

## The heritability and persistence of social class in England

James J. Lee  
Department of Psychology  
University of Minnesota Twin Cities  
75 East River Road  
Minneapolis, MN 55455  
[leex2293@umn.edu](mailto:leex2293@umn.edu)

The extent to which hereditary abilities affect social class has been a matter of interest since classical times (1). In 1918, Ronald Fisher published a framework for answering such questions about the inheritance of continuously varying traits (2) and thereafter applied that framework to the “mental and moral qualities” that he presumed to be at work in the stratification of human populations (3, 4). In this issue Clark (5) reports an application of Fisher’s methods to English pedigree data spanning centuries and achieves a *tour de force* of insight into the inheritance of social status.

The heritability of a trait is the fraction of its variance attributable to genetic differences. Fisher showed that the correlation between relatives induced by genetics is equal to the heritability times a coefficient that depends on the genealogical relationship between the relatives and the extent of assortative mating (spousal resemblance in the trait). By fitting several such correlations to the theoretical equations, one can obtain estimates of the key parameters (heritability, assortative mating). Clark applied this theory to several measures of social class in a remarkable dataset formed by combining the immense labors of amateur genealogists with publicly available records of births, baptisms, marriages, probates, and so forth. It is natural to suspect such data of biases; but Clark found, among other things, that members of different lineages paired at random show no correlations in any of the measured social outcomes. This work is a nice illustration of the statistician David Freedman’s argument that ingenuity and hard work in the collection of informative data—“wearing out the shoe leather”—can prove more scientifically rewarding than the sophisticated technical treatment of data that is easier to find (6).

One can always fit data to theoretical predictions and get out numbers, but Clark’s exercise is not a trivial one. His Figure 2 displays the remarkable closeness of the empirical correlations between nine different types of relatives to theoretical predictions depending on only two adjustable parameters. The outcomes in these two panels are log house value and a measure of occupational status. Not all of Clark’s measures of social class show such a tight fit, but the average correlation is high. Freedman’s essay cited the psychologist Paul Meehl (7), who lamented the rarity in behavioral research of fitting data to narrowly constrained, closely reasoned predictions. Meehl would surely have been delighted by the present work. Specialists in quantitative genetics may be intrigued by Clark finding a use for the correlation between double first cousins, a point once considered theoretically significant but surely too obscure for empirical application (8, 9).

So what is the importance of the numbers that Clark gets out? The estimates of heritability ranged from 0.19 to 0.72, depending on the measure of status and the time period. Although certain measures could not be constructed consistently across time, their heritabilities did decline somewhat over the generations and upon extrapolation are perhaps consistent with modern estimates (10-12). Much more striking are the estimates of the correlations between the genetic values of spouses, which were remarkably constant across measures and averaged a surprising 0.57. As Clark concedes, until recently such an estimate would have been dismissed out of hand “on the grounds that spousal correlations in genetic values could not be so high.”

We might distinguish three possible kinds of models explaining the generational persistence of the abilities underlying social status (13, 14). The first is a simple autoregressive model that has sometimes been used in social science, where the correlation in ability or skill across  $n$  generations is essentially  $r^n$ , the correlation between parent and child to the power of the number of steps in the genealogy (2 for grandparent and grandchild, 3 for great-grandparent and great-grandchild, and so on). Since most single-generation correlations in observable traits are well short of  $1/2$ , this model implies a very rapid decay in resemblance across generations. The second kind of model acknowledges the genetic contribution to ancestor-descendant resemblance but posits random pairing of fathers and mothers, which yields the correlation  $(1/2)^n$  times the heritability. This is still a rapid decay across generations. The third kind of model allows mating to be assortative, in which case the correlation becomes essentially  $[(1 + m)/2]^n$  times the heritability,  $m$  being the correlation between the genetic values of spouses. Strong assortative mating is reasonably regarded as a qualitatively different regime altogether, since with a high enough value of  $m$  (e.g., 0.57) regression to the mean across generations is dramatically slowed and allows the common descendants of a Victorian ancestor to show quantitatively significant genetic resemblance even today. This is because such a scheme of strong spousal matching ensures that the other ancestors of the individual’s descendants tend to transmit DNA of like effect. Such persistent correlations between very distant relatives are what drive Clark’s inferences of strong assortative mating.

