

# Genes are not nature: effects of socio-economic status on genetics in UK Biobank

David Hugh-Jones, Oana Borcan & Abdel Abdellaoui

2020-10-21

## Introduction

It is well known that both an individual’s genetics, and his or her environment, influence a wide range of life outcomes, including physical and mental health, income and social status, spouse choice, family formation and personality. Both popular and scientific discourse often parse these two forces as “nature” and “nurture” (e.g. Plomin 1994; Chakravarti and Little 2003).<sup>1</sup> The implicit logic is that a person’s environment is the product of his or her society; genes, on the other hand, are part of his or her nature.

At the level of the individual, this idea expresses the fact that our genetic endowment is fixed at birth and cannot be influenced by our social environment. While genes may be expressed differently over the life cycle, and importantly may interact with the environment to cause different outcomes, the genes themselves do not change. This feature allows them to be used as an “unmoved mover” to untangle social causality, in techniques like Mendelian randomization.

But as soon as we move beyond individual level, the idea that human genetics are natural becomes highly misleading. Humans inherit their genes from their parents, along with other forms of inheritance such as economic and cultural capital. Human parents, in turn, meet, form spouse pairs, bear children, and raise them, within social institutions including marriage markets and family structures. Therefore, a person’s genetic inheritance is a social and historical fact about them, and not a fact of nature, any more than their inherited wealth or social status is natural. As (Marx 1844) wrote, “History is the true natural history of man”.

Indeed, genetic assets and other kinds may be substitutable in marriage markets. (Marx 1844) also wrote “I am ugly, but I can buy the most beautiful woman.... the effect of ugliness, its repelling power, is destroyed by money”. On the opposite side of the political spectrum, (Herrnstein and Murray 1995) warned of “a merging of the cognitive elite with the affluent”. These quotations suggest that social advantages, such as wealth, income, caste or class, may have genetic consequences. If there is assortative mating between socially and genetically advantaged people, then socio-economic status may be reflected in the DNA of subsequent generations.

Figures 1 and 2 illustrate this idea using data for spouse pairs from UK Biobank, a sample of about 500,000 UK respondents born between 1935 and 1970. The X axis shows a measure of one partner’s socio-economic status: university attendance (Figure 1) or income (Figure 2). The Y axis plots the other partner’s mean polygenic score for educational attainment (PSEA). This measure is derived solely from the partner’s genetic data, and is a predictor of the number of years spent in education, as well as university attendance and IQ.<sup>2</sup> Both males and females who went to university had spouses with higher PSEA. So did males and females with higher income. Since DNA is inherited, these people’s children will also have higher PSEA.

<sup>1</sup>This holds even among scientists who wish to go beyond the nature-nurture dichotomy, e.g. (Rutter 2002).

<sup>2</sup>To minimize concerns about genetic stratification, i.e. correlations between genetics and non-genetic inherited characteristics such as culture, PSEA is residualized on the first 100 principal components of UK Biobank array data.

These figures do not prove that genetic encoding of socio-economic status is taking place. Since an individual’s own PSEA correlates with both their educational attainment, and their income, both figures could be a result of partner selection on a purely genetic basis. In this paper, we test the theory more rigorously, using environmental shocks to socio-economic status that are unlikely to be correlated with own genetics. First, we develop a simple model of assortative mating to show how parents’ social status predicts children’s genetics, and how this varies with social structure.

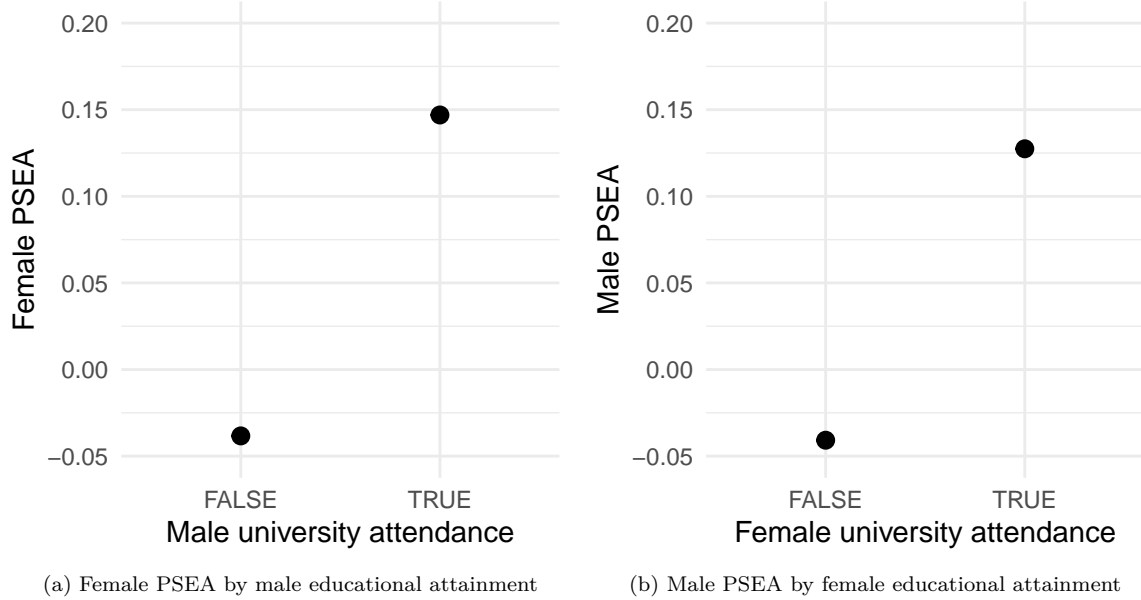


Figure 1: Social and genetic advantage among spouse pairs in UK Biobank

## Model

There is a large population, whose members have a single genetic trait  $g_i$  and a single social trait  $s_i$ , drawn from distributions  $G$  and  $S$ . The genetic trait could be, for instance, a polygenic score, which summarizes the effects of many alleles (genetic variants) at different loci. The social trait is a measure of social status. Broadly conceived, this means any trait that an individual possesses in virtue of his or her position in society, rather than as a natural fact. Caste and class are kinds of social status; so are wealth, income, education and employment.

$G$  and  $S$  are continuously distributed. Without loss of generality,  $EG = ES = 0$  and  $Var(G) = Var(S) = 1$ .<sup>3</sup> People pair according to an attractiveness function

$$A(g_i, s_i) = f((1 - k)g_i, ks_i)$$

where  $f$  is smooth and strictly increasing in both its arguments, and  $0 \leq k \leq 1$ .

The key parameter is  $k$ . This describes the working of the society’s marriage market. If  $k = 0$ , only genetics  $G$  are relevant in marriage markets, and social status  $S$  has no effect. That is, the marriage market is highly egalitarian. Conversely, if  $k = 1$ , only social status matters, to the complete exclusion of genetics. This is the equivalent of a marriage market driven only by “caste” or “class”. Realistic societies are between these extremes.

