Culture Versus Nature in Social Outcomes. A Lineage Study of 285,188 English Individuals, 1750-2018

Gregory Clark, UC Davis and LSE, Neil Cummins, LSE

Economics, Sociology, and Anthropology are all dominated by the belief that while physical traits like height are mainly determined by genetics, children's social outcomes are principally created by parental investment and community socialization. Employing a lineage of 285,188 English people 1750-2018 we test whether there is a clear signal in the patterns of inheritance that culture and institutions dominate in determining social status. Surprisingly we find that status transmission follows a pattern that would not be predicted by cultural transmission, but instead would be predicted by additive genetic transmission. The high persistence of status over multiple generations, however, would require in a genetic mechanism a strong genetic assortative in mating. This has been until recently believed impossible. We show, however, two types of evidence for just such sorting. The first is from a recent whole genome study. The second is evidence in the lineage that marriages show closer matching in an underlying trait than in social phenotypes.

1. Cultural Inheritance

It is widely believed that while social status - measured as occupational status, income, health, or wealth – is correlated between parents and children, this correlation is driven by parental investments in children, or by cultural transmission.[1] This belief has profound influence on peoples' perception of the fairness of social rewards, and of the need for government intervention in the lives of disadvantaged children.

In this paper we test whether culture or genetics offers a better explanation of the inheritance of social attributes. To do so we have to specify both a general model of cultural inheritance, and one of genetic inheritance. There is already a well established model of additive genetic inheritance, formulated by Fisher in 1918. This we test against

---

[1] Studies of adoptions and of twins suggest that this belief is not well founded. Such studies suggest that genetic transmission explains the majority of social outcomes, but leave room for substantial social influences. See, for example, Sacerdote, 2007.

the data below. Specifying a model of cultural transmission as an alternative is more difficult. The ways culture has been hypothesized to operate are many and varied.

However, the lineage data does show some characteristics that any such model of inheritance must have. One important characteristic is illustrated in figure 1. This shows for occupation status (from table – below) the intergenerational correlation of status across four generations. What is notable is the pattern of mobility. In the first generation, going from father to son, there is a period of more rapid mobility. However, after this mobility slows to a significantly lower rate, and maintains this lower rate across all subsequent generations. For social status, for example, the first generation correlation is around 0.6, but the subsequent marginal intergenerational correlations increase to around 0.8.

Figure 1: Occupational Status Correlations across Multiple Generations, England 1780-1930
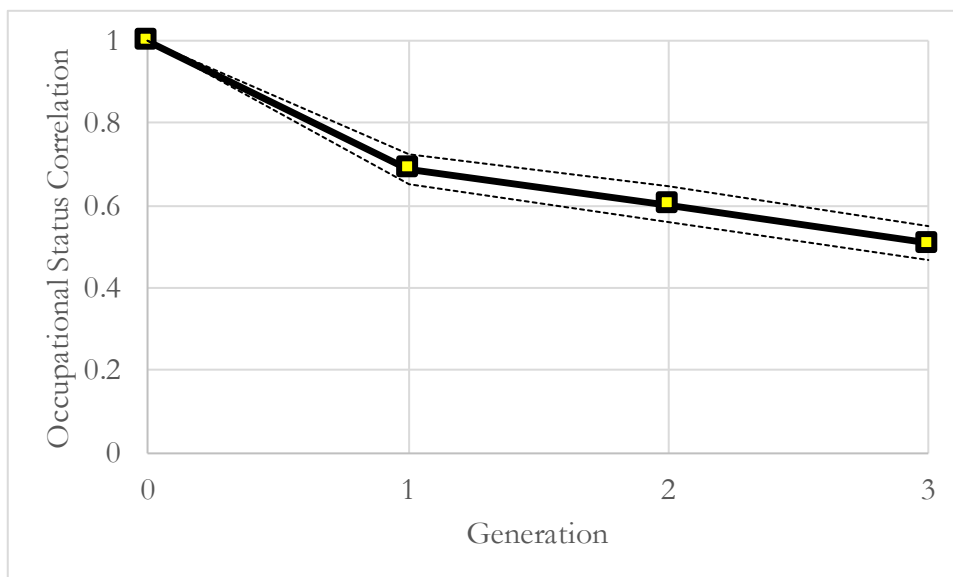
Note: The dotted lines show the 1% confidence intervals around each estimated correlation.

This pattern is very consistent in the lineage data used here, and also in other sets of data on multi-generational social mobility. Thus any model of status inheritance has to capture this pattern: one generation of faster mobility, followed by subsequent generations of greater persistence.

One model that would capture this intergenerational pattern is as below. Suppose that social outcomes, y, are the product of the family culture/environment, z, and some random component, u, so that

$$y_{it} = z_{it} + u_{it} \,.^{2} \tag{1}$$

Suppose also that the family culture/environment is regressing to the mean at rate (1-b). Then

$$z_{it} = b z_{it-1} + e_{it} \,. \tag{2}$$

e_t is a random component that must exist to keep the dispersion of family culture z constant across families across generations.

With this structure the average correlation of social outcomes between parent and child will be

$$\hat{\beta} = b \frac{\sigma_z^2}{\sigma_z^2 + \sigma_u^2} = \theta b \tag{3}$$

If the variance of the random component in equation (1) equals that of the culture/environment then $\hat{\beta} = 0.5b$. The correlation between siblings, sharing a common environment, z_t, will on average be

$$\hat{\rho} = \frac{\sigma_z^2}{\sigma_z^2 + \sigma_u^2} = \theta \tag{4}$$

Thus the sibling correlation will exceed the parent-child correlation on this cultural inheritance model. Note that this occurs even though we have left plenty of room for random influences on the outcomes among siblings exposed to common family culture. Between parent and child there is always the extra element of difference in that culture that does not exist for children. With b = .8, the sibling correlations will exceed the parent-child by 25%.

With the structure of inheritance embodies in equations (1) and (2) we can predict the correlations of any two relatives in a lineage. Thus

---

[2] In all cases in this paper variables are measured with mean 0, so that we can dispense with intercept terms in the equations.

Parent	θb
Sibling	θ
Uncle/Aunt	θb
Grandparent	θb^2
Cousins	θb^2
Great Grandparent	θb^3

In particular the correlation between children and their parents should be the same as that between children and their aunts and uncles. Also the correlation between children and their grandparents should equal that of children and their cousins.

We shall see these predicted correlations do not accord with the lineage data. So below we try and see if there is a modification of this simple model which can fit the data.

The Predictions of Genetic Inheritance

In a famous paper from 1918, "The Correlation between Relatives on the Supposition of Mendelian Inheritance" Ronald Fisher derived the expected correlation between all relatives in a family lineage if a trait was inherited genetically (Fisher, 1918). These predictions are based on a number of assumptions (see Nagylaki, 1978)

1. The traits in question are controlled by many loci in the genome, each of which makes a small contribution.
2. There is an absence of important dominance and epistasis effects.[3]
3. Genes and environment are uncorrelated, or the environment has little independent impact on outcomes.

Assumption 3 may seem unreasonably strong, but we shall see below that there is evidence that family environments have surprisingly small impacts on social outcomes. These assumptions imply that social outcomes would be determined by the weighted sum of the value of alleles at a large set of loci as in table 1.[4]

The predicted correlation between relatives in the Fisher formulas depend only on a very small set of factors. First is the heritability of the trait, h^2, which is just the correlation between the average of the parents and the child. h^2 corresponds to θ in the cultural model above. It determines how much correlation there will be in the phenotype across parents and children. The second key element is the strength and the nature of genetic assortment in marriage. If matching in marriage is based on the phenotype, then correlations between relatives will depend on both the phenotype correlation between partners, r, and the genotype correlation, m. The genotype correlation in this case will be less than the phenotype correlation. If matching is based only on the underlying genetic characteristics that determine social status, this produces a slightly different set of correlations.[5] In particular with matching on the genotype the parent-child and sibling correlations will be the same. With matching on the phenotype the parent-child correlation will exceed the sibling correlation. There is surprising evidence, discussed below, that for social outcomes the matching seems to be based on the genotype.

