Four Fundamental Gaps In Quantitative Genetics *(Version 10 August 2009. Comments welcome)*

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

Abstract: Significant problems in the methods and interpretation of classical quantitative genetics have not been widely recognized or resolved. The problems arise from four fundamental gaps: "Genetic" and "environmental" fractions of variation in traits are distinct from measurable genetic and environmental entities or factors underlying the traits' development; Standard formulas for partitioning variation in human traits are unreliable; Methods for translation from fractions of variation to measurable factors are limited; and Variation within groups is different from variation between averages for separate groups. Attention to the problems associated with these gaps, especially the possibility that genetic and environmental factors underlying a trait are heterogeneous, should help researchers avoid analogous oversights when they advance new methods, such as Genome-Wide Association studies, for analyzing similarity among genealogically related individuals.

* * *

The conventional wisdom is that "[r]esearch into the genetics of complex traits has moved from the estimation of genetic variance in populations to the detection and identification of variants that are associated with or directly cause variation" (Visscher et al. 2007). However, certain significant problems in classical quantitative genetics have not been widely recognized or resolved. Attention to these problems should help researchers avoid similar oversights when they advance new methods for analyzing similarity among genealogically related individuals.

The possibility of "underlying heterogeneity" is one such problem. Consider claims that some human trait, say IQ test score at age 18, show high heritability. These claims can be derived from analysis of data from relatives. For example, the similarity of pairs of monozygotic twins (which share all their genes) can be compared with the similarity of pairs of dizygotic twins (which do not share all their genes). The more that the former quantity exceeds the latter, the higher is the trait's heritability. Researchers and commentators often describe such

comparisons as showing how much a trait is "heritable" or "genetic." However, no genes or measurable genetic factors (that is, entities such as alleles, tandem repeats, chromosomal inversions, etc.) are examined in deriving heritability estimates (or other quantities in quantitative genetics). Nor, as some prominent geneticists have noted (e.g., Rutter 2002, 4), does the method of analysis suggest where to look for them. Moreover, even if the similarity among twins or a set of close relatives is associated with similarity of (yet-to-be-identified) genetic factors, the factors may not be the same from one set of relatives to the next, or from one environment to the next. In other words, the underlying factors may be heterogeneous. It could be that pairs of alleles, say, AAbbcbDDee, subject to a sequence of environmental factors, say, FghiJ, are associated, all other things being equal, with the same outcomes as alleles aabbCCDDEE subject to a sequence of environmental factors FgHiJ (Fig. 1).