<sup>3</sup>Continuous distribution is not strictly required. All that is needed is for a set of pairs of positive measure to have different values of  $G$  and  $S$ , along a set of attractiveness curves of positive measure.

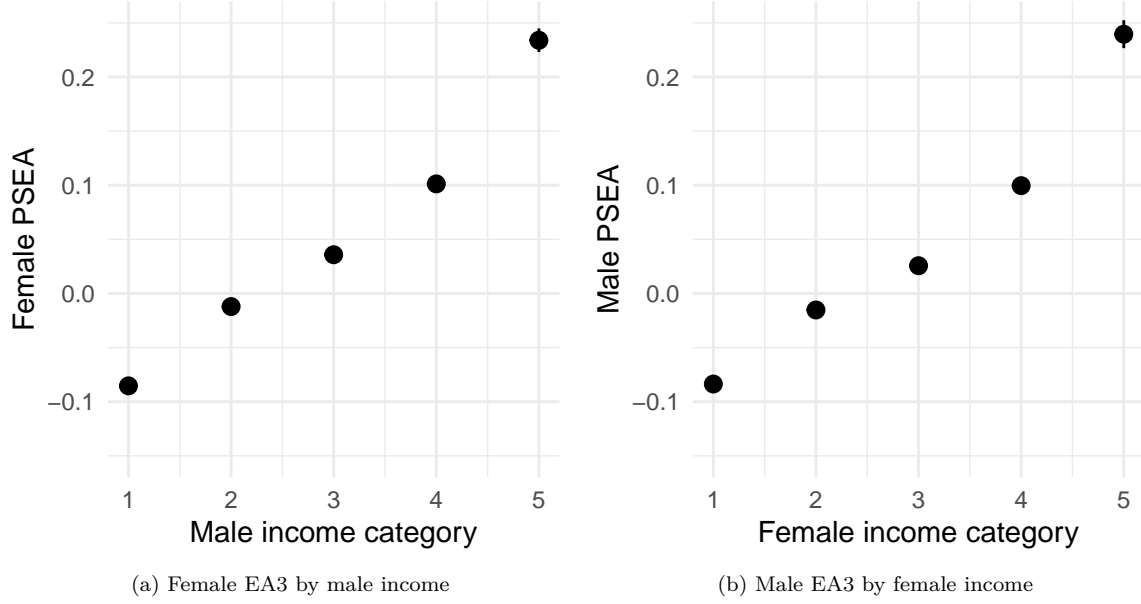


Figure 2: Social and genetic advantage among spouse pairs in UK Biobank

Figure 3 shows the intuition behind our theory. The top row shows a *caste marriage market* with  $k = 1$ . A typical pair is shown: children have intermediate values of  $G$  and  $S$  between their two parents (hollow circle). In this society pairs match only by social status; genetics plays no role. As a result, while the variance of  $G$  shrinks within each status group, genetics remain uncorrelated with social status in the children’s population distribution, shown on the right. The next row shows a purely *egalitarian marriage market* with  $k = 0$ . Parents match only by genetics and ignore social status. Again, as a result there is no correlation between genetics and social status in the children’s generation. The bottom row shows a (more realistic) intermediate society, with an intermediate value of  $k$ . Because both genetics and social status contribute to attractiveness, matched spouses typically trade them off against each other. As a result, the distribution is squeezed along the gradient of  $k$ , and  $G$  and  $S$  are correlated in the children’s generation. We next prove this formally.

If  $k = 0$ , “indifference curves” of attractiveness are vertical lines in  $(G, S)$  space. If  $k = 1$ , they are horizontal lines. If  $k \in (0, 1)$  they are downward sloping curves.

Write  $p(i)$  for  $i$ ’s partner. Pairs always have the same attractiveness.

$$A(g_i, s_i) = A(g_{p(i)}, s_{p(i)}).$$

Each pair has two children. We assume that both children  $c(d)$  of parents  $d, m$  have

$$\begin{aligned} g_{c(d)} &= \frac{g_d + g_m}{2}; \\ s_{c(d)} &= \frac{s_d + s_m}{2}. \end{aligned} \tag{1}$$

This is a strong assumption; we relax it later. For real world examples approximated by it,  $S$  could be wealth which is equally divided between the children;  $G$  could be a highly polygenic trait with many small effects. Write  $G_p, S_p$  to denote the population variables in the parents’ generation;  $G_c, S_c$  for the children’s generation.

**Proposition 1.** (i)  $Cov(G_c, S_c) \geq Cov(G_p, S_p)$ , with strict inequality if and only if  $0 < k < 1$ .

(ii) If  $corr(G_p, S_p) \geq 0$ , then  $corr(G_c, S_c) \geq corr(G_p, S_p)$ , with strict inequality if and only if  $0 < k < 1$  or  $corr(G_p, S_p) > 0$ .

*Proof.* Within each pair  $i, p(i)$  write  $d$  for the person with  $s_d > s_{p(d)}$  and  $m$  for  $p(d)$ . (Think of these as "dukes" and "milkmaids", or if you prefer "duchesses" and "tennis instructors".) If  $k < 1$ , then  $g_d < g_m$ . (If  $k = 0$ , then define  $d$  as the person with  $g_d < g_{p(d)}$ .)

We integrate over the "dukes" to calculate the covariance in the parents' generation:

$$cov(G_p, S_p) = \int \frac{1}{2}(g_d s_d + g_{p(d)} s_{p(d)}) dd.$$

For the children, the equivalent expression is

$$cov(G_c, S_c) = \int g_{c(d)} s_{c(d)} dd,$$

observing that  $EG_c = ES_c = 0$  from (1).

Take an arbitrary pair  $d, m$ . Write

$$\begin{aligned} g_d s_d &= (g_c - \Delta g)(s_c + \Delta s); \\ g_m s_m &= (g_c + \Delta g)(s_c - \Delta s) \end{aligned}$$

where

$$\begin{aligned} \Delta g &= \frac{g_m - g_d}{2} \geq 0, \text{ strictly so if and only if } k > 0; \\ \Delta s &= \frac{s_d - s_m}{2} \geq 0, \text{ strictly so if and only if } k < 1. \end{aligned}$$

Taking the average of the parents gives

$$\frac{1}{2}(g_d s_d + g_m s_m) = g_c s_c - \Delta g \Delta s.$$

This is less than  $g_c s_c$  if  $0 < k < 1$ , and equal to it if  $k = 0$  or  $k = 1$ . Plugging this into the integral shows that

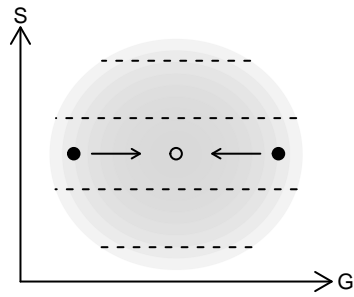
$$cov(G_p, S_p) \leq cov(G_c, S_c)$$

again with strict inequality if and only if  $0 < k < 1$ . This proves the first part. A similar argument, showing  $var(G_c) \leq var(G_p)$  and  $var(S_c) \leq var(S_p)$ , proves the second part (see the appendix). Figure 6 in the appendix shows that the condition in the second part cannot be relaxed further.