One naturally wonders whether the results might be explained by environmental sources of familial resemblance that are confounded with genetic relatedness. Clark reports compelling ancillary evidence, however, against explaining away his findings in such a way. First, he found that the magnitudes of the father-son correlations in occupational status and educational attainment were moderated hardly at all by the son’s age (ranging from zero to maturity) at his father’s death. Second, he found that status was transmitted equally through paternal and maternal lines, exactly as expected if the responsible causal factors are the genes transmitted by fathers and mothers alike. As a negative control, wealth *per se* did show an asymmetric tendency to be transmitted more strongly through paternal lines.

The trait (or composite of traits) with respect to which people are mating assortatively might be regarded as a noisy reflection of the genetic value, and people surely do not have the X-ray vision to discern genetic values. That a correlation between the genetic parts of 0.57 can be attained even after people match with respect to a noisy overall sum (i.e., genetic value + environmental noise) implies a trait-level spousal correlation in excess of the genome-level 0.57

reported by Clark, and some may doubt that any attributes of spouses can be correlated so highly. Clark also has answers to this objection. First, he found that grooms resemble their fathers-in-law in occupational status nearly as much as they resemble their own fathers, suggesting that there is indeed a very strong resemblance between grooms and brides in latent social class. Second, he cites DNA-level studies of educational attainment finding correlations between the estimated genetic scores of fathers and mothers seemingly in excess of what can be explained by partners mating assortatively with respect to observed attainment. There are heritable differences between degree holders of the same nominal rank, it seems, and people care about these differences when looking for partners. This interpretation agrees well with Clark's own hypothesis of a highly heritable latent trait underlying noisy measures of social status, a trait that people reliably assess and seek out in others. Clark actually understates the similarity between his work and Torvik *et al.* (11). This paper applied structural equation modeling to the estimated genetic scores and observed educational attainments of several Norwegian families, each consisting of two siblings and their partners. They posited that partners mate assortatively with respect to a latent trait (of which educational attainment is a noisy reflection) and estimated the correlation between partners in this trait to be 0.68—not so far from Clark's 0.79. Overall, however, Clark's arguments for good quantitative agreement between the DNA-level studies and his own estimates are somewhat overstated at the present time, as a result of lingering uncertainty over possible uncontrolled confounding in the genome-wide association studies used to construct the estimated genetic scores (15).

But more can be added to Clark's case. The near constancy of  $m$  across different measures of social class suggests that they are all noisy estimates of more or less the same trait that couples are matching on, and DNA-level studies do support this. Different measures of social class today show trait-level correlations ranging between 0.3 and 0.4, but their noise-free genetic correlations range between 0.8 and 1.0 (16). Furthermore, is a high degree of matching with respect to the underlying trait really so implausible? A recent study reported a correlation of 0.82 between spouses in a composite of political views (17). This estimate was probably inflated by ascertainment bias, as all couples in this study were required to have at least two children, but there can be no doubt that the true correlation is indeed very high. If people care enough about something, they can assess it accurately without administering a detailed test or questionnaire and sort themselves accordingly.

Fisher anticipated this point when he defined social class for his own purposes in terms of permissible marriage partners.

