Clark (2023) and the Persistence of Hereditarian Fallacies

John W. Benning[1]*, Jedidiah Carlson[2,3]*, Ruth G. Shaw[4], Arbel Harpak[2,3]
[1]Department of Botany, University of Wyoming, Laramie, WY, USA
[2]Department of Integrative Biology, University of Texas at Austin, Austin, TX, USA
[3]Department of Population Health, University of Texas at Austin, Austin, TX, USA
[4]Department of Ecology, Evolution and Behavior, University of Minnesota, St. Paul, MN, USA

*these authors contributed equally
Correspondence should be addressed to: jwbenning@gmail.com, jedidiah.carlson@utexas.edu, shawx016@umn.edu and arbelharpak@utexas.edu

Abstract

Clark (2023) considers the similarity in socioeconomic status between relatives, drawing on records spanning four centuries in England. The paper adapts a classic quantitative genetics model in order to argue the fit of the model to the data suggests that: (1) variation in socioeconomic status is largely determined by additive genetic variation; (2) contemporary English people “remain correlated in outcomes with their lineage relatives in exactly the same way as in preindustrial England”; and (3) social mobility has remained static over this time period due to strong assortative mating on a “social genotype.” These conclusions are based on a misconstrual of model parameters, which conflates genetic and non-genetic transmission (e.g. of wealth) within families. As we show, there is strong confounding of genetic and non-genetic sources of similarity in these data. Inconsistent with claims (2) and (3), we show that familial correlations in status are variable—generally decreasing—through the time period analyzed. Lastly, we find that statistical artifacts substantially bias estimates of familial correlations in the paper. Overall, Clark (2023) provides no information about the relative contribution of genetic and non-genetic factors to social status.

Introduction

Teasing apart genetic and non-genetic contributions to trait variation in humans is notoriously fraught: ancestors transmit not only genes to descendants, but also a wide array of non-genetic factors that influence traits—wealth, place of residence, knowledge, religion, culture, and more. Experiments that could decouple genetic and non-genetic influences, as used to study other organisms, are unethical and infeasible in humans. A long history of work has highlighted the confounding between these contributors to human phenotypic variation, especially when only phenotypic data are considered (Bailey, 1997; Feldman and Lewontin, 1975; Feldman and Ramachandran, 2018; Holtzman, 2002; Lewontin, 1974; Shen and Feldman, 2023; Young et al., 2019).

Recently, (Clark, 2023a) analyzed familial correlations in a dataset of socioeconomic measures (occupational status, house value, literacy, etc.) from a selection of English pedigrees spanning the 18th to 21st centuries. Based on the fits of these observed correlations to a quantitative genetic model of trait inheritance [(Fisher, 1918; Gimelfarb, 1981); **Supplementary Note 1**], (Clark, 2023a) infers that social status persists intergenerationally because parents mate assortatively on a status-determining genotype (or “social genotype” as used by the author in previous work (Clark, 2014)). (Clark, 2023a) then argues that because mates share the genes underlying social status to such a high degree, the persistence of social status within families—and persistence of differences in status among families—have been largely unaffected by changes in social policy in the last four centuries. In a recent commentary about this work (Clark, 2023b), the author presents the results of (Clark, 2023a) as providing strong support for a “hereditarian” (see Mehler, 2015) interpretation and invokes the metaphor of social status being decided by the results of a "genetic lottery," a conceptualization increasingly in vogue in social and behavioral genetics (Harden, 2021) (critiqued in (Coop and Przeworski, 2022a, 2022b; Fletcher, 2022; Panofsky, 2021)).

Here, we discuss three core issues regarding these claims (see discussion of other misinterpretations and errors in (Clark, 2023a) in **Supplementary Notes 2-7**; **Table S1**; **Fig. S1**). First, and most important, we explain the failure in (Clark, 2023a) to confront the confounding of genetic and non-genetic transmission (**Fig. 1**) and demonstrate strong confounding between these two modes of inheritance in the data (**Fig. 2**). Second, we show that the estimated decay in correlations across genealogical relationships is partly attributable to statistical artifacts (**Fig. 3**). Third, we find that familial correlations varied substantially over the

time period examined, generally decreasing (**Fig. 4**). This finding stands in contrast to the paper's claims of familial correlations being insensitive to changes in social policy. In summary, the data and analyses in (Clark, 2023a) provide no information about the contribution of genetics to social status.

The problem of confounding genetic and non-genetic transmission

The claims in (Clark, 2023a) are based on the results of a linear regression model derived from theoretical work by R.A. Fisher (Fisher, 1918; Gimelfarb, 1981) (**Supplementary Note 1**). Fisher (1918) formally derived the heritability (h^2) of a trait with Mendelian inheritance and showed that resemblance between relatives with respect to the trait arises from its heritability times a coefficient that depends on the genealogical relationship between the relatives and the extent of assortative mating (h^2 is the ratio of additive genetic variance to total phenotypic variance, nowadays referred to as "narrow-sense" heritability). Crucially, this model makes the assumption that the covariance between biological relatives in a trait is purely due to a genetically heritable component. In particular, it assumes there are no non-genetic (material, environmental, or cultural) influences on a trait that are systematically shared or transmitted between relatives. As one example, (Clark, 2023a) assumes that similarity in house value (one of the measures of social status analyzed) is strictly due to shared genes, and does not arise from similarity in parental wealth between those relatives, or from the inheritance of wealth or property, or from having learned from one's relatives how to invest.

If non-genetic sources of similarity are present (as is common for many traits in humans), such a model no longer describes genetic heritability, but rather a "total transmissibility" (t^2) comprising both genetic and non-genetic sources of familial resemblance (Rice et al., 1980). Scholars from various disciplines have long appreciated the myriad ways in which phenotypic similarities among family members arise through non-genetic pathways (**Fig. 1**) and may follow the same correlation patterns expected from genetic transmission (Barton et al., 2019; Cavalli-Sforza and Feldman, 1978, 1973a, 1973b; Cloninger et al., 1979a, 1979b; Collado et al., 2023; Feldman et al., 2013; Herzig et al., 2023; Lewontin et al., 1984; Rao et al., 1976, 1974; Rice et al., 1980, 1978; Solon, 2014; Uchiyama et al., 2022; Vilhjálmsson and Nordborg, 2013; Wright, 1931). Mechanisms of non-genetic transmission include "ecological inheritance," where the trait value of an offspring is influenced by the environmental conditions created by their

parents (e.g., familial wealth influencing educational opportunities) (Cavalli-Sforza and Feldman, 1978; Odling-Smee, 1988), and the diffusion of information directly to one's relatives (e.g., literate parents teaching their children how to read) (Cavalli-Sforza and Feldman, 1973b). Non-genetic transmission is ubiquitous for social and behavioral traits and confounds estimation of the genetic sources of phenotypic resemblance between relatives [*cf.* recent work describing such effects on educational attainment (Collado et al., 2023) and occupational status (Akimova et al., 2023)].

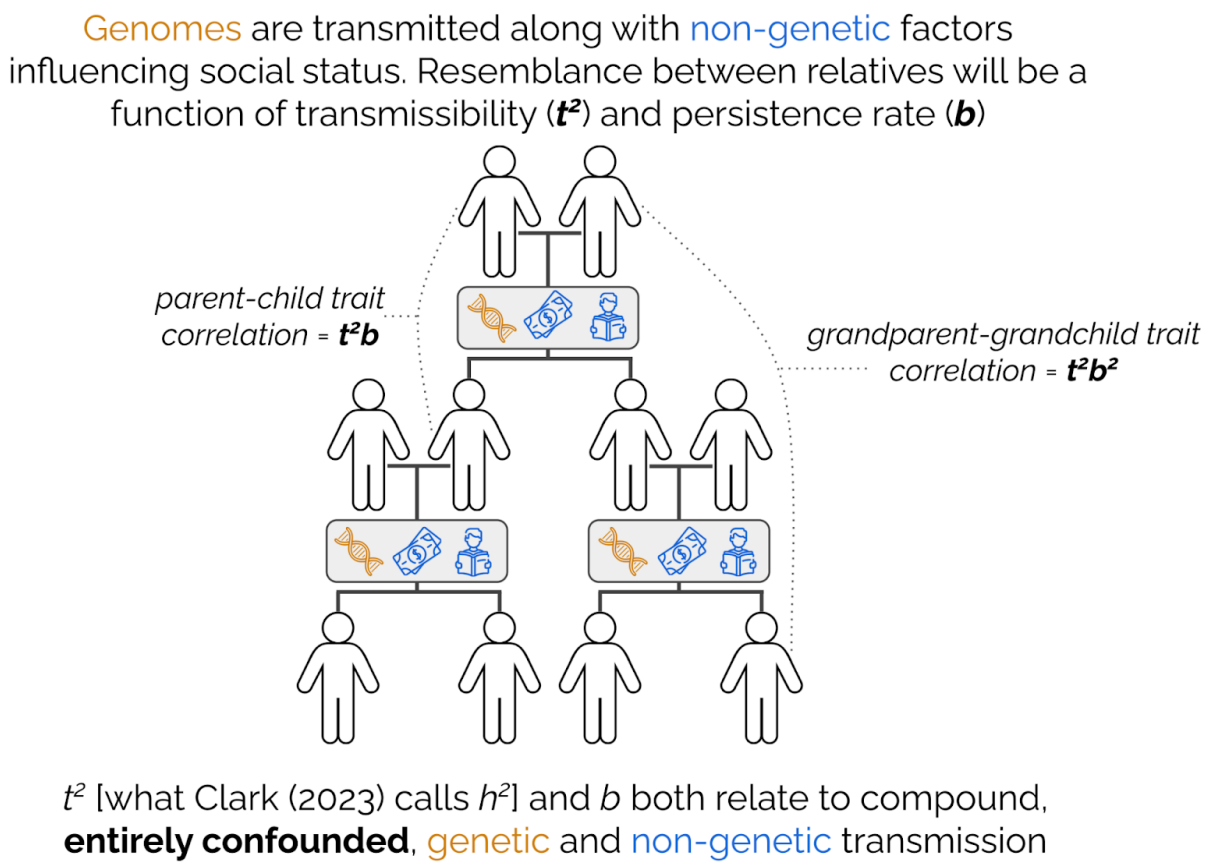


Figure 1. Illustration of how transmissibility (t^2) and rate of decay of correlations across generations (the "persistence rate," b) influence trait resemblance between relatives in the model used by Clark (2023a). b will increase as assortative mating increases. Ignoring the confounding of genetic and non-genetic transmission in the data, Clark (2023a) misassigns all transmission as genetic heritability and all persistence due to assortative mating to be due to spousal correlation with respect to a latent genetic factor.

An extensive body of work has highlighted the confounding between genetic and non-genetic contributors to phenotypic variation when experimental randomization of genotypes over environments is impossible or unethical, and the resulting inability to decouple them using trait correlations alone (Bailey, 1997; Feldman and Lewontin, 1975; Feldman and Ramachandran, 2018; Holtzman, 2002; Lewontin, 1974; Shen and Feldman, 2023). In this context, it is helpful to consider genome-wide association studies (GWAS), aimed at detecting associations between traits and genomic regions. To adjust for confounding, the gold standard method in GWAS is a linear mixed model, which aims to regress out phenotypic resemblance that tracks genetic relatedness. The rationale for this approach is that, when genotypes cannot be randomized over environments, true additive genetic effects are inseparable from other factors underlying phenotypic resemblance between relatives (Barton et al., 2019; Loh et al., 2015; Vilhjálmsson and Nordborg, 2013). In contrast, (Clark, 2023a) assumes *a priori* that this signal of relatedness correlating with phenotypic similarity—well known to geneticists as confounded—reflects only genetic causality.

In point of fact, there are signals of strong confounding between genetic and non-genetic effects on familial resemblance in the data used in (Clark, 2023a). The paper acknowledges the inheritance of material wealth from one's parents as an obvious example of non-genetic transmission—but only when considering wealth itself as the focal trait. For other focal measures, the effect of familial wealth on social status writ large is ignored. Familial wealth can influence a wide range of conditions that offspring experience (e.g., healthcare, place of residence, access to tutors, social circles, etc.). Thus, we were not surprised to discover that all seven status measures analyzed in (Clark, 2023a) are substantially correlated with an individual's father's wealth (Pearson r ranging from 0.19 - 0.66; mean r = 0.36; all $P < 2 \times 10^{-16}$; **Table S2**; **Fig. 2a**; paternal wealth data was available only for 26,813 individuals). These associations of status measures with paternal wealth impede the effort to distinguish the influence of parental wealth and influences of other factors, including genetics, on status. Closer relatives tend to have more similar paternal wealth, and the similarity in paternal wealth between relatives predicts their similarity in occupational status extremely well (Pearson r = 0.91; **Fig. 2b**). This analysis demonstrates the strong confounding in these data between transmission of genes and the effects of parental wealth in their influence on familial similarity in social status. It is important to note that numerous other non-genetic factors, apart from wealth, may contribute to familial correlations (Feldman and Cavalli-Sforza, 1981; Turkheimer, 2000).

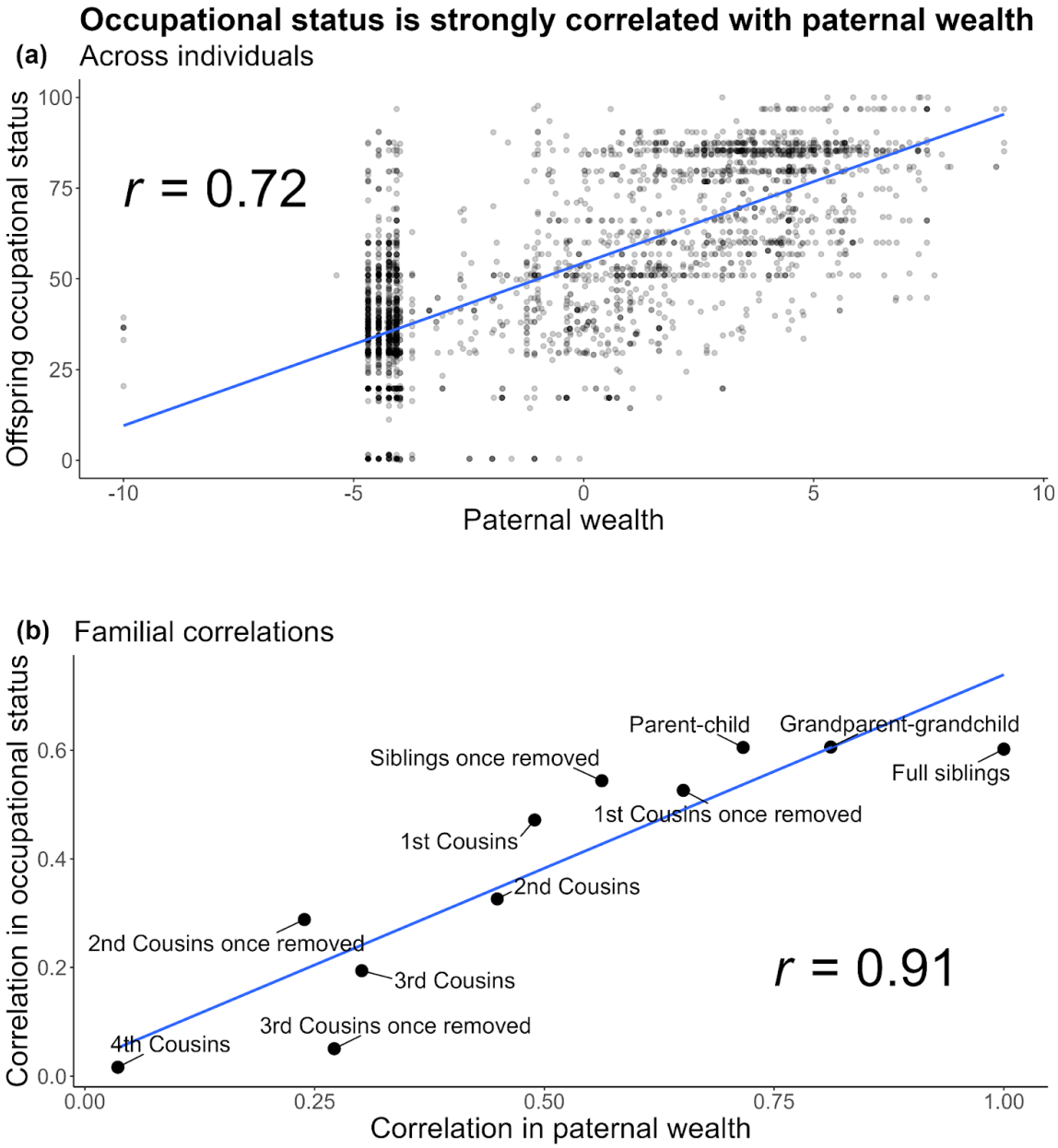


Figure 2. An individual's occupational status is strongly correlated with the wealth of their father. (a) Offspring occupational status is well predicted by paternal wealth. This suggests at least one source of confounding between genetic and non-genetic transmission. **(b)** Relative correlations in occupational status are highly correlated with those relatives' correlation in paternal wealth. Plots show data for individuals born 1780-1859 for which paternal wealth data was available. N = 2,886 individuals in panel (a) and N = 13,030 pairs in panel (b).

This confounding prevents genetic from non-genetic sources of familial resemblance from being distinguished. However, (Clark, 2023a) presents two post hoc analyses as support for genetic effects on social status: (i) the similarity of the correlation of offspring status with maternal status and with paternal status [(Clark, 2023a) Figure 3], and (ii) the relative invariance of father-son correlations to a son's age at his father's death [(Clark, 2023a) Figure 4]. We show that two serious flaws—not confronting the lasting effects of paternal wealth (beyond a father's death) and using disparate datasets—compromise these analyses and stymie inferences based on them (**Supplementary Note 4**; **Fig. S1**). Crucially, patterns in transmission of wealth that (Clark, 2023a) acknowledges are largely due to non-genetic mechanisms are indistinguishable from transmission of other status measures when analyzed validly. More generally, the implied "isolation" of genetic effects is misconceived because it ignores the fact that non-genetic effects of mothers and fathers on their offspring are inseparable from their genetic effects in these data. These analyses are thus uninformative as to the strength of non-genetic effects on resemblance in social status between relatives.