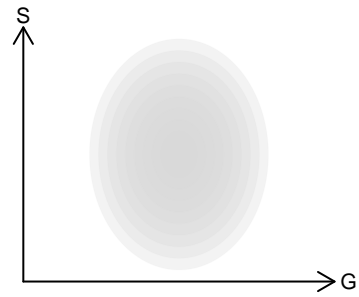
□

We view  $k = 0$  and  $k = 1$  as theoretical "ideal types". Proposition 1 therefore shows that in almost any realistic society, social status will become correlated with genetic traits which are considered attractive in the marriage markets.

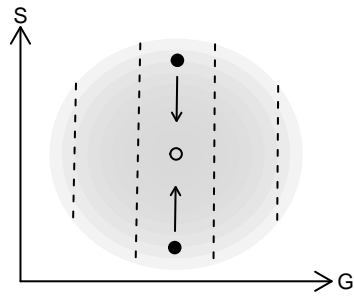
We now relax the condition that children are exactly at the mean of their parents' values for  $G$  and  $S$ . Let



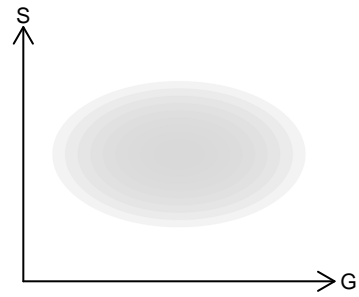
(a) Caste society ( $k = 1$ ): parents



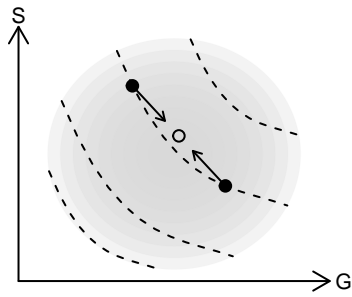
(b) Caste society: children



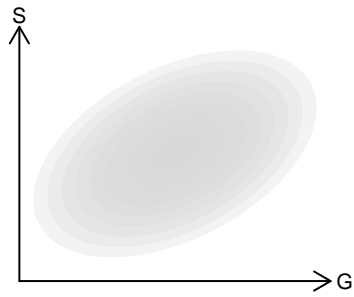
(c) Egalitarian society ( $k = 0$ ): parents



(d) Egalitarian society: children



(e) Intermediate society ( $0 < k < 1$ ): parents



(f) Intermediate society: children

Figure 3: Theory: shaded area is the population distribution. Dotted lines are attractiveness isoquants. Solid dots are example parents, transparent dots are example children. The right hand side shows the children's generation.

$$\begin{aligned} g_{c(i)} &= \bar{g}_i + \varepsilon_i^G \\ s_{c(i)} &= \bar{s}_i + \varepsilon_i^S \end{aligned}$$

where

$$\bar{g}_i = \frac{g_i + g_{p(i)}}{2}; \bar{s}_i = \frac{s_i + s_{p(i)}}{2};$$

$\varepsilon^G$  has mean 0 and variance  $\sigma_G^2$ ; and  $\varepsilon^S$  has mean 0 and variance  $\sigma_S^2$ .

**Proposition 2.** 1. If  $\sigma_G^2$  and  $\sigma_S^2$  are small enough and  $\text{corr}(G_p, S_p) \geq 0$ , then  $\text{corr}(G_c, S_c) > \text{corr}(G_p, S_p)$  for  $k \in (0, 1)$ .  
2. If  $\varepsilon^G$  and  $\varepsilon^S$  are uncorrelated with each other and with  $\bar{G}$  and  $\bar{S}$ ; and if  $G_p$  and  $S_p$  are uncorrelated, then  $\text{corr}(G_c, S_c) \geq 0$ , with strict inequality if and only if  $0 < k < 1$ .

The conditions in Proposition 2 are quite plausible. For  $G$ , they require that either variance in siblings' scores on some summary statistic is not too large, or that it is uncorrelated with the parents' scores. Both of these hold for most polygenic scores, which are additive sums of many small effects of alleles derived randomly from one or other parent. For  $S$ , the conditions would hold, for example, if  $S$  measures wealth, which is inherited not too unequally between siblings; or if wealth is inherited unequally but not in a way that correlates with  $S$  or  $G$ .

It is worth considering what kind of social arrangements would *violate* these conditions. For example, suppose that parents' combined wealth is inherited by the child with the lowest value of  $g_{c(i)}$ . This creates a negative correlation between  $s_{c(i)}$  and  $g_{c(i)}$ .

In the model, intermediate values of  $k$  drive increased covariance between genetics and social status. That naturally raises the question whether the change in covariance and/or correlation is increasing in  $k$  towards some maximum value in  $(0, 1)$ , then decreasing, i.e., whether it is quasiconcave in  $k$ . In general the answer is no, even on the assumption that

$$A(g_i, s_i) = (1 - k)g_i + ks_i$$

i.e. that indifference curves are straight lines in  $(G, S)$  space. However, quasiconcavity in  $k$  does hold when indifference curves are straight and  $(G, S)$  are jointly normally distributed. We conjecture that it also holds whenever the distribution  $(G, S)$  is log-concave or even quasiconcave.

## Discussion

The “marriage market” here is a reduced form mechanism, encompassing that makes a difference to partner choice. For example, if earned income affects attractiveness in the marriage market, then society's level of meritocracy in the labour market will correlate with the value of  $k$ : a more meritocratic labour market will allow people with low social status but high human capital (partly genetically determined) to earn more, and therefore to enter the high group.

Also, the contents of  $G$  – what counts as “good genes” in the marriage market – are themselves likely to vary across societies. For instance, standards of physical attractiveness, and other characteristics which make someone a “good match”, vary both across societies and within a society over time.

The model predicts variation in the strength of genetic lock-in. In particular, in “caste societies” where there is complete endogamy within social status groups, there is no scope for genetic lock-in, because marriage partners do not trade off genetics for social status. The model also assumes that social status is inherited

randomly from one parent, in the same way a genetic allele is inherited. This assumption can be weakened. For example, if social status is inherited deterministically from the father, then the results remain unchanged (for each pair of parents, just assume that one randomly chosen parent is the father).

