Trading social status for genetics in marriage markets: evidence from UK Biobank

David Hugh-Jones, Oana Borcan & Abdel Abdellaoui

2020-11-27

Introduction

There is a well-known association between an individual's socio-economic status (SES) and their genetics. Twin studies show that the heritability of occupational class and educational attainment, i.e. the proportion of variance explained by genetics, is around 50% (Tambs et al. 1989). Genome-wide Complex Trait Analysis (GCTA) shows that even 2-year-old children's family SES can be predicted from their genes (Trzaskowski et al. 2014). Polygenic scores for educational attainment (EA) predict both occupational class and (of course) EA itself (Rimfeld et al. 2018).

A common explanation is that in meritocratic societies, advantageous genes lead to success in education and the labour market. Successful parents may then pass both SES and their genes to their children, leading to an association between the two (Belsky et al. 2018). This mechanism depends on the level of meritocracy in society's institutions (Heath et al. 1985; Branigan, McCallum, and Freese 2013). Indeed, after the fall of communism in Estonia, the heritability of SES increased, presumably because post-communist society was more meritocratic (Rimfeld et al. 2018). In a non-meritocratic society where social status was ascribed rather than earned, this mechanism could not take effect.

We propose an alternative explanation for the association between SES and genetics. It is centred on marriage markets rather than labour markets or educational institutions, and it relies on assortative mating, rather than meritocratic institutions. As a result, it applies to a historically wider range of societies.

Our mechanism is that both social status[1] and genetics contribute to a person's attractiveness in marriage

[1] *Social status* refers to characteristics that an individual has in virtue of their social position. For example, my wealth is a fact about me that holds in virtue of my relationship to certain social institutions (bank deposits, title deeds etc.). Other

markets. For example, suppose that wealth and intelligence are both attractive qualities in a spouse. Then wealthy people will be more likely to marry relatively intelligent people. Children of these marriages will then inherit both wealth (via social institutions), and genetic variants for intelligence (via biology). As a result, wealth and genes for intelligence will become associated in the next generation. More generally, whenever people in marriage markets trade social status for a genetically inherited characteristic, the two will become associated.

We call this mechanism *Social-Genetic Assortative Mating* (SGAM). Note that it does not require meritocracy. Even if social status is inherited from birth, and cannot be earned (as with e.g. titles of nobility, or caste), it will still become associated with attractive genetic variants, so long as there is assortative mating and people trade status off for genetics in marriage markets. By contrast, the ES mechanism, where social status can be earned with the help of advantageous genetic variants, requires meritocracy but not assortative mating.

Another difference is in the chain of causality. In ES, genetic variants cause an increase in social status, and both genetics and status may then be inherited. In SGAM, causality may go from social status to genetics. One person's social status causes their spouse to have different genetic variants; again, both are then inherited. Thus, changes to social status may be encoded in the genetics of subsequent generations.

The two mechanisms are complementary, and both may operate in a given social setting. Table 1 summarizes the conditions for each mechanism.

Table 1: Conditions for Social-Genetic Assortative Mating (SGAM) and Earned Status (ES)

Mechanism	Genetics	Social status	Meritocracy	Assortative mating
ES	Inherited	Inherited	Yes	No
SGAM	Inherited	Inherited	No	Yes

Although SGAM does not require meritocracy, it is affected by a society's marriage market institutions. Surprisingly, there is a non-linear relationship between "egalitarianism" in marriage market institutions, and the strength of the association between social status and genetic variation. When marriage markets are fully egalitarian, or fully inegalitarian, SGAM does not cause any association. In between these extremes, there is a positive association. The intuition is that in a fully "egalitarian" marriage market, social status plays no role in determining attractiveness. As a result, people do not trade off social status against genetics. On

examples include caste, class, income, educational qualifications. *Socio-economic status* (SES) is a specific kind of social status, existing in economically stratified societies, and referring to a combination of educational attainment, occupational class, income and wealth (e.g. White 1982).

the other hand, in a wholly "inegalitarian" marriage market, such as a caste society where mating across castes is forbidden, only social status determines attractiveness and genetics plays no role, so again people do not trade the two off. In between the two extremes, people trade social status for "good genes", and the two become associated.

This paper tests for SGAM in a modern society. To pin down SGAM, we use a causal antecedent of social status which is independent of individual genetics: birth order. Birth order is known to affect outcomes including income, occupational status and educational attainment. We show that people born earlier within a sibling group marry spouses with different genetics, specifically, higher polygenic scores for educational attainment. We present evidence that this effect is mediated by measures of SES, including educational attainment and labour market income. First, we develop a simple model of SGAM to show how parents' social status predicts children's genetics, and how this varies with social structure.