Parental wealth is but one example of non-genetic transmission. More generally, the confounding of genetic and non-genetic transmission means that the parameters of the model used in (Clark, 2023a) are unmoored from their original definitions (**Supplementary Note 1**). The intercept parameter estimated in this model, which in (Clark, 2023a) is interpreted as narrow-sense heritability, h^2 , is in fact an estimate of the "total transmissibility" of a trait, t^2 [i.e., the *total* proportion of variance attributable to transmissible effects, with those transmissible effects including genes, culture, wealth, environment, etc. (Cloninger et al., 1979a; Rice et al., 1980)]. (Clark, 2023a) similarly misrepresents the second key parameter in the model as reflective of a genetic effect. m, which (Clark, 2023a) interprets as the "spousal correlation in the underlying genetics," does not represent a genetic correlation between mates. It is instead the spousal correlation in the transmissible component of the trait. m is derived from the "intergenerational persistence rate," $b = \frac{1+m}{2}$, estimated from the regression model. The expected correlation for a given kinship pair is equal to $t^2 b^n$, where n denotes genealogical distance (**Fig. 1**). [Note that the parameterization of b for father-son and grandparent-grandchild relationships also depends on the degree of assortative mating with respect to the focal trait itself; see **Supplementary Note 1**]. This conflation of genetic and non-genetic transmission helps explain why the model parameters in (Clark, 2023a) that are claimed to represent quantitative genetic parameters, h^2 and m, are much higher than estimates of these parameters

in most other studies that attempt to account for confounding (e.g., Collado et al., 2023; Yengo et al., 2018; Young et al., 2018; Supplementary Note 1).

Beyond this misinterpretation, a source of inflation in t^2 and b may be upstream choices in the collection and preparation of the data. For example, the index of "occupational status" in (Clark, 2023a) was devised by the paper's author and others in a recent preprint [(Clark et al., 2022); **Supplementary Note 5**]. It was derived using an algorithm that maximizes father/son and father/son-in-law correlations in status. Specifically for the data collection and preparation choices (e.g. choice and coding scheme for occupational categories) of (Clark et al., 2022), this method resulted in father/son occupational status correlations ~30% higher than those based on other widely used indices of occupational status (Clark et al., 2022). Given that unusually high correlations of relatives are at the heart of (Clark, 2023a)'s arguments, this appears to be, at least in part, circular reasoning and an unsuitable choice of methodology.

Pseudo-replication biased estimates of familial correlations

The dataset in (Clark, 2023a) consists of hundreds of different family lineages, as inferred by a shared surname. But the trait correlations that form the basis of the paper's claims are "lineage-agnostic," meaning that they are calculated for relative pairs of a given degree from all lineages combined. A key assumption of (Clark, 2023a) (as stated in Appendix 01 of the paper) is that "*whatever sample of the population we start from, estimates of the level of persistence should not be affected.*" That is, if social status is only transmitted via additive genetic mechanisms, first-degree relative pairs should always have the highest correlation, and this correlation should decrease at constant rate as genealogical distance increases, regardless of the cross-section of the population those relative pairs are sampled from. However, this assumption is contradicted by substantial heterogeneity we observe among and within surname lineages (**Supplementary Note 6a**; **Figs. S2, S3**, **S4**, **S5**, **S6**). At the population level, such heterogeneity is consistent with the claim of a log-linear decay of familial correlations with degree of relatedness. However, as we show below, the sampling heterogeneity interacts with statistical errors, driving biases in estimates of familial correlations.

In (Clark, 2023a), a single person is often represented in multiple data points (e.g., as a member of a pair with each one of their sampled cousins) (**Fig. S7**). These pairs are treated as independent observations in (Clark, 2023a), but because the same information is replicated in

multiple relative pairs, such pairwise records are not truly independent from one another. This is commonly known as pseudo-replication. Beyond the widely-appreciated effect of pseudo-replication on the underestimation of statistical uncertainty (Hurlbert, 1984), the estimation of correlations such as that performed in (Clark, 2023a), can be biased. Because the number of relatives typically increases exponentially with genealogical distance, the extent of pseudo-replication for a given relationship type is larger for more distant relatives and for more highly-represented surname lineages (**Supplementary Note 6**; **Figs. S4, S7**, **S8**, **S9**, **S10**). For example, the (1780-1859) occupational status correlation for fourth cousins is calculated from 17,382 pairs, but only 1,878 unique individuals.

We set to eliminate the bias in social status correlation estimates due to pseudo-replication. We randomly sampled pairs of relatives (one pair from each surname lineage), such that no single individual is represented in more than one pair of a given genealogical relationship (**Supplementary Note 6b**), and estimated familial correlations. The revised correlations tended to stay the same or slightly decrease (compared with the uncorrected correlations) among first and second-degree relatives, but increase dramatically for more distant relationships (**Fig. 3**). For example, the correlations in occupational status between fourth cousin pairs increase from ~0.07 to >0.30 in the absence of pseudo-replication (**Supplementary Note 6b**; **Fig. 3**).

We conclude that pseudo-replication had a major effect on the results of (Clark, 2023a). Further, the effect is heterogeneous across levels of relatedness. The variable effect is likely due in part to choices in data collection and preparation—though these effects are difficult to decipher *post hoc* (**Supplementary Note 6c**; **Table S3**; **Fig. S9**). Together, the model misspecification and statistical artifacts further put into question inferences drawn in (Clark, 2023a) based on familial correlations in status.

What does the “persistence rate” measure?

Observed persistence cannot be attributed to genetic mechanisms

Claims about the insensitivity of familial correlations to social interventions in (Clark, 2023a) rest on the paper’s finding that the parameter b (“the persistence rate of the correlation as we move one step down the family tree, or one step across between full siblings”) is similar across status measures and across time. For example, if the correlation in occupational status between first cousins is 80% that of uncles and nephews, which is 80% that of full siblings and so on, then

$b = 0.8$ (**Fig. 1**; also see Table 2 in (Clark, 2023a)). As discussed above and in **Supplementary Note 1**, in (Fisher, 1918)'s generative model, b is a deterministic function of the assortative mating parameter m, the correlation between spouses in the transmitted component of the trait.

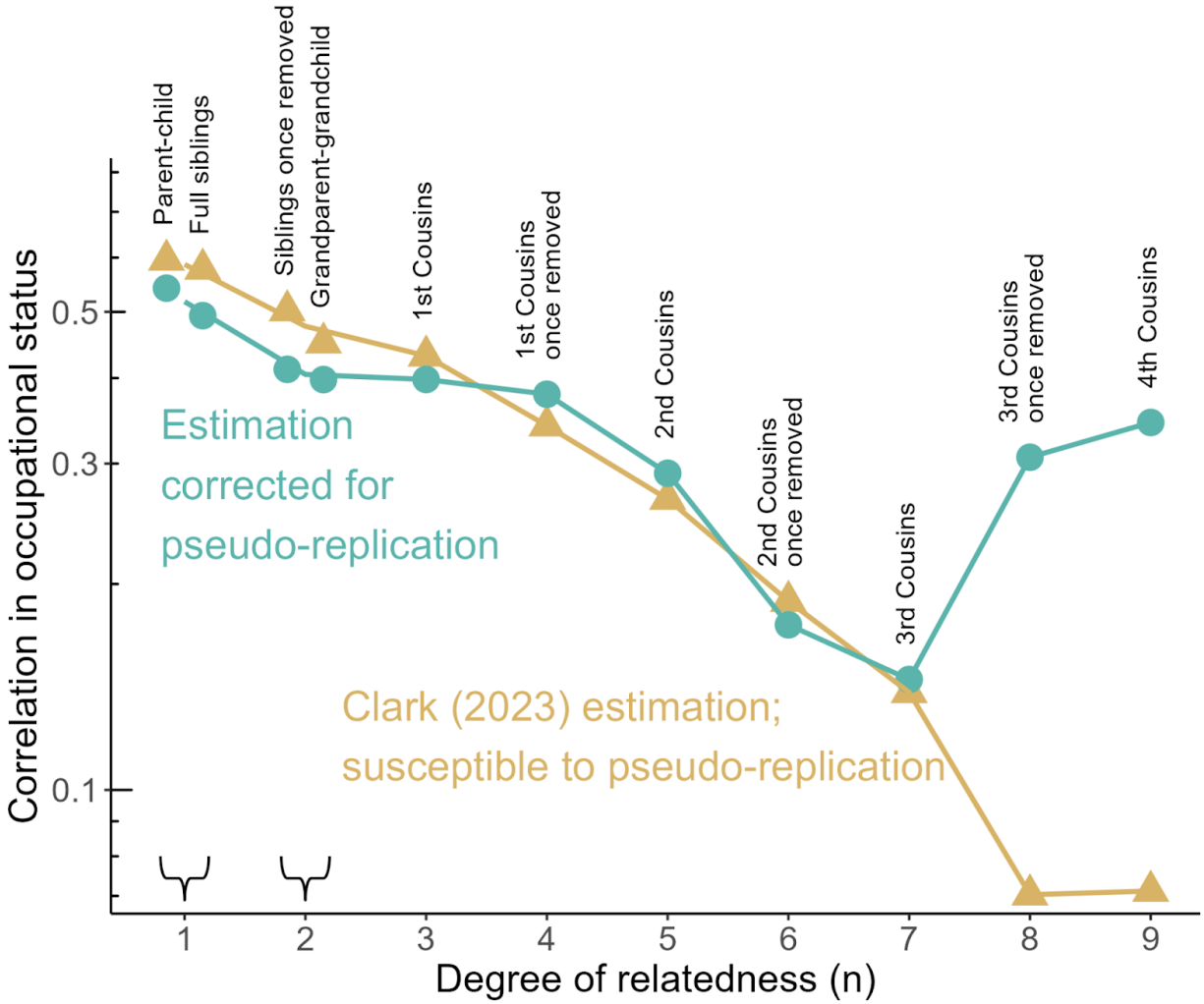


Figure 3. Familial correlations in occupational status (1780-1859) using the approach employed by (Clark, 2023a), which involved pervasive pseudo-replication, versus a conservative estimate that did not, using only a single relative pair per surname (mean estimate over 1000 bootstrap samples is plotted for each familial correlation). Distant cousins show dramatically higher correlations when adjusting for pseudo-replication.

(Clark, 2023a) argues that b is stable across time, traits, and families—and that this stability is due to strong assortative mating on a genetic factor for "social ability", estimated as a genetic correlation of $m = 0.57$ between mates. However, the claim in (Clark, 2023a) that *m* represents the strength of assortative mating by a genetic factor is invalid. Once one acknowledges that both genetic and non-genetic factors are transmitted within families (**Supplementary Note 1**; **Figs. 1, 2**), m tells us nothing about genetic versus non-genetic contributions to assortment, and b tells us nothing about the genetic basis for within-family persistence of social status.

Regardless of the cause, evidence for persistent patterns of familial correlations in status during four centuries in England and across families would be striking. However, as we discuss below, inference based on the (Gimelfarb, 1981) model implemented in (Clark, 2023a) does not provide such evidence.

The persistence rate does not reflect social mobility

We now set aside the misconstrual of the persistence rate, b, as suggestive of a genetic mechanism, along with biases due to statistical artifacts. In (Clark, 2023a), the large value of $b \approx 0.79$, and its stability across social status measures and between two time periods (for two of the measures), are taken as evidence for a persistence of social status and rate of social mobility that has been largely unaffected by societal changes:

> "*People in 2022 remain correlated in outcomes with their lineage relatives in exactly the same way as in preindustrial England.*"
>
> "*The vast social changes in England since the Industrial Revolution, including mass public schooling, have not increased, in any way, underlying rates of social mobility*".

Across time-based comparisons presented in the paper and all degrees of relationship [(Clark, 2023a) Table 2], 16/22 correlations decrease between the two time periods analyzed (on average, decreasing 31%). How could the estimate of b lead to conclusions that contrast what is suggested by the vast majority of data points from which it was inferred?

First, even in the absence of a clear trend of decay, a large estimate of b is likely. For example, fitting the regression model of (Clark, 2023a) to the correlation estimates adjusted for

pseudo-replication shown in **Fig. 3** (where there is no obvious relationship and the linear fit has $R^2 = 0.29$) yields an estimate of $b = 0.93$.

Second, even with a strong linear fit to the data, b is uninformative as to the magnitude of familial correlations and describes only the rate of decrease in the correlation as relatedness declines. For illustration, b approaches 1 as the association between familial correlation and relatedness nears zero [(Clark, 2023a) Eq. 3]. Importantly, familial correlations in a status measure could systematically increase or decrease by any amount, with b remaining unchanged (**Fig. S11**).

Third, the estimate of b can be greatly affected by invalid statistical modeling. We highlight two examples. First, (Clark, 2023a) effectively standardizes trait variance across different relationship types. This procedure obscures systematic trends that we observed in trait mean and variance across these subsets, which are incompatible with (Fisher, 1918) and (Gimelfarb, 1981) model assumptions (**Supplementary Note 1**; **Supplementary Note 6c**; **Figs. S8, S9, S10**). A second example lies in increased sensitivity of the estimate of *b* to small changes in the correlations of distant relatives, but relative insensitivity to the correlations of closer relatives (**Supplementary Note 7**; **Fig. S12**). As we have seen empirically, estimates of status correlations for distant relatives were also most sensitive to pseudo-replication and substructure among and within lineages (due to sampling bias and/or temporal changes in the population distribution) (**Supplementary Note 6**; **Figs. 3**, **S8**, **S9**, **S10**).

Familial correlations decreased over time

Given these characteristics of b, it is worth considering more established metrics of social mobility, such as changes in parent-offspring correlations in social status over time (Causa and Johansson, 2009; Chetty et al., 2014; Longley et al., 2021), in data from (Clark, 2023a). Parent-offspring correlations in occupational status, higher education, and literacy (the only three status measures with data prior to the 20th century) decrease over time (**Fig. 4**; **Fig. S13**). Correlations in occupational status and higher education for sibling, grandparent/grandchild, uncle/nephew, and first cousin relationships also tend to decrease (**Fig. S13**).

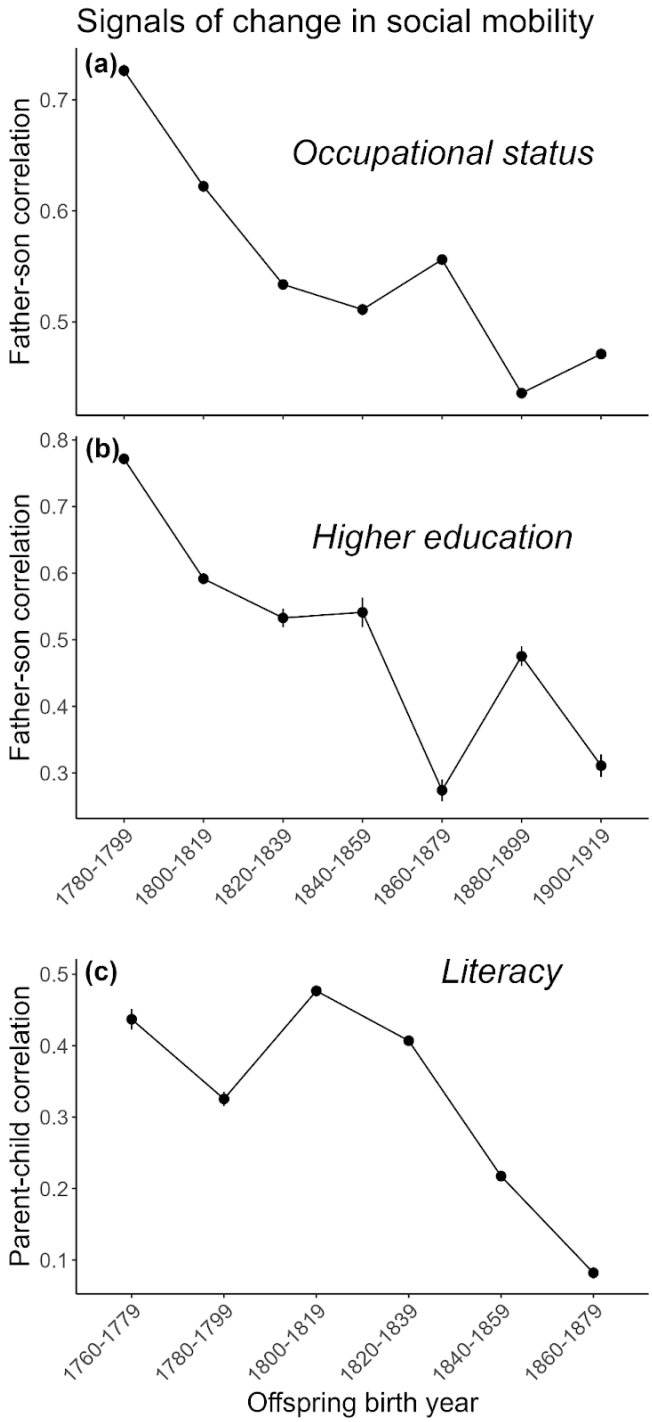


Figure 4. Parent-offspring correlations over time. **(a)** Occupational status 1780-1919; **(b)** higher education 1780-1919; **(c)** literacy 1740-1879. To mitigate pseudo-replication, we calculated correlations using one pair from each surname. Shown are average correlations across 500 bootstrap iterations of correlation estimation. **Fig. S13** shows two complementary analyses where correlations are estimated either without accounting for pseudo-replication, or using percentile ranks—both result in similar trends.

Conclusion

Some readers have already taken arguments in (Clark, 2023a) as compelling evidence that societal variation in social status is largely caused by genetic variation amongst individuals (Cosh, 2023; Lee, 2023; Marks, 2023). Yet the assumptions and interpretations in (Clark, 2023a) ignore decades of quantitative genetic theory, previous evidence for pervasive confounding and the fallacies that arise when it is ignored (Barton et al., 2019; Berg et al., 2019; Coop and Przeworski, 2022a, 2022b; Feldman, 2014; Feldman and Lewontin, 1975; Lewontin, 1974; Lewontin et al., 1984; Mostafavi et al., 2020; Sohail et al., 2019; Young et al., 2019; Young, 2023), as well as conflicting patterns in the paper's own data. (Clark, 2023a) does not merely overstate the findings—the model parameters are misconstrued and the pervasive confounding of genetic and non-genetic transmission in these data is not interrogated.