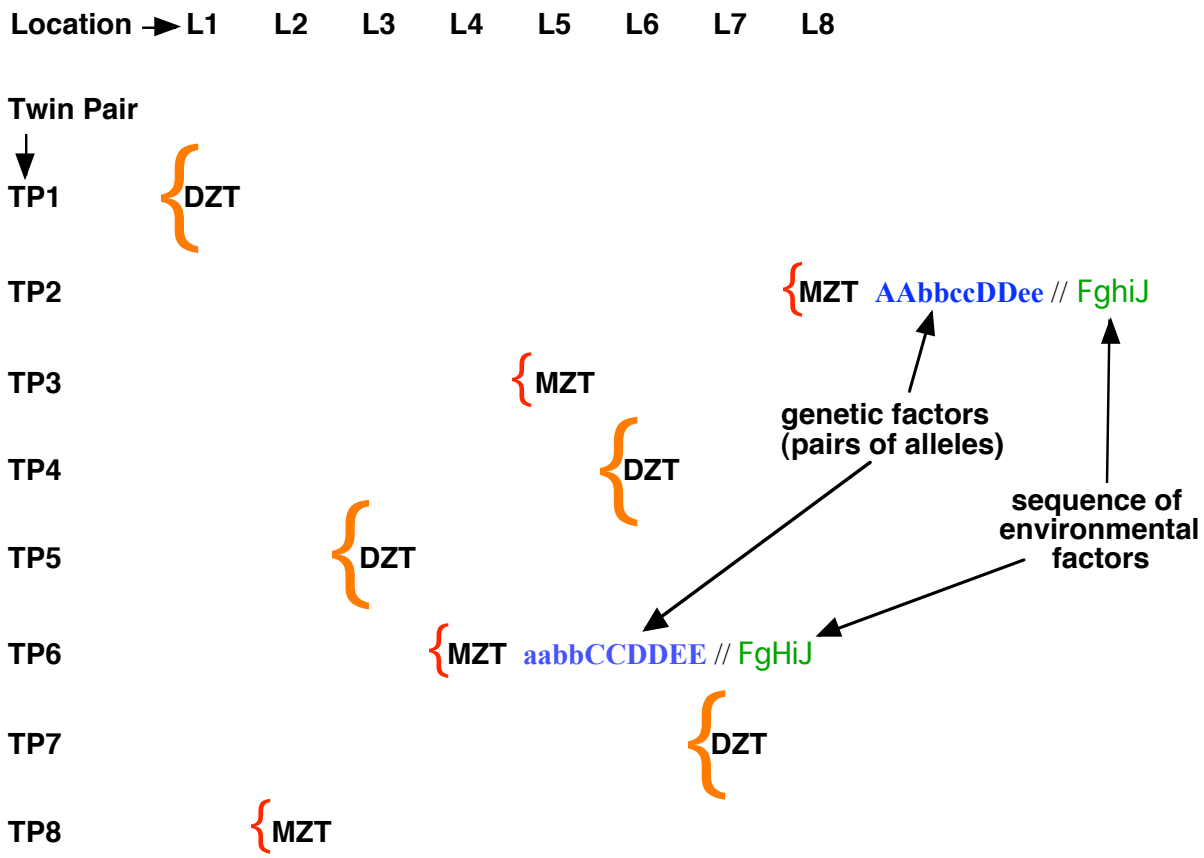

Figure 1. Factors underlying a trait may be heterogeneous even when identical (or monozygotic) twins (MZT) are more similar than fraternal (dizygotic) twins (DZT). The greater similarity is indicated by the smaller size of the curly brackets. The underlying factors for two

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).

It is not, of course, the case that underlying factors are always heterogeneous. Some traits are largely determined by the genes at a single locus more or less independently of the individuals' upbringing (so called "high penetrance major genes"), for example, presence of extra digits (or polydactyly). However, the detection of such traits can be made through examination of family trees; quantitative genetics and heritability estimation need not be involved. On the other hand, there are no obvious grounds to rule out the possibility of heterogeneity in the measurable genetic and environmental factors that underlie patterns in quantitative and other complex traits, such as crop yield, height, human IQ test scores, susceptibility to heart disease, personality type, and so on. The possibility of underlying heterogeneity has yet to be recognized as a significant methodological concern by quantitative geneticists or critical commentators on heritability research (e.g., Downes 2004 and references therein; but see Taylor 2006a,b). To accentuate its significance, let us review four fundamental gaps in quantitative genetics and the problems that follow from each.

Statistical effects are distinct from measurable factors

The first gap lies between quantitative genetics, which deals with the statistical analysis of measurements on a trait for a sample of related and unrelated individuals in a range of situations, and the investigation of measurable genetic and environmental factors influencing the processes through which the trait develops in different individuals. These inquiries are conceptually distinct. This gap needs to be highlighted, not downplayed or obscured.

Conceptual clarity and terminological adjustments can help in this regard. Most notably, the potential for confusion in the varying uses of the term "genetic" diminishes if genetic is reserved as an adjective in reference to factors that are transmitted from parents to offspring and whose presence can, in principle, be observed. In a similar spirit, "environmental" can be taken to refer to measurable factors, which can range widely, say, from average energy intake to maltreatment as a child. Potential for confusion associated with the commonly used nouns "genotype" and "environment" can also be reduced. These terms obscure the first gap by

suggesting, unjustifiably, that the quantities estimated through analysis of data about observed traits have a relationship with measurable genetic and environmental factors influencing the development of the trait. The agricultural nouns "variety" and "location" provide suitable substitutes. A variety can be thought of simply as a group of individuals whose relatedness by genealogy can be characterized, such as offspring of a given pair of parents, or a group of individuals whose mix of genetic factors can be replicated, as in an open pollinated plant variety. A location is the situation or place in which the variety is raised, such as a family or a plot at an agricultural research station. The use of the terms variety and location does not assume that researchers can specify the genetic or environmental factors that influence the trait in the various variety-location combinations.

