material reality rather than a statistical artefact one might question its utility. The practical relevance of claiming that some character in some environment is 80% heritable provides no guidance for how to respond—except in a purely ideological way, by arguing as first Jensen and later Herrnstein and Murray did that the measure indicates that there is a permanent genetically inferior underclass and that no amount of social engineering—to say nothing of social justice—will improve its lot.

Biological systems are complex, non-linear, and non-additive. Heritability estimates are attempts to impose a simplistic and reified dichotomy (nature/nurture) on non-dichotomous processes. Real progress in genetics, developmental and behavioural biology will come from paying attention to Lewontin’s insistence that we attempt to analyse causes, not variances.

References

Commentary: The analysis of variance is an analysis of causes (of a very circumscribed kind)

Peter Taylor

1974—Two publications

The year 1974 saw the publication of two influential works by Richard Lewontin. In different ways, both addressed the measurement and characterization of genetic variation and asked whether this is interesting—what could we explain or do with the resulting knowledge?

The Genetic Basis of Evolutionary Change1 was firmly positioned within the population genetic tradition of viewing evolution as a change of gene frequencies in a population over time. In this light it was obviously important to characterize the amount of genetic variation and account for its maintenance. Lewontin masterfully synthesized research on genetic diversity in laboratory and natural populations in relation to models of selection or its absence. At the same time he drew attention to some troublesome themes for evolutionary biology. It was not variation as such that should count, but variation that resulted in differential fitness among the variants. Yet measurements of the components of fitness—survival and reproduction—were possible only when the phenotypic effect of a single allelic substitution was large not when the effects of gene substitutions make only small differences. This led Lewontin to remark that: ‘What we can measure is by definition uninteresting and what we are interested in is by definition unmeasurable.’ [p. 23 in Ref. (1)]. The problems of relating models of selection to observations become astronomically worse when there are multiple, linked loci [p. 317 in Ref. (1)]. He concluded that population genetics should shift its attention to the fitness effects of long segments of chromosomes; such effects could be measured.

The idea that many genes may contribute small effects to a trait derives from a different research tradition, quantitative genetics, which is the subject of the other publication, ‘The Analysis of Variance and the Analysis of Causes’ (hereon, AVAG).2 Quantitative genetics concerns itself not with any specific genes having discrete (qualitative) effects but with the statistical analysis of continuous (quantitative) traits varying
within populations.\(^3\) Traditionally, the ‘populations’ quantitative genetics deals with have consisted of the varieties manipulated by plant and animal breeders, who use statistics to estimate the predicted rate of improvement in desired traits from possible matings or crosses. However, variation of traits in human populations has also been the subject of quantitative genetic analyses, most notably in behavioural genetics. Lewontin argued that such analyses are definitely not interesting for human genetics—the field addressed in the paper—they provide no basis for effective environmental or clinical interventions.

A proper understanding of the statistical technique at the heart of quantitative genetics, the Analysis of Variance (ANOVA), shows why quantitative genetics can provide little insight into underlying causes. In any ANOVA the ranking of varieties for the trait in question does not stay the same as we move from the environment or location that the varieties develop in to the next. Moreover, the degree of such re-ordering, the ‘variety by location interaction’ (or ‘genotype by environment interaction’), is ‘local’, that is, conditional to the range of locations and sample of varieties in the data under consideration. It is more interesting, Lewontin proposed, to characterize and understand the ‘norms of reaction’—the different responses of genetically replicable varieties across the full range of environments.

Lewontin considers norms of reaction to be an important concept for analysing evolution in changing environments (indeed, many of the abundant citations of AVAC are in that vein). However, the paper was written less from his research on genetic variation in evolution than from his critique of Arthur Jensen’s quantitative genetic work on IQ test scores.\(^4\)\(^-\)\(^7\) Jensen claimed that the heritability of IQ test scores was high within human populations and interpreted the gap in average IQ test scores between racially defined human populations as probably based on genetic differences. Lewontin insisted that heritability—a quantity estimated through the ANOVA (or related statistical techniques)—is conceptually unrelated to how readily IQ test scores (or any other trait) could change when the environment changes. Moreover, as a political matter, the US society had by no means explored the full range of possible environments relevant to the development of IQ test scores.

