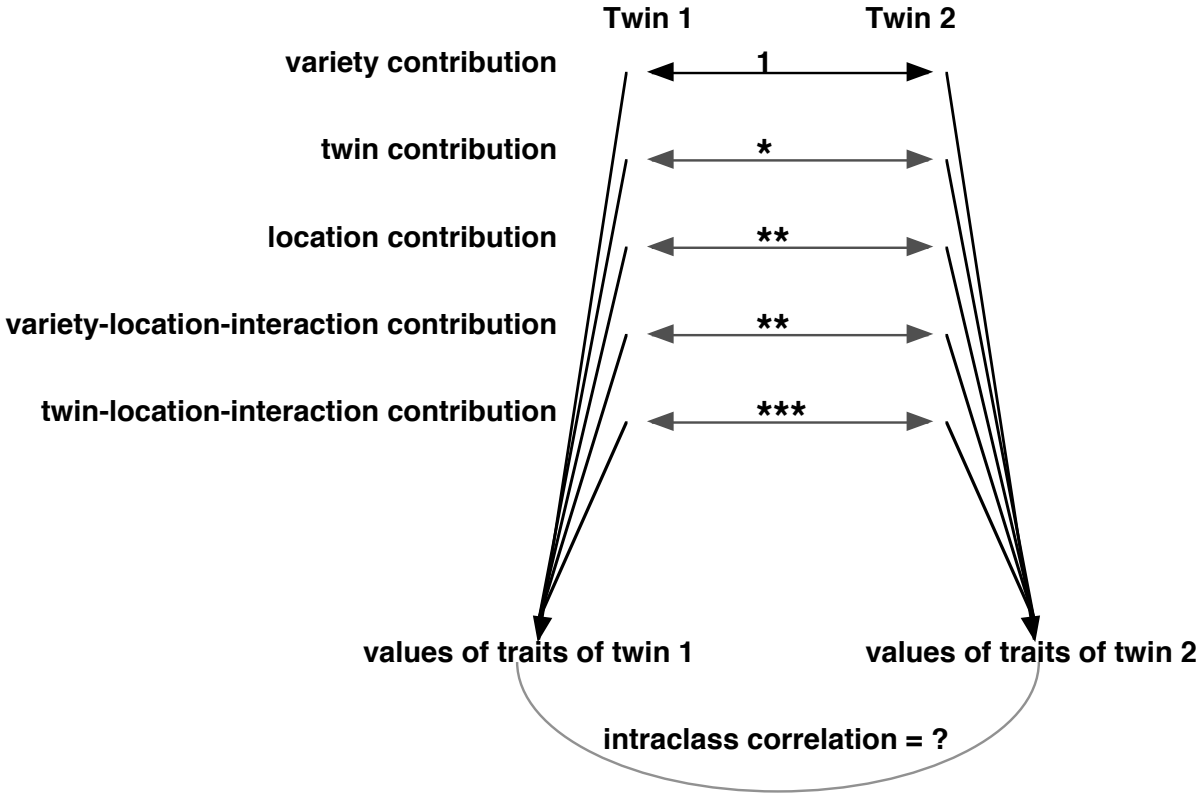

Figure A.1. Path diagram linking trait values of two twins. The path coefficients linking each of the contributions to the values of the traits are the square roots of, respectively, $V^-$, T, L, $VL^-$, TL (i.e., square root [variance ($z_w$) / Y], see equation A.2). The correlation between the variety contributions for any two twins is 1. For the other correlations, * indicates a value of 1 if the twins are MZ, 0 otherwise; ** indicates a value of 1 if the both members of any pair are raised in the same location, 0 otherwise; *** indicates a value of 1 if both conditions hold, 0 otherwise. The noise contributions for the two twins are omitted from this diagram because they are uncorrelated.