Behaviour geneticists often make the point that in meritocratic societies, successful people may transmit relevant genes to their offspring. (TODO: cite relevant papers.) Like the phenomenon described here, meritocracy may therefore lead to a correlation between social status and genetics. However, genetic lock-in is a distinct, though overlapping, mechanism. Under meritocracy, certain genetic variants *cause* higher social status, and are then transmitted along with it. Indeed, genetic variations which lead to high social status will become associated with it, even in the absence of assortative mating. Conversely, this logic would not apply in non-meritocratic societies where social status is ascribed rather than earned. Indeed, (Rimfeld et al. 2018) show that genetic influences on education and occupation were lower in (non-meritocratic) Soviet Estonia than in the more meritocratic post-Soviet period.

By contrast, our mechanism applies to genetic variants that are *associated* with higher social status during the spouse matching process. They do not need to exert any causal effect on an individual’s own social status. Whereas the mechanism above requires meritocracy, but not assortative mating, our mechanism requires assortative mating, but not meritocracy. Our mechanism therefore potentially applies to a much wider range of societies, including societies where social status is wholly ascribed or inherited, such as aristocracies.

In modern societies, both assortative mating and meritocracy are likely to be at play. Genetic variants that cause (e.g.) higher income and wealth will be inherited along with components of social status such as inherited wealth. At the same time, higher social status and “good genes” will assort in the marriage market, even if that higher social status is caused by purely environmental variation. Our empirical analysis shows this latter process at work.

## Data and methods

As mentioned above, simple correlations between one partner’s social status and the other partner’s genetics do not prove that genetic lock-in is taking place. Since social status correlates with one’s own genetics, the correlation could be driven purely by assortative mating for genetics (Hugh-Jones et al. 2016). To demonstrate genetic lock-in, we therefore need a source of social status which is exogenous to genetics. One possibility is *birth order*. It is well known that earlier-born children receive more parental care and have better life outcomes (Booth and Kee 2009; Black, Devereux, and Salvanes 2011). On the other hand, all full siblings have the same *ex ante* expected genetic endowment from their parents, irrespective of their birth order.<sup>4</sup> We can therefore use birth order as a “shock” to social status.

Note that we do not claim that birth order is exogenous to all other variables. It probably isn’t: for example, it naturally correlates with parental age, and it might also be linked with the family’s economic position at the time of birth. We only claim that birth order is exogenous to any genetic variation inherited from the parents.

We use data from the UK Biobank, a study of about 500,000 individuals born between 1935 and 1970. The Biobank contains information on respondents’ genetics, derived from DNA microarrays, along with questionnaire data on health and social outcomes. We use respondents’ exact home location to infer that some are spouse pairs. Our key dependent variable is spouse’s *Polygenic Score for Educational Attainment* (PSEA). A polygenic score is a DNA-derived summary measure of genetic risk or propensity for a particular outcome, created from summing small effects of many common genetic variants, known as Single Nucleotide Polymorphisms (SNPs). We use per-SNP summary statistics from @lee2018gene, recalculated excluding UK Biobank participants, to calculate PSEA.

Our central independent variable is respondents’ birth order, i.e. their number of elder siblings plus one. For controls we use family size, i.e. their total number of siblings including themselves; month of birth; age at

---

<sup>4</sup>This might not be the case, if parents’ choice of whether to have more children is endogenous to the genetic endowment of their earlier children. We check for this below.

interview; respondents’ own PSEA; their father’s and/or mother’s age at their birth (calculated from parent’s current age, only available if the parent was still alive). Our two measures of social status are income, and university attendance.

- TODO: Abdel - details of PSEA calculation and spouse pair inference
- TODO: look at mechanisms by which birth order might affect university
- TODO: get IQ data, control for it
- TODO: overall index of social status?

Ideally, we might prefer to instrument social status with birth order. However, our measures of social status are noisy and incomplete. For example, we know whether subjects went to university, but not which university they went to, and we only have rough categorical data on household income. Birth order likely affects both these measures, and other, unmeasured dimensions of social status. So, an instrumental variables approach would probably fall foul of the exclusion restriction.

Instead, we conduct a mediation analysis, following the strategy of @heckman2013understanding. We first run regressions to confirm that birth order affects our measures of respondents’ social status (income and education). Then, we regress spouse’s PSEA on birth order, with and without controlling for social status. Under the assumption that birth order is exogenous to own genetics, these regressions identify the effect of birth order, plus other environmental variables that correlate with it, on own social status and spouse’s genetics. Also, if the estimated effect of birth order on spouse’s PSEA changes when social status is included, that is evidence that social status mediates the effect of birth order.

- TODO: clarify the empirical model, you may need help....
- TODO: estimate individual income from job SIC codes? ASHE gives data

## Results

Figures 4 and 5 show the relationship between birth order, university education and income, separately for respondents with 1-3 siblings. We test this formally in a linear regression, controlling for family size, which may be correlated with parental characteristics including genetics. Birth order is negatively associated with both university attendance and income (among respondents with 1-7 siblings: university  $p = 8.3e-64$ , income  $p = 4.27e-09$ ).

Next we run regressions of spouse PSEA on birth order. Table 1 reports the results. Subjects with 1-7 siblings are included. Column 1 controls only for family size (using dummies). As expected, higher birth order is negatively associated with spouse’s PSEA, though the estimated effect size is small. Column 2 includes the respondent’s own PSEA, as well as a quadratic in age, and dummies for birth month to control for seasonality effects. The effect size of birth order is not much changed.

Column 3 includes father’s age at birth. Within a family, later children have older parents by definition. Older parents have more life experience and may have higher income, which would presumably help later children. On the other hand, an early explanation for birth order effects was that these could be due to genetic mutations in older fathers, although more recent research has rejected this in favour of “social” explanations [(Kristensen and Bjerkedal 2007); @black2011older].<sup>5</sup> Including father’s age means we can separate the effect of father’s age from birth order. This reduces the N by a lot, since only respondents with live fathers reported the necessary data. However, the effect of birth order jumps in size, to about half the effect of respondent’s own PSEA. Meanwhile, father’s age has a positive effect. This suggests that the previous estimates mixed two opposite-signed effects: having older parents versus being later in birth order.<sup>6</sup>

<sup>5</sup>(Cochran and Harpending 2013) report that mutational load is approximately linear in father’s age, while it is constant in mothers’ age. We observe very similar results if we control for mother’s age at respondent’s birth.

<sup>6</sup>Note that parental age would not be a good independent variable for testing genetic encoding of social advantage, since it is likely to correlate with parents’ genetics.



