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

David Hugh-Jones, Oana Borcan & Abdel Abdellaoui

2020-11-04

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).[1] 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".

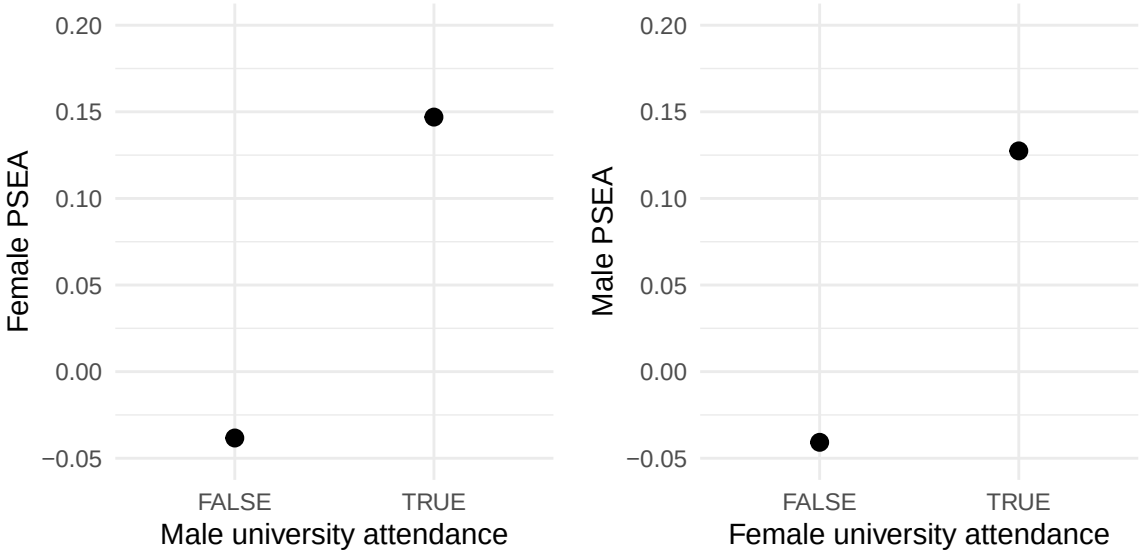
Indeed, genetic assets and other kinds may be substitutable in marriage markets. (Marx 1844) also wrote "I am ugly, but I can buy the most beautiful woman…. the effect of ugliness, its repelling power, is destroyed by money". On the opposite side of the political spectrum, (Herrnstein and Murray 1995) 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.[2] 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.

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

[2] 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.



(a) Female PSEA by male educational attainment

(b) Male PSEA by female educational attainment

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

```
## Warning: Removed 26251 rows containing non-finite values (stat_summary).
```

```
## Warning: Removed 25482 rows containing non-finite values (stat_summary).
```

Model

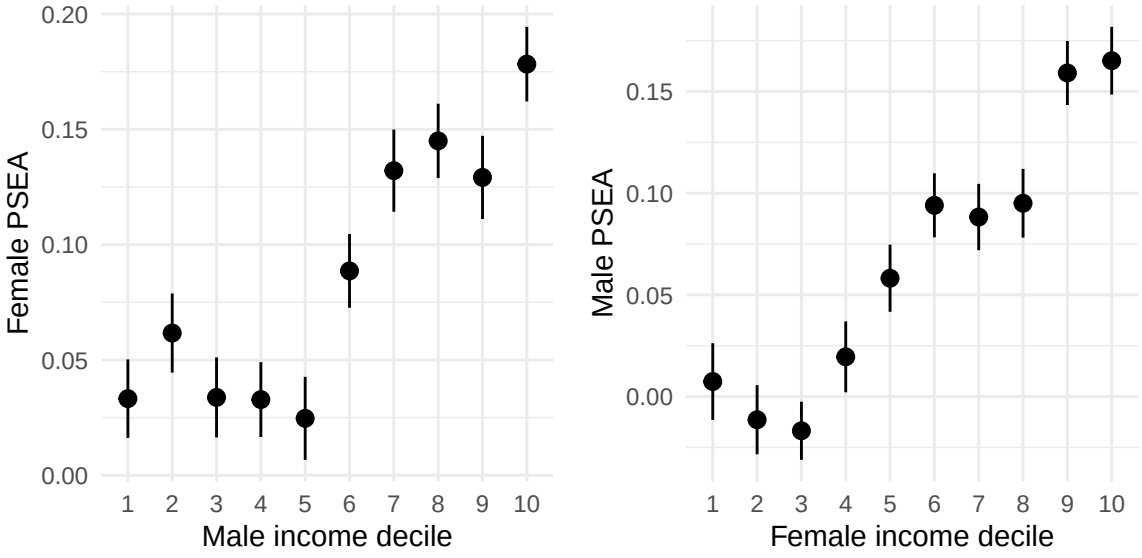
There is a large population, whose members have a single genetic trait g_i and a single social trait s_i, drawn from distributions G and S. The genetic trait could be, for instance, a polygenic score, which summarizes the effects of many alleles (genetic variants) at different loci. The social trait 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 $Var(G) = Var(S) = 1$.[3] 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 $0 \leq k \leq 1$.