---

[3] That is, alleles at a particular loci do not interact, nor do alleles at different loci.
[4] At any locus there are two alleles, which typically have just two alternate values. The value at the loci can thus be coded as 0, 1 or 2.
[5] That is, there is no correlation based on the accidental elements of the phenotype.

Table 1: Determination of Genotypic Value with Additive Inheritance

Locus	Allele Value	Weight	Effect
1	$v_1 = 0, 1, or 2$	w_1	$w_1 \times v_1$
2	$v_2 = 0, 1, or 2$	w_2	$w_2 \times v_2$
3	$v_3 = 0, 1, or 2$	w_3	$w_3 \times v_3$
…..			
n	$v_4 = 0, 1, or 2$	w_n	$w_n \times v_n$
All			$\sum_{i=1}^{n} w_i v_i$

Table 2 shows the predictions under both types of matching. One thing that stands out is that in either variant the additive genetic model meets the requirement of figure 1 of faster mobility in the first generation followed by a slower, but constant, rate of mobility later. The intergenerational correlation in the first generation is

$$h^2 \frac{1+m}{2}$$

while the marginal decline in correlation is

$$\frac{1+m}{2}$$

Thus the additive genetic model is consistent with one important feature of social mobility. Figure 2 shows the expected connection between genetic distance and status correlations.

**Table 2: Phenotype Correlations for a Genetically Inherited Trait**

Relative	Matching on Genotype	Matching on Phenotype
Parental	$h^2 m$	r
Mid-parent - child	h^2	h^2
Single parent – child	$h^2 \dfrac{1+m}{2}$	$h^2 \dfrac{1+r}{2}$
Siblings	$h^2 \dfrac{1+m}{2}$	$h^2 \dfrac{1+m}{2}$
Uncles/Aunts – child	$h^2 \left(\dfrac{1+m}{2}\right)^2$	$h^2 \left(\dfrac{1+m}{2}\right)\dfrac{1+r}{2}$
Grandparent – child	$h^2 \left(\dfrac{1+m}{2}\right)^2$	$h^2 \left(\dfrac{1+m}{2}\right)\dfrac{1+r}{2}$
Cousins	$h^2 \left(\dfrac{1+m}{2}\right)^3$	$h^2 \left(\dfrac{1+m}{2}\right)^2\dfrac{1+r}{2}$
Great Grandparent – child	$h^2 \left(\dfrac{1+m}{2}\right)^3$	$h^2 \left(\dfrac{1+m}{2}\right)^2\dfrac{1+r}{2}$
Second Cousins	$h^2 \left(\dfrac{1+m}{2}\right)^5$	$h^2 \left(\dfrac{1+m}{2}\right)^4\dfrac{1+r}{2}$

Note: m is the correlation of parents in genotype, r the correlation in phenotype.

Figure 2: The Predicted Connection between Status Correlations and Genetic Distance, Genetic Transmission of Status
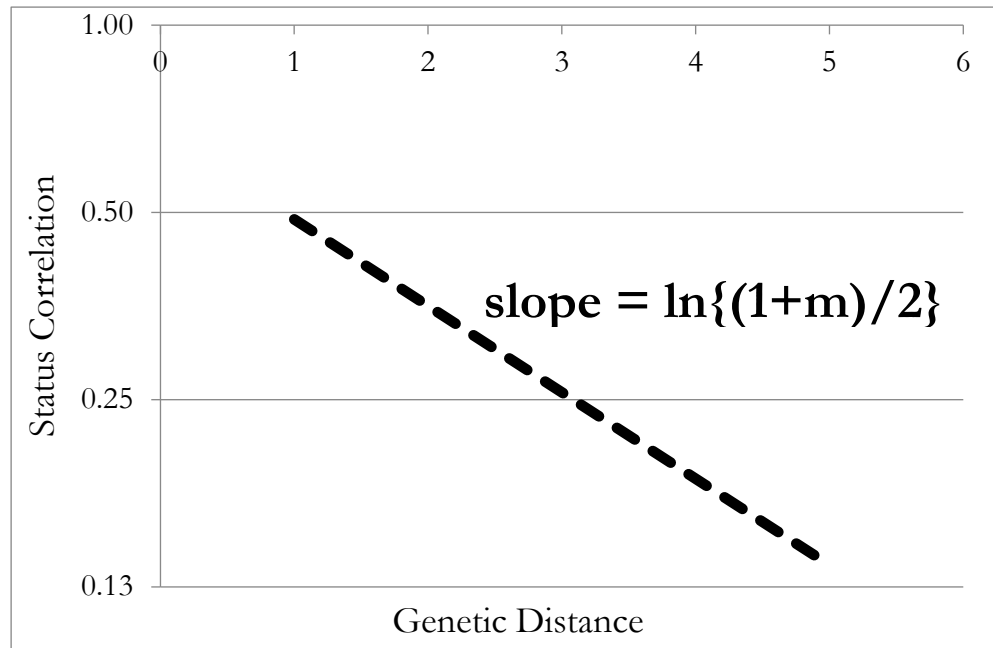
The second thing that stands out in table 2 is that the predictions of cultural transmission deviate from those of genetic transmission in terms of the pattern of correlations. The parent-child correlation is lower than the sibling correlation with cultural transmission, but in table 2 always equals or exceeds the sibling correlation. The cousin correlation in table 2 is lower than the grandparent correlation, but on the cultural model equals this correlation.

The third notable feature is that for the long run correlation $(1+m)/2$ to equal 0.7-0.8 m has to be in the range 0.4-0.5 between marital partners. If mating were random, so that $m = 0$, b would be constrained to be ½.

If parental matching is through the phenotype then the underlying correlation of the parents genotypes, m, will necessarily be low. Empirical evidence on the correlation of phenotypes, r, suggests these are relatively modest. Table 3, for example, shows measured correlations of married couples by a variety of characteristics. These correlations are typically in the range 0.2-0.5, suggesting that m would be in the range 0.1-0.25 only, if matching is through the phenotype.

Table 3: Phenotypic Correlations between Spouses

Characteristics	Correlation	Source
Height	0.29	McManus and Mascie-Taylor, 1984
Education	0.50	Watkins and Meredith, 1981
Income	0.34	Watkins and Meredith, 1981
Occupational Status	0.12	Watkins and Meredith, 1981
IQ	0.20-0.45	Mascie-Taylor, 1989
BMI	0.28	Abrevaya and Tang, 2011
Personality Traits	0.15	Mascie-Taylor, 1989

So a further implication of additive genetic inheritance of social status will be that to produce the observed patterns of very slow mobility, mating must be assortative with respect to the genotype that generates social phenotypes, and not with respect just to the phenotype. In that case the correlation in genotypes will be greater than the observed correlation in phenotypes ($r = h^2 m$). The need for closer genetic correlations of parents implies also that the relevant intergenational correlations will be those of the second column of table 2.

In modern high-income societies height is known to be largely genetically inherited, and is the outcome of at least 300 genes, each of which exerts very modest influence. Height is thus a good model of what we can expect to observe with additive genetic inheritance. The inheritance of height also features another element that would be a characteristics of genetic inheritance, but not necessarily of cultural and social modes of inheritance. This is that if your social status is determined by an underlying score on a weighted set of many genes, as in table 1, then that underlying score will be normally distributed. This means that the process of social mobility will be symmetrical at the top and bottom of the distribution. Specifically the time it takes for the descendants of someone at the top of social ability to regress to the mean will be the same as the time it takes for someone at the bottom to reach the mean.

This property appears when we look at the regression to the mean of height. Figure 3 shows the heights of parents and children in Galton's pioneering study of the inheritance

of heights. The rate of movement to the mean for height is the same for the upper 5% as for the lower 5%. For social processes the underlying differences in genetic endowment may be normally distributed, but the consequences in terms of social status as in occupational status, education or wealth can be highly skewed. However, the argument would be that if we can identify the top 5% versus the bottom 5% of the distribution, the rate of movement to the mean should be identical.