The study of heredity in humans has long been plagued by failures to address the implications of confounding between genetic and non-genetic sources of variation. Even today, when the inherent limitations of observational data are well appreciated, some continue to downplay or ignore such confounding in order to advance claims about an outsized role for genetics. As we have shown, (Clark, 2023a) represents yet another example of this tradition.

Code and Data Availability

All code for reproducing these analyses are available at:
https://github.com/harpak-lab/Clark2023.

Raw data used in these analyses are from Datasets S01-S04 from (Clark, 2023a). A corrected version of the occupational status data was provided to us by the author, but has not been made publicly available.

Acknowledgements

We thank Mark Borrello, Graham Coop, Doc Edge, Sasha Gusev, Kelley Harris, Magnus Nordborg, Molly Przeworski, Noah Rosenberg, and James Schmitz for comments on the manuscript and helpful discussions. We thank Gregory Clark for providing us with a corrected version of the occupational status data. The work was funded by NSF DBI-2010892 to J.W.B and R35GM151108 to A.H.

Supplementary Materials for:

Clark (2023) and the Persistence of Hereditarian Fallacies

Supplementary Notes

Supplementary Note 1: Overview of the model used in Clark (2023)

Here, we reiterate in more detail the inference model used in (Clark, 2023a). As described by (Gimelfarb, 1981), the equations from (Fisher, 1918) (reproduced in Table 1 of (Clark, 2023a)) for non-vertical relative pairs can be generalized into the following equation:

$$log(\rho_n) = log(h^2) + n\, log(\frac{1+m}{2}). \qquad \text{[equation 1 from (Clark, 2023a)]}$$

Here, ρ_n is the observed phenotypic correlation between relatives of a given degree, n is the degree of relatedness, h^2 is the heritability of the trait, and m is the genetic correlation between parents. Note that in (Gimelfarb, 1981), ρ_n is parameterized as $cov(x, y)/\sigma^2$, where σ^2 is the population variance in the trait and it is assumed that the same variance estimator is used for all relationship types. In (Clark, 2023a)'s implementation of the model, the sample correlation, $\rho_n = cov(x, y)/\sigma_x\sigma_y$ is used, where σ_x is the sample standard deviation of the trait among the set of records listed first in each relative pair, and σ_y is that of the records listed second in each pair; if these standard deviations differ across relationship types, the correlation coefficients will be based on different denominators. Also note that (Clark, 2023a) uses ρ_n to represent the Pearson correlation coefficient, and not Spearman's rank correlation coefficient (which is customarily denoted as ρ); we follow this convention to differentiate from the phenotypic assortative mating parameter, r.

When ρ_n measures the phenotypic correlation between vertical relationships (e.g., father-son), (Gimelfarb, 1981) finds that

$$log(\rho_n) = log(h^2) + n\, log(\frac{1+m}{2}) + log(\frac{1+r}{2}). \qquad \text{[equation 2 from (Clark, 2023a)]}$$

This equation includes an additional term that is a function of r, the phenotypic correlation between parents. Note that equation 2 as published in (Clark, 2023a) is incorrect: based on the Fisher equations presented in Table 1, the third term on the right hand side of this equation should be $log(\frac{1+r}{2})$, not $log(\frac{1+r}{1+m})$.

Combining these two equations, the h^2, b, and m parameters can be estimated by fitting a linear regression model:

$$log(\rho_n) = a + n\,log(b) + c\,d_{lin}$$ [equation 3 from (Clark, 2023a)]

Where d_{lin} is an indicator variable for single parent-child or single grandparent-grandchild relationships.

For clarity, we rewrite this model as:

$$log(\rho_n) = \beta_0 + \beta_1 n + \beta_2 d_{lin} + \epsilon.$$

Under this generative model, (Clark, 2023a) obtains the following parameter estimates:

$$\widehat{h^2} = exp(\widehat{\beta_0})$$

$$\widehat{b} = exp(\widehat{\beta_1})$$

$$\widehat{m} = 2\widehat{b} - 1$$

The model presented in (Clark, 2023a) assumes *all* transmissibility is narrow sense heritability. For sociobehavioral traits in humans, we must consider some non-genetic transmissibility component, $c^2 > 0$ (following the notation of (Cloninger et al., 1979a), who refer to this parameter as “cultural heritability;” we extend this definition to encompass material transmission and other non-genetic sources of similarity). In the model of (Fisher, 1918), phenotypic correlations between relatives can then be modeled as a function of the total transmissible variance component,

$$t^2 = h^2 + c^2 + 2whc,$$

where whc is the covariance between additive genetic and transmitted cultural factors within individuals [see equation 2 in (Cloninger et al., 1979a), for example]. Thus, allowing that non-genetic inheritance is at play, $\widehat{h^2}$ of (Clark, 2023a) is not an estimator of the narrow sense heritability, h^2, but of t^2. This partly explains unusually high estimates, such as an estimate of $\widehat{h^2} = 0.94$ for occupational status (Clark, 2023a) Figure 1). Similarly, $\widehat{m}$ is taken to estimate the genetic correlation between parents that are mating assortatively on some latent "social ability" phenotype, and proposes that this latent phenotype plausibly has a narrow-sense heritability of 0.72. However, under the generative model of (Fisher, 1918), the true value of m is a function of t^2 of the latent phenotype (as given by equation 4 of (Clark, 2023a)), so it follows that the description of m as a genetic parameter is mistaken. m in (Clark, 2023a) can only be interpreted as the correlation between spouses in a transmitted latent variable; *b* in turn can be interpreted as the log-additive decay of phenotypic correlations with degree of relatedness. (Collado et al., 2023) interpret *b* as representing "*how strongly* [socioeconomic] *advantages are transmitted from one generation to the next*." If we consider (Fisher, 1918) as the generative process, this persistence of transmission is a result of assortative mating (regardless of whether the factor by which mates assort is fully, partly, or not at all genetic).

Supplementary Note 2: Heritability and assortative mating estimates in other literature

The suggestion in (Clark, 2023a) that social inequality emanates from near-deterministic *genetic* inheritance of social status are at odds with recent literature (Akimova et al., 2023; Collado et al., 2023; Domingue et al., 2014; Yengo et al., 2018). While estimates of narrow-sense heritability and the degree of assortative mating greatly depend on the population considered (geographically, genetically, socially and otherwise), definition of the trait, time, study characteristics and more, we discuss below some recent estimates of these parameters in genetic studies to contextualize estimates in (Clark, 2023a).

(Collado et al., 2023) applied a model of genetic and cultural inheritance under assortative mating to a genealogical dataset from Sweden, similar in principle to that used in (Clark, 2023a) and estimated m to be ~0.025 for educational attainment (EA). Similarly, (Yengo et al., 2018)

estimated m~0.027 for educational attainment using genomic data from ~400,000 residents of the United Kingdom.

The estimate of $m = 0.57$ in (Clark, 2023a) is noted as "surprising" but is claimed to be in accord with two recent genomics studies of educational attainment, referring to (Okbay et al., 2022; Robinson et al., 2017)). First, (Robinson et al., 2017) is cited in (Clark, 2023a) as having found a "correlation [between spouses] at trait-associated loci for educational attainment of 0.654," deemed compatible with $m = 0.57$. However, (Robinson et al., 2017) states that 0.654 is the correlation between the observed phenotype of an individual and the phenotype that is predicted by their partner's genotypes at EA-associated SNPs, and acknowledge that this approach "cannot differentiate between direct assortment on a phenotype and assortment on a genetically correlated trait."

A more recent paper that reported the correlation in polygenic scores (PGS) for educational attainment to be 0.175 (Okbay et al., 2022) is also cited in (Clark, 2023a), with the assertion that "*since the polygenic index is a noisy measure of the full genetic educational potential*, *the full correlation will be significantly higher than this measured correlation*." There follows an attempt to derive what this "full correlation" should be for educational attainment by pivoting to the "analogous case of height" (it is unclear in what sense height should be considered analogous to educational attainment). Using Okbay et al.'s estimate of $r = 0.29$ as the observed phenotypic correlation in height between spouses and an assumed $h^2 = 0.8$ for height, (Clark, 2023a) estimates that the "true" spousal genetic correlation (m) for height must be $rh^2 = 0.236$. Okbay et al. estimated the spousal correlation in PGS for height to be 0.106. From this, the spousal PGS correlation (0.106) is claimed by (Clark, 2023a) to be an underestimate of the true spousal genetic correlation (0.236) by a factor of 1.65-3.27. This same adjustment is then applied by (Clark, 2023a) to Okbay et al.'s spousal PGS correlation for EA (0.175) leading to the conclusion that "*the implied actual* [genetic] *correlation averages 0.39* [$0.175 \times \frac{0.236}{0.106}$], *with a 95% CI* [confidence interval] *of 0.29 to 0.57…so, the 0.175 genetic correlation observed between partners for educational attainment is potentially consistent with a true genetic correlation of 0.57*."

This numerological reasoning has no basis in quantitative genetic theory. The concordance of results from (Collado et al., 2023) and (Yengo et al., 2018)—which developed orthogonal

methods applied to entirely different types of data and estimated m to be more than 20 times lower than 0.57—further underscores flaws in the estimation method deployed in (Clark, 2023a). We also note that the 95% confidence interval given for the implied spousal genetic correlation point estimate—[0.29, 0.57]—is not derived from the probability distribution of the parameter of interest (i.e. the spousal PGS correlation for educational attainment), and is difficult to interpreted as a measure of statistical uncertainty in this parameter estimate.

Supplementary Note 3. Inconsistency of parameter estimates

The model summary statistics (b and h^2 for 9 different traits) presented in (Clark, 2023a) Figure 1 are claimed to be the same as those found in Supplementary Table S2, but these values are different (for example, the heritability of occupational status among men born 1780-1859 appears to be ~0.90-0.95 in Figure 1, but is given as 0.72 in Table S2). Our attempts to reproduce the results in (Clark, 2023a) (using a weighted least-squares regression, where the weights for each ρ_n are equal to $1/\sigma_n$) tend to agree more closely with values in Figure 1 of (Clark, 2023a), suggesting that the parameter estimates in Table S2 of (Clark, 2023a) may have been derived from a different model or used a different version of the dataset (**Table S1**).

Supplementary Note 4: Flawed tests for non-genetic influences

The main analyses in (Clark, 2023a) assume all transmissibility is due to narrow-sense heritability. Then, Clark uses two analyses to rule out “environmental influences” on social status post-hoc.

First, it is shown that the father-son correlations in occupational status and education are relatively unaffected by the son’s age at his father’s death [(Clark, 2023a) Figure 4]. The high correlation between father and son traits even when the father is largely absent from the son’s life is taken in (Clark, 2023a) to indicate the effect of a father on the social status of their child being largely limited to genetic heritability. But many pertinent non-genetic factors and environmental conditions can be transmitted regardless of a father’s presence, such as familial wealth, place of residence, interactions with father’s family, and familial traditions of occupation. Assortative mating is also likely to buffer any decrease in paternal environmental effects, as mothers tend to be correlated with fathers in attributes (literacy, etc.) that shape child outcomes.

We illustrate the predictiveness of non-genetic paternal effects in the absence of fathers by examining whether the association between paternal wealth and offspring status changes when fathers die early in a son's life. In these data, there is no significant effect of a son's age at father's death on the correlation between paternal wealth and son's educational attainment or occupational status, even when the father dies early in his son's life (**Fig. S1a**). This accords with the father-son correlations in education and occupational status that are also high for fatherless boys, as paternal wealth is strongly associated with these status measures (**Fig. 2**, **Table S2**). The absent/present father analysis in (Clark, 2023a) therefore does not help disentangle genetic from non-genetic transmission.

Second, status traits were shown to be transmitted equally through maternal and paternal lines, whereas wealth was transmitted more strongly through paternal lines [(Clark, 2023a) Figure 3]. It is unclear why similar statistical associations of social status with mothers and fathers is taken as evidence for genetic underpinnings. Furthermore, the models for higher education, occupational status, and wealth are fit to different data sets with different distributions of wealth and status (**Fig. S1b**), such that a comparison of coefficients across models is uninformative. We performed the same analysis on the subset of (Clark, 2023a)'s data that had complete information (higher education, occupational status, and wealth) for each individual (N = 817). Using this analysis that rules out differences due to underlying data sets, the maternal wealth and paternal wealth effects on an individual's own wealth are statistically indistinguishable (**Fig. S1c**).

Supplementary Note 5: Construction of the occupational status index

The occupational status index used in (Clark, 2023a) was devised by the paper's author and others in a recent preprint (Clark et al., 2022). The data underlying the index are 1.6 million marriage records in England across years 1837-1939, which include data on occupations for brides, grooms, and both of their fathers. (Clark et al., 2022) condense the more than 100,000 occupation description strings in those data into 442 occupational categories [listed in the Appendix Table A.3 in (Clark et al., 2022). For comparison, a standard occupational status index for this period, the HISCAM-GB, uses 1,300 occupational categories (Clark et al., 2022; Lambert et al., 2013). After individuals were assigned an occupation, Goodman's RCII association model (Goodman, 1979) was used to generate a status index for each occupation. Though the specific details of model specification are not presented in (Clark et al., 2022), RCII models generally proceed as follows. The core idea in such analyses is that social stratification

will be reflected in patterns of occupational interactions. The analysis takes as input a contingency table, where rows represent one individual's occupation, and columns represent the paired individual's occupation, with cell counts corresponding to the number of cases where this combination of occupations is observed. The RCII approach then fits a log-multiplicative model to these categorical data, of the general form

$$log(F_{ij}) = \mu + \mu_i^R + \mu_j^C + \beta\phi_i\varphi_j,$$

where F_{ij} is the expected cell frequency, μ is the main effect, μ^R is the row effect, μ^C is the column effect, β is the association parameter measuring the association between row and column variables, and ϕ_i and φ_j are the unknown row and column scores (indices) to be estimated (Lewis-Beck et al., 2011). When used to estimate status indices, row and column scores of a given occupation are constrained to be equal. An algorithmic procedure that iteratively assigns scores to occupational categories is used to estimate the occupation index scores that maximize the fit between the observed counts (f_{ij}) in each cell and the model-predicted counts (F_{ij}). Thus, this procedure maximizes the correlation in occupational status index for whatever pair of individuals is analyzed (bride-groom, father-son, etc.). In (Clark et al., 2022), index estimation was performed according to this approach separately for the father/son occupation associations and for the father-in-law/son-in-law associations; the average of these indices was used as the overall index of occupational status in (Clark, 2023a).

Supplementary Note 6: Statistical artifacts influencing familial correlation estimates

a. Within-lineage correlations are highly attenuated

The core results in (Clark, 2023a) are based on "lineage-agnostic" correlations in social status, which are calculated between individuals in all pairs of a given genealogical relationship, ignoring the surname lineage to which each pair belongs and ignoring heterogeneity among and within these lineages.

We find that within-surname distributions of correlations are highly variable across surnames. This is true for all 11 genealogical relationships and for all nine measures of status (**Figs. S2a, S3**). The central tendencies of these surname-specific correlations are substantially lower than the corresponding lineage-agnostic correlations reported in Table 2 of (Clark, 2023a). Beyond

the attenuation of correlations that is to be expected when conditioning on a given family, there is remarkable heterogeneity among lineages: the medians and modal values of the surname-specific correlations are near zero for all genealogical relationships beyond first cousins (**Figs. S2a, S3**), but some surnames show high correlations, even among distant cousins (**Figs. S2, S3, S4a**).

This finding may also be partly explained by the transmissibility of social status varying among families, reminiscent of the Scarr-Rowe effect, where the heritability of cognitive ability varies by socioeconomic status (Giangrande and Turkheimer, 2022; Rowe et al., 1999; Scarr-Salapatek, 1971; Tucker-Drob and Bates, 2016; Turkheimer et al., 2009; Turkheimer and Horn, 2014), and is consistent with arguments that when families vary in their access to, use of, and transmission of cultural tools, resources, and norms, the transmissibility of phenotypes impacted by these non-genetic sources of variation will also vary among families (Feldman et al., 2013; Kolodny et al., 2022).

An illustrative example is surname lineage number 1436 (**Fig. S4b**). Applying the regression model of (Clark, 2023a) to the occupational status correlations calculated within this surname lineage (among men born 1780-1859), we obtain an estimate of $b = 0.99$—there is no decay in correlation, even out to fourth cousin pairs. The reason for this result is strong structure in this surname lineage. For instance, the fourth cousin pairs come from four different sublineages (i.e., descended from a different great-great-great grandfather, as we have identified using the *pidf* variable in the data). The correlation within each sublineage is 0, but because sublineages differ in the average occupational status of their members, the sublineage-agnostic correlation among all fourth cousin pairs with this surname is 0.63 (**Fig. S4b**). Importantly, similar substructure is what drives the lineage-agnostic correlation when fourth cousin data from multiple surnames are combined (**Fig. S4c**).

For first-degree relatives, within-surname heterogeneity tends to be more pervasive in surnames with more observed relative pairs, resulting in surname-specific correlations that tend to be dramatically higher (Clark, 2023a) (**Fig. S2b-c**, **Fig. S5**). As genealogical distance increases, however, the relationship between the number of surname pairs and surname-specific correlation gradually dissipates (**Fig. S5**). At a more granular level, when we condition on sublineages, the correlations (and thus the covariances) are almost always near 0, across all genealogical relationships (**Fig. S6**).