Related literature

It is widely known that assortative mating may affect social inequality (Greenwood et al. 2014), including variance in polygenic scores. But the mechanism is usually genetic assortative mating (GAM), i.e. marriage partners having similar genetic variants (Hugh-Jones et al. 2016). A closely-related concept to SGAM is cross-trait assortative mating (Sundet et al. 2005; Beauchamp et al. 2010). This refers to people with (genes for) e.g. height marrying people with (genes for) e.g. intelligence. As a result, the two kinds of variation become associated. Our model simply extends this idea to the case of a socially-inherited trait and one or more genetic variants.

The basic idea behind SGAM is not new. It is a cliché that e.g. wealth and beauty tend to assort in marriage markets. However, few papers has analysed this mechanism or drawn out its consequences explicitly. Halsey (1958) showed in a two-class model that social mobility combined with assortative mating might increase the association between genetics and social class. Belsky et al. (2018) offer three reasons for the association between education-linked genetics and SES, but do not consider SGAM.

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 $Var(G) = Var(S) = 1$.[2] 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.

Figure 1 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:

$$A(g_i, s_i) = A(g_{p(i)}, s_{p(i)}), \tag{1}$$

[2]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.

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

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

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 (2).

Take an arbitrary pair d, m. Write

$$g_d s_d = (g_c - \Delta g)(s_c + \Delta s);$$
$$g_m s_m = (g_c + \Delta g)(s_c - \Delta s)$$

where

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

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.

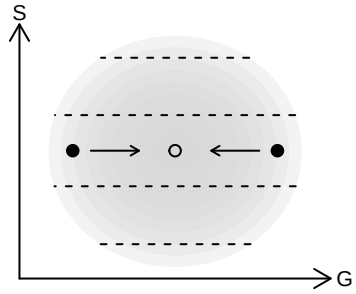
□

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.

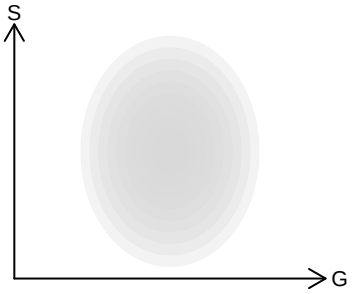
We now relax the condition that children are exactly at the mean of their parents' values for G and S. Let

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

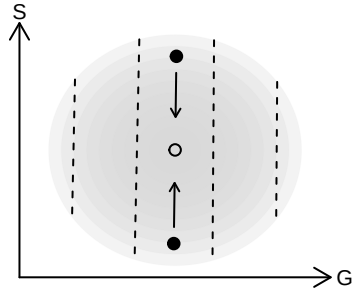
where



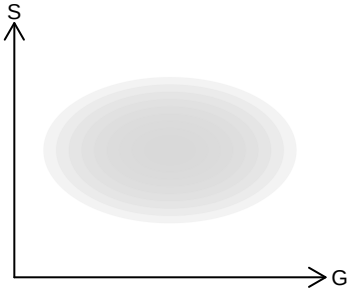
(a) Caste society ($k = 1$): parents



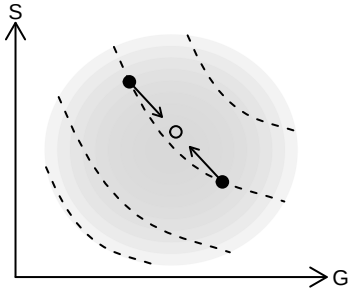
(b) Caste society: children



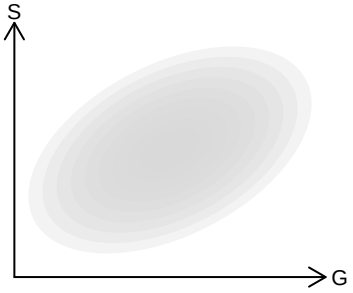
(c) Egalitarian society ($k = 0$): parents



(d) Egalitarian society: children



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



(f) Intermediate society: children

Figure 1: 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.

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

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

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 (genetically determined in part) to earn more, and therefore to match with more attractive partners.

The contents of both S, social status, and G – "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, SES. 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 SGAM. In particular, in "caste societies" where there is complete endogamy within social status groups, there is no scope for SGAM, 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).

In modern societies, both SGAM and ES 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

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.

The Biobank does not contain explicit information on spouse pairs. We categorize respondents as pairs if they:

- had the same home postcode on at least one occasion[3]
- both reported living in the same type of accommodation, homeownership/renting status, length of time at the address, number of inhabitants in the household, number of vehicles in the household, and number of children;

[3]Typically a UK postcode contains about 15 properties.