Unreliable or questionable partitioning of variation

The second gap lies between the values generated by the methods commonly used in analysis of human twin studies and the actual heritability and the shared environmental fraction of the overall variance. The former values do not reliably estimate the latter. The standard methods need to be repaired and to be shorn of unsupported or unnecessary assumptions.

Given the abundance and sophistication of publications analyzing human twin studies, this gap needs more explication than the others. The classical quantitative genetic analysis of variation among related and unrelated individuals centers on partitioning variation into fractions according to simple additive models. In these models the value of the trait for a given individual is a sum of separate elements, including ones associated with the individual's variety and location as well as noise or unsystematic influences (e.g., measurement error). ("Element" is used here in place of the technical term "effect," whose causal connotations are unwarranted and thus confusing.) The overall variation in the trait becomes a sum of the variances of the elements in the additive model. Misestimation and misinterpretation of the elements, and the associated challenges of repairing the second gap, can be best understood by teasing out six steps:

1. Terminological clarification necessitated by the first gap. When heritability is described as the fraction of variation in a trait associated with "genetic differences" or "genetic variance," this does not refer to variation among the genes possessed by the individuals. The descriptions are loose expressions for the variance of the variety elements, where each variety's

element can be estimated by the average or mean of the trait for the variety across all locations and replicates minus the overall mean. Similarly, the "shared environmental" variance is the variance of the location elements. The variance of the "variety-location interaction" elements, unjustifiably omitted from most analyses of human variation (Jacquard 1983), is the variance of the means for each variety-location combination after removing the variety element, the location element, and the overall mean of the trait. The label "non-shared environmental" variance is given to what remains after the preceding systematic variation has been taken into account and all that is left is the variation between replicates within variety-location combinations. A better label for this variation would be simply the residual variance. (Figure 2 depicts the partitioning of variation for an agricultural evaluation trial, where it is possible to raise or grow a set of animal or plant varieties in each of a set of locations and to raise replicates for each variety-location combination. As will become evident in the steps ahead, it is often helpful to consider agricultural studies and to contrast what can be known through those studies with what can be known through analyses of data from humans.)

2. Assessment of the reliability of standard formulas for estimating the different fractions (e.g., Rijsdijk and Sham 2002). In a situation where there is as much data on a trait as could be needed, namely, the agricultural evaluation trial (Figure 2), it is possible to examine how well the methods recover the correct fractions of the variation (Taylor 2007). In order to translate from the agricultural situation back to research on humans, the replicates need to be twins—some MZ and some DZ (i.e., Figure 2 where the curly brackets are taken to denote a twin pair, as is the case in Figure 1). (Having twins as replicates is equivalent to defining a variety as the group of offspring of a pair of parents.) It turns out that only the residual fraction (i.e., the fraction corresponding to the inappropriately labeled non-shared "environmental" elements) is recovered directly. The other formulas are difficult to interpret except under a certain empirical condition, namely, the variation among the means of trait for the DZ twin pairs is equal to the variation of the trait within the DZ twin pairs (see step 5 below). When that condition is met, the formula for the location fraction recovers the correct values on average, but the heritability formula systematically overestimates the correct value by incorporating the variety-location-interaction fraction of the total variation (see steps 3 and 4 below). If the empirical condition does not hold, so that variation among DZ pairs exceeds variation within, then the systematic inflation of heritability is reduced and the estimate of the location fraction is increased above the correct

value. The opposite is case if the ratio favors within over among. In summary, the standard formulas do not reliably separate the required fractions of the variation in the trait (Taylor 2007).