In the years since 1974 many researchers have not heeded Lewontin’s suggestion to ‘stop the endless search for better methods of estimating useless quantities’.\(^2\) Indeed, heritability estimates have continued to fuel policy and popular debates about the source of differences between averages for racial groups in IQ test scores and other traits. In this sense there has been a lot that researchers and policy makers have been able to do with the knowledge claims emerging from quantitative genetics of human traits. Accounting for this history would require sociological analysis beyond the scope of this commentary (or commentator). Let me note, however, that a significant, oppositional strand in this history has often referred to Lewontin’s arguments in AVAC and other places \[e.g. \text{Ref. (8)}\]. As a contribution to the ongoing efforts of philosophers, biologists, and others who criticize genetic explanations of differences in intelligence and other human behavioural traits, let me identify ways that Lewontin’s arguments could be strengthened, clarified, and adjusted so as to better illuminate why the knowledge that many human traits have high heritability has not shown the way to discoveries of their actual genetic basis. Ironically, this will depend on arguing that the ANOVA is an Analysis of Causes, but of a very circumscribed kind.

2006—The analysis of variance and analysis of causes, revisited

[Mis]understanding about the relationship between heritability and plasticity [of traits]... arises from the entire system of analysis of causes through linear models... I will begin by saying some very obvious and elementary things about causes, but I will come thereby to some very annoying conclusions.\(^2\)

Lewontin’s argument in AVAC centres on four themes:

1. Different kinds of genetic causes need to be distinguished. The influence of rare deleterious genes—alleles at a single locus—that show a clear effect on individuals that are homozygous for the allele is a distinct kind of causality from the interaction of the environment and many genes of small effect.

2. High heritability of a trait does not imply that it is hard to change through environmental changes. Lewontin invokes the case of curable ‘inborn errors of metabolism’, referring presumably to the dietary amelioration of the effects of homozygosity for the gene for phenylketonuria (PKU).

3. The ANOVA analyses values of the observed trait; its underlying model does not refer to any measurable factors—either genetic factors (e.g. the presence or absence of specific alleles at one or more loci, tandem repeats, or chromosomes) or environmental.

4. The ANOVA partitioning of variation in a trait is ‘local,’ that is, conditional to the particular set of genetically replicable varieties and locations or environments for which the trait is observed. As plots of possible ‘norms of reaction’ make clear, the differences among varieties and their relative ranking can change from a subset of the locations to the full set.

How would I modify Lewontin’s argument? To the distinction in theme #1, we might add the possibility of not-so-rare genes at a single locus whose effect on individuals depends on the environment.\(^9\) Indeed, with respect to difficulty of change (theme #2), even in the classic case of PKU, the ‘cure’ is modulated by many social complexities.\(^10\) But AVAC is not concerned with identification of single genes of major effect, so a more relevant illustration of changeability would be of changes in a trait not governed by a single gene. However, it is the distinction between observed traits and measurable factors (theme #3) that I want to focus on, because this means that the observation of traits that differ across varieties and locations provides no direct information about measurable genetic and environmental factors that correspond to these differences. When combined with knowledge from other sources, the ANOVA can, at best, help us to hypothesize about what specific measurable genetic and environmental factors might be worth further investigation.
The analysis of variance and the analysis of causes

To explore the implications of the observed trait/measurable factor distinction, let us consider the analysis of agricultural crop trials, where, as will emerge, the ‘best’ case can be achieved, before returning to human genetics, where it cannot. Lewontin’s themes do not depend on the data subject to ANOVA being of human traits. Moreover, unlike the terms ‘genotype’ and ‘environment,’ the agricultural terms ‘variety’ and ‘location’ do not suggest what needs, in fact, to be established, namely, that the quantities estimated through an ANOVA have a relationship with measurable genetic and environmental factors. (Similar thinking leads me to refer to ‘trait’ not ‘phenotype’.)

In a typical crop trial, a number of different plant varieties are grown in multiple plots or ‘replicates’ in one or more locations and some trait, say, yield, is measured. For any trial there will be an overall mean for the trait, say, yield, from which any specific observation will deviate:

\[
\text{Deviation from the overall mean yield of the yield of variety } \ i \ \text{'grown in replicate 'k' of location 'j'} = \text{amount associated with variety } i \\
+ \text{amount associated with location } j \\
+ \text{amount associated with the variety-location combination } i, j \ \\
+ \text{residual}
\]

or, in statistical terminology,

\[
\text{variety-effect + location } j \ \text{effect} \\
+ \text{variety-location } i, j \ \text{interaction effect} \\
+ \text{residual}
\]

The relative sizes of the different kinds of effects can be assessed by examining the relative sizes of their variances if the values of the effects minimize the variance of the residual values and sum to zero for each kind of effect. (Conditionality of the ANOVA, theme #4, follows because the values of the effects will not be the same if they are estimated only from a subset of the varieties or locations.)