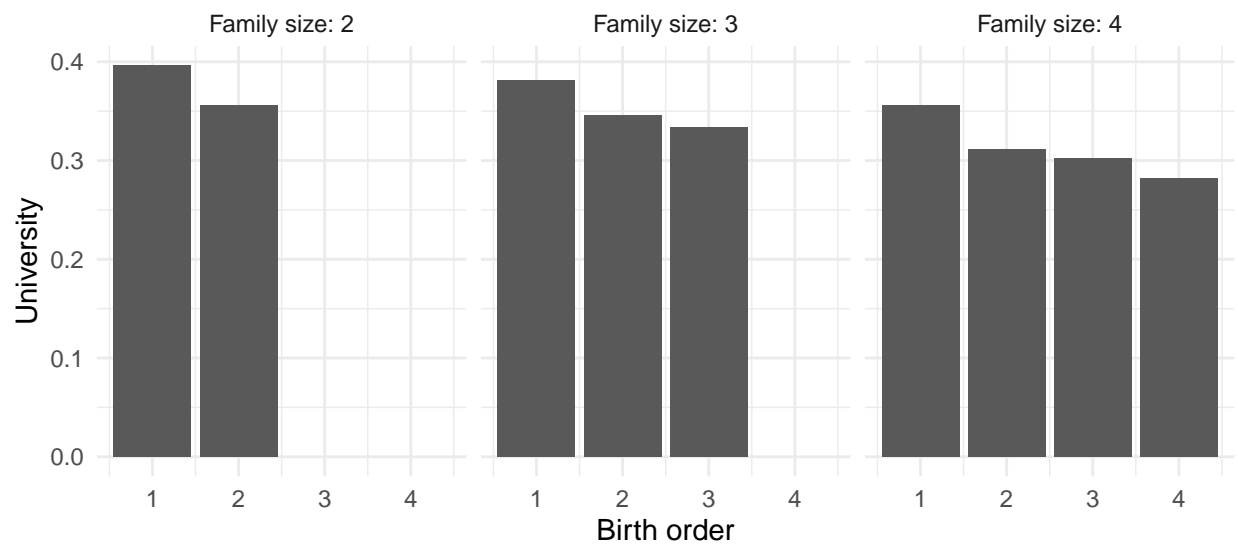


Figure 4: University attendance by birth order and family size

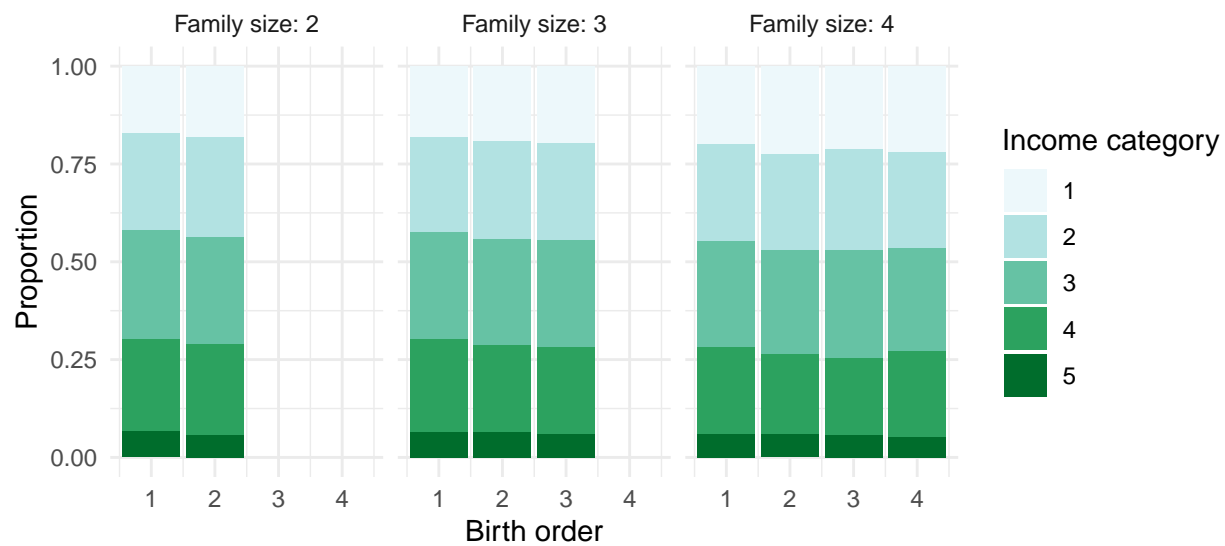


Figure 5: Income by birth order and family size

Table 1: Regressions of spouse PSEA

	(1)	(2)	(3)
Birth order	-0.0049 ** (0.0019)	-0.0046 * (0.0019)	-0.0236 *** (0.0048)
Own PSEA		0.0471 *** (0.0022)	0.0413 *** (0.0037)
Father's age at birth			0.0091 *** (0.0008)
Family size dummies	Yes	Yes	Yes
Birth month dummies	No	Yes	Yes
Age quadratic	No	Yes	Yes
N	340676	340061	88510
R2	0.001	0.004	0.004

\*\*\*  $p < 0.001$ ; \*\*  $p < 0.01$ ; \*  $p < 0.05$ . Standard errors clustered by spouse pair. Respondents with 1-7 siblings included.

- TODO: include age FTE as well as university - to check university has an “extra effect” - this slightly suggests that uni is a “marriage market” and not just granting extra skills
- TODO: enquire about job history data - does it predict current profession well?
  - otherwise we risk the critique that current profession is endogenous to spouse PSEA
- TODO: consider alternative exogenous shocks to income. For example, some professions are more “cyclical” than others wrt recessions. If we could do predicted income at age 21-25 from business cycle X profession, that might count as exogenous. (Could use an independent source to estimate evolution of incomes, e.g. GHS or BHPS)
- TODO: get IQ data from Abdel - again this controls for skills
- TODO: get 2178 “overall health” and 2188 “longstanding illness/disability/infirmity”
  - control for effect on health
- Overall idea here is to unpack the different effects of birth order, separate them from university/income.

Having tested that birth order affects spouse's PSEA, we now look for potential mediators of this effect. Despite the lower N, we continue to control for respondents' fathers' age, since this removes a confound which would bias our results towards zero.<sup>7</sup>

Table 2 shows the results. Column 1 shows the effect of birth order, using the same specification as column 3 of the previous table. (We exclude respondents without data on income or university attendance, to make comparison easier across the columns: this reduces the N slightly.) The remaining columns control for our

<sup>7</sup>The appendix reports results without controlling for father's age.