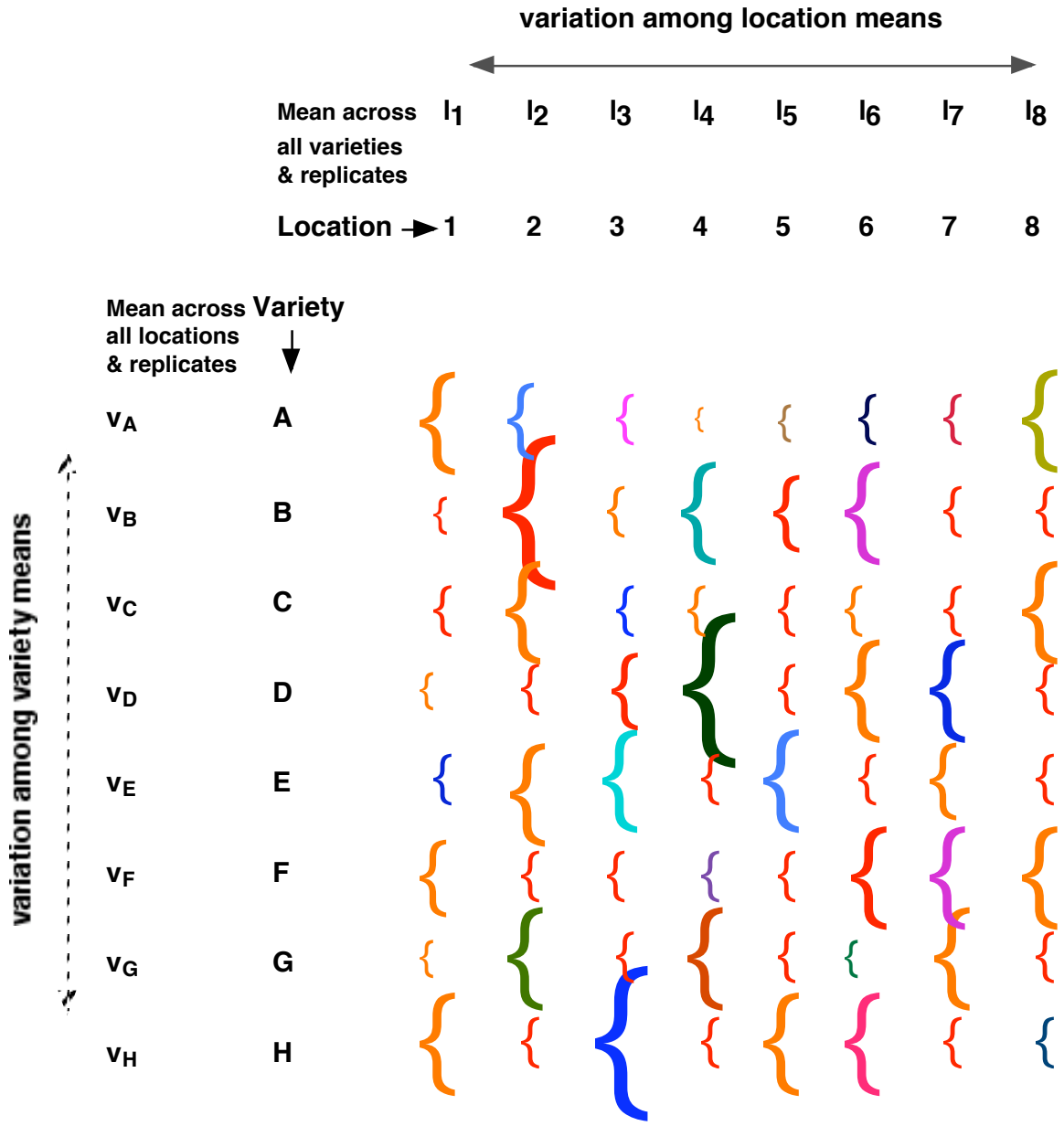

Figure 2. Partitioning of variation in an agricultural trial where each of a set of varieties is raised or grown in each of set of locations, and there are two or more replicates in each variety-location combination. Heritability corresponds to the variation among the variety means (v_A, v_B, etc.) and the "shared environmental" fraction to the variation among the location means (l_1, l_2,

etc.). The variation between replicates within variety-location combinations is indicated by the curly brackets and the variation among variety-location combinations by their color.