- both reported living with their spouse;
- consisted of one male and one female.

We also eliminated any pairs where either spouse appeared more than once in the data.

Some of these pairs could be false positives, i.e. people who are not each others' spouse but simply live in the same postcode. To validate the accuracy of our measures, we used genetic relationships. Some respondents in the Biobank sample have a child who is also in the sample, as inferred from genetic data. Out of our 36142 spouse pairs, 510 have a genetic child of at least one partner in the sample. For 444 of these, the child is the genetic child of both partners. If this subsample is representative of the whole, then at least 87.06% of the pairs who have had a child, have had a child together. This is a lower bound because those who had a child with someone else may also have had a child with the presumed partner in our data.

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 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). We use per-SNP summary statistics from (Lee et al. 2018), recalculated excluding UK Biobank participants, to calculate PSEA.[4]

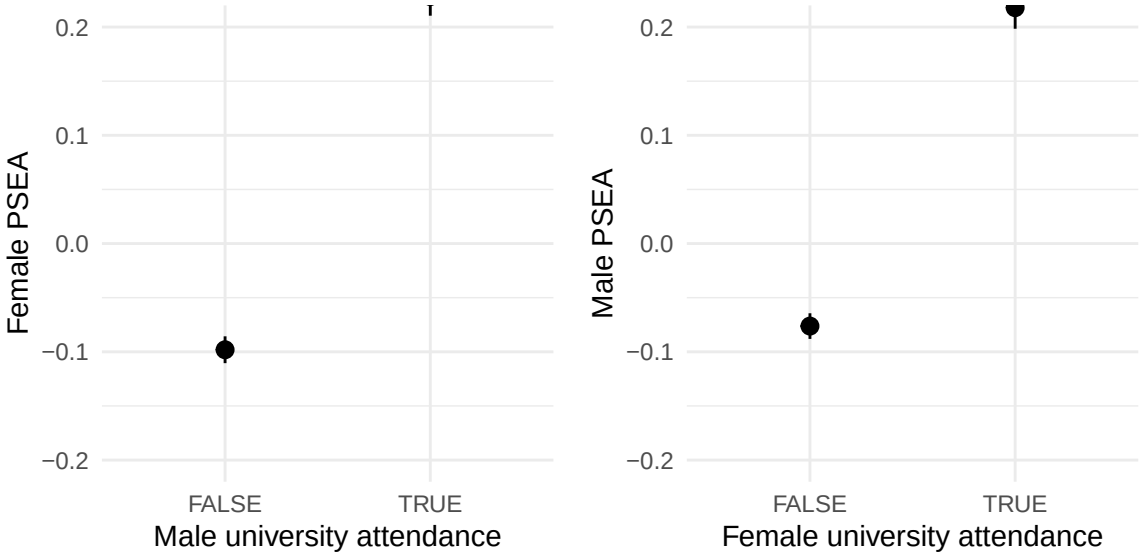
- TODO: if our N gets small, we may have to control, rather than residualize, to get accurate p-values.

Figures 2 and 3 illustrate the possible consequences of SGAM. The X axis shows a measure of one partner's socio-economic status: university attendance (Figure 2 or income (Figure 3). The Y axis plots the other partner's mean PSEA. 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.

```
## Warning: Removed 20 rows containing non-finite values (stat_summary).
```

```
## Warning: Removed 17 rows containing non-finite values (stat_summary).
```

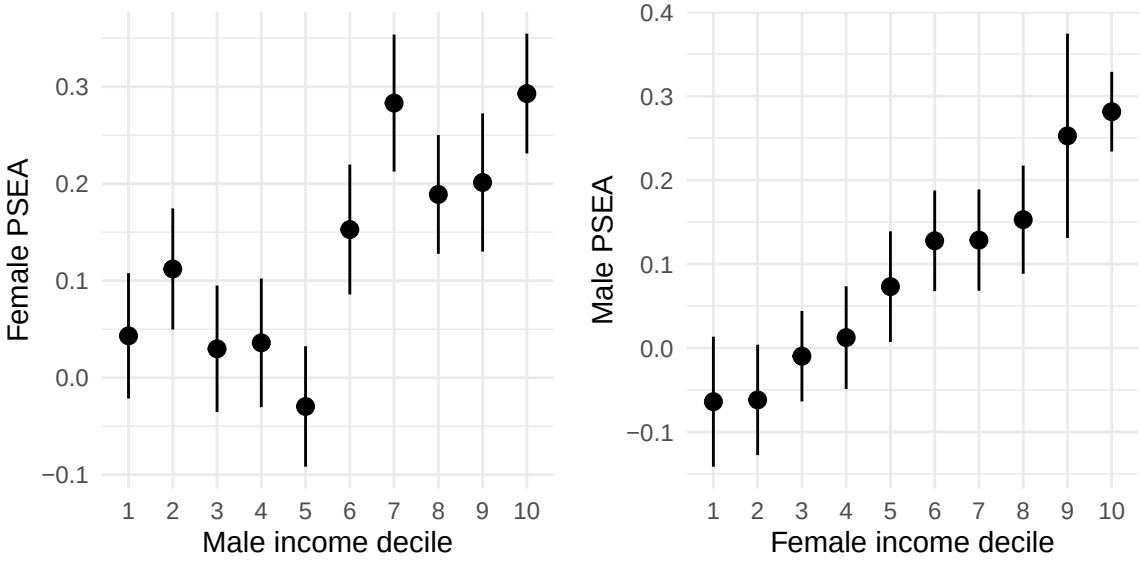
[4]PSEA is residualized on the first 100 principal components of the SNP array data.