This trend can also be illustrated through the dependency of within-lineage correlations on sample size among first-degree relatives: surnames with more represented members (≥30 father-son or full sibling pairs) are highly correlated (median $r = 0.46$ for full siblings and $r = 0.52$ for father-son pairs) and surnames with fewer represented members (<30 father-son or full sibling pairs) are less correlated (median $r = 0.01$ for full siblings and $r = 0.18$ for father-son pairs). These differences cannot be attributed to differences in estimation noise due to varying sample sizes.

b. Inference about decay in correlations is substantially affected by pseudo-replication artifacts

The dataset used in (Clark, 2023a) consists of all observed pairs of a given relationship, and each individual will typically be represented in multiple records. For example, one family of ten brothers (all sons of individual 80207) is represented by $\binom{10}{2} = 45$ sibling pairs in the data, all of which are treated as independent observations. This means that even though the pairwise records are all distinct from one another, when data from the same individual is repeated in multiple pairwise records, these records are considered "pseudo-replicates" of one another, because they are not mutually independent. Pseudo-replication can lead to a multitude of problematic statistical artifacts and inference errors, some of which are documented in (Lazic, 2010) (see (Rosenberg and Vanliere, 2009) for a related discussion of how genealogical pseudo-replication impacts inference in the context of genetic association studies). In (Clark, 2023a), the effects of pseudo-replication are increasingly pervasive for more distant relationships, where there are generally exponentially more pairwise relationships within a given lineage (**Fig. S7**).

Pseudo-replication among relative pairs can affect the lineage-agnostic correlation estimates in two ways: by altering the covariance between relatives and by altering the sample variance. For close relationships (out to first cousins), pseudo-replication causes both the covariance and variance of occupational status to increase, but for more distant relatives, the covariance and variance of occupational status decrease (**Fig. S8**). Though the sample variances change at most by ~20%, the changes in covariance can be dramatic: for third cousins once removed, the covariance decreases nearly 80% when pseudo-replicated records are included. To mitigate this effect of pseudo-replication on relative correlation estimates, we sampled a single pair of

relatives of each relationship type from each surname. While mitigating bias, this down-sampling reduces the sample size significantly. We therefore performed 1,000 random bootstrap samples and took the averages of the relative correlations as an estimate. Note that this approach also removes some higher-order pseudo-replication effects; for example, suppose we have a pair of brothers who are fourth cousins with another pair of brothers, producing four pairwise fourth cousin relationships. Even though we can downsample these into two pairs that contain mutually exclusive individuals (removing the first-order effects of pseudo-replication), each pair of brothers share the same father, which will presumably have a strong influence on their status, so there is some degree of second-order pseudo-replication caused by resemblance between siblings. This sampling strategy does not account for other forms of pseudo-replication (e.g., the same individual being represented in multiple pairs across different relationship types; or relatedness via a recent maternal common ancestor), and we are unable to resolve potential biases that might arise due to genealogical or social relationships across different surname lineages in these data. As shown in **Fig. 3**, in light of substructure in the data (**Supplementary Note 6a),** pseudo-replication drives a bias in lineage-agnostic occupational status correlations estimated by (Clark, 2023a). When applied to the revised correlation estimates, the log-linear regression model of (Clark, 2023a) explains only 29% of the variation in relative correlations.

c. Other sources of bias on trait correlations

The distributions of each trait differ across the subsets of different genealogical relationships, indicating these subsets do not represent the same cross-sections of the population (**Fig. S9**). For instance, among men in the dataset born 1780-1859, the mean occupational status among unique individuals represented in father-son pairs (37.5) is ~10% lower than among members of fourth cousin pairs (mean occupational status = 41.6; $p = 3.58 \times 10^{-23}$) (**Fig. S10**). As genealogical distance increases, the individuals comprising each subset also tend to be wealthier, born more recently, and have greater variance in occupational status (**Fig. S10**). One potential explanation for these differences in sample distributions is that social or demographic changes have altered the true population distributions of status measures over this time period. Specifically, temporal changes to the population variance of occupational status are important to consider: because the denominator of the familial correlations modeled by (Clark, 2023a) is assumed to be an estimator of the population variance [see (Gimelfarb, 1981)], and because more distant relatives in the data tend to have been born more recently, the changes in

correlation across genealogical distance may partly reflect changes in the population distribution over time.

Another partial driver of the differences in familial correlation may be sampling and ascertainment biases that render the data unrepresentative of the broader sample/population. The selection criteria for lineages to be included in the dataset, described in Appendix 01 of (Clark, 2023a), acknowledges that this dataset is largely a convenience sample, stating: "*lineages were chosen for inclusion based on their completeness, and either the public posting of the lineages or their creators' willingness to share the data*," but the potential biases that might be induced by these particular selection criteria are not addressed. In addition, we find that the average wealth and occupational status are significantly higher (and the year of birth significantly earlier) for fathers from large lineages compared to those of small lineages (**Table S3**), indicating that the lineage-agnostic correlations for first-degree relatives were biased upward due to the preferential inclusion of larger, more affluent surname lineages in the dataset.

Supplementary Note 7: Sensitivities of the persistence rate parameter

We used simulations to understand how estimates of the persistence rate parameter, b, vary across a wide range of possible patterns of familial correlations (from full sibling to fourth cousins). We did not include direct descendant relationships (parent-child, grandparent-grandchild) for simplicity. For each instance of the simulation, we generated correlations for each relative type as follows: for full siblings, a correlation was drawn from $U(0, 0.7)$. For 4th cousins, a correlation was drawn from $U(0, 0.1)$, with the constraint that its value was less than that of the full sibling correlation. We then set *b* as the exponential of the slope describing the relationship between the logarithms of those two correlations:

$$log(b) = \frac{log(\rho_{4th\ cousin}) - log(\rho_{full\ sibling})}{8}$$

We then set correlations for the remaining relationships (from siblings once removed to 3rd cousins once removed) as

$$log(\rho_n) = log(h^2) + n\, log(b) \qquad \text{[Equation 1 from (Clark, 2023a)]}$$

where n is the degree of relatedness (i.e., $n = 1$ for full siblings and $n = 9$ for fourth cousins; **Supplementary Note 1**). Each correlation was then adjusted by random deviate, drawn from a Normal distribution with zero mean and a standard deviation of 0.2. Values were exponentiated to return to the original scale and then bounded between 0 and 1 to ensure validity. Using the set of generated correlations, the same linear regression model described above was fitted, and b was estimated. The simulation results (**Fig. S12**) show that *b* is most sensitive to distant relative correlations and is nearly guaranteed to exceed 0.7 for cases having correlation between 4th cousins exceeding 0.04.

Supplementary Tables and Figures

Table S1. Parameter estimates reported in (Clark, 2023a) Figure 1 and Table S2 are inconsistent with one another. Attempts to reproduce these results are consistent with (Clark, 2023a) Figure 1, but not (Clark, 2023a) Table S2. In (Clark, 2023a) t^2 is referred to as h^2.

	Clark (2023) Figure 1		Clark (2023) Table S2		WLS from raw data		OLS from correlations in Clark (2023) Table 2	
	t^2	b	t^2	b	t^2	b	t^2	b
Modern Status	0.39	*0.81*	0.47	*0.77*	0.39	*0.80*	0.39	*0.81*
log(House Value)	0.36	*0.82*	0.41	*0.80*	0.37	*0.81*	0.36	*0.82*
Company Director	0.11	*0.82*	0.19	*0.78*	0.13	*0.86*	0.11	*0.82*
Index Mult. Deprivation	0.36	*0.72*	0.36	*0.73*	0.39	*0.70*	0.36	*0.72*
Literacy	0.41	*0.84*	0.43	*0.85*	0.47	*0.82*	0.47	*0.83*
Occ. Status (1780-1859)	0.94	*0.75*	0.72	*0.81*	0.93	*0.75*	0.94	*0.75*
Occ.Status (1860-1919)	0.66	*0.8*	0.65	*0.79*	0.66	*0.80*	0.66	*0.80*
Higher Ed (1780-1859)	0.57	*0.72*	0.63	*0.77*	0.36	*0.83*	0.58	*0.76*
Higher Ed (1860-1919)	0.58	*0.76*	0.42	*0.77*	0.46	*0.75*	0.58	*0.71*

721 **Table S2**. Correlation matrix of individual status measures, including correlations with an
722 individual's paternal wealth.

	Higher education	Occupational status	Index multiple deprivation	Log (house value)	Modern status	Literacy
Higher education						
Occupational status	0.59					
Index multiple deprivation	0.11	0.24				
Log (house value)	0.26	0.45	0.54			
Modern status	0.25	0.47	0.81	0.85		
Literacy	0.09	0.31	*N/A*	*N/A*	*N/A*	
Paternal wealth	**0.48**	**0.66**	**0.19**	**0.36**	**0.35**	**0.26**

723

Table S3. Testing for equal mean status measures between fathers from surname lineages with <30 father-son pairs versus fathers from lineages with ≥30 father-son pairs (considering only unique individuals).

Variable	Small sample group (<30 father-son pairs)		Large sample group (≥30 father-son pairs)		T-test p-value
	Mean	Standard deviation	Mean	Standard deviation	
Occupational status	35.4	12.1	37.9	16.0	1.66e-6
Wealth	-3.61	2.06	-2.50	2.65	5.34e-5
Birth year	1807	17.8	1803	20.4	3.88e-6

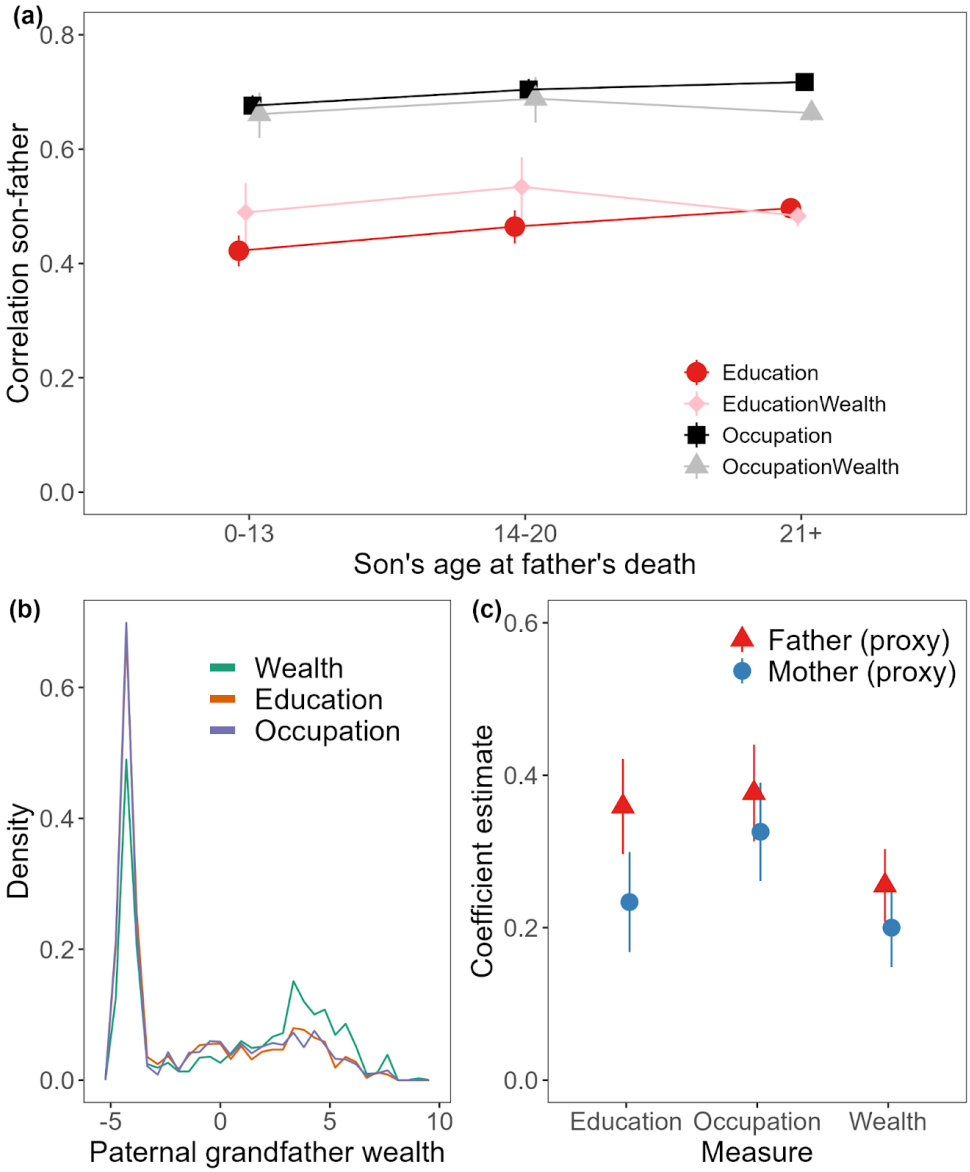


Figure S1. (a) Father-son correlations (95% CI) in education (red), occupation (black), son education with father wealth (pink), and son occupation with father wealth (grey), across different bins of the son's age at his father's death. Compare to Figure 4 in (Clark, 2023a). **(b,c)** maternal and paternal effects on offspring status measures and wealth are statistically indistinguishable when data used across these three regressions are identical (Supplementary Note 4). Compare to Figure 3 in (Clark, 2023a). (b) shows distributions (30 bins) of log(paternal grandfather wealth) for individuals comprising each data set in the original analysis of (Clark, 2023a); models of wealth were fit to a data set with more high wealth individuals. In (c), coefficients (95% CI) are shown for linear regressions of 'grandchild measure ~ maternal grandfather measure + paternal grandfather measure', using a data set of complete records. Here and in the analyses of (Clark, 2023a), maternal and paternal grandfather measures are used as proxies for maternal and paternal measures, due to no information on mothers directly.

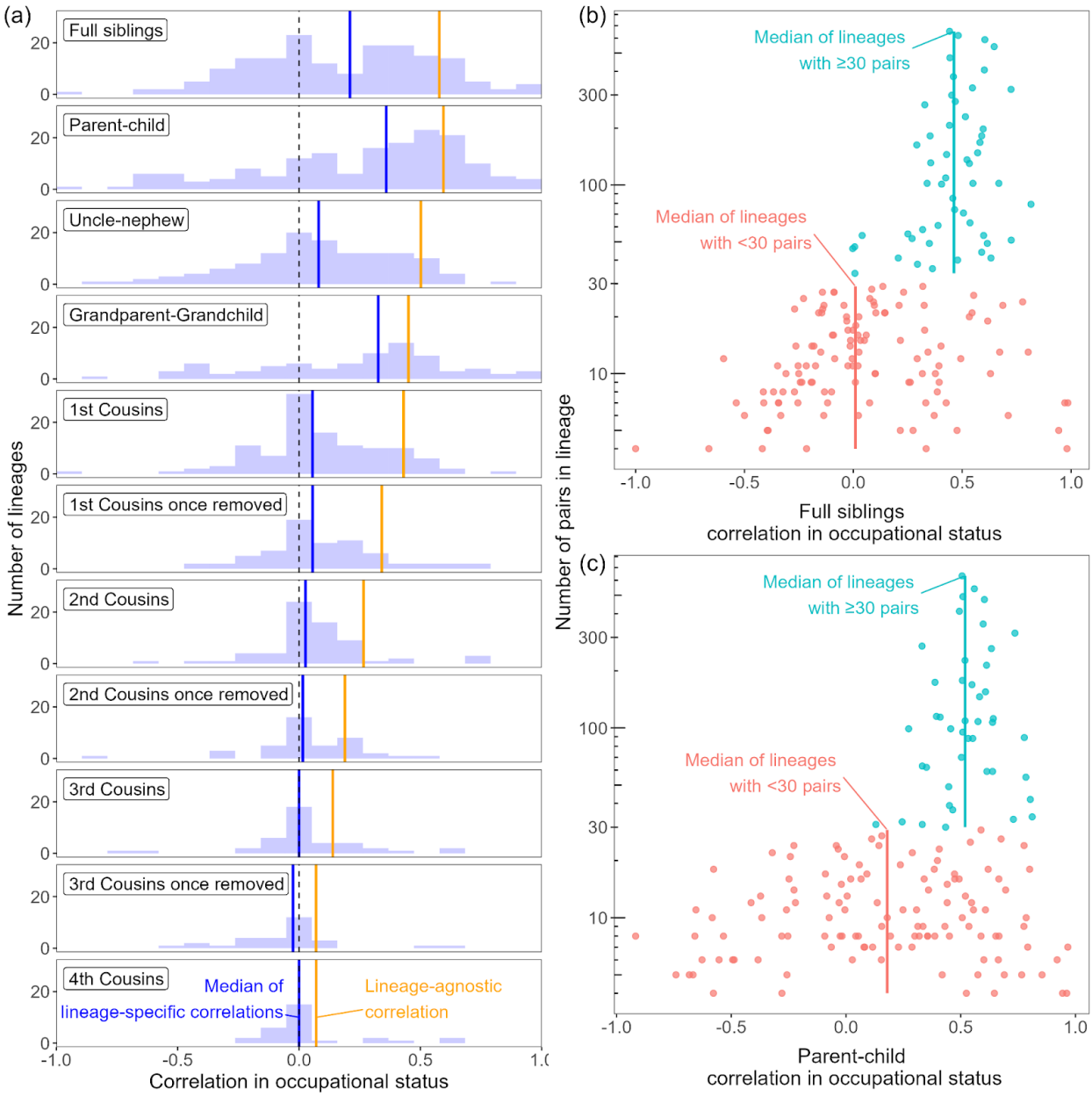


Figure S2. (a) Comparing the "lineage-agnostic" correlations in occupational status (1780-1859) that (Clark, 2023a) relies on (orange lines) with the distributions of lineage-specific correlations and their medians (blue lines). **(b,c)** Lineage-specific correlations in occupational status for first-degree relatives are associated with the representation of the lineage in (Clark, 2023a). Lineages with ≥30 full sibling pairs **(b)** or father-son pairs **(c)** have a median trait correlation of 0.5, only slightly below the overall lineage-agnostic correlations shown in **(a)**. Lineages with <30 pairs in the data show much lower median correlations (~0.01 for full siblings, ~0.18 for father-son pairs).