To show that the Fisher model correctly predicts the correlation of relatives in a lineage, table 4 shows the correlation of heights between relatives recorded in a modern health study of a district in Norway 1984-6 (Tambs et al., 1992). The spousal correlation, measuring only 0.18, was taken as correct. From this, and the parent-child correlation, h^2 is estimated at 0.73. If sorting was on the phenotype for height this implies an underlying correlation of height genotype between parents of only 0.13. That in turn implies that the long run intergenerational correlation $\left(\frac{1+m}{2}\right) = 0.56$.

Knowing r, h^2, and m we can predict the other correlations between relatives – siblings, grandparent, avuncular, and cousins – and compare this with the measured correlation. Except for cousins the model predictions correlations are close to the actual. But for cousins the sample size is very small, and the correlation consequently measured with much potential error. Figure 4 shows the fit of the Fisher model to the data, if we just run OLS in the log of the height correlation against genetic distance.

Another human trait which is almost entirely genetically inherited is the finger Total Ridge Count, which is the number of ridges, measured in a standardized way, on all 10 digits. Table 5 shows the familiar correlations for this measure for a sample of 200 husbands and wives and their children, including disproportionately twins. Here since this is a completely unobserved trait we would expect no assortment in marriage. Indeed the spousal correlation in the phenotype is not significantly different from 0. This trait is also highly hereditable. The h^2 here will be the correlation between the monozygotic twins, which is 0.95. When we apply the Fisher formulas the familial correlations fit very well with the Fisher predictions.

**Figure 3: Symmetry of Regression to the Mean with Height**
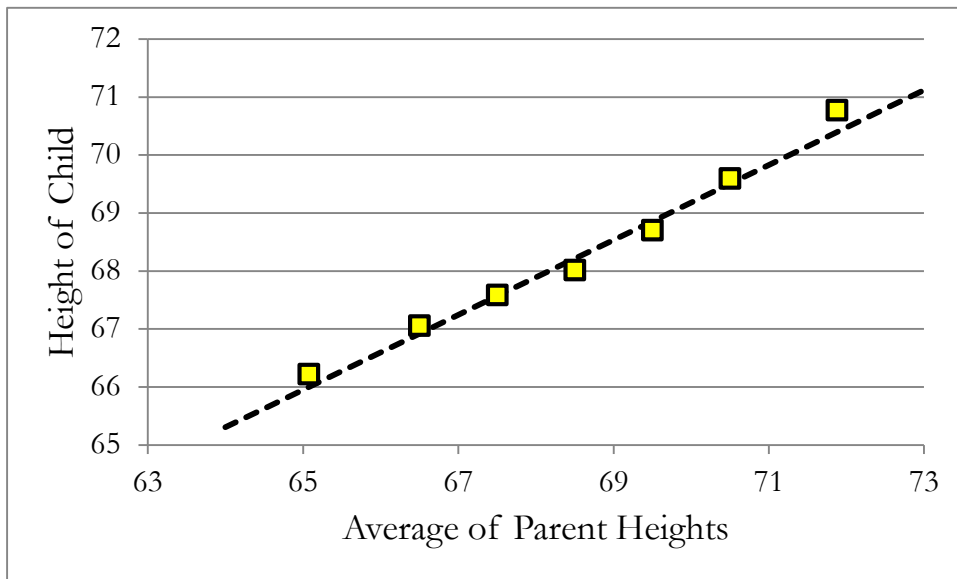
Source: Clemons, 2000.

Table 4: Height Correlations in Norway, 1984-6

Relation	Number	Measured Correlation	Predicted Value	Fitted Value
Spouses	24,281	0.179	R	(0.179)
Parent-Child	43,613	0.430	$h^2 \dfrac{1+r}{2}$	(0.430)
Siblings	19,168	0.453	$h^2 \left(\dfrac{1+m}{2} \right)$	0.412
Grandparent-Child	1,318	0.250	$h^2 \left(\dfrac{1+m}{2} \right) \dfrac{1+r}{2}$	0.243
Avuncular	1,218	0.217	$h^2 \left(\dfrac{1+m}{2} \right) \dfrac{1+r}{2}$	0.243
Cousins	112	0.209	$h^2 \left(\dfrac{1+m}{2} \right)^2 \dfrac{1+r}{2}$	0.137

Source: Tambs et al., 1992. Values in parentheses assumed correct.

Figure 4: Fitted Height Correlations
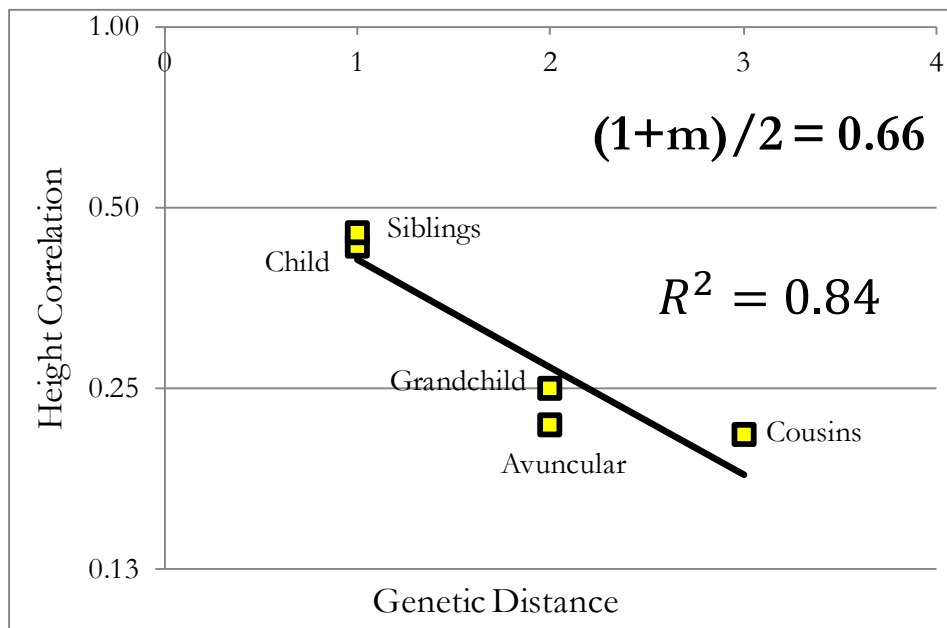
Table 5: Inheritance of Total Ridge Count

Relationship	Number of Pairs	Correlation (s.e.)	Predicted
Mother-Child	405	0.48 (.02)	$h^2\left(\frac{1+m}{2}\right)= 0.50$
Father-Child	405	0.49 (.02)	$h^2\left(\frac{1+m}{2}\right)= 0.50$
Husband-wife	200	0.05 (.03)	$m =0.05$
Sibling-Sibling	642	0.50 (.02)	$h^2\left(\frac{1+m}{2}\right)= 0.50$
Monozygotic Twins	80	0.95 (.01)	$h^2=0.95$
Dizygotic Twins	92	0.49 (.04)	$h^2\left(\frac{1+m}{2}\right)= 0.50$

Source: Holt, 1961.

The Nature of Inheritance in an English Lineage

We propose to test whether the cultural or additive genetic model of inheritance works predicts better using a lineage under construction for English families 1750-2018, that shows all familial links, plus a variety of social outcomes. So far we have the familial connections of all the people in the lineage (263,414), but only social outcomes for a subgroup of people. Figure 5 shows a sample lineage for one couple and some of their descendants from the database. The figure illustrates the richness of the set of family links that the database contains. In this case the lineage covers 7 generations. But what matters is the set of social outcomes we can associate with the members of the lineage. Table 5 summarizes the data currently available.[6] The social status indicators we have are wealth at death, occupation, educational attainment, schooling and training 11-20, and age at death.[7] The ones we employ here are wealth at death, occupational status, and higher educational attainment. Because of the time period covered by these measures, which is births 1750-1929, we include only men in the sample.

Wealth at Death: For England and Wales the Principle Probate Registry records whether someone was probated, and the value of their estate for all deaths in England 1858-2018. For 1799-1857 we also get from the Canterbury and York courts estate values for those higher in the wealth distribution (top 4% of men). For this measure we have women also, and so could potentially extend the study to study women also.