3. Recognition of the importance of knowing the variety-location-interaction fraction of the variation. For agricultural breeders, a large value for this fraction means that the ranking of varieties (or at least of differences between them) varies among locations. The most favorable variety to raise may not be the same for all locations and so recommendations to farmers have to be made for a delimited set of locations (or, for animals, for defined conditions of husbandry). For human research, the breeders' concerns are not relevant, but the size of the interaction variance remains important to anyone wanting to claim that the "shared environmental" fraction is of small importance (or smaller importance than had been believed). To support such a claim requires showing not only that the location variance is a small fraction of the total variation, but the variety-location-interaction variance is as well.

4. Separation of the variety-location-interaction fraction. Incorporating this fraction in an inflated heritability estimate is unavoidable in studies of human twins raised together (i.e., both members of any pair raised in the same location). Such studies are like a special form of the agricultural trial in which each variety is observed only in a single, randomly chosen location (one location for each variety; this partial data set is depicted in Figure 1). Such a trial would not allow the varieties to be ranked in any one location, nor would it allow examination of how such a ranking would differ among locations. The separate estimation of the variety-location-interaction fraction is possible, but only under special conditions: It has to be possible to undertake additional forms of trial in which the twins (replicates) are grown apart in randomly chosen locations and in which groups of replicates from different varieties are raised together so that each group occurs in one randomly chosen location. For studies of humans this means two things: 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; and the locations of twins raised apart are no more similar to each other on average than any two of the possible locations. Human studies that make use of data from the equivalent of these additional kinds of trial are reliable only to the extent that these two conditions have been shown to hold. Whether that is so in any actual case remains in question (Richardson and Norgate 2005). If observations under those disputed conditions are put aside, only a crude estimate for the variety-

location-interaction variance fraction is available (Taylor 2007). Subtracting this estimate from the heritability formula has the effect of reducing human heritability estimates (e.g., data cited in Falconer 1960, 185 and Plomin et al. 2000, 187) to values below the fractions for location ("shared environment"), variety-location-interaction, and residual variation ("non-shared environment") (Taylor 2007) or producing negative estimates that are difficult to interpret.

5. Determining the ratio of variation among the means of the DZ twin pairs to the variation within the DZ twin pairs. Empirical determination of this ratio requires one of the additional forms of trial above (step 4), in which groups of replicates from different varieties are raised together, and requires that the same strict conditions hold (Taylor 2009). If the necessary conditions are in question, a theoretical determination of the ratio is available: Suppose MZ and DZ twins were raised in the one location and differences among and within the varieties were the only source of total variation. The MZ twins will be identical because they are genetically identical. Because DZ twins share half the genes that vary in that population or species it is said that they should be half as similar (Kendler and Prescott 2006, 42). (Technically, "half as similar" means has an intraclass correlation of .5, and is the same as the technical condition mentioned under step 2, namely, variation among twin pairs is equal to variation within twin pairs.) The last step in this reasoning, however, is heuristic only: the relevant similarity (intraclass correlation) is based on observed traits and, as such, is not directly given by the number of shared genes involved in the development of those traits. (Such heuristic coefficients are ubiquitous when the methods of path analysis and structural equation modeling are used to partition variation.) The unreliability of the heuristic is illustrated by simulations (available from the author) showing an intraclass correlation often, but not always, above .5 for a disease trait that is modeled in the following, biologically plausible 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.

6. Analysis of variation without using Mendelian models and a direct contribution of genes to the trait. In general, not only in human twin studies, quantitative genetic analysis has relied on models constructed through a series of steps that build on the case of a trait governed by hypothetical alleles at a single locus in a single location. Questionable assumptions, such as the intraclass correlation of .5 in twin studies, are made as the construction moves from a single

locus and single location to situations in which multiple hypothetical loci, multiple locations, and unsystematic effects are involved (Taylor 2009). In any case, because the analysis is of variation in traits, it must always be possible to partition variation into fractions without recourse to models of hypothetical alleles (Taylor 2009). To build models up from Mendelian models of alleles at a single locus may be algebraically convenient, but it cannot be necessary for analysis of the data.

Translation from fractions of variation to measurable factors

The third gap lies between the available methods to translate from fractions of variation to hypotheses about the underlying measurable genetic and environmental factors and the methods needed for that task. This gap needs to be reckoned with. If it cannot be bridged, the second gap becomes moot and methods of analysis of variation among relatives need a basis quite different from that of classical quantitative genetics.