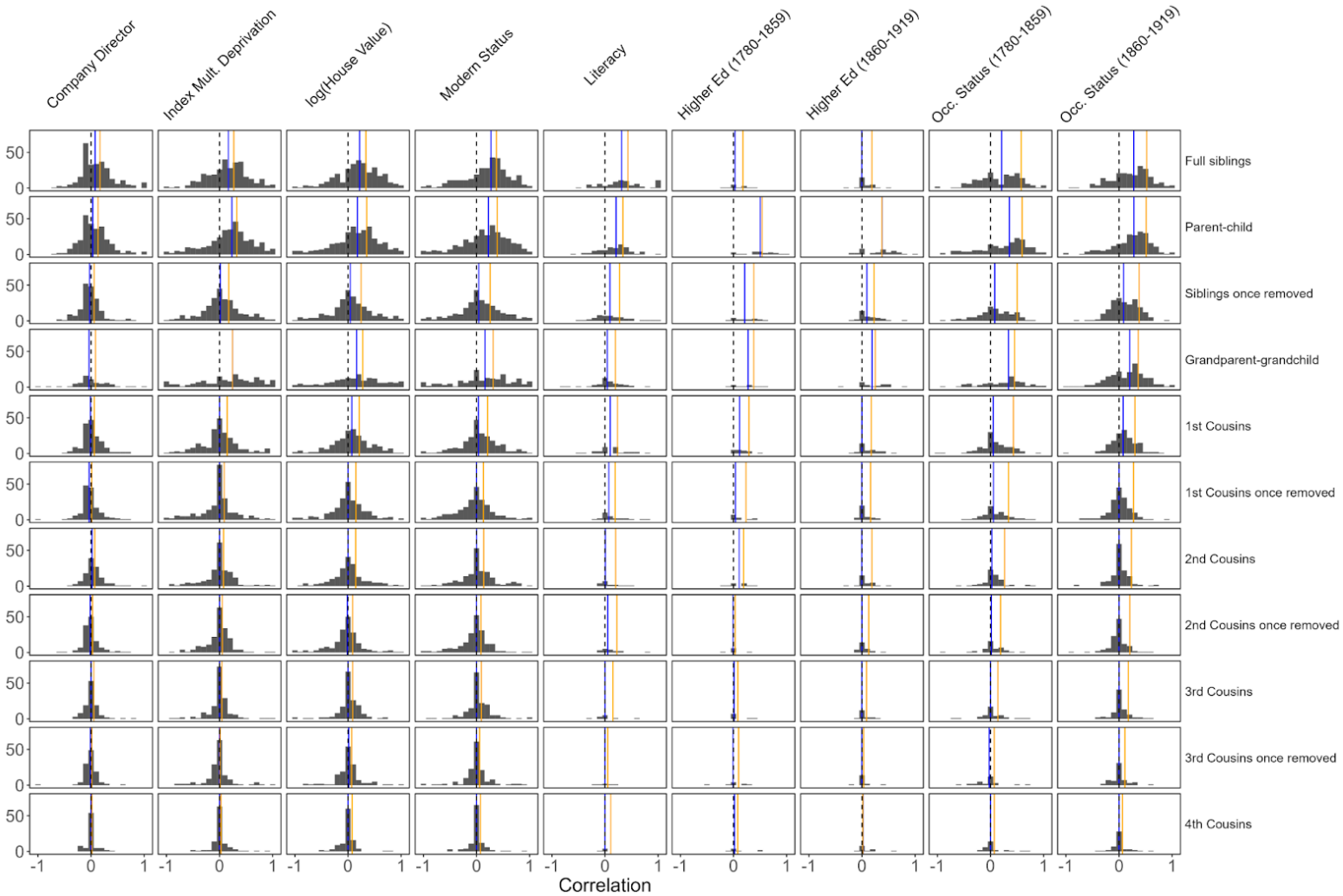


Figure S3. Comparing the "lineage-agnostic" correlations presented in (Clark, 2023a) (orange lines) with the distributions of lineage-specific correlations and their medians (blue lines), for measures of status analyzed in (Clark, 2023a). Dashed black line marks zero. Many lineage-specific correlations for literacy and higher education could not be calculated (and are thus not plotted) due to there being no variation within the lineage. This is because most individuals did not go to university, and most individuals were literate.

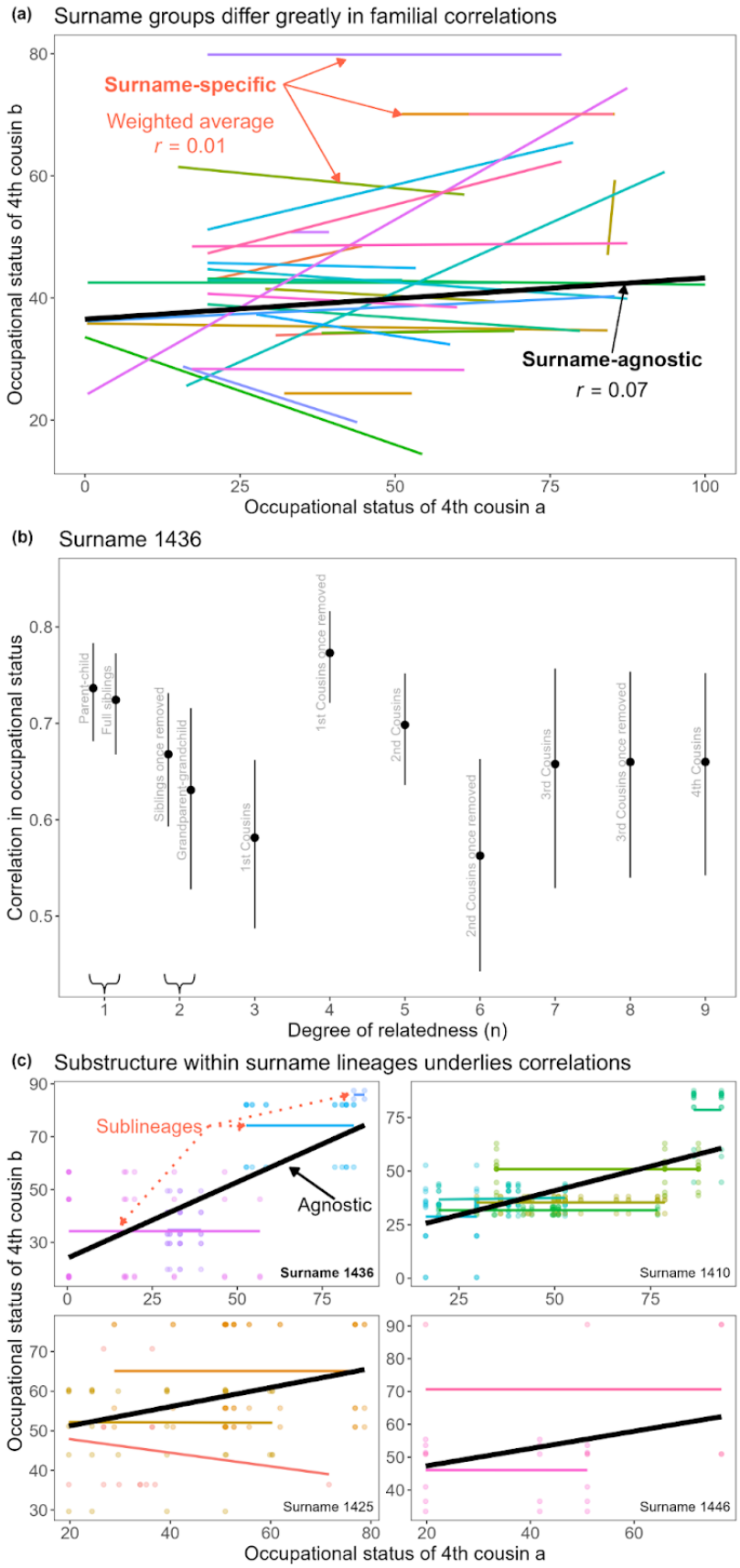


Figure S4. (a) When conditioning on surname lineage, the association in occupational status between fourth cousins is highly variable. The weighted average of these correlations (where weights are determined by the number of pairs per lineage) is lower than the surname-agnostic correlation. **(b)** In one surname lineage (surname ID 1436), the surname-specific correlations are >0.5 for all relationship types, and there is no decay with genealogical distance, indicating that the model parameters estimated using lineage-agnostic correlation estimates cannot be considered representative of a given surname lineage. **(c)** Strong positive correlations within a surname lineage are attributable to substructure among those lineage members. Here we show fourth cousin data for four surname lineages. For each of these lineages, there are two or more sublineages (identified by their last common male ancestor) that differ in their average occupational status, but there is no within-sublineage correlation. All data shown in this figure are for individuals born 1780-1859.

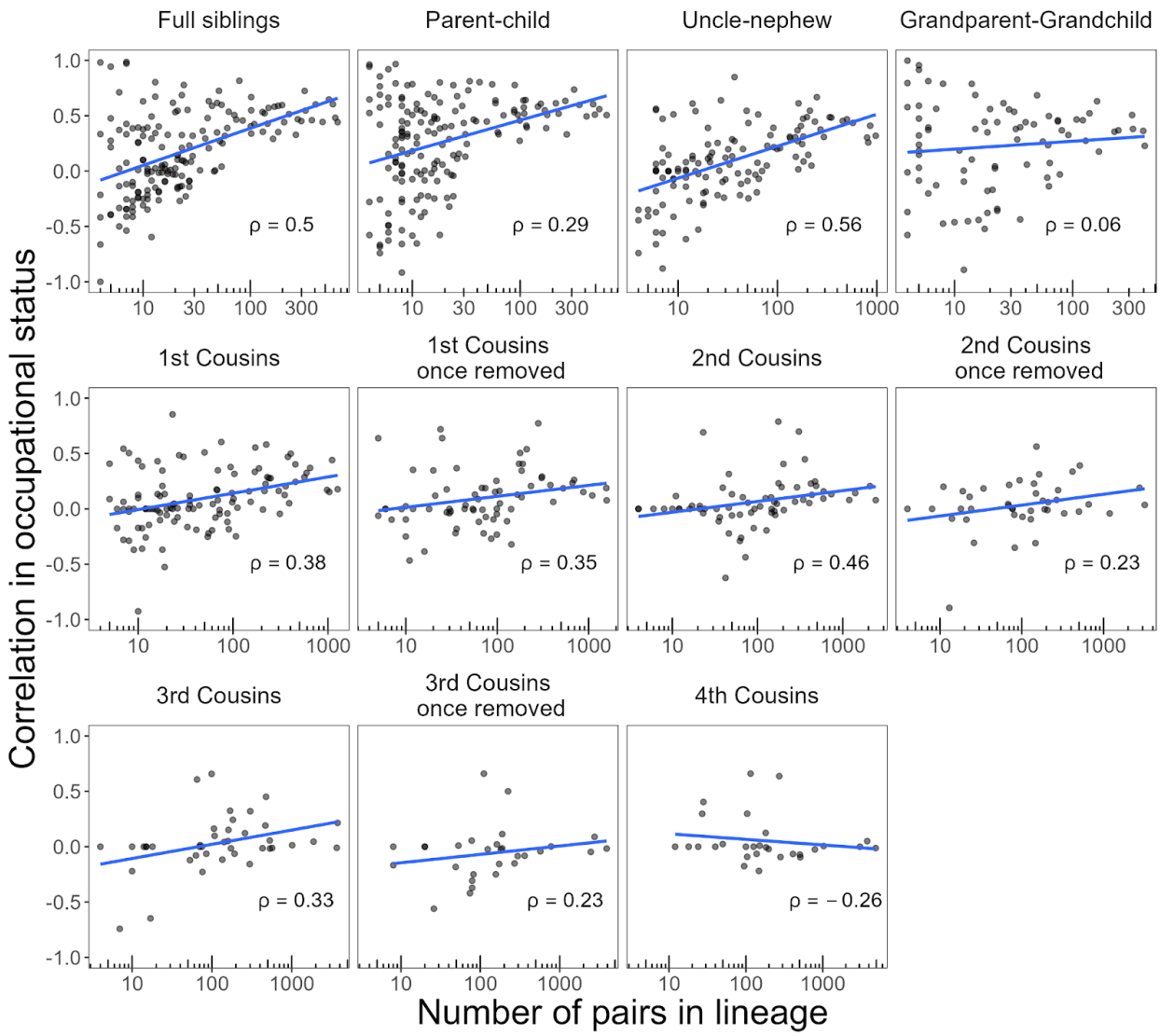


Figure S5. Surname lineages with more relative pairs tend to have higher correlations in occupational status, with the exception of 4th cousins. The Spearman rank correlation (ρ) between these two variables is given in the bottom right corner of each panel. Blue lines show a linear fit to the data.

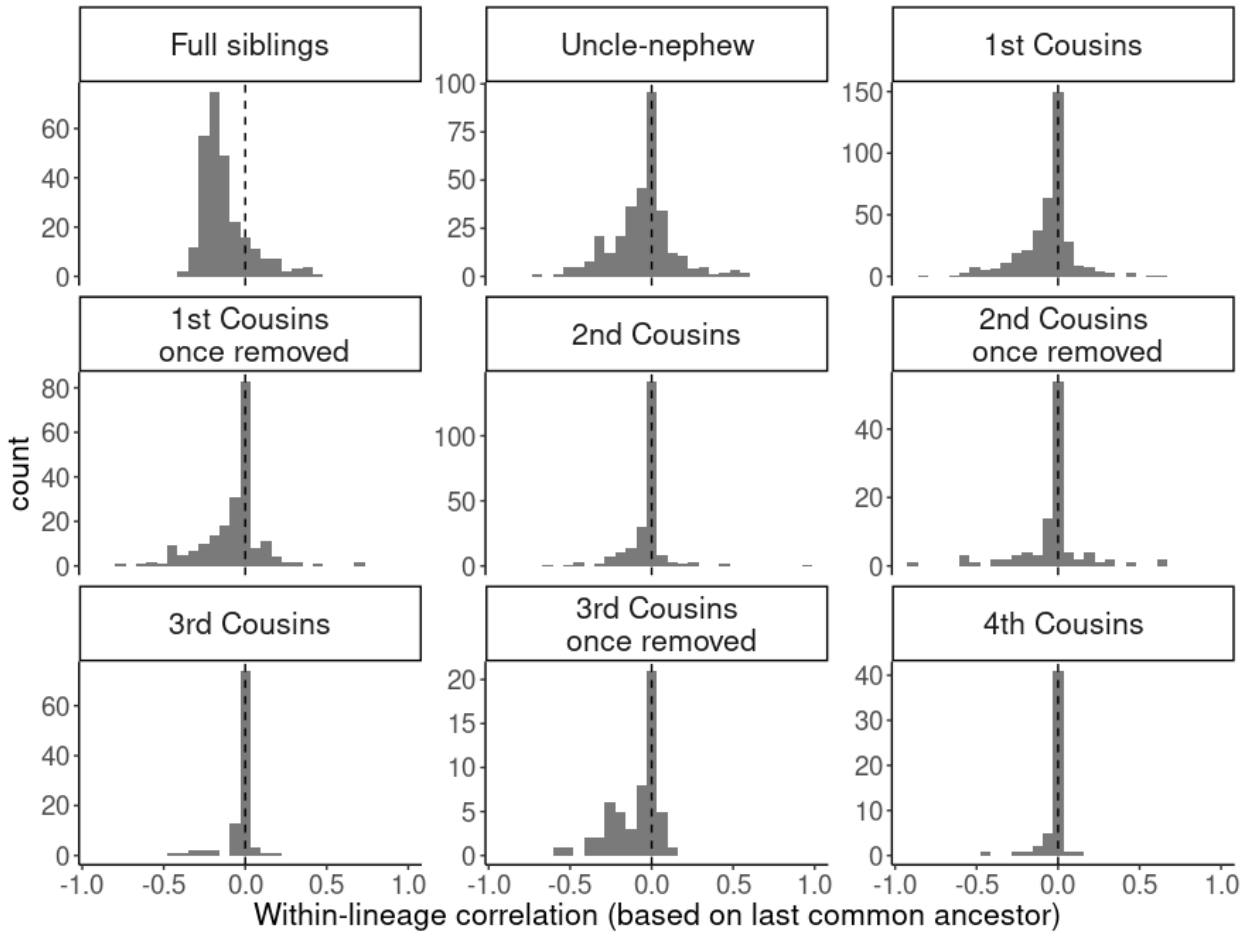


Figure S6. When conditioning on a single sublineage (defined by the last common ancestor of a set of relatives), the within-sublineage correlation in occupational status is typically near 0, for all genealogical relationships. Data are shown for sublineages in each relationship subset with at least 10 relative pairs. Note that vertical relationships are excluded because each relative pair includes an observation for the same father or grandfather, so there is no variance and a correlation cannot be calculated.