Occupation Status: Occupations are given in the censuses of 1841-1911 as well as the population register of 1939. There are also occupation statements in some marriage registers for both grooms and the fathers of the marriage parties, for fathers in birth registers, for the deceased in death registers, and also in some years for the deceased or for executors in probate records. We translated these various occupational statements into 242 occupational categories – carpenter, laborer, solicitor, dealer, stockbroker etc. We gave these occupations a social status score between 0 and 100. That score was created as an equally weighted average of three elements: average normalized ln wealth at death by occupation, average fraction of people in each occupation with a university degree or equivalent, and average fraction of males in each occupation who were in school or in training when observed ages 11-20 in the censuses of 1811-1911, and the population register of 1939.

---

[6] We expect to be able to add much more information on occupations and schooling.
[7] In recent years in England first names are a strong indicator of social status.

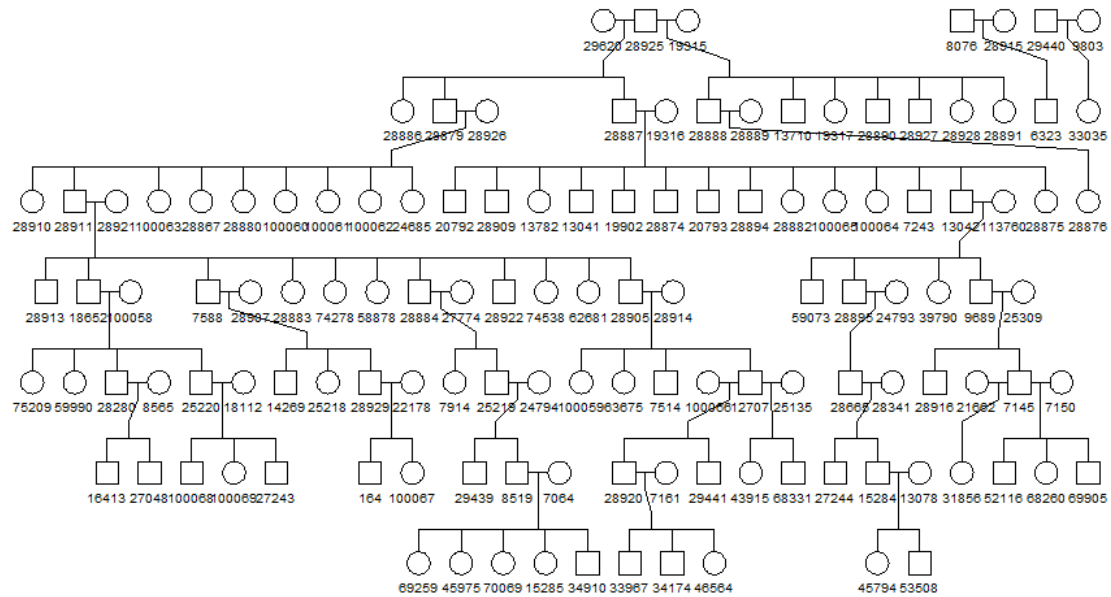**Figure 5: An Illustrative Portion of a Family Lineage, Lineage Database**

Table 6: Data Availability in the Families of England Lineage

Relationship	Identical by Descent	All	Higher Ed.	Wealth at Death	Occupation	Age at Death
Father	1/2	83,573	18,347	13,241	15,151	34,656
Brother	1/2	95,874	19,611	13,769	15,486	43,587
Grandfather	1/4	56,272	13,607	9,534	10,678	26,137
Uncle	1/4	146,806	33,072	24,304	26,307	73,943
Gt Grandfather	1/8	40,778	8,935	5,265	6,619	18,414
Gt Uncle	1/8	69,422	13,205	8,493	9,920	31,710
Cousin	1/8	86,026	20,134	15,504	17,430	47,391
Gt Gt Grandfather	1/16	29,061	4,633	1,779	3,219	11,502
Gt Gt Uncle	1/16	52,280	7,339	3,839	5,244	20,336
Gt Gt Gt Grandfather	1/32	20,318	2,178	353	1,594	6,975
2nd Cousin	1/32	84,238	18,169	14,443	16,561	45,033
Gt Gt Gt Uncle	1/32	36,103	3,341	981	2,222	11,362
Gt Gt Gt Gt Grandfather	1/64	14,287	1,025	46	885	4,497
Gt Gt Gt Gt Uncle	1/64	23,839	1,428	102	1,050	6,563
3rd Cousin	1/128	71,687	12,596	10,378	12,657	35,229
4th Cousin	1/512	63,102	10,009	8,338	11,895	30,738
5th Cousin	1/2048	53,769	8,643	7,886	11,383	29,248

Higher Education: This is an indicator variable with a value 1 if the person achieved a higher educational status. Complete records are available for attendees Oxford and Cambridge (1750-2018), the Royal Military Academy Woolwich (1790-1839) and and the Royal Military College Sandhurst (1800-1946). Complete records are available for the UK Medical Registers, 1859-2017, UK, Civil Engineer Lists, 1818-1930, UK, Electrical Engineer Lists, 1871-1930, UK, Mechanical Engineer Records, 1847-1930, UK, Articles of Clerkship (attorneys), 1756-1874.

Table 6 shows the information available on each category of relative in our database currently. It also shows the share of genes that on average will be identical by descent between the various relatives. As can be seen, if there is not significant genetic assortment in marriage the share of genes relations like 3rd to 5th cousins will share will be extremely small, and thus any correlation in outcomes explained by genetics trivially small.

Table 7 shows the correlations of relatives of different degrees on three of the status characteristics, as well as the closeness of the genetic connection. If the degree of genetic distance is n, then the correlation in genetics will be

$$\left(\frac{1+m}{2}\right)^n$$

If m = 0 then that correlation for 5th cousins will be .0005. If, however, m = 0.6, then the correlation would be 0.09. As noted above the Fisher equation implies that the logarithm of the intergenerational correlation of status on any measure will be linear with respect to genetic distance, as was shown in figure 2. In particular where ρ_n is the correlation between relatives n steps apart genetically

$$\rho_n = \ln(h^2) + nln\left(\frac{1+m}{2}\right)$$

The most dramatic contrast between the cultural and the genetic model of transmission concerns the relative correlation of siblings versus parent-child in outcomes. These correlations are shown in the second and third lines of table 7. Figure 7 shows the relative father-son and brother correlations for the three attributes in table 7, as well as for lifespan, age at first marriage and wife's age at first marriage. As can be seen these correlations are near identical. This outcome is inconsistent both with cultural transmission, and also even with genetic transmission, but with assortment based on the phenotype (assuming strong assortment). Something is causing an unexpected degree in variation in the outcomes of siblings (from a cultural or parental investment perspective). Additive genetic transmission has a build in explanation of this pattern.

Table 7: Intergenerational Correlations, Males

Relationship	Genetic Distance	Wealth	Occupation	Education
Father	1	.592 (.007)	.692 (.006)	.397 (.007)
Brother	1	.531 (.007)	.684 (.006)	.398 (.007)
Grandfather	2	.501 (.009)	.598 (.008)	.305 (.008)
Uncle	2	.447 (.006)	.620 (.005)	.326 (.005)
Gt Grandfather	3	.423 (.012)	.506 (.011)	.219 (.010)
Gt Uncle	3	.402 (.010)	.553 (.008)	.287 (.008)
Cousin	3	.380 (.007)	.605 (.006)	.313 (.007)
Gt Gt Grandfather	4	.391 (.022)	.354 (.016)	.084 (.015)
Gt Gt Uncle	4	.324 (.015)	.449 (.012)	.193 (.011)
Gt Gt Gt Grandfather	5	.379 (.049)	.182 (.025)	.041 (.021)
2nd Cousin	5	.260 (.008)	.446 (.007)	.274 (.007)
Gt Gt Gt Uncle	5	.302 (.030)	.319 (.020)	.087 (.017)
3rd Cousin	7	.136 (.010)	.283 (.009)	.217 (.009)
4th Cousin	9	.096 (.011)	.143 (.009)	.145 (.010)
5th Cousin	11	.030 (.011)	.032 (.009)	.036 (.011)

Note: Pearson Correlations. Standard errors in parentheses.