[3]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.



(a) Female EA3 by male income

(b) Male EA3 by female income

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

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.

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:

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

For the children, the equivalent expression is

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

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

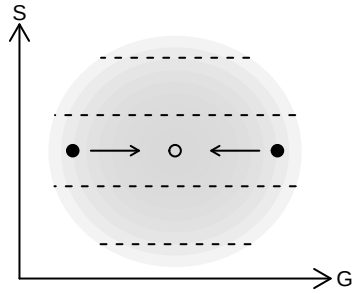
$$\frac{1}{2}(g_d s_d + g_m s_m) = g_c s_c - \Delta g \Delta s.$$

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

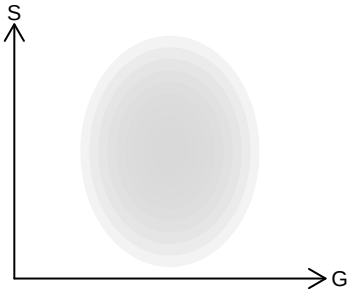
$$cov(G_p, S_p) \leq cov(G_c, S_c)$$

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.

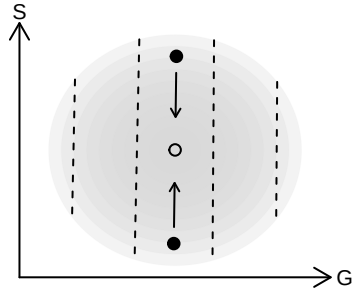
□



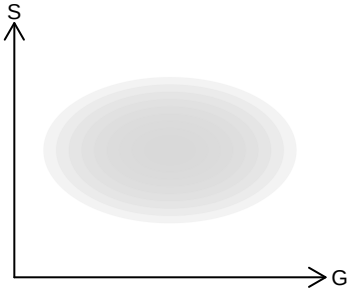
(a) Caste society ($k = 1$): parents



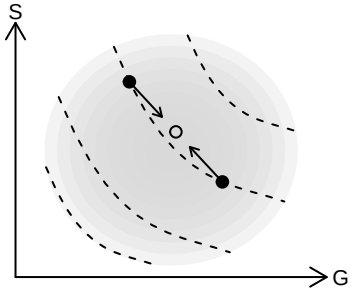
(b) Caste society: children



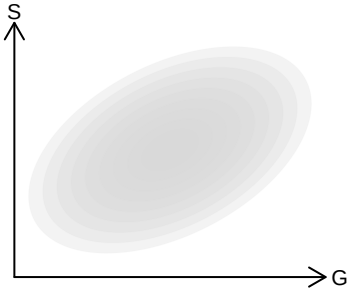
(c) Egalitarian society ($k = 0$): parents



(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.

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

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

where

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

ε^G has mean 0 and variance σ_G^2; and ε^S has mean 0 and variance σ_S^2.

Proposition 2. *1. If σ_G^2 and σ_S^2 are small enough and $corr(G_p, S_p) \geq 0$, then $corr(G_c, S_c) > corr(G_p, S_p)$ for $k \in (0, 1)$.*

2. If ε^G and ε^S are uncorrelated with each other and with $\bar{G}$ and $\bar{S}$; and if G_p and S_p are uncorrelated, then $corr(G_c, S_c) \geq 0$, 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

- TODO: discuss relationship with linkage disequilibrium between genes, caused by assortative mating. Hunt for any pre-existing theory in the genetics literature.

The "marriage market" here is a reduced form mechanism, encompassing everything 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 match with more attractive partners.