Table 2: Regressions of spouse PSEA

	(1)	(2)	(3)	(4)
Birth order	-0.0256 *** (0.0050)	-0.0136 ** (0.0050)	-0.0185 *** (0.0050)	-0.0106 * (0.0050)
University		0.1551 *** (0.0073)		0.1212 *** (0.0076)
Income			0.0707 *** (0.0033)	0.0557 *** (0.0034)
Own PSEA	0.0421 *** (0.0038)	0.0266 *** (0.0039)	0.0344 *** (0.0038)	0.0239 *** (0.0039)
Father's age at birth	0.0092 *** (0.0009)	0.0063 *** (0.0009)	0.0079 *** (0.0009)	0.0059 *** (0.0009)
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Age quadratic	Yes	Yes	Yes	Yes
N	81838	81838	81838	81838
R2	0.004	0.010	0.010	0.013
logLik	-116303.477	-116078.343	-116072.942	-115945.248
AIC	232652.953	232204.686	232193.884	231940.496

\*\*\*  $p < 0.001$ ; \*\*  $p < 0.01$ ; \*  $p < 0.05$ . Standard errors clustered by spouse pair.

social status variables, which are potential mediators of birth order effects. Column 2 includes university attendance. Column 3 includes income. Column 4 includes both. We estimate the percentage decrease in the effect of birth order across the columns, along with 95% confidence intervals for this figure, by running bootstraps ( $N = 199$ ).<sup>8</sup> Including university attendance alone reduces the effect of birth order by 50.4% (CI 35.2% – 78.5%). Including income alone reduces the effect of birth order by 29.6% (CI 20.1% – 46.0%). Including both decreases the effect by 62.7% (CI 43.7% – 98.3%).

Our next regressions split up the data into subsets. Cultural stereotypes often assume that the link between status and genes is not symmetric across the genders, for example, that males with high social status are particularly likely to marry attractive spouses. To test this, we separately regress female spouses' PSEA on male birth order, and male spouses' PSEA on female birth order. We also rerun regressions among the subset of individuals who had children. A significant result here will confirm that the association between status and genetics is carried over into the next generation.

Table 3 shows the results.

<sup>8</sup>The sample percentage decrease calculated from the figures in Table 2 is not the correct estimator, since  $E(X/Y)$  is not equal to  $EX/EY$ .

- TODO: describe results
- TODO: very few couple pairs (< 50%) report the same number of children. Why?
  - equally likely to be males or females reporting more.

Table 3: Regressions of spouse PSEA: subsets

	Females	Females	Males	Males	With children	With children
Birth order	-0.019 *** (0.006)	-0.008 * (0.004)	-0.030 *** (0.007)	-0.018 ** (0.006)	-0.032 *** (0.006)	-0.018 ** (0.006)
University		0.160 *** (0.012)		0.149 *** (0.014)		0.169 *** (0.009)
Father's age at birth	0.010 *** (0.001)	0.006 *** (0.001)	0.009 ** (0.003)	0.006 * (0.002)	0.011 *** (0.001)	0.008 *** (0.001)
Own PSEA	0.040 *** (0.004)	0.025 *** (0.004)	0.044 *** (0.002)	0.028 *** (0.003)	0.044 *** (0.005)	0.027 *** (0.005)
N	40362	40362	41476	41476	53001	53001
R2	0.004	0.010	0.005	0.010	0.006	0.012
logLik	-57284.814	-57164.809	-59002.723	-58898.765	-75378.737	-75208.098
AIC	114615.628	114377.619	118051.445	117845.530	150803.473	150464.197

\*\*\*  $p < 0.001$ ; \*\*  $p < 0.01$ ; \*  $p < 0.05$ . Standard errors for columns 5 and 6 clustered by spouse pair.

## Robustness

Although all children of the same parents have the same polygenic scores in expectation, it could still be possible that genetics correlates with birth order within the sample. This could happen in three ways.

First, siblings with high birth order will typically come from larger families than those with low birth order, and parents of different-sized families are likely to differ systematically on many dimensions, including genetics. We controlled for this by including a full set of family size dummies in the regression.

Second, there could be ascertainment bias. For example, if later siblings with high PSEA, and earlier siblings with low PSEA, are more likely to enter the sample, then this would bias our results.

Thirdly, parents might select family size on the basis of genetics. For example, suppose that if the first child had a phenotype reflecting a high PSEA, parents are more likely to have a second child. Then within the subset of two-child families, first children would have higher-than-average PSEA while second children would not.

We check this by regressing 33 different polygenic scores on birth order, controlling for family size.<sup>9</sup> No scores were significant at  $p < 0.10/33$ . 3 scores were significant at  $p < 0.10$  (body mass index, conscientiousness,

<sup>9</sup>Polygenic scores were residualized on the first principal components of the genetic data. Scores were for: ADHD, age at menarche, age at menopause, agreeableness, age at smoking initiation, alcohol use, Alzheimer's, autism, bipolarity, BMI,

and neuroticism). Coefficients were never greater than 0.01 of a standard deviation. Table 5 in the appendix reruns regressions controlling for these scores. Results are almost unchanged.

## Conclusion

Behaviour geneticists have pointed out that in meritocratic societies, genetics and social status will be correlated, because some genetic variants will lead to success in the labour market. We argue that causality can also go the other way. If social status and genetics both matter in marriage markets, then people with high social status will attract partners with “good genes”, and the two will become associated in the next generation.

Our empirical analysis shows that in a modern Western democracy, earlier-born children had spouses with higher PSEA. Thus, an environmental shock has effects on the genetics of people’s spouses and becomes encoded in the genetics of their children. We also provided evidence that these effects are mediated by social status: income and education.

Social scientists have tended to assume that social status is unrelated to “natural” differences between individuals, including genetic differences. This view dates back at least to (Hobbes 1651). By contrast, it has been widely believed in many societies that innate traits do vary across social classes. The ancient Greeks described the social elite as *kaloi kagathoi* (“fine and good”), while the Roman nobility were the *optimates* (“best”).<sup>10</sup> This belief has been explained by the human tendency to believe in a just world (Furnham 1993), or as an ideology promoted by the dominant class (e.g. Gramsci 1971). However, it may also simply have been a recognition of (social) reality. In other words, the belief that elites are taller, stronger, better-looking, etc. is not much different from the belief that elites are richer and more powerful, and is likely to be held for fairly similar reasons. The belief that human traits are inherited across generations is probably quite widespread across different cultures (Moya, Boyd, and Henrich 2015): this, plus the observation that status matters in marriage markets, provides a cognitive basis for reasoning to the conclusion of class differences. Alternatively, the belief might simply come from empirical observation.

In our model, the association between social status and genetic variation depends on the structure of the society’s marriage market. The association is weaker when marriage markets are very socially egalitarian, with marriage pairing driven only by genetics, or very inegalitarian, with pairings driven only by status. Testing this, by comparison across societies or over time, is beyond our scope here, but we see it as a good area for future work. Marriage markets may play an important role in explaining how social inequality varies across different societies. More broadly, the correlation between genetics and social status does not undermine our concern with social inequality; rather, since it is a social outcome, it is itself one dimension of social inequality. Genes, like other kinds of resource, are distributed unequally by social systems.