(a) Female PSEA by male educational attainment

(b) Male PSEA by female educational attainment

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



(a) Female EA3 by male income

(b) Male EA3 by female income

Figure 3: Social and genetic advantage among spouse pairs in UK Biobank

These figures do not prove that SGAM 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 genetic assortative mating (GAM) alone (Hugh-Jones et al. 2016). To demonstrate SGAM, 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, 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 changes to the school leaving age (Davies et al. 2018).

We use *birth order*. It is known that earlier-born children receive more parental care and have better life outcomes, including measures of SES such as educational attainment and occupational status (Lindahl 2008; 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.[5] 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 may also relate to 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
- TODO: look at mechanisms by which birth order might affect university
- 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

[5] 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.

these measures, and other, unmeasured dimensions of SES. So, an instrumental variables approach would probably fall foul of the exclusion restriction.

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

Linearizing our model so that $A(g,s) = (1-k)g + ks$ and applying (1) shows that:

$$\frac{dA(g_{p(i)}, s_{p(i)})}{ds_i} = k$$

We wish to test whether $k \in (0,1)$, i.e. whether GSAM is taking place. If $k > 0$ then an increase in i's social status s_i will increase i's attractiveness A; if $k < 1$ then an increase in A will be associated (in expectation) with an increase in i's partner's genetic endowment $g_{p(i)}$. We therefore wish to estimate the effect of i's status on their partner's genetics, while controlling for i's own genetics g_i. Since our measures of genetic endowment (e.g. PSEA) are noisy and incomplete, it is not enough to include them in the regression. Instead, we use birth order as a source of variation in s_i which is orthogonal to g_i.

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 SES (as an effect of birth order) on spouse PSEA.

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

$$Y = \delta + \alpha \cdot BO + \beta \cdot \mathbf{X} + \varepsilon, \tag{3}$$

where $Y \equiv g_{p(i)}$ and $\mathbf{X}$ 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:

$$Y = BO \cdot Y_1 + (1 - BO) \cdot Y_0 \tag{4}$$

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 (4) 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:

$$Y_d = \kappa_d + \alpha_d \cdot \theta_d + \beta_d \cdot \mathbf{X} + \tilde{\varepsilon_d}, \tag{5}$$

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: bold the alpha and theta and beta vectors in eq 3.

$$Y_d = \tau_d + \sum_{j \in J_m} \alpha_d^j \cdot \theta_d^j + \beta \cdot \mathbf{X} + \varepsilon_d, \tag{6}$$

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 BO 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 (6) into (4) we obtain:

$$Y = \tau_0 + \tau \cdot BO + \sum_{j \in J_m} \alpha^j \cdot \theta^j + \beta \cdot \mathbf{X} + \varepsilon, \tag{7}$$

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 (7) 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 (7), 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.

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 $p =$ 1.04e-13).