The contents of both S, social status, and G – what counts as "good genes" in the marriage market – are likely to vary across societies. S could encompass variables like social class or caste; ethnic identity in "ranked" ethnic systems; or in modern societies, *socio-economic status* (SES), i.e. variables affecting a person's position in the labour market. Regarding G, 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).

It is well known that genetics, e.g. PSEA, correlate with SES in modern societies (e.g. Krapohl and Plomin 2015). A standard explanation for this is that in meritocratic societies, successful people may transmit relevant genes to their offspring, along with their status (Belsky et al. 2018). Our mechanism is different from this, though not mutually exclusive. In the standard model, certain genetic variants *cause* higher SES, and are then transmitted along with it. Indeed, genetic variations which lead to high 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 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.

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.

Related literature

- Idea that assortative mating might increase inequality is widely known, but it is usually put in terms of assortation by IQ or by genetics alone.
- Could the result be thought of as a kind of population stratification?
- (Halsey 1958) showed in a two-class model that social mobility combined with assortative mating might increase the association between genetics and social class. But NB, the assortation is completely random within classes. So effectively this is a society with $k = 1$, but some social mobility.
-
- (Belsky et al. 2018) offer three reasons for the association between education-linked genetics and SES, but do not consider our mechanism.

Data and methods

To test the theory, 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. Using respondents' exact home location, we categorize some respondents as 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 (Lee et al. 2018), recalculated excluding UK Biobank participants, to calculate PSEA. We focus on PSEA, rather than other polygenic scores, because educational attainment plays a key role in human mate search. People are attracted to educated potential partners (Buss and Barnes 1986; Belot and Francesconi 2013). Spouse pairs often have similar levels of educational attainment, as well as similar PSEA (Vandenberg 1972; Schwartz and Mare 2005; Greenwood et al. 2014; Hugh-Jones et al. 2016).

As we said above, simple correlations between one partner's social status and the other partner's genetics do not prove that genetic encoding 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 need a source of social status which is exogenous to genetics. Also, the link between social status and spouse genetics is likely to be noisy, for three reasons: first, measures of genetic variation such as polygenic scores contain a large amount of error; second, causal mechanisms behind variation in social status are likely to be noisy; third, to paraphrase (Shakespeare 1595), the spouse matching process is notoriously unpredictable. So, we need a large N to give us sufficient power. This rules out time-limited shocks such as e.g. changes to the school leaving age.

One possibility is *birth order*. It is 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. For example, siblings' expected polygenic score is equal to the mean of their parents' polygenic scores.[4] We can therefore use birth order as a "shock" to social status. We do not claim that birth order is exogenous to all other variables. 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 genetic variation.

Our main 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 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). To test whether birth order effects are mediated by SES, we use two measures: income, and university attendance. Current income is a direct measure of SES, while university attendance is a predictor of income over the entire life course.

- 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 use birth order as an instrument for SES. 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 SES. So, an instrumental variables approach would probably fall foul of the exclusion restriction.

[4] Although genetic variation is randomly assigned to children at birth, genetics and birth order could be dependent if parents' choice of whether to have more children is endogenous to the genetic endowment of their earlier children. We check for this below.

Instead, we conduct a mediation analysis, following the strategy of (Heckman, Pinto, and Savelyev 2013). We first confirm statistically that birth order affects our measures of respondents' SES (income and education). Then, we regress spouse's PSEA on birth order, with and without controlling for SES. 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 SES is controlled for, that is evidence that SES mediates the effect of birth order.

Decomposition of the Birth Order effect on spouse genetics

We follow Heckman, Pinto, and Savelyev (2013) to decompose the aggregate treatment effect into components due to observed and unobserved proximate channels affected by the treatment. Our aim is to estimate the effect of socio-economic status (as an effect of the random assignment of birth order) on spouse PSEA.

The reduced-form regression of spouse PSEA on birth order (henceforth BO) is:

$$Y = \delta + \alpha \cdot BO + \beta \cdot \mathbf{X} + \varepsilon, \tag{2}$$

where $\mathbf{X}$ is a vector of predetermined controls (own year of birth, own PSEA, father or mother's year of birth).

Assume BO is a binary treatment variable (e.g. indicator for first born). Then, the observed outcome is:

$$Y = BO \cdot Y_1 + (1 - BO) \cdot Y_0 \tag{3}$$

Y_d with $d \in \{0,1\}$ are the counterfactual outcomes for the first and second born, respectively. Given d, spouse PSEA is assumed to be independent across observations, conditional on predetermined controls, which are assumed not to be affected by BO.

