sociological science

Supplement to:

Larrègue, Julien, Frédéric Lebaron, Hervé Perdry, and Nicolas Robette. 2023. “Eurythmics or Xenakis? Cultural Tastes (Are Not Made of Genes): Comment on Jæger and Møllegaard, ‘Where Do Cultural Tastes Come From? Genes, Environments, or Experiences.’ ” Sociological Science 10: 454-466.

Some limitations of the concept of heritability

Supplementary material to
Eurythmics or Xenakis? Cultural tastes (are not made of genes)

Julien Larrègue, Frédéric Lebaron, Hervé Perdry and Nicolas Robette

The ADE model

The genetic model in which heritability is defined has been proposed by Ronald Fisher (1919) to model the correlation between relatives. In this model, a phenotype P is supposed to be the sum $P = G + E$ of genetic effects G (themselves obtained as the sum of a large number of independent genetic loci) and environmental effects E; G and E are assumed to be independent. Genetic effects are further decomposed in $G = A + D$, where A is the result of additive allelic effects; the deviation D from additivity is dubbed "the dominance effect". In this simple model, $P = A + D + E$, one denotes $\text{var}(A) = a^2$, $\text{var}(D) = d^2$ and $\text{var}(E) = e^2$; all the terms are independent (by hypothesis for E and A, E and D ; by construction for A and D), and one has $\text{var}(P) = a^2 + d^2 + e^2$. One let

$$h^2 = \frac{a^2}{a^2 + d^2 + e^2} \text{ and } H^2 = \frac{a^2 + d^2}{a^2 + d^2 + e^2};$$

h^2 and H^2 are the narrow sense and broad sense heritability, respectively.

These quantities have not been constructed and proposed as measures of causality; Fisher (1919), p. 399-400 himself clearly stated that "loose phrases about the 'percentage of causation,' which obscure the essential distinction between the individual and the population, should be carefully avoided." Their interest is that they allow to compute the phenotypic correlation between any pair of relatives, as shown in table 1 – assuming that the model holds, and that the environments of the relatives are indeed independent.

Table 1: Phenotypic correlations between pairs of relatives in the ADE model

Relatives pair	**Phenotypic correlation**	**Alternative expression**
parent / offspring	$r_{PO} = \frac{1}{2}h^2$	$r_{PO} = \frac{1}{2}a^2$
grandparent / grandchild	$r_{GPGC} = \frac{1}{4}h^2$	$r_{GPGC} = \frac{1}{4}a^2$
siblings / dizygotic twins	$r_{Sibs} = \frac{1}{2}h^2 + \frac{1}{4}\left(H^2 - h^2\right)$	$r_{Sibs} = \frac{1}{2}a^2 + \frac{1}{4}d^2$
monozygotic twins	$r_{MZ} = H^2$	$r_{MZ} = a^2 + d^2$

The mathematical expressions in the last column are by obtained assuming, without loss of generality, that $\text{var}(P) = a^2 + d^2 + e^2 = 1$. In this case, $h^2 = a^2$ and $H^2 = a^2 + d^2$.

The limitations of this concept are well-known (Lewontin, 1974). Departures from the theoretical model, including gene-environment interactions, non-linearity, non-additivity, make it impossible to interpret heritability estimates. A classic example is the high heritability of some infectious diseases such as leprosy or tuberculosis, which cannot be interpreted as evidence for the genetic origins of these diseases; it is impossible to interpret this heritability without acknowledging the primary role of the infectious agent and the presence of gene-environment interaction.

Even if one assumes that the model holds, in the presence of a non-zero gene-environment correlation ρ, the variance of the trait decomposes in $a^2 + 2\rho ae + e^2$: the $2\rho ae$ term makes the computation of h^2 meaningless, and if one wants to analyze the trait's variance it is necessary to find ways to estimate ρ.

Common environment and twin studies: the $ADCE$ and ACE models

Another natural objection to this model is that the hypothesis of independence of environments of relatives is very unlikely to hold. For a given level of relatedness, for example parent/offspring, one writes that the parent's phenotype is

$$P = A + D + C_{PO} + E$$

while the children's phenotype is

$$P' = A' + D' + C_{PO} + E',$$

where C_{PO} is the common environment to the parent and the offspring, and E and E' are independent environments. Assuming here and in the sequel that $\text{var}(P) = \text{var}(P') = 1$, standard computations lead to $r_{PO} = \text{cor}(P, P') = \frac{1}{2}a^2 + c^2_{PO}$ where $c^2_{PO} = \text{var}(C_{PO})$. Table 2 summarizes the correlations for various degrees of relatedness, with the convention $\sigma^2 = 1$.