Over history, the marriage market mechanism has probably operated in a much greater variety of societies than the labour market mechanism, because societies in which status can be earned are relatively rare compared to those in which status is inherited. (TODO: cite WEIRD? Or “status vs contract”?) As a result, we would predict long-standing correlations between social status and causal genetic variants. This prediction could be tested on ancient DNA data.

Our analysis also has implications for the practice of controlling polygenic scores by partialling out principal components of genetic data. This is done so as to avoid confounding the effects of genetic variation with social stratification. However, insofar as the geneticists’ concept of stratification (mating which is non-random with respect to genetics) overlaps with the sociological concept of stratification (a hierarchical ranking), stratification will predictably associate with causally relevant genetic variants. For this reason, we

---

body fat, caffeine consumption, cannabis (ever vs. never), cognitive ability, conscientiousness, coronary artery disease, smoking (cigarettes per day), type II diabetes, drinks per week, educational attainment (EA2 and EA3), anorexia, extraversion, height, hip circumference, major depressive disorder, neuroticism, openness, smoking cessation, schizophrenia, smoking initiation, waist circumference, and waist-to-hip ratio.

<sup>10</sup>The appendix contains a selection of relevant historical quotations.

would expect principal components to contain real information about causally relevant variants. So, it is an empirical question whether controlling for principal components improves or weakens the predictive power of polygenic scores. Within-family analyses could resolve this question.

In popular media, and sometimes by social scientists, genetics are often thought of as an external, “natural” constraint on social structure. This is one reason why behaviour genetics remains a controversial topic. The broadest message of our research is that individual genetics are a social outcome like any other. They can even be viewed as another form of capital, alongside human, social and cultural capital: a resource to be sought, accumulated and competed over. The analysis of this kind of capital is an exciting area for further research, which collaborating teams of social scientists and geneticists will be well-placed to exploit.

## Appendix

### Second part of Proposition 1

*Proof.* Write

$$\text{corr}(G_j, S_j) = \frac{\text{cov}(G_j, S_j)}{\sqrt{\text{var}(G_j)\text{var}(S_j)}} \text{ for both generations } j \in \{p, c\}. \quad (2)$$

where

$$\begin{aligned} \text{var}(G_p) &= \frac{1}{2} \int g_d^2 + g_{p(d)}^2 dd; \\ \text{var}(G_c) &= \int g_{c(d)}^2 dd. \end{aligned}$$

Much as before,

$$\begin{aligned} g_d^2 + g_m^2 &= (g_c - \Delta g)^2 + (g_c + \Delta g)^2 \\ &= 2g_c^2 + 2(\Delta g)^2 \\ &\geq 2g_c^2. \end{aligned}$$

This shows that  $\text{var}(G_c) \leq \text{var}(G_p)$  and a similar argument shows  $\text{var}(S_c) \leq \text{var}(S_p)$ . Thus the covariance is higher (and positive) in the children’s generation, while the variances are lower. Combining these ensures that

$$\text{corr}(G_c, S_c) \geq \text{corr}(G_p, S_p).$$

Since for any  $k$ , either  $\text{var}(G_c) < \text{var}(G_p)$  or  $\text{var}(S_c) < \text{var}(S_p)$ , the only way to get strict equality for the above is if  $k \in \{0, 1\}$  and  $\text{cov}(G_c, S_c) = \text{cov}(G_p, S_p) = 0$ .

□

To show that the condition in the second part cannot be relaxed further, consider the distribution in Figure 6. There is negative correlation in the parents’ generation (the shaded area). If  $k = 1$  or is close enough to 1, then assortative mating along the dotted lines will reduce the variance of  $S$  along those lines, pushing the distribution towards the darker central area, without affecting the covariance. This will make the correlation more negative. After repeated generations the horizontal variance within values of  $G$  will almost disappear and the correlation will approach -1.

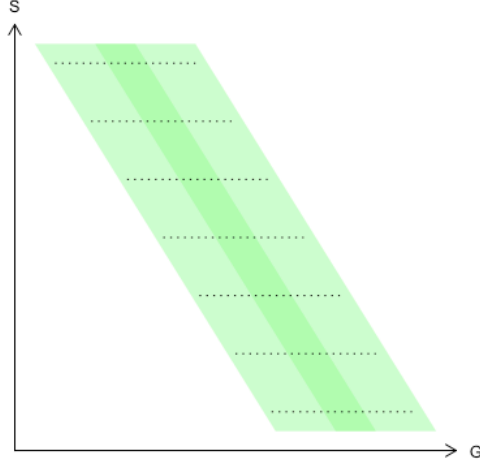


Figure 6: Correlation counterexample

## Proposition 2

*Proof.* Note that in proposition 1, we took  $g_{c(i)} = \bar{g}_i$  and  $s_{c(i)} = \bar{s}_i$ . Write

$$\begin{aligned} \text{cov}(G_c, S_c) &= \text{cov}(\bar{G} + \varepsilon^G, \bar{S} + \varepsilon^S) \\ &= \text{cov}(\bar{G}, \bar{S}) + \text{cov}(\varepsilon^G, \bar{S}) + \text{cov}(\bar{G}, \varepsilon^S) + \text{cov}(\varepsilon^G, \varepsilon^S). \end{aligned} \quad (3)$$

For any  $X$  and  $Y$ ,  $\text{cov}(X, Y)$  is bounded by  $\sqrt{\text{var}(X)\text{var}(Y)}$ . Plugging  $\sigma_G^2$  and  $\sigma_S^2$  into this formula shows that under condition 1,  $\text{cov}(G_c, S_c)$  will be arbitrarily close to  $\text{cov}(\bar{G}, \bar{S})$ . Similarly, writing

$$\text{var}(G_c) = \text{var}(\bar{G}) + \text{var}(\varepsilon^G) + 2\text{cov}(\bar{G}, \varepsilon^G)$$

shows that  $\text{var}(G_c)$  will approach  $\text{var}(\bar{G})$  as  $\sigma_G^2$  grows small, and similarly for  $\text{var}(S_c)$ . Plugging these facts into (2) shows that  $\text{corr}(G_c, S_c)$  approaches  $\text{corr}(\bar{G}, \bar{S})$  as  $\sigma_G^2$  and  $\sigma_S^2$  grow small. Proposition 1 then shows  $\text{corr}(\bar{G}, \bar{S}) < \text{corr}(G_p, S_p)$  for  $k \in (0, 1)$ .

Under condition 2,  $\text{cov}(G_c, S_c) = \text{cov}(\bar{G}, \bar{S})$  since the last three terms of the sum in (3) are zero. Then since

$$\text{cov}(\bar{G}, \bar{S}) \geq \text{cov}(G_p, S_p) = 0$$

with strict inequality iff  $k \in (0, 1)$ , the covariance signs the correlation.