Let θ_d be a set of proximate outcomes determined by BO, which account (at least in part) for the BO treatment on spouse PSEA. We can think of θ_d as all the channels by which BO affects attractiveness in the marriage market, including socio-economic status (SES), health, cognitive and non-cognitive skills. Similar to (3) we define $\theta = BO \cdot theta_1 + (1 - BO) \cdot \theta_0$. We are mainly interested in estimating the effect of SES on spouse PSEA.

Our linear model is:

$$Y_d = \kappa_d + \alpha_d \cdot \theta_d + \beta_d \cdot \mathbf{X} + \tilde{\varepsilon_d}, \tag{4}$$

We can simplify the model if we assume that the effects of $\mathbf{X}$ do not differ by treatment, i.e. $\beta_0 = \beta_1$. $\tilde{\varepsilon_d}$ is a mean-zero residual assumed independent of θ_d and $\mathbf{X}$.

We can break down the set of investments θ_d into SES and other measured investments, and investments or skills we cannot measure.

- TODO: Find adequate terminology for the set of proximate outcomes/investments. Maybe gains?
- TODO: bold the alpha and theta and beta vectors in eq 3.

$$Y_d = \tau_d + \sum_{j \in J_m} \alpha_d^j \cdot \theta_d^j + \beta \cdot \mathbf{X} + \varepsilon_d, \tag{5}$$

Where $\tau_d = \kappa_d + \sum_{j \in J_u} \alpha_d^j \cdot E(\theta_d^j)$ and $\varepsilon_d = \tilde{\varepsilon_d} + \sum_{j \in J_u} \alpha_d^j \cdot (\theta_d^j - E(\theta_d^j))$ and J_m and J_u are the index sets of mediators of attractiveness which are measured and unmeasured.

We assume differences in unmeasured investments due to BO are independent of $\mathbf{X}$. We also assume that $\alpha_0 = \alpha_1$ (we can test this if the measured and unmeasured investments gains from BO are independent in both treatment regimes).

With these assumptions, substituting equation (5) into (3) we obtain:

$$Y = \tau_0 + \tau \cdot BO + \sum_{j \in J_m} \alpha^j \cdot \theta^j + \beta \cdot \mathbf{X} + \varepsilon, \tag{6}$$

Where $\tau = \tau_1 - \tau_0$ is the contribution of unmeasured variables to average treatment effects, $\varepsilon = BO \cdot \varepsilon_1 + (1 - BO) \cdot \varepsilon_0$ is a zero-mean error term, and $\theta^j = BO \cdot \theta_1^j + (1 - BO) \cdot \theta_0^j$, $j \in J_m$ denote the investments that we can measure.

Estimating (6) by OLS will generate unbiased estimates of α^j, $j \in J_m$ if θ^j is measured without error and is uncorrelated with the error term ε. Since ε contains both individual disturbances $\tilde{\varepsilon}$ and differences in unmeasured investments due to BO, the identifying assumptions that need to hold for unbiased OLS estimates are:

1. The measured investments (specifically SES) should be independent of unmeasured investments generated by BO. Failing this, the estimates α^j will be conflated with the effects of unmeasured investments.
2. The measured investments should be uncorrelated with other shocks $\tilde{\varepsilon}$.

The overall treatment effect can then be decomposed as follows:

$$E(Y_1 - Y_0) = \tau_1 - \tau_0 + \sum_{j \in J_m} \alpha^j \cdot E(\theta_1^j - \theta_0^j), \tag{6}$$

Where $\tau_1 - \tau_0$ is the unmeasured component of the treatment effect and $\sum_{j \in J_m} \alpha^j \cdot E(\theta_1^j - \theta_0^j)$ is the treatment effect due to measured investments.

By running a least square regression of (6), we can estimate $\tau_1 - \tau_0$. If assumption 1. above holds, the part of the BO treatment effect on spouse PSEA due to measured investments can be constructed using the estimated α^j and the effects of treatment on measured investments (from an OLS regression of measured variables on the BO treatment).

- TODO:clarify how this ports to the case of discrete treatment with multiple values. OR do regressions with first born vs later born (as first-born is anyway the most different in terms of university and income)
- TODO:We need to discuss the plausibility of these assumptions and what we can do to test it.
- TODO: estimate individual income from job SIC codes? ASHE gives data
- TODO: add year-of-birth dummies for cohort effects

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 $p = 8.3\text{e-}64$, income $p = 1.04\text{e-}13$).