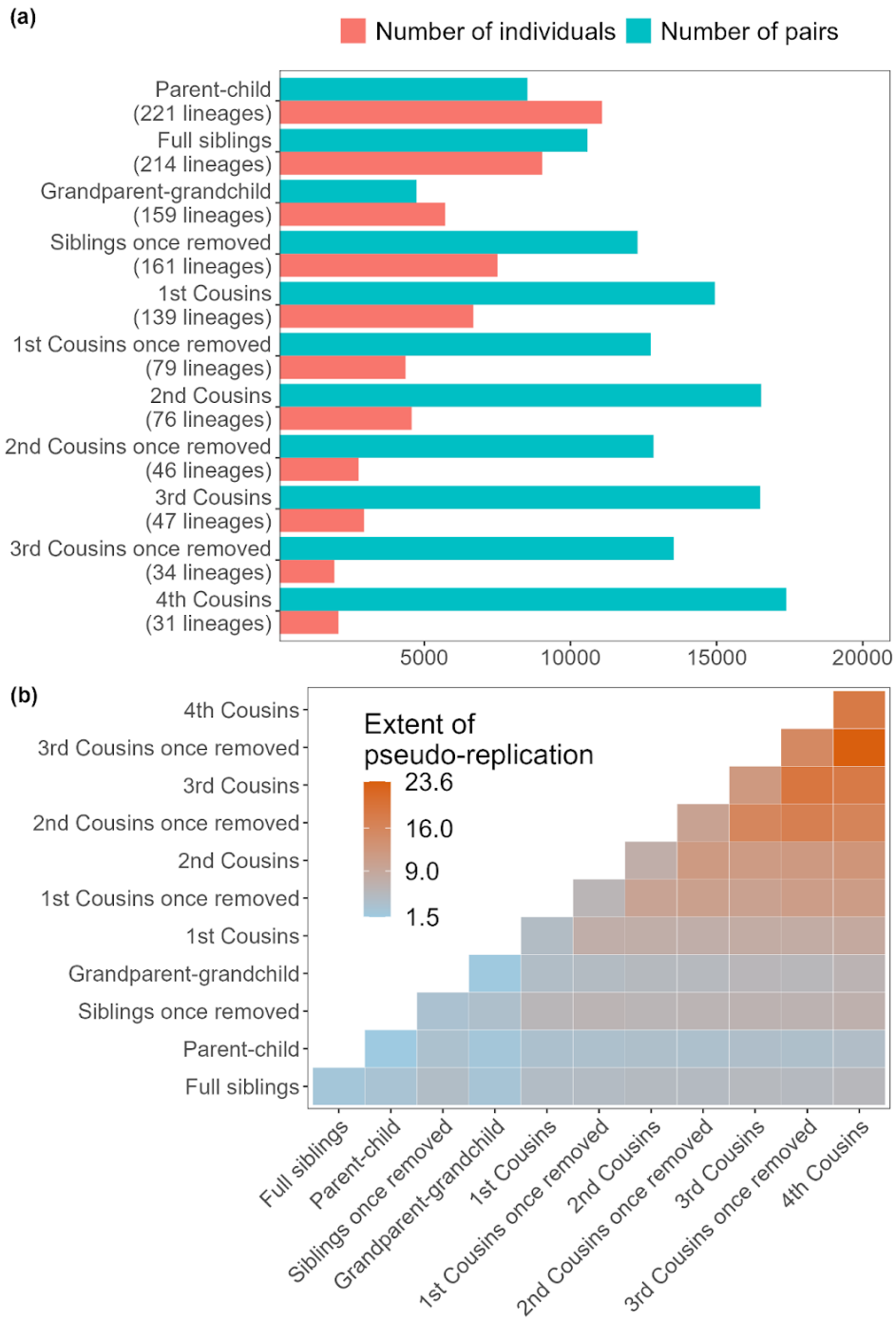


Figure S7. Pseudo-replication of individuals is widespread in (Clark, 2023a), illustrated here for the measure of occupational status (1780-1859). **(a)** the number of relative pairs increases with genealogical distance, even though the number of unique individuals and the number of surname lineages decreases, indicating increasing pseudo-replication among distant relatives. **(b)** We measure the "extent of pseudo-replication" as the average number of times an individual is included as a member of a pair in either the row category or the column category, across all individuals included in either of the two categories; this statistic ranges from 1.5 to 23.6 for these data. In the absence of pseudo-replication, this number would be 1.

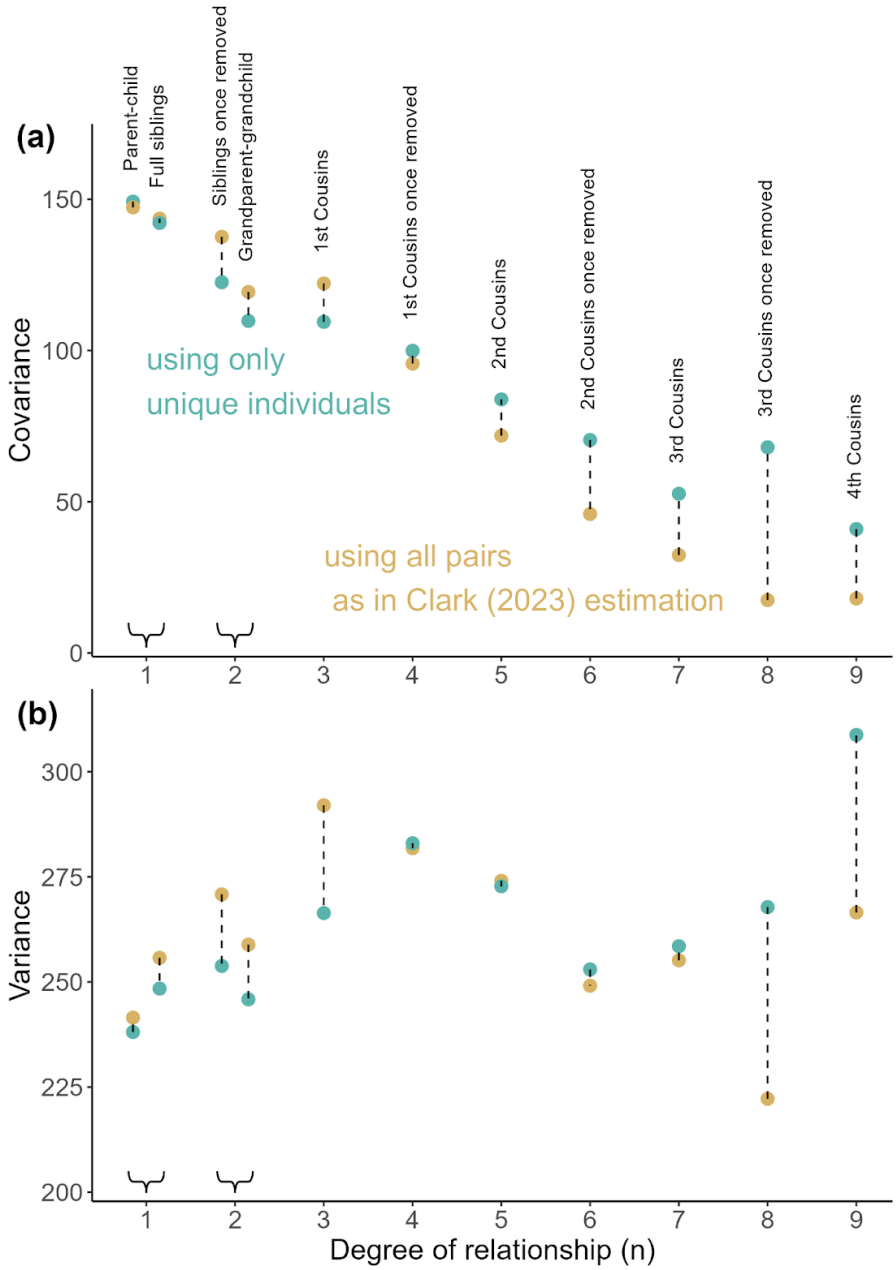


Figure S8. Pseudo-replication among relative pairs alters both the covariance **(a)** and variance **(b)** of the data, which will lead to biased correlation estimates. Note that pseudo-replication generally increased the covariance and variance for closer relatives, but decreased covariance and variance for more distant relatives.

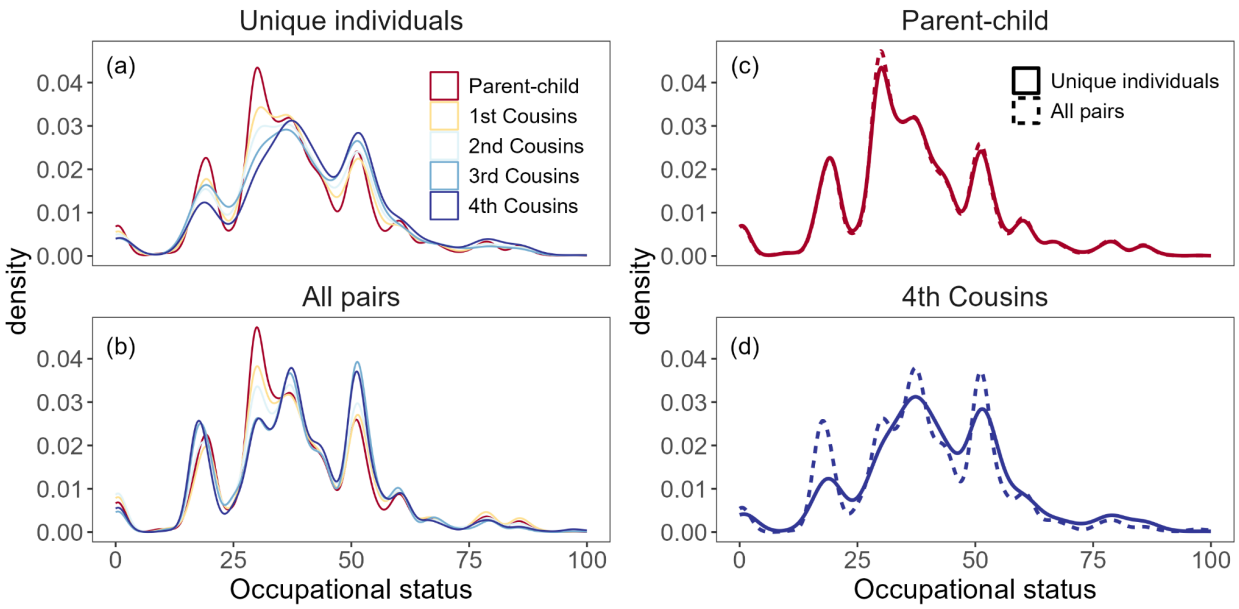


Figure S9. Empirical distributions of occupational status in each relationship subset (data shown for men born 1780-1859), considering **(a)** only unique individuals or **(b)** data from individuals included in all relative pairs considered, including replicated records (as used in (Clark, 2023a)). The distributions exhibit differences in abundance around different modal values (e.g. distant cousins tend to consist of relatively more individuals with occupational status ~50), and the overall mean and variance of these distributions tend to increase with genealogical distance, suggesting the different relative subsets are susceptible to sampling bias and/or come from different temporal populations (**Supplementary Note 6c**; **Fig. S10**). Direct comparisons of the distributions for unique individuals versus replicated records in father-son pairs **(c)** and fourth cousin pairs **(d)** show that pseudo-replication among distant relatives amplifies these distributional differences and induces bias in the final estimates of familial correlations.

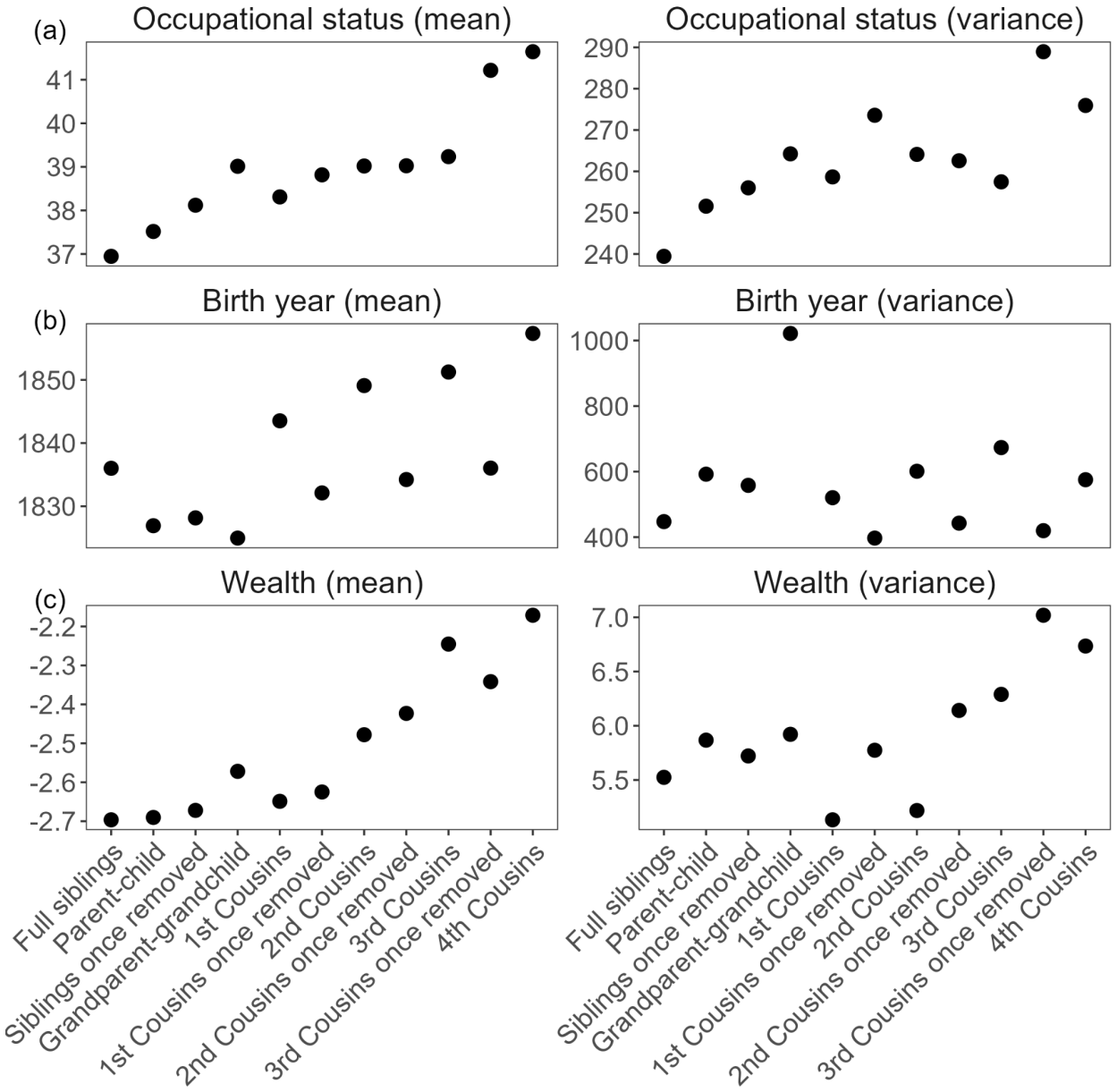


Figure S10. (a) The mean and variance of occupational status (considering only values from unique individuals) increases with genealogical distance in the dataset of (Clark, 2023a). Two potential explanations for this systematic variation are: **(b)** more distant relatives tend to have been born more recently so the changes in sample distribution of occupational status may reflect unaccounted changes in the population distribution, or **(c)** biases in sampling or completeness/availability of genealogical data mean that more distant relatives are increasingly represented by more affluent lineages.

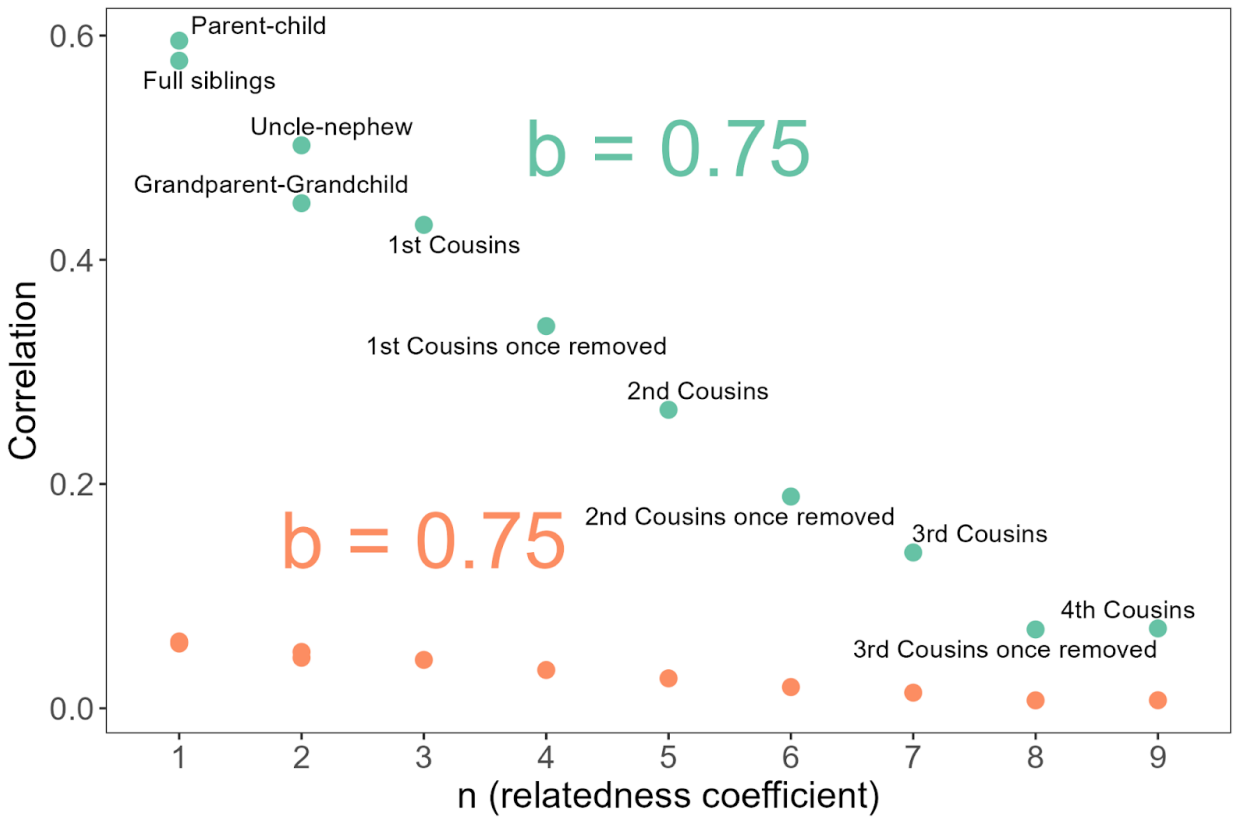


Figure S11. b is insensitive to the magnitude of familial correlations and describes only the *rate* at which they change as relatedness decreases. Thus, familial correlations in a status measure could increase or decrease by any degree with b remaining unchanged (here, orange points are 10% of their corresponding points in green). For both, $R^2 = 0.96$.