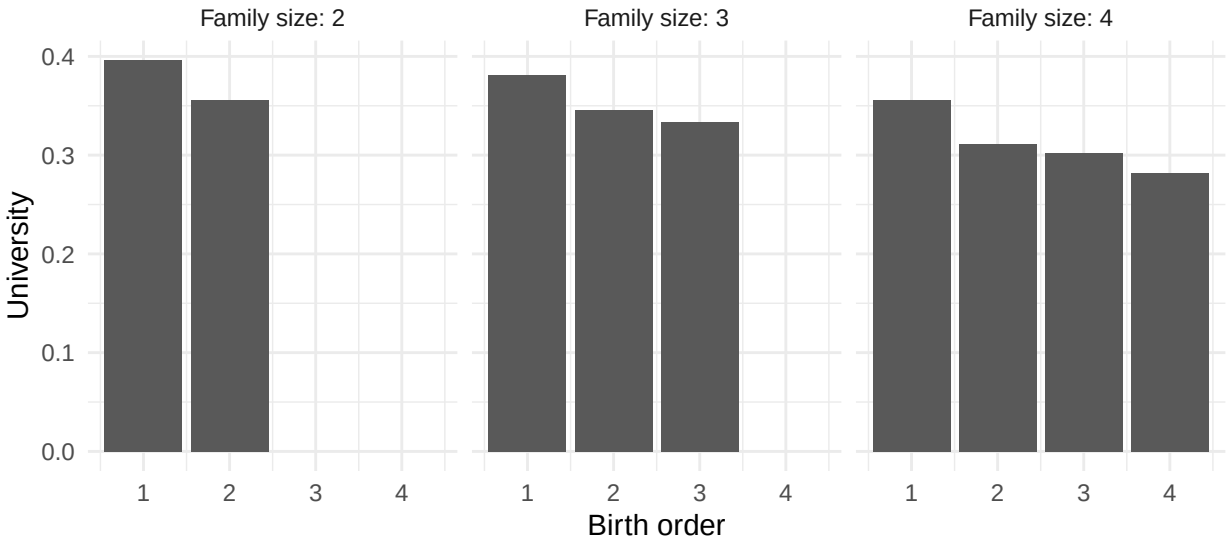


Figure 4: University attendance by birth order and family size

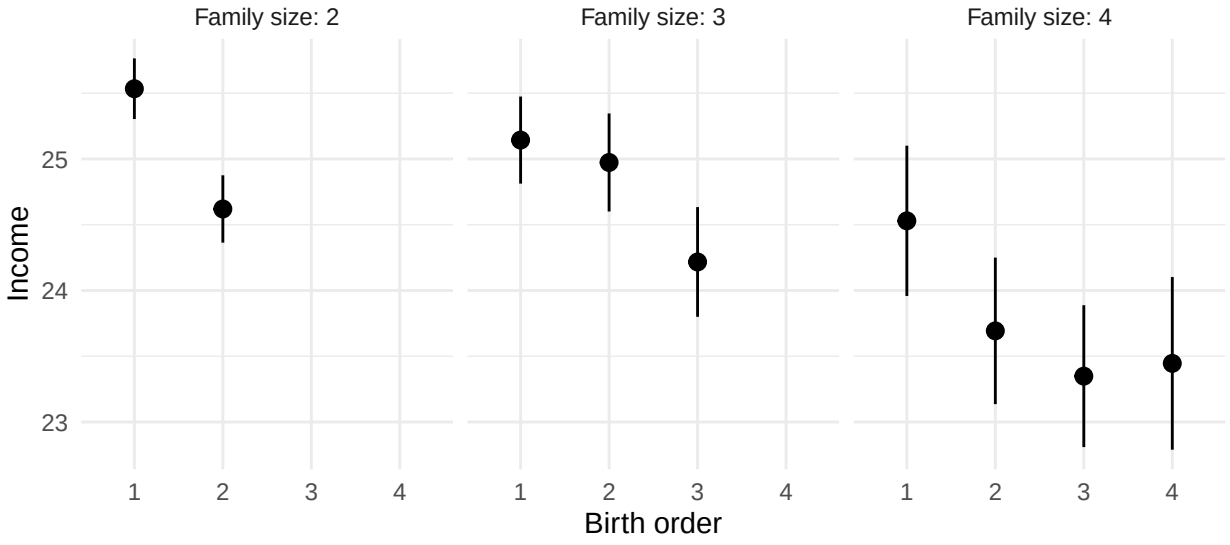


Figure 5: Median income of first job by birth order and family size

Next we run regressions of spouse PSEA on birth order. Table 2 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).[6] 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.[7]

- TODO: maybe - 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: maybe - 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: number of elder *brothers*? (We don't have this info but we have total number of brothers, so we can interact this with the birth order effect)

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.[8]

Table 3 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

[6](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.

[7]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.

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

Table 2: Regressions of spouse PSEA on birth order

	(1)	(2)	(3)
Birth order	−0.0133 *	−0.0124	−0.0491 *
	(0.0066)	(0.0066)	(0.0194)
Own PSEA		0.0682 ***	0.0540 ***
		(0.0084)	(0.0156)
Father's age at birth			0.0126 ***
			(0.0034)
Family size dummies	Yes	Yes	Yes
Birth month dummies	No	Yes	Yes
Birth year dummies	No	Yes	Yes
N	25452	25412	5194
R2	0.003	0.010	0.021

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

comparison easier across the columns: this reduces the N.) The remaining columns add potential mediators of birth order effects. Column 2 includes fluid IQ and self-rated health, both of which could be affected by birth order and affect spouse matching. Columns 3 to 5 then add our measures of SES. Column 3 includes university attendance. Column 4 includes income. Column 5 includes both.