Figure 7: Comparative Father-Son and Brother Correlations

16
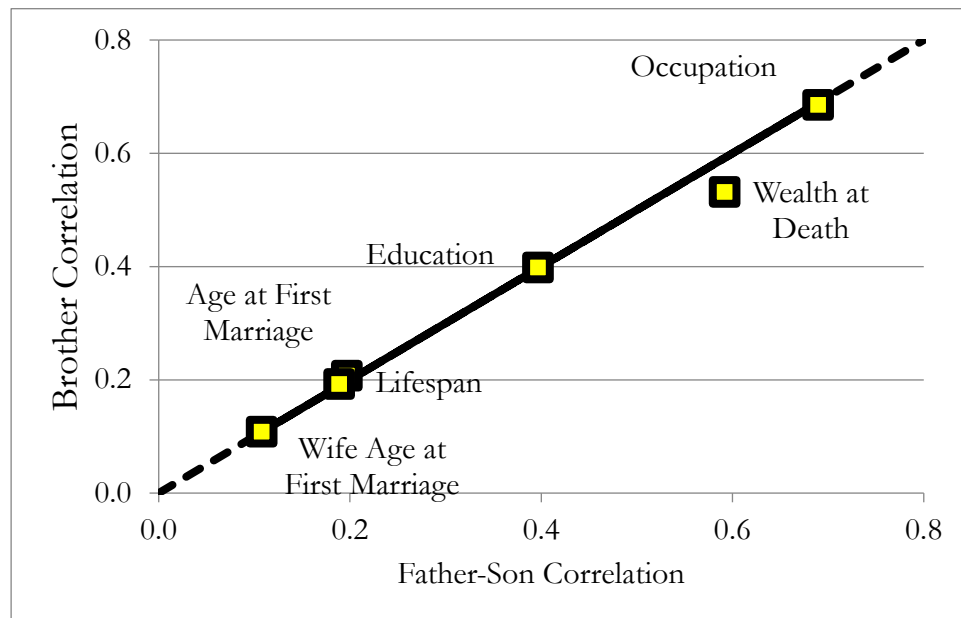
Another implication of the cultural model outlined above is that the correlation of children in status with their fathers should equal that with their uncles. With genetic transmission the uncle correlation is lower by a factor of $\left(\frac{1+m}{2}\right)$. Figure 8 shows these comparative correlations compared with the prediction of the cultural model, which is that the correlations will fall along the 45° line. The pattern of correlations is again systematically at variance with the simple cultural model, and in line with the genetic model. The uncle-nephew correlations are all smaller than those of father-son.

Figure 8: Father versus Uncle Correlations
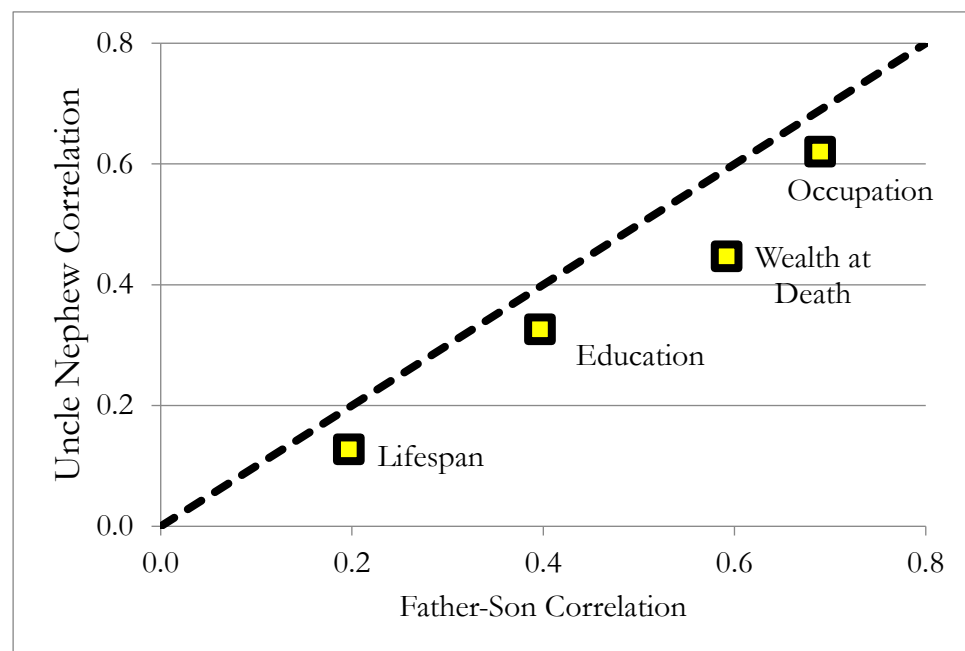
This suggests that we need a modified cultural transmission model, which increases the correlation between father and son, relative to brothers. Suppose that we have, as before,

$$y_{it} = z_{it} + u_{it} \tag{1}$$

But now,

$$z_{it} = b(z_{it-1} + u_{it}) + e_{it} . \tag{5}$$

The idea here is that any deviation in brother outcomes from that predicted by their family environment as children gets embedded in the environment of their children. Thus now

the outcomes of children are more closely linked to their fathers than to their brothers. Now the correlation between father and son rises from θb to b, while that between brothers remains at θ. With the right parameters we can have $\theta = b$, so that the father and sibling correlations are the same.

However, while this model will capture this observed feature of social mobility, it fails completely to capture other significant features. One is that whatever the short run mobility rates is for any aspect of social status, the long run mobility is the same. Now the long run mobility rate has to equal that in the short run. So this crucial aspect of the social mobility process is missing.

Figure 9, for example, plots all the correlations in table 7 for wealth at death, against the expected genetic distance between relatives. Since genetic distance has a predicted multiplicative effect in reducing correlations the correlation is shown on a logarithmic scale on the vertical axis. As can be seen the correlations for the entire set of relatives fall close to the predicted linear pattern, with *(1+m)/2* = 0.81 (the standard error on this estimate is 0.02). The R^2 of the fit is 0.91. Wealth persists across the lineage to a remarkable degree after the first generation, when the correlation is significantly lower. Thus even 5th cousins, who shared a common ancestor only 6 generations in the past, still have a small but statistically significant correlation in wealth.

Wealth involves for richer families a physical transfer of property. But even here the recipients have choices about whether to spend this transferred wealth, or accumulate it. And their decisions will be influences by their current wage earnings. So it is not impossible for there to be important genetic components underlying wealth at death.

Figure 10 shows the same graph, but now for occupational status. Now the correlations again fall relatively closely to the predicted linear pattern. The R^2 for the fitted line is 0.77. The estimated value for *(1+m)/2* = 0.82 (the standard error on this slope estimate is 0.031). Notice that the estimate of (1+m)/2 using occupations is very close to that using wealth, despite the short run father-son correlation being different.

Figure 11 shows again the pattern of correlations, this time for educational status (measured as an indicator variable for attaining higher educational qualifications. Here the estimated value for *(1+m)/2* = 0.83 (the standard error on this estimate is, however, much higher at 0.057). The R^2 for the fitted line, estimated as above by least squares, is 0.48. The outliers observed for this graph make the fit much less precise. Again, despite the short run father-son correlation being different, long run persistence is very similar.
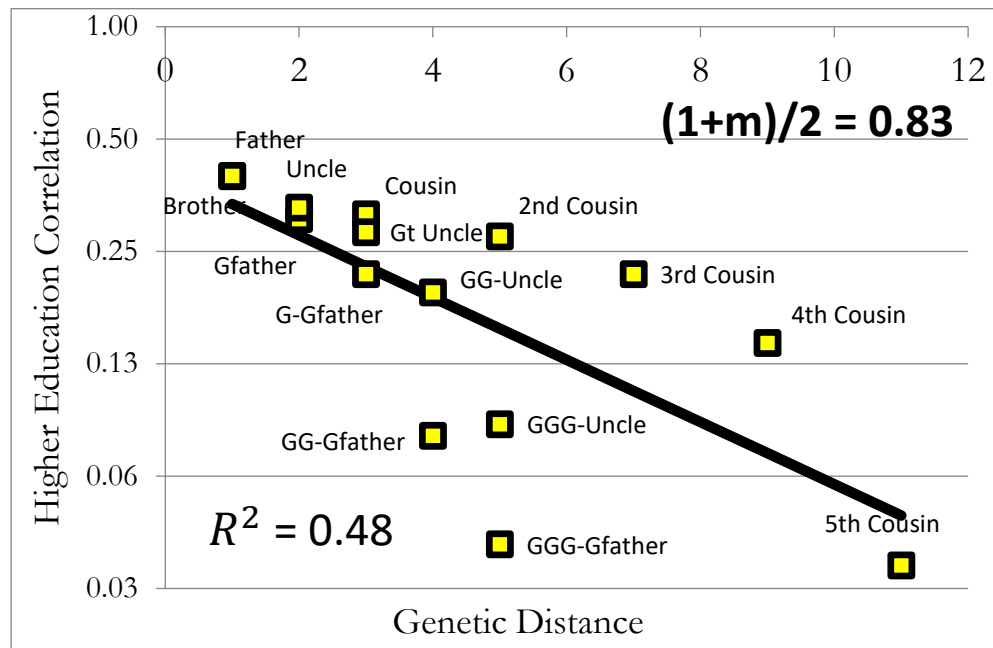
Figure 9: Wealth at Death Correlations and Genetic Distance