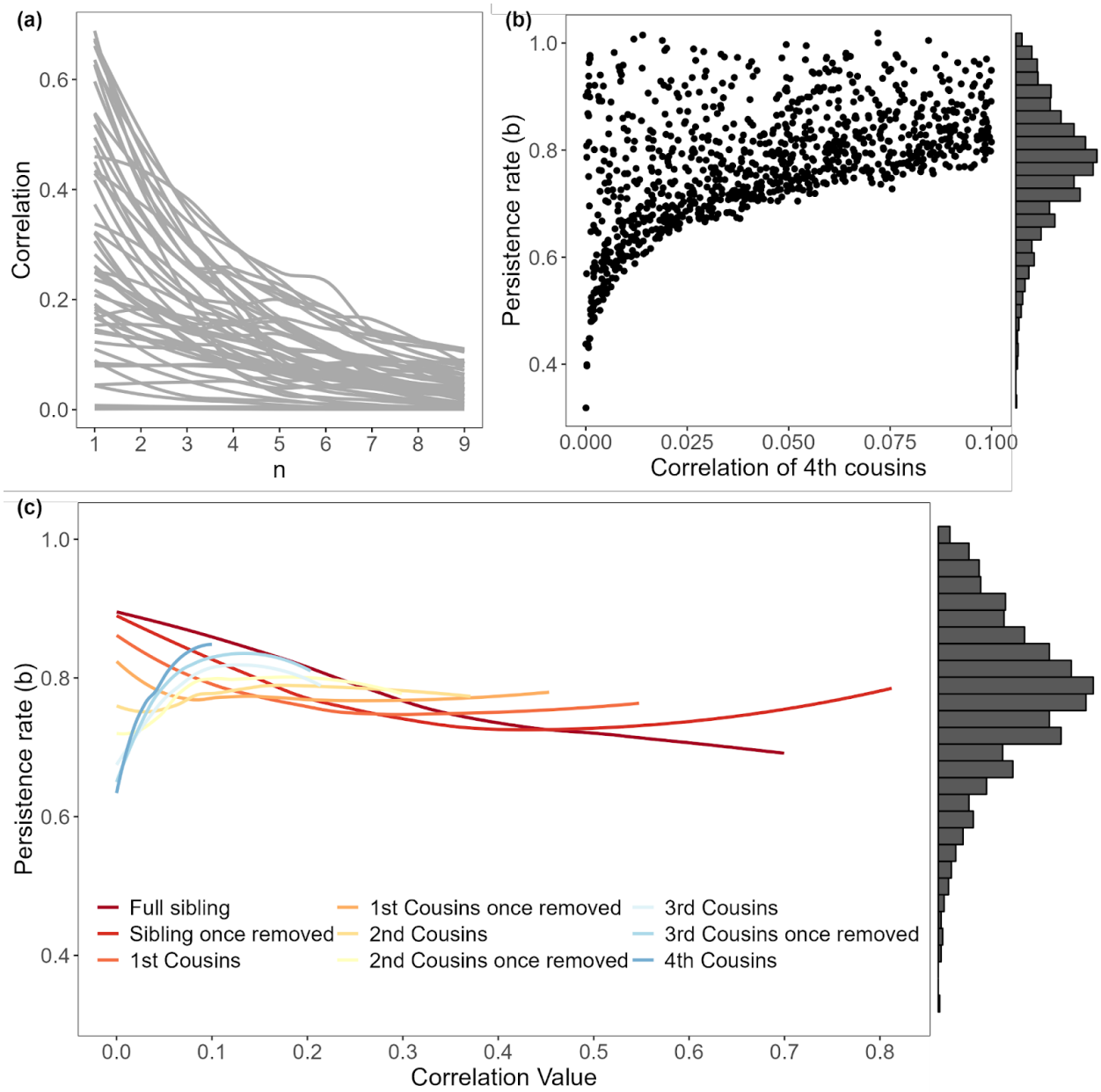


Figure S12. Estimation of the persistence parameter (*b*) in simulations shows increased sensitivity to distant relative correlations. See Supplementary Note 7 for simulation details. **(a)** smoothed patterns of simulated familial correlations (only 50 out of 1000 simulations shown for clarity). **(b)** As the correlation of fourth cousins increases from zero, the minimum possible estimate of *b* increases sharply. **(c)** The sensitivity of the estimated *b* to different familial correlations. LOESS smoothed patterns of *estimated b ~ correlation value*, across all 1000 simulations, are plotted for each relationship. Small changes in the value of distant relative correlations result in relatively large changes in estimated *b*.

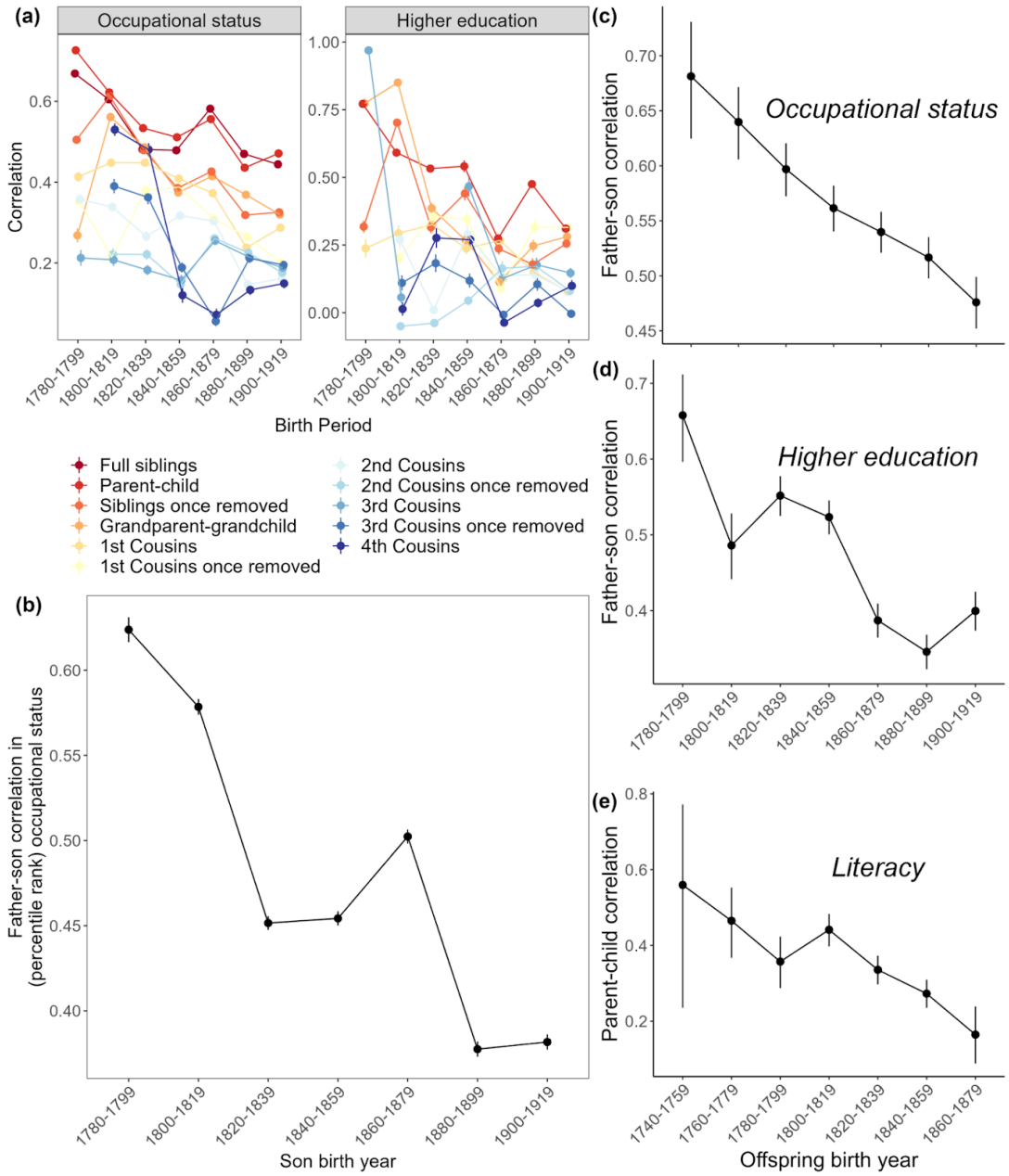


Figure S13. Complementary analyses of temporal trends in familial correlations. **(a)** Correlations (95% CI) in occupational status and higher education for all relationship types. Data were binned according to the birth year of the 2nd relative in each kin pair (for relationships that cross generations, this will be the younger generation). Higher education data for full siblings is not plotted, as correlations calculated from the raw data do not match those published in (Clark, 2023a); we were not able to obtain corrected data from the author at time of this publication. **(b)** Change in father-son occupational status correlations across 1780-1919, using percentile ranks of each individual within each time period. Compare to **Fig. 4a**. As in the main analysis, to mitigate pseudo-replication, we calculated correlations in (a) and (b) using one pair from each surname. Shown are average correlations across 500 bootstrap iterations of correlation estimation. **(c-e)** Parent-offspring correlations (95% CI) over time, calculated from raw data and not corrected for pseudo-replication. (c) Occupational status 1780-1919; (d) higher education 1780-1919; (e) literacy 1740-1879.

References

Akimova ET, Wolfram T, Ding X, Tropf FC, Mills MC. 2023. Polygenic predictions of occupational status GWAS elucidate genetic and environmental interplay for intergenerational status transmission, careers, and health. *bioRxiv*. doi:10.1101/2023.03.31.534944

Bailey RC. 1997. Hereditarian scientific fallacies. *Genetica* **99**:125–133.

Barton N, Hermisson J, Nordborg M. 2019. Why structure matters. *Elife*. doi:10.7554/eLife.45380

Berg JJ, Harpak A, Sinnott-Armstrong N, Joergensen AM, Mostafavi H, Field Y, Boyle EA, Zhang X, Racimo F, Pritchard JK, Coop G. 2019. Reduced signal for polygenic adaptation of height in UK Biobank. *Elife* **8**. doi:10.7554/eLife.39725

Causa O, Johansson Å. 2009. Intergenerational Social Mobility, OECD Economics Department Working Papers. Organisation for Economic Co-Operation and Development (OECD). doi:10.1787/223106258208

Cavalli-Sforza LL, Feldman MW. 1978. The Evolution of Continuous Variation. III. Joint Transmission of Genotype, Phenotype and Environment. *Genetics* **90**:391–425.

Cavalli-Sforza LL, Feldman MW. 1973a. Models for cultural inheritance I. Group mean and within group variation. *Theor Popul Biol* **4**:42–55.

Cavalli-Sforza LL, Feldman MW. 1973b. Cultural versus biological inheritance: phenotypic transmission from parents to children. (A theory of the effect of parental phenotypes on children's phenotypes). *Am J Hum Genet* **25**:618–637.

Chetty R, Hendren N, Kline P, Saez E, Turner N. 2014. Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility. *Am Econ Rev* **104**:141–147.

Clark G. 2023a. The inheritance of social status: England, 1600 to 2022. *Proc Natl Acad Sci U S A* **120**:e2300926120.

Clark G. 2023b. As a hereditarian, I strongly support economic redistribution. *Quillette*. https://quillette.com/2023/08/21/hereditarianism-and-economic-redistribution/

Clark G. 2014. The Son Also Rises. Princeton University Press.

Clark G, Cummins N, Curtis M. 2022. The Mismeasure of Man: Why Intergenerational Occupational Mobility is Much Lower than Conventionally Measured, England, 1800-2021.

Cloninger CR, Rice J, Reich T. 1979a. Multifactorial inheritance with cultural transmission and assortative mating. II. a general model of combined polygenic and cultural inheritance. *Am J Hum Genet* **31**:176–198.

Cloninger CR, Rice J, Reich T. 1979b. Multifactorial inheritance with cultural transmission and assortative mating. III. Family structure and the analysis of separation experiments. *Am J Hum Genet* **31**:366–388.

Collado MD, Ortuño-Ortín I, Stuhler J. 2023. Estimating Intergenerational and Assortative Processes in Extended Family Data. *Rev Econ Stud* **90**:1195–1227.

Coop G, Przeworski M. 2022a. Lottery, luck, or legacy. A review of "The Genetic Lottery: Why DNA matters for social equality." *Evolution* **76**:846–853.

Coop G, Przeworski M. 2022b. Luck, lottery, or legacy? The problem of confounding. A reply to Harden. *Evolution* **76**:2464–2468.

Cosh C. 2023. Colby Cosh: Is socioeconomic status hereditary? *National Post*.

Domingue BW, Fletcher J, Conley D, Boardman JD. 2014. Genetic and educational assortative mating among US adults. *Proc Natl Acad Sci U S A* **111**:7996–8000.

Feldman M. 2014. Echoes of the past: hereditarianism and A Troublesome Inheritance. *PLoS Genet* **10**:e1004817.

Feldman M, Cavalli-Sforza LL. 1981. Cultural Transmission and Evolution: A Quantitative Approach. Princeton University Press.

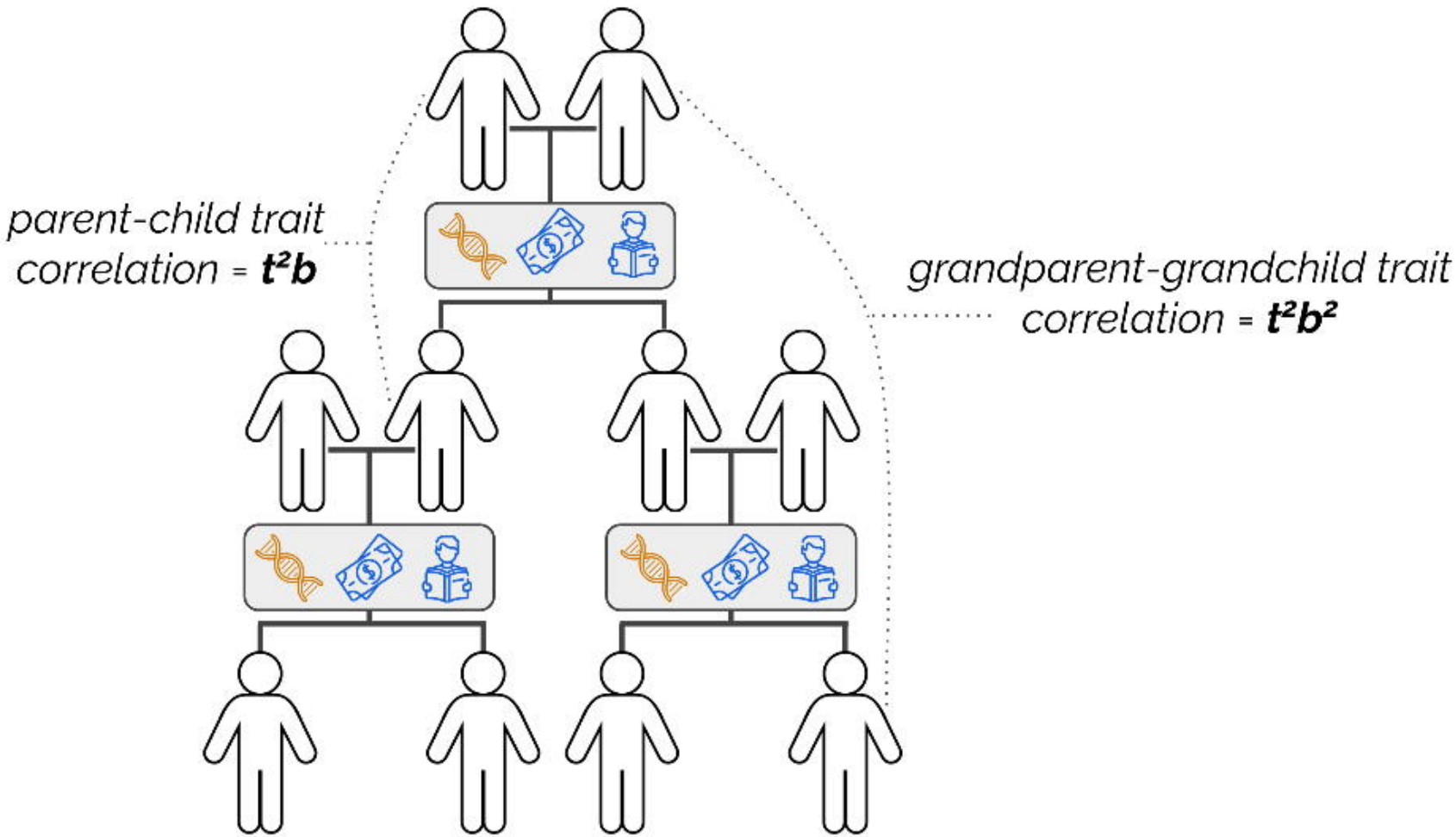
Feldman MW, Christiansen FB, Otto SP. 2013. Gene-culture co-evolution: teaching, learning,

and correlations between relatives. *Isr J Ecol Evol* **59**:72–91.
Feldman MW, Lewontin RC. 1975. The heritability hang-up. *Science* **190**:1163–1168.
Feldman MW, Ramachandran S. 2018. Missing compared to what? Revisiting heritability, genes and culture. *Philos Trans R Soc Lond B Biol Sci* **373**. doi:10.1098/rstb.2017.0064
Fisher RA. 1918. The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Trans R Soc Edinb* **52**:399–433.
Fletcher J. 2022. Backdoor to a dead end: A review essay. *Popul Dev Rev* **48**:253–258.
Giangrande EJ, Turkheimer E. 2022. Race, Ethnicity, and the Scarr-Rowe Hypothesis: A Cautionary Example of Fringe Science Entering the Mainstream. *Perspect Psychol Sci* **17**:696–710.
Gimelfarb A. 1981. A general linear model for the genotypic covariance between relatives under assortative mating. *J Math Biol* **13**:209–226.
Goodman LA. 1979. Multiplicative Models for the Analysis of Occupational Mobility Tables and Other Kinds of Cross-Classification Tables. *Am J Sociol* **84**:804–819.
Harden KP. 2021. The Genetic Lottery: Why DNA Matters for Social Equality. Princeton University Press.
Herzig AF, Noûs C, Pierre AS, Perdry H. 2023. A model for co-occurrent assortative mating and vertical cultural transmission and its impact on measures of genetic associations. *bioRxiv*. doi:10.1101/2023.04.08.536101
Holtzman NA. 2002. Genetics and social class. *J Epidemiol Community Health* **56**:529–535.
Hurlbert SH. 1984. Pseudoreplication and the Design of Ecological Field Experiments. *Ecol Monogr* **54**:187–211.
Kolodny O, Feldman MW, Lotem A, Ram Y. 2022. Differential application of cultural practices at the family and individual levels may alter heritability estimates. *Behav Brain Sci* **45**:e167.
Lambert PS, Zijdeman RL, Van Leeuwen MHD, Maas I, Prandy K. 2013. The Construction of HISCAM: A Stratification Scale Based on Social Interactions for Historical Comparative Research. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* **46**:77–89.
Lazic SE. 2010. The problem of pseudoreplication in neuroscientific studies: is it affecting your analysis? *BMC Neurosci* **11**:5.
Lee JJ. 2023. The heritability and persistence of social class in England. *Proc Natl Acad Sci U S A*.
Lewis-Beck MS, Bryman A, Liao TF. 2011. The SAGE Encyclopedia of Social Science Research Methods. Sage Publications, Inc.
Lewontin RC. 1974. Annotation: the analysis of variance and the analysis of causes. *Am J Hum Genet* **26**:400–411.
Lewontin RC, Rose SPR, Kamin LJ. 1984. Not in Our Genes: Biology, Ideology, and Human Nature. New York: Pantheon Books.
Loh P-R, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, Chasman DI, Ridker PM, Neale BM, Berger B, Patterson N, Price AL. 2015. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* **47**:284–290.
Longley PA, van Dijk J, Lan T. 2021. The geography of intergenerational social mobility in Britain. *Nat Commun* **12**:6050.
Marks GN. 2023. Has Cognitive Ability Become More Important for Education and the Labor Market? A Comparison of the Project Talent and 1979 National Longitudinal Survey of Youth Cohorts. *J Intell* **11**. doi:10.3390/jintelligence11080169
Mehler B. 2015. Hereditarianism. *The Wiley Blackwell Encyclopedia of Race, Ethnicity, and Nationalism*. doi:10.1002/9781118663202.wberen430
Mostafavi H, Harpak A, Agarwal I, Conley D, Pritchard JK, Przeworski M. 2020. Variable prediction accuracy of polygenic scores within an ancestry group. *Elife* **9**. doi:10.7554/eLife.48376