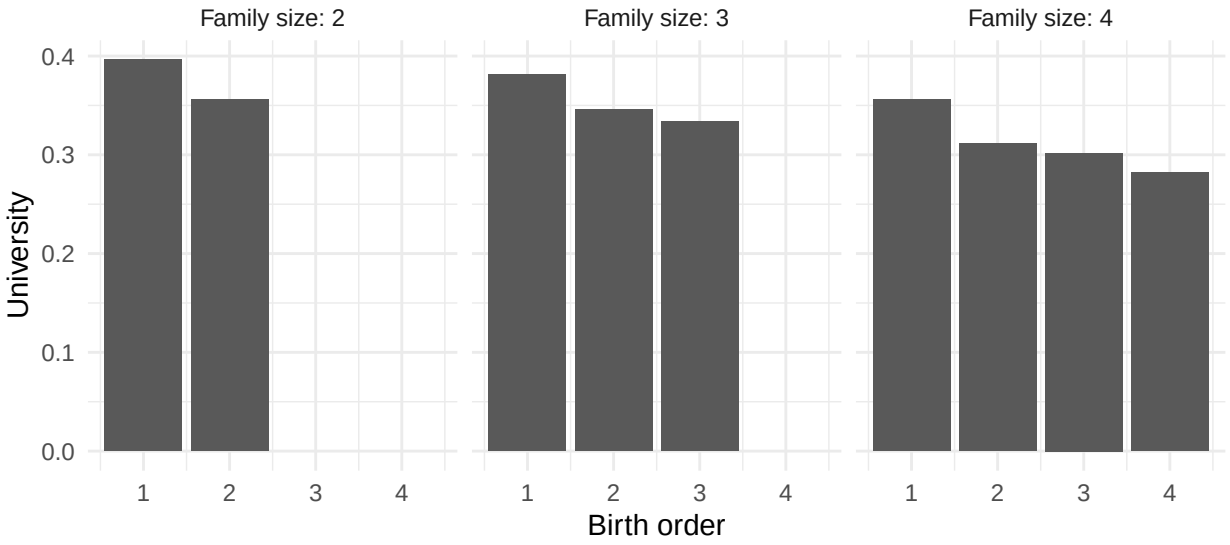


Figure 4: University attendance by birth order and family size

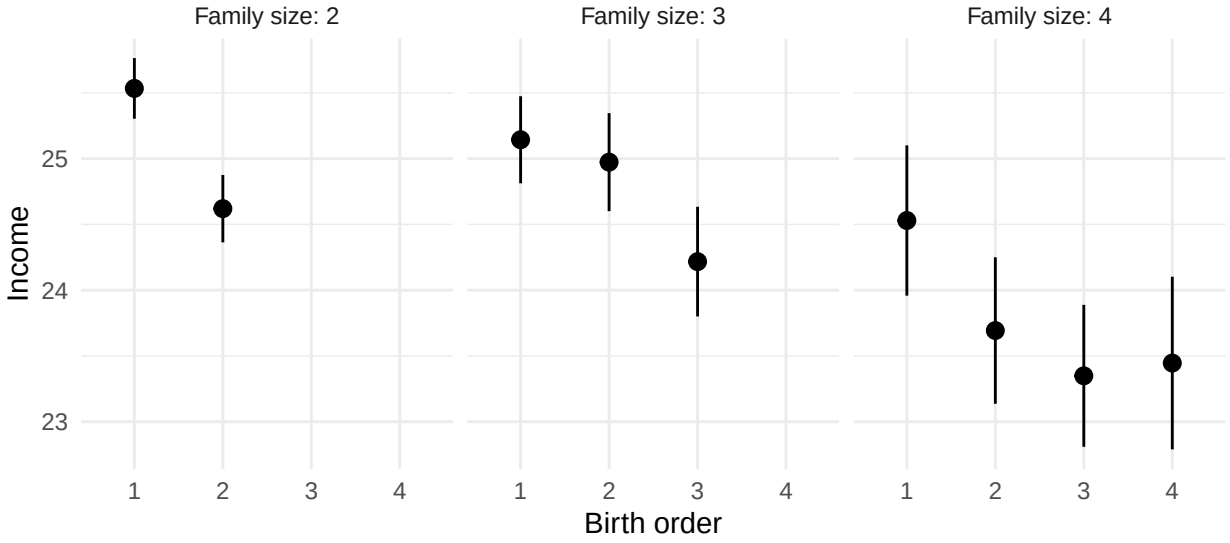


Figure 5: Median income of first job by birth order and family size

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 dummies for birth year to control for cohort effects, 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; Black, Devereux, and Salvanes 2011).[5] 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.[6]