$$(1+m)/2 = 0.81$$

$$R^2 = 0.91$$
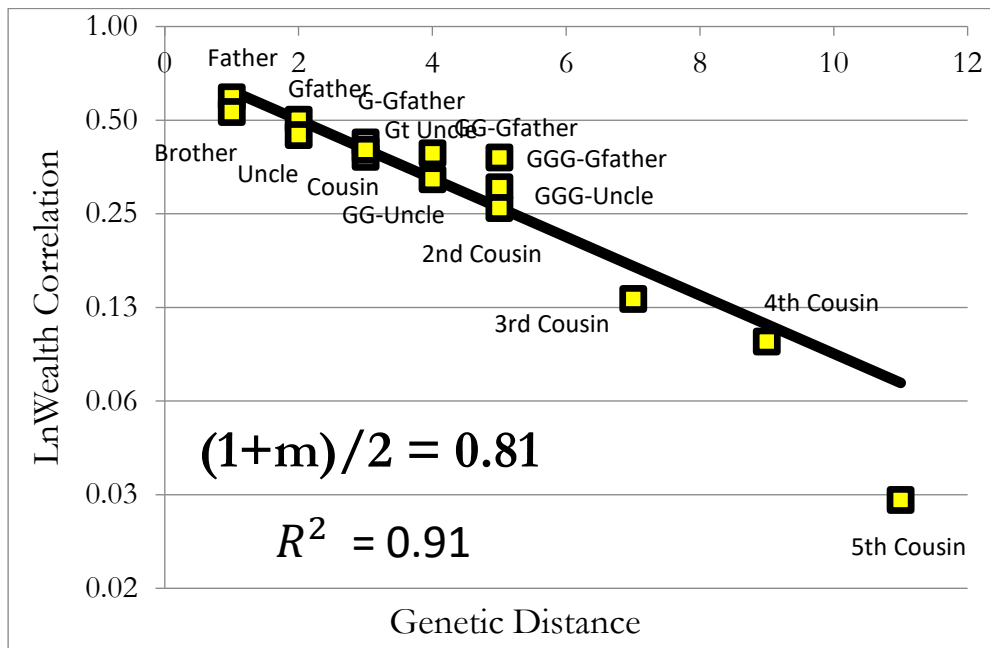
Figure 10: Occupational Status Correlations

$$(1+m)/2 = 0.82$$

$$R^2 = 0.77$$
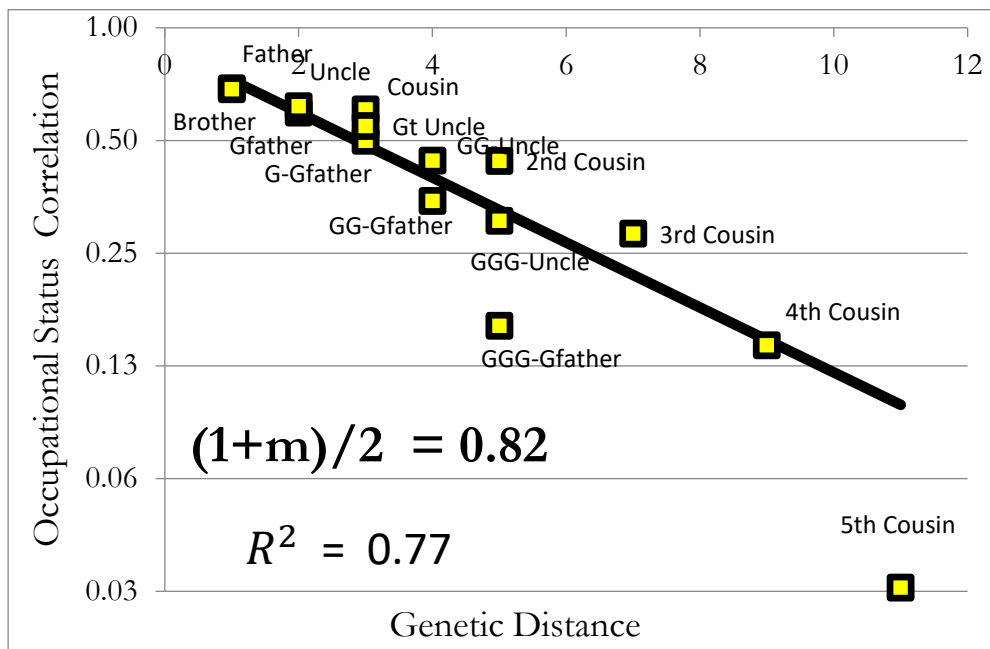
Figure 11: Educational Status Correlations

Thus the empirical evidence here from the correlations is all consistent with genetic transmission of social status. But it requires what would be regarded as an impossibly high degree of marital genetic assortment, with m required to be in the range 0.6-0.65, in light of observed phenotype correlations as in table 3.

Marital Assortment

There is recent evidence that marital assortment for social traits can be much stronger for the genotype than for the phenotype. Thus Robinson et al., 2017 look at the phenotype and genotype correlations for a variety of traits – height, BMI, blood pressure, years of education using recent full genome data for England. As we saw in table 3, for most traits the phenotype correlation is modest. For height, for example, 0.2. Also the genotype correlation for height is no higher than the phenotype correlation. For height it is also around 0.2, where now the correlation is between a sum of scores as in table 1.

But there is one notable exception. For years of education, the phenotype correlation across spouses is 0.41 (0.011 SE). However, the correlation across the same couples for the genetic predictor of educational attainment is significantly higher 0.654 (0.014 SE) (Robinson et al., 2017, 4). Thus couples are actually sorting on the genotype as opposed to the phenotype when it comes to educational status.

How can we test empirically if mating is assortative based potentially on underlying genetics as opposed to individual phenotype characteristics? Suppose, for example, that the various characteristics associated with status – education, wealth, occupation – all derive from the same underlying genetics, so that for an individual i and trait k

$$y_{ik} = x + u_{ik},\qquad(1)$$

Let M indicate males and F females. Then if we regress for couples

$$y_{ikM} = \lambda y_{ikF} + \varepsilon_{ik},\qquad(6)$$

we can measure the phenotype correlation, λ, but not the genotype correlation. Suppose, however, we have multiple measures of the phenotype of the parents, wealth and education for example, y_{i1} and y_{i2}. In this case if we regress

$$y_{i1M} = \lambda y_{i1F} + \varepsilon_{i1},\qquad(7)$$

but instrument for y_{i1} with y_{i2}, then $E(\hat{\lambda}_{IV})$ = correlation of the underlying characteristic, x.

Our data for spouse status is for marriages 1780-1959, largely in a period before women had good independent measures of status. We have to proxy for the status of women using either that of their brothers, or the status of their father. This will be a noisy measure of the underlying status of women, with an expected correlation of the underlying genetics of father and brother with daughter being $(1+m)/2$. So in this case if we use an IV estimate, the coefficient we would expect to recover would be $m(1+m)/2$ rather than $(1+m)/2$.