□

## Robustness checks

Table 4: Regressions of spouse PSEA, without controls for father's age at respondent's birth

	(1)	(2)	(3)	(4)
Birth order	-0.0009 (0.0045)	0.0036 (0.0044)	0.0027 (0.0044)	0.0054 (0.0044)
University		0.1633 *** (0.0072)		0.1284 *** (0.0075)
Income			0.0729 *** (0.0033)	0.0564 *** (0.0034)
Own EA3	0.0462 *** (0.0038)	0.0286 *** (0.0039)	0.0376 *** (0.0038)	0.0257 *** (0.0039)
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Age quadratic	Yes	Yes	Yes	Yes
N	81838	81838	81838	81838
R2	0.003	0.009	0.009	0.013
logLik	-116360.234	-116104.671	-116114.267	-115968.117
AIC	232764.468	232255.342	232274.533	231984.234

\*\*\*  $p < 0.001$ ; \*\*  $p < 0.01$ ; \*  $p < 0.05$ . Standard errors clustered by spouse pair.



Table 5: Regressions of spouse PSEA with controls for polygenic scores

	(1)	(2)	(3)	(4)
Birth order	-0.0253 *** (0.0050)	-0.0134 ** (0.0050)	-0.0183 *** (0.0050)	-0.0104 * (0.0050)
University		0.1550 *** (0.0073)		0.1211 *** (0.0076)
Income			0.0706 *** (0.0033)	0.0556 *** (0.0034)
Father's age at birth	0.0091 *** (0.0009)	0.0062 *** (0.0009)	0.0078 *** (0.0009)	0.0058 *** (0.0009)
Own PSEA	0.0393 *** (0.0039)	0.0239 *** (0.0039)	0.0317 *** (0.0039)	0.0213 *** (0.0039)
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Age quadratic	Yes	Yes	Yes	Yes
Polygenic score controls	Yes	Yes	Yes	Yes
N	81838	81838	81838	81838
R2	0.005	0.010	0.010	0.013

\*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05. Standard errors clustered by spouse pair.

## Quotations on natural inequality

...your face and figure have nothing of the slave about them, and proclaim you of noble birth.

– *Odyssey*, Odysseus to Laertes

Citizens, we shall say to them in our tale, you are brothers, yet God has framed you differently. Some of you have the power of command, and in the composition of these he has mingled gold, wherefore also they have the greatest honour; others he has made of silver, to be auxiliaries; others again who are to be husbandmen and craftsmen he has composed of brass and iron; and the species will generally be preserved in the children. But as all are of the same original stock, a golden parent will sometimes have a silver son, or a silver parent a golden son.

– Plato *Republic*

Nature would like to distinguish between the bodies of freemen and slaves, making the one strong for servile labor, the other upright, and although useless for such services, useful for political life in the arts both of war and peace. But the opposite often happens – that some have the souls and others have the bodies of freemen.

– Aristotle *Politics*

Sons have no richer endowment than the quality

A noble and brave father gives in their begetting.

– Euripides *Heracleidae*

His looks are full of peaceful majesty,

His head by nature fram'd to wear a crown,

His hands to wield a sceptre....

– Shakespeare *Henry VI Part 3*

A daughter of a green Grocer, walks the Streets in London dayly with a baskett of Cabbage Sprouts, Dandelions and Spinage on her head. She is observed by the Painters to have a beautiful Face, an elegant figure, a graceful Step and a debonair. They hire her to Sitt. She complies, and is painted by forty Artists, in a Circle around her. The Scientific Sir William Hamilton outbids the Painters, Sends her to Schools for a genteel Education and Marries her. This Lady not only causes the Tryumphs of the Nile of Copenhagen and Trafalgar, but Seperates Naples from France and finally banishes the King and Queen from Sicilly. Such is the Aristocracy of the natural Talent of Beauty.

– John Adams to Thomas Jefferson, on Emma Hamilton

## References

- Black, Sandra E, Paul J Devereux, and Kjell G Salvanes. 2011. “Older and Wiser? Birth Order and Iq of Young Men.” *CESifo Economic Studies* 57 (1): 103–20.
- Booth, Alison L, and Hiau Joo Kee. 2009. “Birth Order Matters: The Effect of Family Size and Birth Order on Educational Attainment.” *Journal of Population Economics* 22 (2): 367–97.
- Chakravarti, Aravinda, and Peter Little. 2003. “Nature, Nurture and Human Disease.” *Nature* 421 (6921): 412–14.
- Cochran, Gregory, and Henry Harpending. 2013. “Paternal Age and Genetic Load.” *Human Biology* 85 (4): 515–28.
- Furnham, Adrian. 1993. “Just World Beliefs in Twelve Societies.” *The Journal of Social Psychology* 133 (3): 317–29.

- Gramsci, Antonio. 1971. *Selections from the Prison Notebooks*. Lawrence; Wishart London.
- Herrnstein, Richard, and Charles Murray. 1995. "The Bell Curve: Intelligence and Class Structure in American Life." In *National Forum*, 75:46–48. 2. Honor Society of Phi Kappa Phi.
- Hobbes, Thomas. 1651. *Leviathan*.
- Hugh-Jones, David, Karin JH Verweij, Beate St Pourcain, and Abdel Abdellaoui. 2016. "Assortative Mating on Educational Attainment Leads to Genetic Spousal Resemblance for Polygenic Scores." *Intelligence* 59: 103–8.
- Kristensen, Petter, and Tor Bjerkedal. 2007. "Explaining the Relation Between Birth Order and Intelligence." *Science* 316 (5832): 1717–7.
- Marx, Karl. 1844. "Economic and Philosophical Manuscripts." *Early Writings* 333.
- Moya, Cristina, Robert Boyd, and Joseph Henrich. 2015. "Reasoning About Cultural and Genetic Transmission: Developmental and Cross-Cultural Evidence from Peru, Fiji, and the United States on How People Make Inferences About Trait Transmission." *Topics in Cognitive Science* 7 (4): 595–610.
- Plomin, Robert. 1994. *Genetics and Experience: The Interplay Between Nature and Nurture*. Sage Publications, Inc.
- Rimfeld, Kaili, Eva Krapohl, Maciej Trzaskowski, Jonathan R. I. Coleman, Saskia Selzam, Philip S. Dale, Tonu Esko, Andres Metspalu, and Robert Plomin. 2018. "Genetic Influence on Social Outcomes During and After the Soviet Era in Estonia." *Nature Human Behaviour* 2 (4): 269–75. <https://doi.org/10.1038/s41562-018-0332-5>.
- Rutter, Michael. 2002. "Nature, Nurture, and Development: From Evangelism Through Science Toward Policy and Practice." *Child Development* 73 (1): 1–21.