Table 1: Regressions of spouse PSEA on birth order

	(1)	(2)	(3)
Birth order	-0.0049 **	-0.0046 *	-0.0237 ***
	(0.0019)	(0.0019)	(0.0048)
Own PSEA		0.0471 ***	0.0410 ***
		(0.0022)	(0.0037)
Father's age at birth			0.0091 ***
			(0.0008)
Family size dummies	Yes	Yes	Yes
Birth month dummies	No	Yes	Yes
Birth year dummies	No	Yes	Yes
N	340676	340061	88510
R2	0.001	0.004	0.005

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. 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: 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

[5] (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.

[6] 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.

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.[7]

Table 2: Regressions of spouse PSEA on birth order and potential mediators

	(1)	(2)	(3)	(4)
Birth order	-0.0188 *	-0.0064	-0.0152	-0.0053
	(0.0084)	(0.0084)	(0.0084)	(0.0084)
University		0.1583 ***		0.1444 ***
		(0.0120)		(0.0124)
Income			0.0036 ***	0.0022 ***
			(0.0005)	(0.0005)
Own PSEA	0.0471 ***	0.0305 ***	0.0417 ***	0.0287 ***
	(0.0061)	(0.0062)	(0.0061)	(0.0062)
Father's age at birth	0.0070 ***	0.0044 **	0.0062 ***	0.0042 **
	(0.0014)	(0.0014)	(0.0014)	(0.0014)
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes
N	30352	30352	30352	30352
R2	0.006	0.011	0.007	0.012
logLik	-43106.208	-43019.835	-43076.985	-43009.444
AIC	86318.417	86147.671	86261.971	86128.888

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

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

[7]The appendix reports results without controlling for father's age.

comparison easier across the columns: this reduces the N slightly.) The remaining columns control for our SES 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).[8] Including university attendance alone reduces the effect of birth order by 48.8% (CI 31.6% – 69.6%). Including income alone reduces the effect of birth order by 29.1% (CI 18.9% – 42.7%). Including both decreases the effect by 61.0% (CI 40.4% – 87.0%).

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 SES are particularly likely to marry attractive spouses. To test this, we separately regress male spouses' PSEA on female birth order, and female 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. Columns 1 and 2 use birth order of male respondents to predict female spouses' PSEA. Columns 3 and 4 use female respondents to predict male spouses' PSEA. Effect sizes are larger for female respondents. This might be because social status matters more for females in the marriage markets (contrary to a common stereotype), or because males place less value on phenotypes related to PSEA. For both sexes, controlling for university attendance reduces the effect of birth order. Columns 5 and 6 use only couples with children. The effect of birth order is slightly larger than for the sample as a whole, and university attendance still seems to mediate the birth order effect.

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

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.[9]

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.

To check this, we run balance tests on 33 different polygenic scores.[10] We regress each score on birth order, controlling for family size. No scores were significant at $p < 0.10/33$. 3 scores were significant at $p < 0.10$ (body mass index, conscientousness, and neuroticism). Coefficients were never greater than 0.01 of a standard deviation. Table 6 in the appendix reruns regressions controlling for these scores. Results are

[8] 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.

[9] Table 5 in the appendix estimates a separate birth order coefficient within each family size.

[10] 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, 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.

Table 3: Regressions of spouse PSEA on birth order: subsets

	Male respondents	Male respondents	Female respondents	Female respondents	With children	With children
Birth order	-0.006	0.006	-0.029 ***	-0.016	-0.032 **	-0.019
	(0.019)	(0.019)	(0.007)	(0.009)	(0.011)	(0.011)
University		0.164 ***		0.150 ***		0.172 ***
		(0.027)		(0.005)		(0.015)
Father's age at birth	0.006 *	0.003	0.007 **	0.005	0.008 ***	0.005 **
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Own PSEA	0.057 ***	0.039 **	0.039 ***	0.024 ***	0.043 ***	0.025 **
	(0.011)	(0.013)	(0.004)	(0.003)	(0.008)	(0.008)
Family size dummies	Yes	Yes	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	13613	13613	16739	16739	19048	19048
R2	0.010	0.016	0.007	0.012	0.006	0.013

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

almost unchanged. To test whether polygenic scores might vary for particular birth orders within particular family size, we also regress each score on a full set of birth order dummies, interacted with a full set of family size dummies. None of the 924 birth order coefficients were significant at $p < 0.001$. Of course, there could still be unmeasured genetic variants which correlate with birth order in our sample. Nevertheless, a wide set of polygenic scores shows no large or significant correlation. This increases our confidence that birth order is indeed exogenous to genetics.