Table 2: Phenotypic correlations between pairs of relatives in the $ADCE$ model

Relatives pair	Alternative expression
parent / offspring	$r_{PO} = \frac{1}{2}a^2 + c^2_{PO}$
grandparent / grandchild	$r_{GPGC} = \frac{1}{4}a^2 + c^2_{GPGC}$
siblings	$r_{Sibs} = \frac{1}{2}a^2 + \frac{1}{4}d^2 + c^2_{Sibs}$
dizygotic twins	$r_{DZ} = \frac{1}{2}a^2 + \frac{1}{4}d^2 + c^2_{DZ}$
monozygotic twins	$r_{MZ} = a^2 + d^2 + c^2_{MZ}$

In this model with shared environment, not only the correlation between pairs of a certain type of relatives provides no longer an unbiased estimate of a^2, but it does not allow to predict the correlation between pairs with another degree of relatedness. A common solution is to assume that $d^2 = 0$ and that all pairs of relatives share the same amount of common environment (or, at least $c^2_{MZ} = c^2_{DZ} = c^2$, known as the "equal environment assumption" or EEA): this is the ACE model, $P = A + C + E$. If these assumptions hold, then

$$2(r_{MZ} - r_{DZ}) = a^2$$
$$2r_{DZ} - r_{MZ} = c^2$$

These formulas (the first one being known as Falconer's formula) lead to simple moment estimates of a^2 and c^2. However, if these simplifying assumptions do not hold, that is if $d^2 > 0$ and $c^2_{MZ} = c^2_{DZ} + \kappa^2$ (thus, κ^2 is the excess of environment shared by monozygotic twins, as compared to dizygotic twins), then

$$2(r_{MZ} - r_{DZ}) = a^2 + \frac{3}{2}d^2 + 2\kappa^2$$
$$2r_{DZ} - r_{MZ} = c^2_{DZ} - \left(\frac{1}{2}d^2 + \kappa^2\right)$$

and the estimates of a^2 and c^2 obtained through these formulas are hopelessly biased, respectively upward, and downward. These results can of course be found easily in the literature (see e.g. table 10.5 in Falconer and Mackay (1996)).

Note that more sophisticated approaches are used, such as maximum likelihood estimates in structural equation models, allowing, when all the model assumptions hold, to obtain slightly more precise estimates (and, in particular, to constraint a^2 and c^2 to be non-negative), and more importantly to take covariates into account (such as the twins' ages), etc. Yet, these approaches present the same limits that the moment estimates we just described.

Assortative mating with cultural transmission

Assortative mating for a given trait occurs when there is a positive correlation between the trait values in two mates of a pair. It is observed for lots of traits, including height and many behavioral traits.

The effect of assortative mating has been studied by many authors, from Fisher (1919) himself to Wright (1921), to Crow and Felsenstein (1968) and Nagylaki (1978). Here we want to draw the readers' attention to the fact that the genetic variance in a given population is modified by assortative mating. Specifically, let us consider a population in which there is no assortative mating for the trait of interest, and denote a^2_0 the genetic variance in this population. Assume now that at some point in time, the phenomenon of assortative mating starts, and that the trait

values of mates in a pair have a correlation of $r = \text{cor}(P_1, P_2)$; then the value of the genetic variance will increase over time, until reaching an equilibrium point where one has $a^2(1-rh^2) = a_0^2$. Thus, assortative mating affects the value of the genetic variance itself, and therefore the total variance of the traits. In the case of behavioral traits, like cultural practices, assortative mating for the trait modifies the value of its genetic variance, and thus of its heritability. This makes it hard to make sense of the partition of the trait's total variance.

Moreover, cultural transmission induces a correlation between parental and offspring environments. This leads, in the offspring, to a gene-environment correlation; as for genetic variance, this gene-environment correlation will increase over time after assortative mating has begun, until reaching an equilibrium point. Classical studies of assortative mating do not provide formulas to compute this and have not explored the consequences of this phenomenon; this is done in detail in (Noûs and Perdry, 2022). In the presence of such a gene-environment correlation ρ, the total variance of the traits can be decomposed as $a^2 + 2\rho ae + e^2$. There is no longer a partition of variance as the sum of a genetic part and an environmental term.