In conventional interpretations, a high heritability value indicates a strong genetic contribution to the trait, which makes it "a potentially worthwhile candidate for molecular research" that might identify the specific genetic factors involved (Nuffield Council on Bioethics 2002, chap. 11). The finding that the variance of location elements ("shared environmental" variance) is a small fraction of the variation in human traits relative to the residual ("non-shared environmental") variance—a finding called into question by the second gap—is typically interpreted as the shared environment (e.g., socioeconomic status of the family) is less important (strictly: is associated with less variation in the trait) than social or "environmental" influences that vary for siblings within a family.

Unless the first gap (fractions of variation in a trait are distinct from measurable factors underlying the trait's development) is overlooked, the conventional interpretations of the size of the fractions of variation presume the existence of some method to expose the measurable genetic and environmental factors. The method might not be explicit, but the obvious initial step is to assume that the variety elements in the additive models used for partitioning (see Figure 2) are related to the level of some genetic factor or composite of genetic factors (which remain to be exposed). Similarly, it is assumed that the location elements are related to the level of some composite of environmental factors, and that the residuals are related to some factors not

captured by either of these relations. These assumptions are questionable. Such genetic-factor gradients need not exist, as is obvious in the case where the varieties are drawn from different species. Even when all varieties are from the same species, the genetic factors that influence the trait need not be the same for all varieties. Indeed, as mentioned earlier, the combinations of underlying genetic and environmental factors may be heterogeneous. Notice, also, that the calculation of the variety elements involves averaging over a particular set of locations, which means that the variety elements, and thus the variance of these elements, are not properties of the varieties alone. (Similarly for location elements and their variance.)

Agricultural trials allow generation of hypotheses about the genetic and environmental factors, but noting how this is done accentuates the difficulty of bridging the third gap in human research. Where a number of varieties or animals or plants can be raised or grown in multiple replicates over many locations, varieties can be grouped by similarity in responses across all locations (using techniques of cluster analysis; Byth et al. 1976). Varieties in any resulting group tend to be above average for a location in the same locations and below average in the same locations. The wider the range of locations in the measurements on which the grouping is based, the more likely it is that the ups and downs shared by varieties in a group are produced by the same conjunctions of underlying genetic and environmental factors. This gives researchers some license to discount the possibility of underlying heterogeneity within a group, allowing them to hypothesize about the group averages—about what factors in the locations elicited basically the same response from varieties in a particular variety group, a response that distinguishes them from other groups. (It should be noted that data analysis is never self-sufficient; knowledge from other sources is always needed to help researchers generate their hypotheses about genetic and environmental factors.) However, clustering becomes infeasible when analyzing measurements from studies of human twins because such studies have only two replicates (twins) in one or at most two locations (families).

In short, in agricultural research there is a path to bridge the third gap, but it is not one that research on human variation can follow. Fortunately, it is now possible to undertake research to identify the specific, measurable genetic and environmental factors without reference to the trait's heritability or the other fractions of the total variance (e.g., Moffitt et al. 2005, Davey Smith and Ebrahim 2007, Khoury et al. 2007).

Differences within groups and among averages for separate groups

The fourth gap lies between "within-group" variation and "between-group" differences (i.e., variation among the means of the groups). The two kinds of variation have no logical or methodological relationship. This gap is widely acknowledged, but then sometimes hedged when, in the contentious debates about differences among the averages for racial and other groups, writers propose that high heritability confers plausibility on hypothesizing a role for genetic factors in explaining those differences (e.g., Jensen in Miele 2002, 111ff). The within-group/between-group gap is, however, firm and its deep implications need to be kept always in mind.

If the third gap is not bridged, statistical analysis of variation among traits and heritability estimates provide little or no guidance in hypothesizing about measurable factors underlying observations of human traits within one group of varieties. It follows logically that such analysis can provide little or no guidance about measurable factors associated with differences between the means of two groups.