Controlling for fluid IQ and self-rated health (column 2) does not reduce the effect of birth order, although both variables are significant in the expected direction. When we add a control for university attendance (column 3), the effect of birth order drops and becomes insignificant. The fluid IQ variable also loses size and significance, suggesting that this effect too may work via itse effect on university attendance. By contrast, self-rated health is unaffected. Controlling for income alone (column 4), birth order again becomes insignificant though its size is unchanged. Lastly, when we control for both university and income, again birth order's effect size drops and the coefficient becomes insignificant.

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 4 shows the results. Columns 1 and 2 use birth order of male respondents to predict female spouses' PSEA. Column 1 runs the regression of birth order plus controls; column 2 adds the full set of mediators, including university and income. Columns 3 and 4 repeat the exercise for female respondents, using their birth order 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. Columns 5 and 6 use only couples with children. The effect of birth order is larger than for the sample as a whole, and university attendance and income still seem 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.

Table 3: Regressions of spouse PSEA on birth order and potential mediators

	(1)	(2)	(3)	(4)	(5)
Birth order	−0.0624	−0.0586	−0.0406	−0.0514	−0.0381
	(0.0353)	(0.0432)	(0.0428)	(0.0429)	(0.0427)
University			0.2783 ***		0.2304 ***
			(0.0627)		(0.0653)
Income				0.0099 ***	0.0077 **
				(0.0026)	(0.0027)
Father's age at birth	0.0092	0.0110	0.0078	0.0091	0.0069
	(0.0060)	(0.0073)	(0.0073)	(0.0074)	(0.0073)
Own PSEA	0.0488	0.0481	0.0270	0.0371	0.0221
	(0.0269)	(0.0339)	(0.0340)	(0.0335)	(0.0337)
Fluid IQ		0.0124	−0.0011	0.0055	−0.0041
		(0.0149)	(0.0151)	(0.0152)	(0.0152)
Self-rated health		0.0547	0.0369	0.0442	0.0318
		(0.0430)	(0.0424)	(0.0427)	(0.0422)
Family size dummies	Yes	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes	Yes
N	1765	1219	1219	1219	1219
R2	0.029	0.038	0.054	0.051	0.062
logLik	−2477.763	−1710.716	−1700.371	−1702.202	−1695.465
AIC	5055.526	3525.431	3506.741	3510.404	3498.929

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

Table 4: Regressions of spouse PSEA on birth order: subsets

	Male respondents	Male respondents	Female respondents	Female respondents	With children	With children
Birth order	−0.126 **	−0.113 *	−0.026	0.018	−0.053	−0.027
	(0.047)	(0.047)	(0.056)	(0.069)	(0.039)	(0.048)
University		0.254 *		0.203 **		0.219 **
		(0.115)		(0.065)		(0.071)
Income		0.009 ***		0.007 ***		0.008 **
		(0.003)		(0.002)		(0.003)
Father's age at birth	0.015	0.008	0.005	0.004	0.009	0.006
	(0.010)	(0.009)	(0.003)	(0.004)	(0.007)	(0.008)
Own PSEA	0.068	0.070	0.026	−0.007	0.056	0.033
	(0.047)	(0.075)	(0.022)	(0.023)	(0.030)	(0.038)
Fluid IQ		−0.021		0.008		−0.001
		(0.047)		(0.018)		(0.017)
Self-rated health		0.014		0.016		0.017
		(0.025)		(0.026)		(0.046)
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	744	507	1021	712	1499	1034
R2	0.075	0.136	0.038	0.073	0.035	0.070

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. 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.[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 for these two problems, 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 $p < 0.10/33$. 4 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 7 in the appendix reruns regressions controlling for these scores. Results are 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 693 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

[9] Table 6 in the appendix estimates a separate birth order coefficient within each family size.

[10] Polygenic scores were residualized on the first 100 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.

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.

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.

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, and thus lead to monotonically stronger gene-environment correlations (rGE) with 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.

Over history, SGAM has probably operated in a much greater variety of societies than earned status, 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 residualizing on 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

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

with respect to genetics) overlaps with the sociological concept of stratification (a hierarchical ranking of social status), SGAM predicts that stratification will be associated 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.