Thus, even if one chooses to embrace the additive polygenic model, without gene-environment interactions, trying to estimate heritability values from twin studies and neglecting the analysis of cultural transmission (that can be modeled as correlations between parent and offspring environments) is nothing but a dead-end.

Consequences of assortative mating with vertical transmission on intra-familial correlation

To complement the previous section, we give here a simple numeric example of the implications of assortative mating in the ACE model. In table 3, we consider a quantitative trait which has a heritability of 0.5 in a population without assortative mating. The parameter r is the correlation between mates for this trait, and v is the correlation between each of the parents' environment, and the offspring's environment. From these parameters, one can compute ρ, the correlation between genetic and environmental components. The total variance of the trait is σ^2, which can be broken done as the sum of three terms, $\sigma^2 = a^2 + 2\rho ae + e^2$.

Table 3: Example of implications of assortative mating in the *ACE* model

Parameters			Variance decomposition			Total variance	Parent-offspring
r	ν	ρ	a^2	$2\rho ae$	e^2	σ^2	correlation
0	0	0	0.5	0	0.5	1	0.25
0.6	0	0	0.79	0	0.5	1.29	0.49
0.6	0.4	0.15	0.85	0.30	0.5	1.55	0.68

When $r = \nu = 0$ (first line of the table), which is the situation in which the *ACE* model can be easily interpreted, the gene-environment correlation ρ is null and the total variance decomposes nicely in $\sigma^2 = a^2 + e^2$ with, for this example, $a^2 = e^2 = 0.5$, corresponding to a heritability $h^2 = 0.5$; the parent-offspring correlation is then 0.25, corresponding to $\frac{1}{2}h^2$.

If one assumes that there is some assortative mating in the population, by letting $r = 0.6$ (second line of the table), still neglecting to take into account cultural transmission ($\nu = 0$), the gene-environment correlation ρ remains equal to 0, but the genetic variance is increased by 58% and the total trait variance by 29% (the environmental variance remaining constant). The trait heritability, if one defines it by $h^2 = \frac{a^2}{a^2+e^2}$, would then be 0.61. However, the parent-offspring correlation is now equal to 0.49, an estimate that could not be deduced from the value of h^2 without knowledge of the value of r.

In the third line of the table, we introduce a correlation $\nu = 0.4$ between each of the parents' environment, and the children's environment. This induces a gene-environment correlation $\rho = 0.15$, and further increases the genetic variance. The total trait variance is now increased by 55% as compared to the ideal situation of the first line. It cannot be decomposed in the sum of a genetic and an environmental variance anymore, as the term $2\rho ae$ is neither purely genetic, nor purely environmental; it is not possible to compute a value for h^2 – should it incorporate $2\rho ae$ or not? The parent-offspring correlation is now as high as 0.68, a value which, again, only makes sense if one considers the presence of assortative mating and of cultural transmission.

References

Crow, James F and Joseph Felsenstein. 1968. "The effect of assortative mating on the genetic composition of a population." *Eugenics Quarterly* 15:85–97. https://doi.org/10.1080/19485565.1968.9987760.

Falconer, Douglas S. and Trudy F.C. Mackay. 1996. "Introduction to quantitative genetics. Essex." *Longman* .

Fisher, Ronald A. 1919. "The Correlation between Relatives on the Supposition of Mendelian

Inheritance." *Transactions of the Royal Society of Edinburgh* 52:399–433. https://doi.org/10.1017/S0080456800012163.

Lewontin, Richard C. 1974. "The analysis of variance and the analysis of causes." *American Journal of Human genetics* 26:400.

Nagylaki, Thomas. 1978. "The correlation between relatives with assortative mating." *Annals of Human Genetics* 42:131–137.

Noûs, Camille and Hervé Perdry. 2022. "The additive polygenic model with assortative mating and shared parent-offspring environment." *bioRxiv* https://doi.org/10.1101/2022.11.08.515653.

Wright, Sewall. 1921. "Systems of mating. III. Assortative mating based on somatic resemblance." *Genetics* 6:144.