This model can be used to summarize or redescribe the data without making reference to the causes or dynamics that generated the data. However, researchers might want to construe the difference between variety i and j effects as the cause of the difference between varieties i and j in their mean yield over all locations and replications. This construal of causes in terms of differences in effects does not depend on researchers accepting the unrealistic assumption that traits result simply from the adding up of effects and residuals (see AVAC, 408–9). Rather, it follows the philosophical principle that, if researchers want model-based associations to illuminate causes (in some sense of the term), these associations must be construed in relation to some class of changes or interventions that could, in principle, be made. 11-13 For the ANOVA, the class of changes or interventions in which it makes sense to construe differences in effects as causes can be visualized as follows.

Suppose that the data are generated by unknown dynamics that include some unsystematic variation or ‘noise’. Imagine that the same set of varieties and locations is observed again and the only change in this ‘rerun’ is noise at the same level as the original noise, but uncorrelated with it. If the noise is small so the rerun remains close to the original situation, an additive model can serve as a good approximation for a wide range of actual, but unknown, dynamics. (In technical terms, such an approximation is analogous to a truncated Taylor series expansion of dynamical equations around an equilibrium point.) The stipulation that everything remains close to the original situation means that when differences in effects are construed as causes, such causes are specific to the combination of varieties and locations that make up the observed data. Conditionality (theme #4) thus extends to ‘difference-in-effects’ causes and should apply to any hypotheses about the actual dynamics that researchers draw from analysis of the data using additive models.

Provided researchers recognize the ‘close to the original situation’ rerun assumption entailed by a difference-in-effects construal of causes, they could use the ANOVA as a starting point in looking for specific, measurable genetic factors whose differences between the varieties corresponds to the differences in effects. As a reiteration of the observed trait/measurable factor distinction (theme #3), let us note that, although effects associated with varieties are often called ‘genetic’ effects, this label is potentially misleading because the differences between variety effects cannot be translated in any direct fashion into hypotheses about specific genetic factors. Similarly, there is a conceptual gap between ANOVA and analyses of environmental factors. (This gap is obscured when Lewontin invokes norms of reaction in AVAC, because norms of reaction link the observed trait to some measured environmental factor. 14)

For agricultural crop trials, hypothesis generation is enhanced by the use of cluster analysis to group varieties by similarity in responses across all locations. 15 Varieties in any resulting group tend to be above average for a location in the same locations and below in the same [Figure 3 in Ref. (15)]. The wider the variety of locations in the data 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 measurable factors. If researchers can discount the possibility of ‘heterogeneity,’ i.e. that similar responses have been produced by different conjunctions of measurable factors, they can hypothesize about the group means—about what factors in the locations elicited basically the same response from varieties in a particular variety group that distinguishes them from other groups. Of course, knowledge from sources other than the data analysis is always needed to help researchers generate any (variety-group-specific and location-specific) hypotheses about genetic and environmental factors. [See note 16 in Ref. (14) or further discussion of heterogeneity, grouping, and generation of hypotheses.]

If the genetic and environmental factors hypothesized as underlying the responses of varieties have been measured for the different varieties and locations, it is possible to use regression analysis to associate the yield with those factors and to undertake experimental trials that probe the associations by varying the factors. Insights from these studies can contribute to research on the ways in which pathways of growth and development are affected by the genetic make up of varieties and the environmental factors in the locations—assuming such research has been taking place. This research might, in turn, provide a basis for interventions outside the typically well-controlled conditions in which research on causes in
growth and development is undertaken. In summary, in agricultural research it is possible (at least in principle) to move beyond the circumscribed, ‘rerun’ notion of cause.

In human quantitative genetic research, however, genetically replicable varieties can at most be replicated in two locations (i.e. identical twins separated at birth) and these locations differ from one variety (twin pair) to the next. This means that grouping of varieties by similarity of responses across locations is impossible and the ANOVA cannot contribute to generating hypotheses about specific genetic and environmental factors [note 16 in Ref. (14)]. Moreover, there is no way of discounting the possibility that any such factors could be heterogeneous. It could be that alleles, say, AbcDe, subject to a sequence of environmental factors, say, FghHI, are associated, all other things being equal, with the same outcomes as alleles abCDE subject to a sequence of environmental factors FgHI. This implies that even if the similarity in the observed trait of a set of close relatives is associated with similarity of genetic factors, the factors may not be the same from one set of relatives to the next. Moreover, even if the trait of two genetically replicable varieties has been produced by similar conjunctions of measurable factors in one location, this need not be the case for the values observed for the trait in the next location. While subsequent investigation may show that the factors underlying similar responses are the same, no method of data analysis can establish this if the method assumes it in advance.