The broadest message of this paper is that *genetics are a social outcome*. Both popular and scientific discourse often parse genetics as "nature", in opposition to "nurture" or "society" (e.g. Plomin 1994; Chakravarti and Little 2003). This idea expresses the fact that our genetic endowment is fixed at birth and cannot be influenced by our social environment (though genes may interact with the environment to cause different outcomes). But the idea that human genetics are natural can be 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, form spouse pairs, bear children, and raise them, within social institutions. 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."[12] Genetic endowments 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 will require the contributions of both social scientists and geneticists.

[12]He also wrote (ibid.) "I am ugly, but I can buy the most beautiful woman.... the effect of ugliness, its repelling power, is destroyed by money".

Appendix

Second part of Proposition 1

Proof. Write

$$corr(G_j, S_j) = \frac{cov(G_j, S_j)}{\sqrt{var(G_j)var(S_j)}} \text{ for both generations } j \in \{p, c\}. \tag{8}$$

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.

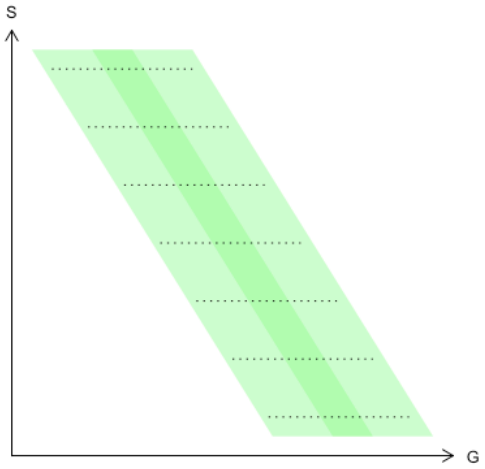


Figure 6: Correlation counterexample

Proposition 2

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

$$\begin{aligned} cov(G_c, S_c) &= cov(\bar{G} + \varepsilon^G, \bar{S} + \varepsilon^S) \\ &= cov(\bar{G}, \bar{S}) + cov(\varepsilon^G, \bar{S}) + cov(\bar{G}, \varepsilon^S) + cov(\varepsilon^G, \varepsilon^S). \end{aligned} \tag{9}$$

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 (8) 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 (9) 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 5: Regressions of spouse PSEA, without controls for father's age at respondent's birth

	(1)	(2)	(3)	(4)	(5)
Birth order	−0.0136	−0.0005	0.0061	0.0030	0.0075
	(0.0124)	(0.0148)	(0.0146)	(0.0147)	(0.0146)
University			0.2538 ***		0.2185 ***
			(0.0283)		(0.0295)
Income				0.0075 ***	0.0051 ***
				(0.0012)	(0.0012)
Own PSEA	0.0625 ***	0.0653 ***	0.0461 **	0.0574 ***	0.0435 **
	(0.0138)	(0.0168)	(0.0168)	(0.0168)	(0.0167)
Fluid IQ		0.0180 **	0.0036	0.0118	0.0014
		(0.0067)	(0.0068)	(0.0067)	(0.0069)
Self-rated health		0.0489 *	0.0353	0.0449 *	0.0345
		(0.0201)	(0.0199)	(0.0200)	(0.0199)
Family size dummies	Yes	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes	Yes
N	8225	5829	5829	5829	5829
R2	0.010	0.013	0.026	0.020	0.030
logLik	−11616.779	−8266.024	−8225.174	−8243.866	−8215.794
AIC	23337.558	16636.048	16556.347	16593.731	16539.589

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

Table 6: Regressions of spouse PSEA, separate birth order coefficients for each family size

	(1)	(2)	(3)	(4)	(5)
Birth order X family size 2	−0.1677 *	−0.1971 *	−0.1881 *	−0.1846	−0.1800
	(0.0777)	(0.0956)	(0.0954)	(0.0952)	(0.0952)
Birth order X family size 3	−0.0213	−0.0129	0.0110	−0.0007	0.0164
	(0.0601)	(0.0727)	(0.0723)	(0.0719)	(0.0718)
Birth order X family size 4	−0.0377	−0.0686	−0.0597	−0.0653	−0.0587
	(0.0746)	(0.0943)	(0.0925)	(0.0936)	(0.0921)
Birth order X family size 5	0.0450	0.0806	0.1194	0.0871	0.1180
	(0.0727)	(0.0914)	(0.0917)	(0.0910)	(0.0918)
Birth order X family size 6	−0.2064 *	−0.1584	−0.1497	−0.1440	−0.1401
	(0.0930)	(0.1075)	(0.1096)	(0.1106)	(0.1115)
Birth order X family size 7	−0.2586	−0.2523	−0.2652	−0.2729	−0.2789
	(0.1489)	(0.2210)	(0.2099)	(0.2195)	(0.2096)
Birth order X family size 8	−0.0002	0.0030	0.0431	−0.0292	0.0120
	(0.0944)	(0.1169)	(0.1155)	(0.1004)	(0.1037)
University			0.2903 ***		0.2428 ***
			(0.0631)		(0.0657)
Income				0.0099 ***	0.0076 **
				(0.0026)	(0.0027)
Own EA3	0.0482	0.0474	0.0252	0.0365	0.0205
	(0.0268)	(0.0339)	(0.0339)	(0.0334)	(0.0336)
Family size dummies	Yes	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes	Yes
N	1765	1219	1219	1219	1219
R2	0.034	0.043	0.061	0.057	0.068
logLik	−2473.576	−1707.330	−1696.063	−1698.764	−1691.278
AIC	5059.151	3530.660	3510.126	3515.528	3502.556