[P]revailing opinion, mutual interest, and the opportunities for social intercourse, have proved themselves sufficient, in all civilized societies, to lay on the great majority of marriages the restriction that the parties shall be of approximately equal social class. In this statement social class should, of course, be taken to comprehend, not merely income or wealth, but also the prestige attaching to occupation, personal talents, and family associations. Its meaning is thus somewhat different from, though closely correlated with, the purely economic use of the term. But the factor of intermarriage is so important in its social and biological consequences that it will be best to use the term

‘social class’ solely in this sense and to lay down that the social class of an individual or his family shall be defined by the aggregate of persons or families, intermarriage with whom will encounter no social obstacles. (4, pp. 210-211)

If Clark is correct, then social class can be defined more conventionally so as to render Fisher’s condition regarding marriage an empirical fact rather than a tautology.

Finally, we should keep in mind that the abilities posited to underlie the attainment of status are just that—abilities useful in getting to the top. There is no implication that the mix of relevant characteristics includes virtue, which even in ancient times (*areté*) was recognized as distinct from mere status. When we examine what the people at the top have wrought in our country today, we see clear signs of their inadequacies (18-20). We need a new elite. Let us hope that the surprisingly lawlike behavior of movement in and out of elite status that Clark seems to be revealing does not preclude a badly needed renewal.

1. Plato, *Republic* (Hackett, ca. 370 BC/1997).
2. R.A. Fisher, The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* **52**, 399-433 (1918).
3. R.A. Fisher, The biometrical study of heredity. *Eug. Rev.* **16**, 189-210 (1924).
4. R.A. Fisher, *The Genetical Theory of Natural Selection* (Oxford University Press, 1930).
5. G. Clark, The inheritance of social status: England, 1600-2022. *Proc. Natl. Acad. Sci. U.S.A.* (2023).
6. D.A. Freedman, Statistical models and shoe leather. *Sociol. Methodol.* **21**, 291-313 (1991).
7. P.A. Meehl, Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *J. Consult. Clin. Psychol.* **46**, 806-834 (1978).
8. T. Nagylaki, *Introduction to Theoretical Population Genetics* (Springer, 1992).
9. L. Yengo, P. Visscher, Assortative mating on complex traits revisited: Double first cousins and the X chromosome. *Theoret. Pop. Biol.* **124**, 51-60.
10. E.F. Eifler, R. Riemann, The aetiology of educational attainment: A nuclear twin family study into the genetic and environmental influences on school leaving certificates. *B. J. Educ. Psychol.* **92**, 881-897.
11. F.A. Torvik *et al.*, Modeling assortative mating and genetic similarities between partners, siblings, and in-laws. *Nat. Commun.* **13**, 1108 (2022).
12. J. Erola, H. Lehti, T. Baier, A. Karhula, Socioeconomic background and gene-environment interplay in social stratification across the early life course. *Eur. Sociol. Rev.* **38**, 1-17 (2022).
13. G. Clark, *The Son Also Rises: Surnames and the History of Social Mobility* (Princeton University Press, 2014).
14. A. Rustichini, W.G. Iacono, J.J. Lee, M. McGue, Educational attainment and intergenerational mobility: A polygenic score analysis. *J. Polit. Econ.* (2023).
15. A.I. Young *et al.*, Mendelian imputation of parental genotypes improves estimates of direct genetic effects. *Nat. Genet.* **54**, 897-905.

16. E.T. Akimova, T. Wolfram, X. Ding, F.C. Tropf, M.C. Mills, Genome-wide association study of occupational status and prestige identifies 106 genetic variants and defines their role for intergenerational status transmission and the life course. *bioRxiv* (2023).
17. E.A. Willoughby *et al.*, Parent contributions to the development of political attitudes in adoptive and biological families. *Psychol. Sci.* **32**, 2023-2034 (2021).
18. A. Case, A. Deaton, Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21<sup>st</sup> century. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15078-15083 (2015).
19. S.H. Preston, Y.C. Vierboom, Excess mortality in the United States in the 21<sup>st</sup> century. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2024850118 (2021).
20. Bollen, J. *et al.* Historical language records reveal a surge of cognitive distortions in recent decades. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2102061118 (2021).