Even if the third gap were bridged, hypothesizing about the difference between the mean values for varieties replicated within, but not across, locations is subject to the limitations of any nested analysis of variation. A textbook example (following Lindman 1992) illustrates these limitations. Consider high school students' test scores in algebra viewed in relation to their teacher and school. The students within a school can be randomly assigned to a teacher in their usual school. A significant difference among the means scores for the schools might, at first sight, be interpreted in terms of differences among the schools' facilities or organization. However, the influences of the teachers in the different schools and the capacity of the students to be taught are also part of the differences among the schools' mean scores. The observed differences between schools could be due to some characteristic of the school as a whole, or to the fact that some schools have better teachers and/or more teachable children, or to combinations of factors, such as students responding worse to teachers whose attention is distracted because their school's administrators insist more on detailed documentation of student performance, and so on. In short, analysis of variation cannot help researchers hypothesize about the difference in the mean scores from one school to the next when the teachers are replicated (in their students' test scores) only within schools, not across schools. To translate this into the

concerns here, nested analysis of variation cannot help researchers hypothesize about the difference in the mean scores from one location to the next when the varieties are replicated only within locations. Researchers might just as well conduct a separate analysis for each subset of varieties and location—or, in the context of racial differences, for each combination of group of individuals and experience of membership in different racial groups. (To respect this methodological limitation of nested analysis is not to make the claim that disjunct kinds of causes must be operating in the different racial groups.)

Implications for past and future analysis of human variation

The four fundamental gaps in quantitative genetics pose challenges to the common interpretations or key results of classical quantitative genetic analyses of variation, especially analyses of human variation. Recognition of the gaps might lead researchers as well as historians and philosophers of science to revisit studies that have interpreted heritability and "genetic variance" as measuring the contribution of the genetic factors in influencing variation in outcomes of the process through which the trait develops. It is possible that key results and interpretations from many decades of human quantitative genetics are not justified or, at best, are unreliable.

When the four gaps are considered together with the possible heterogeneity of measurable genetic and environmental factors that underlie patterns for traits, further implications for the understanding of the analysis of human variation follow. First, underlying heterogeneity may be hard to visualize when the gap between fractions of variation in a trait and measurable factors underlying the trait's development is obscured by ambiguous terms such as, "variation associated with genetic differences." Second, the translation from patterns in variation to hypotheses about measurable factors is possible in agricultural trials when, through clustering (as discussed above), groups be defined within which underlying heterogeneity is minimized. This cannot be done in defining human groups. Third, consider what happens if researchers put aside the search for measurable factors and focus on deriving reliable estimates of heritability as a fraction of the variation for the trait, as is common in agricultural and laboratory breeding. If the actual advance under selective breeding is less than predicted, one source of the discrepancy could be the underlying heterogeneity of genetic factors and their re-assortment through mating.

Whether this is the case matters little for breeders, because they can compensate for discrepancies: they discard the undesired offspring, breed the desired ones, and continue. Yet, this kind of selective breeding is not an acceptable option for humans. Finally, if measurable factors underlying variation are identified for one group (presumably, for humans, by some means other than classical quantitative genetics), the possibility of underlying heterogeneity tempers any impulse to hypothesize that the same factors apply within other groups and to the difference between their means.

This last observation extends the relevance of this account of problems in classical quantitative genetics to other realms of biomedical and social science. Consider, as a pertinent illustration, the phenomenon of the large mean differences on IQ test scores between generations, which still lacks a satisfactory explanation. Dickens and Flynn (2001) propose "reciprocal causation" models, which involve two key features: a matching of environments to differences that may initially be small (e.g., children who show an earlier interest in reading will be more likely to be given books and receive encouragement for their reading and book-learning); and a social multiplier through which society's average level for the attribute in question influences the environment of the individual (e.g., if people grow up and are educated with others who, on average, have higher IQ test scores, this will stimulate their own development). Such models open up further challenges. Once it is recognized that the potency of social multipliers depends on different groups' capacity to capitalize on historical changes in society, there is no reason to assume that the multipliers apply uniformly across individuals despite their differences in age, gender, geographical location, culture, and so on, or even that the multipliers move different individuals in the same direction but at different speeds. To adapt a basketball analogy that Dickens and Flynn employ, the onset of TV coverage of basketball acted as a social multiplier by eliciting greater participation in basketball while, at the same time, it elicited more "couch potato" spectatorship. Now, once researchers envisage developmental pathways whose heterogeneous components differ among individuals at any given point of time, they have opened up the challenge of developing methods to collect and analyze the data so as to discriminate among many possible models of those pathways. The same challenge applies to explaining longstanding differences between mean IQ test scores for racially defined groups in the United States.