Now, heritability, in the technical sense of the term, is calculated from effects or variances estimated from an ANOVA (or equivalent analyses based on related additive models [note 5 in Ref. (3)]). It follows, therefore, that heritability also: (i) is conditional to the particular set of varieties and locations observed (i.e. ‘local’ in AVAC); (ii) has causal significance that rests on the same rerun conditions under which differences in effects could be construed as a form of causality; and (iii) in the case of humans, offers no guidance in generating hypotheses about measurable genetic or environmental factors. If one wonders why a quantity having such limitations was ever invented, notice that it was first used in selective breeding in agricultural and laboratory settings where researchers do have the ability to replicate varieties and locations (give or take some variability of weather from season to season in the field). Indeed, when agricultural researchers compare varieties and make recommendations to farmers, and when they select among varieties for the next round of crop trials, they do so on the assumption that the environmental factors will remain unchanged. In short, heritability, like the ANOVA, from which it is derived, has a relationship to ‘genetic’ causes, but only in the very circumscribed rerun sense of causality. The fact that human heritability estimation is based on data that are less ideal than agricultural crop trials does not somehow, miraculously, allow researchers to support claims about more general (global) notions of genetic or environmental causality.

The limitations of heritability may be seen, to borrow Lewontin’s phrase, as ‘very annoying conclusions’. To make these conclusions more digestible, it is helpful to note what heritability is not. It is common to refer to a trait as ‘heritable’ or ‘genetic’ if differences in the trait are caused by differences in specific genetic factors in the gene-based dynamics of the organisms’ reproduction. When quantitative or behavioural geneticists describe a trait with high heritability as highly heritable [e.g. Ref. (16)], they allow their audience (and themselves) to envision a connection with genetic factors. Heritability, however, is defined on the basis of observed traits without reference to any information about measurable genetic (or environmental) factors. Admittedly, in some situations heritability can be calculated as a ratio of ‘genetic’ variance to total variance for the trait (phenotypic variance), but genetic variance in this context refers to the variance of trait effects (using the agricultural terminology of this commentary) not to the variance of specific measurable genetic factors. (Moreover, if heritability is estimated in a single location only, the variety-in-location effects subsume the variety–location i,j interaction effects from the full model stated earlier.)

Elsewhere, I have extended this line of thinking about the limited relevance of heritability in exposing and investigating genetic and environmental factors to argue that heritability can have no relevance in explaining differences between means for different human groups (e.g. ‘races’) or different generations. These conclusions disturb Paren’s even-handed overview of past and potential contributions of human behavioural genetics to discussions of social importance, but they do not discount the persistent ‘black-white test score gap’ in the US. However, if we are going to learn useful things about the changeability of this gap and other human behavioural traits, we need to leave behind the persistent conflation of the ideas of heritability, genetic, and resistance to alteration by changing environmental factors. At the same time, we need to open up more conceptual space for deriving empirically validated models of developmental pathways whose components are heterogeneous and differ among adults at any one time and over generations. Understanding what change in outcome would people would see if factor f were changed at time t in development is an interesting challenge not only for human quantitative genetics but also for social epidemiology. [Impressive work in this direction is offered in Refs. (21,22).]

Of course, if we ask what people could do with that knowledge, we have to consider whether factor f could be changed in the situation in question and whether social and personal conditions could be changed to make the change in factor f realizable. There are, indeed, ‘plenty of real problems’ to engage us 30 years after Lewontin’s provocative intervention.

Acknowledgements
This paper draws on material from Refs. (14,23) and comments by Michael Bradie, Richard Lewontin, Diane Paul, and Hamish Spencer on drafts. Research and writing were, in part, supported by the National Science Foundation under Grant No. SES-0327696.

References
Commentary: The attainability of causal knowledge of genetic effects in complex human traits

Gert-Jan Vreeke

Lewontin\(^1\) points to a central issue in heritability analysis: can heritability analysis provide clues to the impact of genetic factors on behaviour?

Lewontin's verdict does not lack clarity: he thinks that heritability analysis is not helpful in any way, it merely presents results that are 'a unique function of the present distribution of environment and genotypes' (1974, p. 409).

In this comment I shall reconstruct Lewontin's arguments, outlining broadly the basic tenets of the debate, which still surrounds heritability and behavioural genetics. The task of finding out if information about variation in genotypes and environments can provide causal knowledge and, if so, in what way, is, as we shall see, far from being resolved.


\(^7\) Lewontin RC. Further remarks on race and the genetics of intelligence. *Bull At Sci* 1970;23–25.


\(^9\) Moffitt TE, Caspi A, Rutter M. Strategy for investigating interactions or behavioural traits (such as IQ).


\(^17\) Taylor PJ. Heritability and heterogeneity: II. On the irrelevance of heritability in explaining differences between means for different human groups or generations. *Biol Theory Integr Dev Evol Cognit* 2006;1 (In Press).