Odling-Smee FJ. 1988. Niche-constructing phenotypes In: Plotkin HC, editor. The Role of Behavior in Evolution , (pp. Cambridge, MA, US: The MIT Press, viii. pp. 73–132.
Okbay A, Wu Y, Wang N, Jayashankar H, Bennett M, Nehzati SM, Sidorenko J, Kweon H, Goldman G, Gjorgjieva T, Jiang Y, Hicks B, Tian C, Hinds DA, Ahlskog R, Magnusson PKE, Oskarsson S, Hayward C, Campbell A, Porteous DJ, Freese J, Herd P, 23andMe Research Team, Social Science Genetic Association Consortium, Watson C, Jala J, Conley D, Koellinger PD, Johannesson M, Laibson D, Meyer MN, Lee JJ, Kong A, Yengo L, Cesarini D, Turley P, Visscher PM, Beauchamp JP, Benjamin DJ, Young AI. 2022. Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nat Genet* **54**:437–449.
Panofsky A. 2021. Biology meets public policyThe Genetic Lottery: Why DNA Matters for Social Equality Kathryn Paige Harden Princeton University Press, 2021. 312 pp. *Science* **373**:1449.
Rao DC, Morton NE, Yee S. 1976. Resolution of cultural and biological inheritance by path analysis. *Am J Hum Genet* **28**:228–242.
Rao DC, Morton NE, Yee S. 1974. Analysis of family resemblance. II. A linear model for familial correlation. *Am J Hum Genet* **26**:331–359.
Rice J, Cloninger CR, Reich T. 1980. Analysis of behavioral traits in the presence of cultural transmission and assortative mating: Applications to IQ and SES. *Behav Genet* **10**:73–92.
Rice J, Cloninger CR, Reich T. 1978. Multifactorial inheritance with cultural transmission and assortative mating. I. Description and basic properties of the unitary models. *Am J Hum Genet* **30**:618–643.
Robinson MR, Kleinman A, Graff M, Vinkhuyzen AAE, Couper D, Miller MB, Peyrot WJ, Abdellaoui A, Zietsch BP, Nolte IM, van Vliet-Ostaptchouk JV, Snieder H, Medland SE, Martin NG, Magnusson PKE, Iacono WG, McGue M, North KE, Yang J, Visscher PM. 2017. Genetic evidence of assortative mating in humans. *Nature Human Behaviour* **1**:1–13.
Rosenberg NA, Vanliere JM. 2009. Replication of genetic associations as pseudoreplication due to shared genealogy. *Genet Epidemiol* **33**:479–487.
Rowe DC, Jacobson KC, Van den Oord EJCG. 1999. Genetic and environmental influences on vocabulary IQ: Parental education level as moderator. *Child Dev* **70**:1151–1162.
Scarr-Salapatek S. 1971. Race, Social Class, and IQ. *Science* **174**:1285–1295.
Shen H, Feldman MW. 2023. Drowning in shallow causality. *Behav Brain Sci* **46**:e199.
Sohail M, Maier RM, Ganna A, Bloemendal A, Martin AR, Turchin MC, Chiang CWK, Hirschhorn J, Daly MJ, Patterson N, Neale B, Mathieson I, Reich D, Sunyaev SR. 2019. Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *Elife* **8**:e39702.
Solon G. 2014. Theoretical models of inequality transmission across multiple generations. *Res Soc Stratif Mobil* **35**:13–18.
Tucker-Drob EM, Bates TC. 2016. Large Cross-National Differences in Gene × Socioeconomic Status Interaction on Intelligence. *Psychol Sci* **27**:138–149.
Turkheimer E. 2000. Three Laws of Behavior Genetics and What They Mean. *Curr Dir Psychol Sci* **9**:160–164.
Turkheimer E, Harden KP, d'Onofrio B, Gottesman II. 2009. The Scarr--Rowe interaction between measured socioeconomic status and the heritability of cognitive ability. *Experience and development: A festschrift in honor of Sandra Wood Scarr* **2**:81–97.
Turkheimer E, Horn EE. 2014. Interactions Between Socioeconomic Status and Components of Variation in Cognitive Ability In: Finkel D, Reynolds CA, editors. Behavior Genetics of Cognition Across the Lifespan. New York, NY: Springer New York. pp. 41–68.
Uchiyama R, Spicer R, Muthukrishna M. 2022. Cultural evolution of genetic heritability. *Behav Brain Sci* **45**:e152.
Vilhjálmsson BJ, Nordborg M. 2013. The nature of confounding in genome-wide association

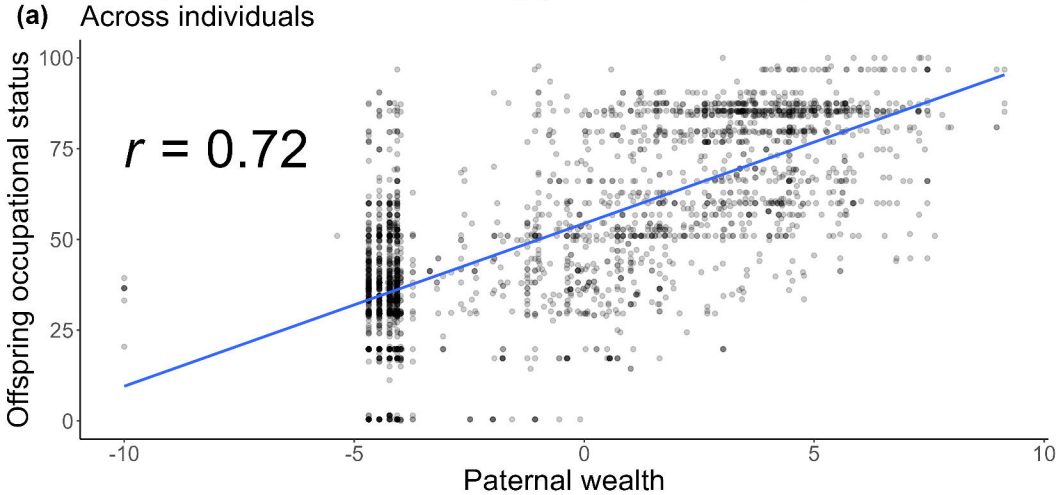
studies. *Nat Rev Genet* **14**:1–2.
Wright S. 1931. Statistical Methods in Biology. *J Am Stat Assoc* **26**:155–163.
Yengo L, Robinson MR, Keller MC, Kemper KE, Yang Y, Trzaskowski M, Gratten J, Turley P, Cesarini D, Benjamin DJ, Wray NR, Goddard ME, Yang J, Visscher PM. 2018. Imprint of assortative mating on the human genome. *Nat Hum Behav* **2**:948–954.
Young AI, Benonisdottir S, Przeworski M, Kong A. 2019. Deconstructing the sources of genotype-phenotype associations in humans. *Science* **365**:1396–1400.
Young AI, Frigge ML, Gudbjartsson DF, Thorleifsson G, Bjornsdottir G, Sulem P, Masson G, Thorsteinsdottir U, Stefansson K, Kong A. 2018. Relatedness disequilibrium regression estimates heritability without environmental bias. *Nat Genet* **50**:1304–1310.
Young AS. 2023. Estimation of indirect genetic effects and heritability under assortative mating. *bioRxiv*. doi:10.1101/2023.07.10.548458

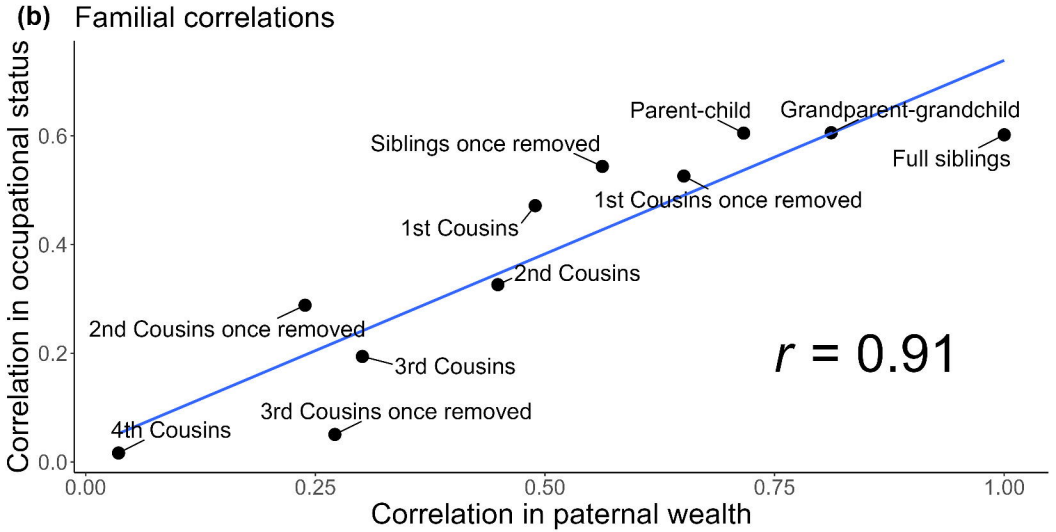
Genomes are transmitted along with non-genetic factors influencing social status. Resemblance between relatives will be a function of transmissibility (***t²***) and persistence rate (***b***)

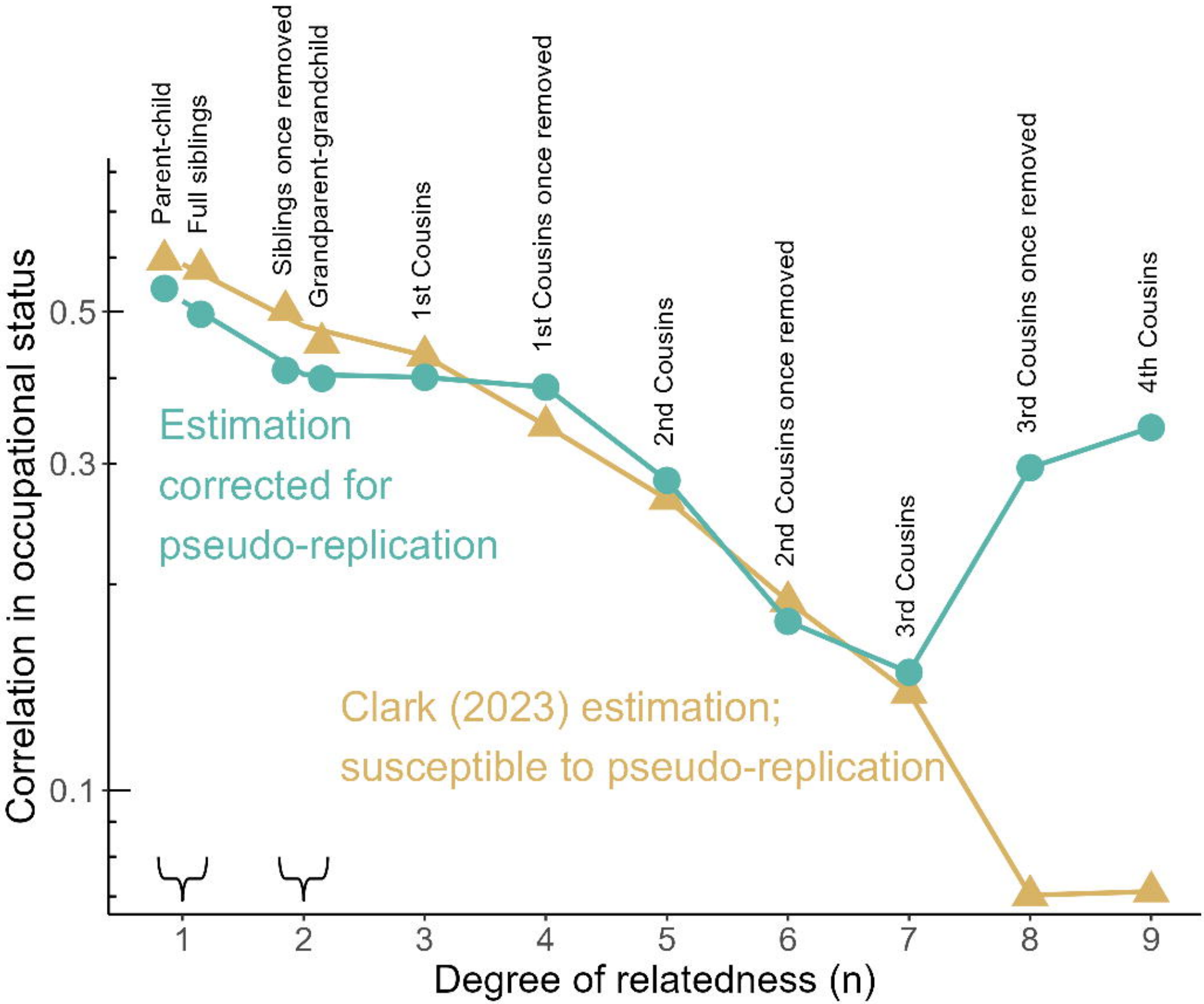


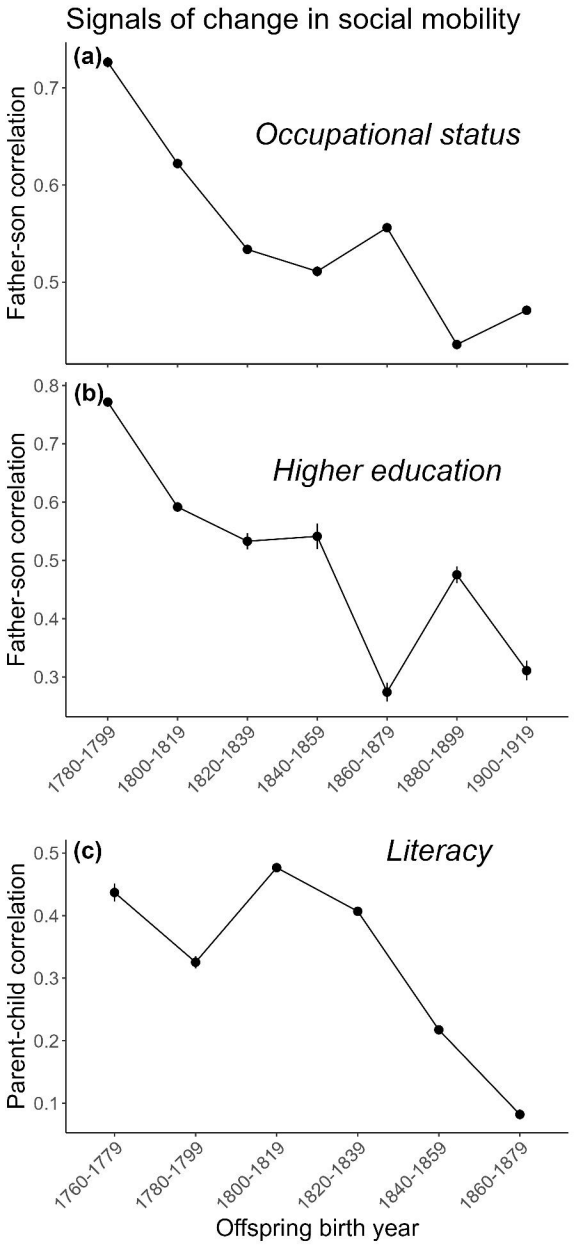
t^2 [what Clark (2023) calls h^2] and b both relate to compound, **entirely confounded**, genetic and non-genetic transmission

Occupational status is strongly correlated with paternal wealth






Parent-child
Full siblings
Siblings once removed
Grandparent-grandchild
1st Cousins
1st Cousins once removed
2nd Cousins
2nd Cousins once removed
3rd Cousins
3rd Cousins once removed
4th Cousins
Estimation corrected for pseudo-replication
Clark (2023) estimation; susceptible to pseudo-replication
Correlation in occupational status
0.5
0.3
0.1
1
2
3
4
5
6
7
8
9
Degree of relatedness (n)


Signals of change in social mobility
(a)
Occupational status
Father-son correlation
0.7
0.6
0.5
(b)
Higher education
Father-son correlation
0.8
0.7
0.6
0.5
0.4
0.3
1780-1799
1800-1819
1820-1839
1840-1859
1860-1879
1880-1899
1900-1919
(c)
Literacy
Parent-child correlation
0.5
0.4
0.3
0.2
0.1
1760-1779
1780-1799
1800-1819
1820-1839
1840-1859
1860-1879
Offspring birth year