Research on complex human traits now applies more powerful tools than the formulas of classical quantitative genetics. Genome-Wide Association (GWA) studies, most notably, can identify variants at large numbers of genetic loci that confer small but statistically significant increases in risk for diseases such as diabetes, heart disease, and cancers for defined populations (Khoury et al. 2007). However, the fundamental gaps in quantitative genetics discussed in this review suggest that high heritability value for a trait is not a reliable guide for choosing which traits to explore at the molecular genetic level. Even if it were, it remains possible that the genetic and environmental factors underlying the risk patterns detected by GWA are heterogeneous. This possibility puts an exclamation point on the emerging consensus that most medically significant traits are associated with many genes of quite small effect (McCarthy et al. 2008). Because the implications of underlying heterogeneity diminish the utility for medical research and potential treatment of the results of quantitative genetics and GWA, the four fundamental gaps should be taken carefully into consideration.

Acknowledgements

This article is based on research supported by the National Science Foundation under grant SES–0634744. A visiting fellowship at the Konrad Lorenz Institute for Evolution and Cognition Research provided conditions conducive to the completion of the manuscript.

References

Byth, D.E., Eisemann, R.L. and DeLacy, I.H.: 1976, Two-Way Pattern Analysis of a Large Data Set to Evaluate Genotypic Adaptation. Heredity **37**(2), 215-230.

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

Dickens, W.T. and Flynn, J.R.: 2001, Heritability Estimates Versus Large Environmental Effects: The IQ Paradox Resolved. Psychological Review, **108**(2), 346-369.

Downes, S.M.: 2004, Heredity and Heritability, in E. N. Zalta (ed.), The Stanford Encyclopedia of Philosophy, (http://plato.stanford.edu/entries/heredity/ (viewed 11 May 2006).

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

Jacquard, A.: 1983, Heritability: One Word, Three Concepts. Biometrics, **39**, 465-477.

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

Khoury, M.J., Little, J., Gwinn, M. and Ioannidis, J.P.: 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.

Lindman, H.R.: 1992, Analysis of Variance in Experimental Design, Springer-Verlag, New York.

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P. A., et al.: 2008, Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nature Reviews Genetics, **9**, 356-369.

Miele, F.: 2002, Intelligence, Race, and Genetics: Conversations with Arthur Jensen, Westview Press, Boulder, CO.

Moffitt, T. E.: 2005, "The New Look of Behavioral Genetics in Developmental Psychopathology: Gene-Environment Interplay in Antisocial Behaviors." Psychological Bulletin **131**(4): 533-554.

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

Plomin, R., Defries, J.C., McClearn, G.E., McGuffin, P., 2000. Behavioral genetics. Worth Pubs., New York.

Richardson, K. and Norgate, S.: 2005, The Equal Environments Assumption of Classical Twin Studies May Not Hold. British Journal Educational Psychology, **75**(3), 339-350.

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

Rutter, M.: 2002, Nature, Nurture, and Development: From Evangelism through Science toward Policy and Practice, Child Development **73**(1), 1-21.

Taylor, P.J.: 2006a, 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.: 2006b, Heritability and Heterogeneity: On the Irrelevance of Heritability in Explaining Differences between Means for Different Human Groups or Generations.

Biological Theory: Integrating Development, Evolution and Cognition **1**(4), 392-401.

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.: 2009, Gene-free quantitative genetics: A thought experiment (manuscript under review).

Visscher, P.M., Macgregor, S., Benyamin, B., Zhu, G., Gordon, S., Medland, S., Hill, W.G., Hottenga, J-J., Willemsen, G., Boomsma, D. I., Liu, Y-Z., Deng, H-W., Montgomery, G. W., Martin, N.G.: 2007, Genome Partitioning of Genetic Variation for Height from 11,214 Sibling Pairs, The American Journal of Human Genetics, **81** (5): 1104-1110.