Table 8 shows the brother-brother in law correlations in wealth, occupation and higher education attainment. In all cases the instrumented coefficient is significantly higher than the OLS coefficient. Indeed the average of the instrumented correlations is 0.8, compared to 0.49 for the OLS correlations. These estimates are consisted with marriage being highly assortative on some underlying factor, which could include a genetic score,

A similar set of estimates to table 8, looking at the instrumented father to son-in-law correlation finds that to average 0.70, again very high. These would imply, if the underlying factor is genetics, an m that lies in the range 0.78-0.86.

Table 8: Brother versus Brother-in-Law Correlations

	OLS	IV$_{occupation}$	IV$_{education}$
	(1)	(2)	(3)
Ln(Wealth)	.449***	.920***	.835***
	(.013)	(.028)	(.041)
Observations	4,602	3,693	4,464
R^2	.218	.020	.036

	OLS	IV$_{wealth}$	IV$_{education}$
	(1)	(2)	(3)
Occupational Status	.705***	.986***	.739***
	(.010)	(.022)	(.014)
Observations	6,662	3,814	5,808
R^2	.418	.444	.445

	OLS	IV$_{wealth}$	IV$_{occupation}$
	(1)	(2)	(3)
Higher Education	.297***	.767***	.543***
	(.010)	(.037)	(.014)
Observations	8,871	5,835	7,238
R^2	.090	$-.201$	.064

Note:	*p<0.1; **p<0.05; ***p<0.01

Genetics and Environment Interactions

The assumption above that underlies the Fisher formula for the correlation of relatives that "*Genes and environment are uncorrelated, or the environment has little independent impact on outcomes*" would seem to be obviously violated in the case of social traits. The familial environment of families does clearly vary, and that variation will be correlated with the genetics that help determine family social status. However, there is good information from the *Families of England* database that the second part of the condition largely holds, and that family environments do indeed have *little independent impact on outcomes.*

The lineage used in this paper also allows us to test for the effects of elements of family environment on social outcomes, because for part of the period covered by the lineage, marriages 1780-1880, family size was largely random. In this period there was great variation in completed family size, numbers of children reaching age 21, with the size range in the sample for men ranging from 1 to 18. There was no correlation between family size and any measure of social status for fathers. There was also very weak correlation between brothers, and between fathers and sons, in terms of either births or completed family size. That correlation was in the range 0.03-0.05. Since brothers and fathers and sons correlate very strongly on an underlying latent variable for social status, which would correlate with lifestyles and choices on family size, this implies that both the number of births, and also childhood mortality, were mainly random in this interval, and not the product of individual decisions.

We just summarize the effects of the family size and birth order on social outcomes for marriages here, since we have another paper devoted to this substantial topic (Clark and Cummins, 2017). The families in the lineage can be separated into those lines where average wealth at death circa 1850 was high, and those where wealth at death then was average or non-existent. In the high wealth families servants anyway provided much of child care, so the effects of size might be expected to be less. In poorer families, this was a period where there was mostly no compulsory education. A legal requirement of school attendance to age 10 was only introduced in 1881. Thus in poorer families parents had to make an important decision about whether to support the children in schooling ages 11-20 that would be affected potentially by family size.

Table 9 summarizes the effect of family size, measured as either births (N0) or children reaching age 21 (N21) per father, on various social outcomes for marriages 1780-1879. Mostly results for sons are shown, except for the case of wealth at death, since in this period only sons have occupations and educational attainments. In each case the elasticity of the measure with respect to family size is given. In most cases there is a negative effect of size on outcomes. But if we look at adult outcomes – wealth at death, occupational status at

Table 9: Elasticities of Outcomes with respect to Family Size, Marriages 1780-1879

Variable	Rich	.	Poor/Avg.	.
	N0	N21	N0	N21
At School, 11-20, Boys	−.07	−.05	−.24***	−.13*
At School, 11-20, Girls	−.04	−.04	−.26***	−.14*
Higher Education	−.17***	−.12**	.32	.25
Occupational Status	−.09***	−.08***	−.05**	−.03
Child Survival, Children	.01	.01	−.01	.01
Wealth	−.56***	−.50***	−.25***	−.10

Note: *p<0.1; **p<0.05; ***p<0.01

age 40, attainment of higher educational qualifications, and child mortality for children – then only in the case of wealth are there significant effects, and here only for families from wealthy lineages. For occupational status, for example, the elasticities range from -0.03 to -0.09, implying that a 10% increase in family size reduces occupational status by 0.3%-0.9%. Even for wealth there is evidence that the shock to wealth created by larger family size tends to dissipate across generations. Even for wealthy family lineages by the time we come to grandchildren there is no significant effect of family size at the grandparent generation on wealth at death of the grandchildren.

Thus in the long run wealth seems to depend more on underlying abilities and attitudes whose inheritance is not affected by shocks to family size. So while in the course of two generations wealth is dependent on non-genetic factors, in the longer run wealth dynamics are possibly still genetic.

In all cases birth order has no significant effects on outcomes. Even in families of 12 the oldest son had the same life outcomes as the youngest.

Conclusions

It is generally assumed that the elements that define social status – occupational status, educational attainment, wealth, and even health – are transmitted across generations in important ways by the family environment. Above we show that the patterns of correlation of social status attributes in an extended lineage of 263,456 people in England are mainly those that would be predicted by simple additive genetic inheritance of social status in the presence of highly assortative mating around status genetics. Parent-child correlations for a trait equal those of siblings, and the patterns of correlation of relatives of different degrees

of genetic affinity is mainly consistent with that predicted by additive genetics. Further family size and birth order, elements that would significantly affect the family environment for children, have modest effects on adult outcomes. The underlying persistence of traits is such that people who have likely never interacted socially, such as second to fifth cousins, remain surprisingly strongly correlated in terms of occupational status and wealth. The patterns observed imply that marital sorting must be strong in terms of the underlying genetics. But we find evidence, through IV estimates, that there is indeed such sorting, with an underlying correlation in latent social status estimated to be in the range 0.6-0.9.

If this interpretation is correct then aspirations that by appropriate social design, rates of social mobility can be substantially increased will prove futile. We have to be resigned to living in a world where social outcomes are substantially determined at birth.

Data Sources

Births, Deaths, Marriages

General Register Office. England and Wales Civil Registration Indexes. London, England: General Register Office.

FreeBMD. England & Wales, FreeBMD Death Index: 1837-1915 [database on-line]. Provo, UT, USA: Ancestry.com Operations Inc, 2006. Available at www.ancestry.com (last accessed: 01 Apr 2016).

FreeBMD. England & Wales, FreeBMD Birth Index, 1837-1915 [database on-line]. Provo, UT, USA: Ancestry.com Operations Inc, 2006. Available at www.ancestry.com (last accessed: 01 Apr 2016).

Ancestry.com. England & Wales, Birth Index: 1916-2005 [database on-line]. Provo, UT, USA: Ancestry.com Operations Inc, 2008. Available at www.ancestry.com (last accessed: 01 Apr 2016).

Ancestry.com. England & Wales, Death Index: 1916-2006 [database on-line]. Provo, UT, USA: Ancestry.com Operations Inc, 2007. Available at www.ancestry.com (last accessed: 01 Apr 2016).

Ancestry.com. Public Member Trees [database on-line]. Provo, UT, USA: www.ancestry.com Operations. Inc., 2006. Original data: Family trees submitted by Ancestry members.

Social Outcomes

Censuses, 1841-911 and Population Register 1939

England and Wales, Censuses, 1841-1901. Available online at

http://www.nationalarchives.gov.uk/records/census-records.htm.

England and Wales, Population Register, 1939. London, UK. 1939 Original Data: 1939 Register, 1939. Kew, Surrey, England: The National Archives of the UK (TNA). Available at http://www.findmypast.com/1939register.

Wealth at Death

England and Wales, Index to Wills and Administrations, 1858-2012. Principal Probate registry, London (available online 1858-1966 at Ancestry.co.uk).