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").[11] 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. This logic is different from the standard reason for association between genes and SES. There, more meritocratic societies allow a greater role for genetic variants linked to labour market success. Thus, there is a monotonic relationship: more meritocracy leads to a stronger gene-environment correlation (rGE) between genetics and SES. Here, the relationship is non-monotonic: rGE is strongest at intermediate levels of "meritocracy" in marriage markets. Comparing rGE 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 of social status), stratification will predictably associate with causally relevant genetic variants. For this reason, we 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.

[11] The appendix contains a selection of relevant historical quotations.

Appendix

Second part of Proposition 1

Proof. Write

$$corr(G_j, S_j) = \frac{cov(G_j, S_j)}{\sqrt{var(G_j)var(S_j)}} \text{ for both generations } j \in \{p, c\}. \tag{7}$$

where

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

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

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

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

□

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.

Proposition 2

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

$$\begin{aligned} cov(G_c, S_c) &= cov(\bar{G} + \varepsilon^G, \bar{S} + \varepsilon^S) \\ &= cov(\bar{G}, \bar{S}) + cov(\varepsilon^G, \bar{S}) + cov(\bar{G}, \varepsilon^S) + cov(\varepsilon^G, \varepsilon^S). \end{aligned} \tag{8}$$

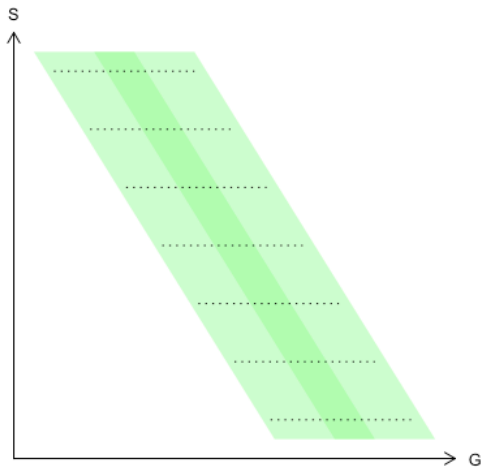


Figure 6: Correlation counterexample

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

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

shows that $var(G_c)$ will approach $var(\bar{G})$ as σ_G^2 grows small, and similarly for $var(S_c)$. Plugging these facts into (7) shows that $corr(G_c, S_c)$ approaches $corr(\bar{G}, \bar{S})$ as σ_G^2 and σ_S^2 grow small. Proposition 1 then shows $corr(\bar{G}, \bar{S}) < corr(G_p, S_p)$ for $k \in (0,1)$.

Under condition 2, $cov(G_c, S_c) = cov(\bar{G}, \bar{S})$ since the last three terms of the sum in (8) are zero. Then since

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

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	-0.0017	0.0031	-0.0002	0.0035
	(0.0034)	(0.0034)	(0.0034)	(0.0034)
University		0.1434 ***		0.1265 ***
		(0.0061)		(0.0064)
Income			0.0039 ***	0.0025 ***
			(0.0002)	(0.0003)
Own EA3	0.0439 ***	0.0296 ***	0.0385 ***	0.0278 ***
	(0.0034)	(0.0034)	(0.0034)	(0.0034)
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes
N	112010	112010	112010	112010
R2	0.003	0.008	0.005	0.009
logLik	-158652.280	-158380.574	-158527.874	-158331.165
AIC	317410.560	316869.148	317163.748	316772.330

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

Table 5: Regressions of spouse PSEA, separate birth order coefficients for each family size