The relationship between the path diagram above and standard diagrams for analysis of twins raised together can be shown if the correlation between the variety contributions for any two DZ twins is set at $\gamma$ (typically assumed to be .5; see assumption 2) and the correlation for any two MZ twins remains at 1, because this assumption allows the separate twin contribution to be eliminated. Similarly for the variety-location interaction contribution (which is typically omitted from published path diagrams). Figure A.2 gives the resulting path diagram, which generates:

$$I_{MZ} = (V + L + VL) / Y \quad \text{(A.4)}$$

$$I_{DZ} = (\gamma V + L + \gamma VL) / Y \quad \text{(A.5)}$$

If $\gamma$ is replaced by f / (1+f), equations A.4 and A.5 can be rearranged into equations 24 and 25.

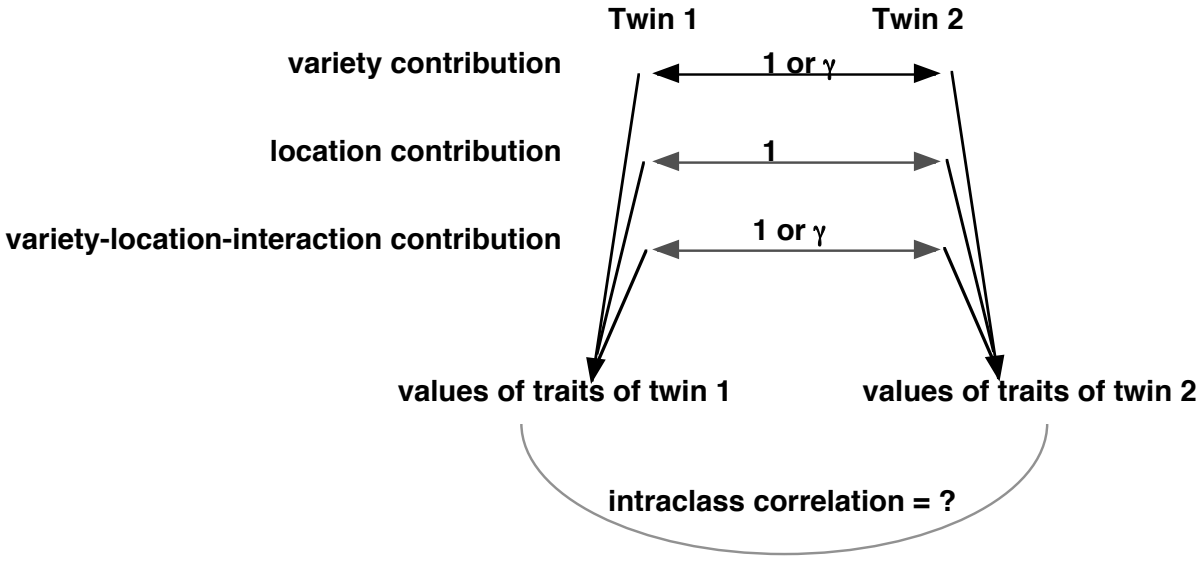


Figure A.2. Path diagram linking trait values of two twins raised together for a fixed correlation

between the variety contributions and between the variety-location-interaction for DZ twins. The path coefficients linking each of the contributions to the values of the traits are the square roots of, respectively, V, L, VL. The correlation between the variety contributions for any two MZ twins is 1 and for any two DZ twins is $\gamma$ (typically assumed to be .5); similarly for the variety-location interaction contribution (which is typically omitted from published path diagrams). The correlation between location contributions is 1 because they are raised together. The noise contributions are omitted from this diagram because they are uncorrelated.

---

Since separate values of V and VL cannot be estimated in this case, i.e., where the twins are raised in the same locations, $V^+$ can be substituted for V+VL and A4 and A5 can be replaced by:

$$I_{MZ} = (V^+ + L) / Y \quad \text{(A.6)}$$

$$I_{DZ} = (\gamma V^+ + L) / Y \quad \text{(A.7)}$$

which are the same as equations 5 and 6 in the special case of $\gamma = .5$.