Gov.UK. Wills and Probate 1858-1996 [database on-line]. Original data: Principal Probate

Registry. Calendar of the Grants of Probate and Letters of Administration made in the Probate Registries of the High Court of Justice in England. London, England © Crown copyright. Available at https://probatesearch.service.gov.uk/Calendar#calendar (last accessed: 01 Apr 2016).

General Register Office. 1861. Annual Report of the Registrar General 1859. (pp. 173-181) Available online at http://www.histpop.org (last accessed: 01 Apr 2016).

Prerogative Court of Canterbury and Related Probate Jurisdictions: Probate Act Books. Volumes: 1850-57. Held at the National Archives, Kew. (Catalogue Reference: PROB 8/243-250.)

Educational Status

Ancestry.com. UK, Articles of Clerkship, 1756-1874 [database on-line]. Provo, UT, USA: www.ancestry.com Operations, Inc., 2012. Original data: Court of King's Bench: Plea Side: Affidavits of Due Execution of Articles of Clerkship, Series I, II, III (KB 105-107). The National Archives, Kew, Richmond, Surrey. Registers of Articles of Clerkship and Affidavits of Due Execution (CP 71). The National Archives, Kew, Richmond, Surrey. Available at www.ancestry.com(last accessed: 01 Apr 2016).

Ancestry.com. Cambridge University Alumni, 1261-1900 [database on-line]. Provo, UT, USA: Ancestry.com Operations Inc, 1999. Original data: Venn, J. A., comp.. Alumni Cantabrigienses. London, England: Cambridge University Press, 1922-1954. Available at www.ancestry.com (last accessed: 01 Apr 2016).

Ancestry.com. Oxford University Alumni, 1500-1886 [database on-line]. Provo, UT, USA: Ancestry.com Operations Inc, 2007. Original data: Foster, Joseph. Alumni Oxonienses: The Members of the University of Oxford, 1715-1886 and Alumni Oxonienses: The Members of the University of Oxford, 1500-1714. Oxford: Parker and Co., 1888-1892. Available at www.ancestry.com (last accessed: 01 Apr 2016).

Ancestry.com. UK, Crockford's Clerical Directories [database on-line]. Provo, UT, USA: Ancestry.com Operations Inc, 2009. Original data: Crockford Clerical Directory.

England: 1868-1932. See title page image of each directory for original source information. Available at www.ancestry.com (last accessed: 01 Apr 2016).

Ancestry.com. UK, Electrical Engineer Lists, 1871-1930 [database on-line]. Provo, UT, USA: Ancestry.com Operations, Inc., 2014. Original data: Society of Telegraph Engineers (later, Institution of Electrical Engineers) Membership Lists, 1887–1930. Institution of Engineering and Technology, Savoy Place, London, England. UK, Civil Engineer Lists, 1818-1930. Available at www.ancestry.com (last accessed: 01 Apr 2016).

Ancestry.com. UK, Mechanical Engineer Records, 1847-1930 [database on-line]. Provo, UT, USA: Ancestry.com Operations, Inc., 2013. Original data: Mechanical Engineering Records, 1847-1930. London, UK: Institution of Mechanical Engineers. Available at v (last accessed: 01 Apr 2016).

Ancestry.com. UK Medical Registers, 1859-1959 [database on-line]. Provo, UT, USA: Ancestry.com Operations Inc, 2008. Original data: General Medical Council, comp. UK Medical Registers, 1859-1959. London: General Medical Council, 1859-1959. This data is provided in partnership with the General Medical Council. Available at www.ancestry.com (last accessed: 01 Apr 2016).

Royal Military Academy Sandhurst. The Sandhurst Collection [database on-line]. Available at http://archive.sandhurstcollection.co.uk/ (last accessed: 01 Apr 2016).

Senate House Library, University of London. 2016. University of London Students 1836-1934 [database on-line]. Available at http://www.senatehouselibrary.ac.uk/our-collections/special-collections/archives-manuscripts/university-of-london-students-1836-1934 (last accessed: 01 Apr 2016).

Bibliography

Abrevaya, J and H. Tang. 2011. "Body mass index in families: spousal correlation, endogeneity, and intergenerational transmission." *Empirical Economics, 41*: 841–864.

Clark, Gregory. 2017. "Estimating Social Mobility Rates from Surnames: Social Group versus Family Effects." Working Paper, March.

Clark, Gregory and Neil Cummins. 2017. "The People, not the Place. The Decline of the North of England 1918-2017: A Surname Investigation." Working Paper, February.

Clark, Gregory and Neil Cummins. 2016. "Family Matters? Do Relatives other than Parents Matter to Social Outcomes, England 1780-2016?" Working Paper, November.

Clark, Gregory and Neil Cummins. 2016. "The Child Quality-Quantity Tradeoff, England, 1780-1880: A Fundamental Component of the Economic Theory of Growth is Missing." Working Paper, August.

Clark, Gregory and Neil Cummins. 2015. "Intergenerational Wealth Mobility in England, 1858-2012. Surnames and Social Mobility." *Economic Journal, 125*(582): 61-85.

Clark, Gregory and Neil Cummins. 2014. "Surnames and Social Mobility: England, 1170-2012." 2014. *Human Nature, 25*(4), 517-537.

Clark, Gregory, Cummins, Neil et al. 2014 *The Son Also Rises: Surnames and the History of Social Mobility.* Princeton: Princeton University Press.

Clemons, Traci. 2000. "**A Look at the Inheritance of Height Using Regression toward the Mean.**" *Human Biology*, *72*(3): **447-454**.

Falconer, D. S. 1981. *Introduction to Quantitative Genetics.* 2nd ed. London: Longman.

Fisher, R. A. 1918. "The Correlation between Relatives on the Supposition of Mendelian Inheritance." *Transactions of the Royal Society of Edinburgh, 52:* 399-433.

Hill, William G., Michael E. Goddard, and Peter M. Visscher. 2008. "Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits." Plos Genetics, 4(2), e1000008.

Holt, Sarah B. 1961. "Quantitative Genetics of Fingerprint Patterns." *British Medical Bulletin*, 17: 247-50.

Mascie-Taylor, C. G. Nicholas. 1987. "Assortative mating in a contemporary British population." *Annals of Human Biology*, *14*:1: 59-68. DOI: 10.1080/03014468700008841

Mascie-Taylor, C. G. N. 1989. "Spouse Similarity for IQ and Personality and Convergence." *Behavior Genetics, 19(2)*: 223-7.

McManus, I. C. and C. G. N. Mascie-Taylor. 1984. "Human Assortative Mating for Height: Non-Linearity and Heteroscedasticity." *Human Biology, 56(4)*: 617-623.

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

Plomin, R 2011. "Commentary: Why are children in the same family so different? Non-shared environment three decades later." *International Journal of Epidemiology* 40: 582–592.

Plomin, R., Fulker, D.W., Corley, R., and DeFries, J. C. 1997. "Nature, nurture, and cognitive development from 1-16 years: A parent-offspring adoption study." *Psychological Science*, *8*: 442-447.

Robinson, Matthew R., et al., 2017 "Genetic evidence of assortative mating in humans," *Nature Human Behaviour*, 2016, 1. Article number: 0016 (2017).

Sacerdote, Bruce. 2007. "How Large Are the Effects from Changes in Family Environment? A Study of Korean American Adoptees." *Quarterly Journal of Economics, 122*(1): 119-157.

Silventoinen, Karri, Jaakko Kaprio, Eero Lahelma, Richard J. Viken, and Richard J. Rose. 2003. "Assortative Mating by Body Height and BMI: Finnish Twins and Their Spouses." *American Journal of Human Biology 15*:620–627.

Tambs, Kristian, Torbjørn Moum, Lindon J. Eaves, Mike C. Neale, Kristian Midthjell, Per

G. Lund-Larsen and Siri Næss. 1992. "Genetic and Environmental Contributions to the Variance of Body Height in a Sample of First and Second Degree Relatives." *American Journal of Physical Anthropology 88: 285-294.*

Watkins MP, Meredith W. 1981. "Spouse similarity in newlyweds with respect to specific cognitive abilities, socioeconomic status, and education." *Behavior Genetics, 11(1):* 1-2