	(1)	(2)	(3)	(4)
Birth order X family size 2	-0.0298	-0.0140	-0.0238	-0.0118
	(0.0186)	(0.0186)	(0.0186)	(0.0186)
Birth order X family size 3	-0.0339 *	-0.0217	-0.0290 *	-0.0198
	(0.0142)	(0.0142)	(0.0142)	(0.0142)
Birth order X family size 4	-0.0196	-0.0059	-0.0180	-0.0061
	(0.0160)	(0.0160)	(0.0160)	(0.0160)
Birth order X family size 5	0.0167	0.0302	0.0185	0.0302
	(0.0195)	(0.0195)	(0.0195)	(0.0195)
Birth order X family size 6	0.0036	0.0163	0.0085	0.0182
	(0.0233)	(0.0232)	(0.0233)	(0.0232)
Birth order X family size 7	-0.0741 *	-0.0737 *	-0.0699 *	-0.0712 *
	(0.0344)	(0.0343)	(0.0345)	(0.0344)
Birth order X family size 8	0.0293	0.0289	0.0262	0.0270
	(0.0380)	(0.0385)	(0.0382)	(0.0386)
University		0.1588 ***		0.1450 ***
		(0.0120)		(0.0124)
Income			0.0036 ***	0.0022 ***
			(0.0005)	(0.0005)
Own EA3	0.0475 ***	0.0309 ***	0.0421 ***	0.0291 ***
	(0.0061)	(0.0062)	(0.0061)	(0.0062)
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes
N	30352	30352	30352	30352
R2	0.006	0.011	0.008	0.012
logLik	-43101.371	-43014.613	-43072.598	-43004.524
AIC	86320.742	86149.226	86265.195	86131.049

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

Table 6: Regressions of spouse PSEA with controls for polygenic scores

	(1)	(2)	(3)	(4)
Birth order	-0.0188 *	-0.0063	-0.0152	-0.0052
	(0.0084)	(0.0084)	(0.0084)	(0.0084)
University		0.1584 ***		0.1444 ***
		(0.0120)		(0.0124)
Income			0.0036 ***	0.0022 ***
			(0.0005)	(0.0005)
Father's age at birth	0.0069 ***	0.0043 **	0.0061 ***	0.0041 **
	(0.0014)	(0.0014)	(0.0014)	(0.0014)
Own PSEA	0.0438 ***	0.0272 ***	0.0384 ***	0.0253 ***
	(0.0062)	(0.0063)	(0.0062)	(0.0063)
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	30352	30352	30352	30352
R2	0.006	0.011	0.008	0.012

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. 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

Belot, Michèle, and Marco Francesconi. 2013. "Dating Preferences and Meeting Opportunities in Mate Choice Decisions." *Journal of Human Resources* 48 (2): 474–508.

Belsky, Daniel W, Benjamin W Domingue, Robbee Wedow, Louise Arseneault, Jason D Boardman, Avshalom Caspi, Dalton Conley, et al. 2018. "Genetic Analysis of Social-Class Mobility in Five Longitudinal Studies." *Proceedings of the National Academy of Sciences* 115 (31): E7275–E7284.

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.

Buss, David M, and Michael Barnes. 1986. "Preferences in Human Mate Selection." *Journal of Personality and Social Psychology* 50 (3): 559.

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.

Greenwood, Jeremy, Nezih Guner, Georgi Kocharkov, and Cezar Santos. 2014. "Marry Your Like: Assortative Mating and Income Inequality." *American Economic Review* 104 (5): 348–53.

Halsey, AH. 1958. "Genetics, Social Structure and Intelligence." *The British Journal of Sociology* 9 (1): 15–28.

Heckman, James, Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review* 103 (6): 2052–86.

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.

Krapohl, E, and R Plomin. 2015. "Genetic Link Between Family Socioeconomic Status and Children's Educational Achievement Estimated from Genome-Wide SNPs." *Molecular Psychiatry* 21 (3): 437–43. https://doi.org/10.1038/mp.2015.2.

Kristensen, Petter, and Tor Bjerkedal. 2007. "Explaining the Relation Between Birth Order and Intelligence." *Science* 316 (5832): 1717–7.

Lee, James J, Robbee Wedow, Aysu Okbay, Edward Kong, Omeed Maghzian, Meghan Zacher, Tuan Anh Nguyen-Viet, et al. 2018. "Gene Discovery and Polygenic Prediction from a Genome-Wide Association Study of Educational Attainment in 1.1 Million Individuals." *Nature Genetics* 50 (8): 1112–21.

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.

Schwartz, Christine R, and Robert D Mare. 2005. "Trends in Educational Assortative Marriage from 1940 to 2003." *Demography* 42 (4): 621–46.

Shakespeare, William. 1595. *A Midsummer Night's Dream.*

Vandenberg, Steven G. 1972. "Assortative Mating, or Who Marries Whom?" *Behavior Genetics* 2 (2-3): 127–57.