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

Table 7: Regressions of spouse PSEA with controls for polygenic scores

	(1)	(2)	(3)	(4)	(5)
Birth order	−0.0605	−0.0565	−0.0388	−0.0496	−0.0365
	(0.0353)	(0.0432)	(0.0427)	(0.0429)	(0.0426)
University			0.2768 ***		0.2286 ***
			(0.0627)		(0.0653)
Income				0.0099 ***	0.0077 **
				(0.0026)	(0.0027)
Father's age at birth	0.0092	0.0111	0.0079	0.0092	0.0070
	(0.0060)	(0.0074)	(0.0073)	(0.0074)	(0.0073)
Own PSEA	0.0392	0.0439	0.0228	0.0337	0.0185
	(0.0276)	(0.0347)	(0.0347)	(0.0342)	(0.0343)
Fluid IQ		0.0124	−0.0011	0.0055	−0.0041
		(0.0149)	(0.0151)	(0.0152)	(0.0152)
Self-rated health		0.0535	0.0360	0.0436	0.0313
		(0.0431)	(0.0425)	(0.0428)	(0.0423)
Family size dummies	Yes	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes	Yes
Age quadratic	Yes	Yes	Yes	Yes	Yes
Polygenic score controls	Yes	Yes	Yes	Yes	Yes
N	1765	1219	1219	1219	1219
R2	0.033	0.041	0.057	0.054	0.064

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Standard errors clustered by spouse pair.
Polygenic scores: BMI, conscientiousness, neuroticism.

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

Beauchamp, Jonathan P., David Cesarini, Magnus Johannesson, Erik Lindqvist, and Coren Apicella. 2010. "On the Sources of the Heightintelligence Correlation: New Insights from a Bivariate ACE Model with Assortative Mating." *Behavior Genetics* 41 (2): 242–52. https://doi.org/10.1007/s10519-010-9376-7.

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.

Branigan, Amelia R, Kenneth J McCallum, and Jeremy Freese. 2013. "Variation in the Heritability of Educational Attainment: An International Meta-Analysis." *Social Forces* 92 (1): 109–40.

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.

Davies, Neil M., Matt Dickson, George Davey Smith, Gerard J. van den Berg, and Frank Windmeijer. 2018. "The Causal Effects of Education on Health Outcomes in the UK Biobank." *Nature Human Behaviour* 2 (2): 117–25. https://doi.org/10.1038/s41562-017-0279-y.

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.

Heath, A. C., K. Berg, L. J. Eaves, M. H. Solaas, L. A. Corey, J. Sundet, P. Magnus, and W. E. Nance. 1985. "Education Policy and the Heritability of Educational Attainment." *Nature* 314 (6013): 734–36. https://doi.org/10.1038/314734a0.

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.

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.

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.

Lindahl, Lena. 2008. "Do Birth Order and Family Size Matter for Intergenerational Income Mobility? Evidence from Sweden." *Applied Economics* 40 (17): 2239–57. https://doi.org/10.1080/00036840600949421.

Marx, Karl. 1844. "Economic and Philosophical Manuscripts." *Early Writings* 333.

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.

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.*

Sundet, Jon Martin, Kristian Tambs, Jennifer R. Harris, Per Magnus, and Tore M. Torjussen. 2005. "Resolving the Genetic and Environmental Sources of the Correlation Between Height and Intelligence: A

Study of Nearly 2600 Norwegian Male Twin Pairs." *Twin Research and Human Genetics* 8 (4): 307–11. https://doi.org/10.1375/twin.8.4.307.

Tambs, Kristian, Jon Martin Sundet, Per Magnus, and K re Berg. 1989. "Genetic and Environmental Contributions to the Covariance Between Occupational Status, Educational Attainment, and IQ: A Study of Twins." *Behavior Genetics* 19 (2): 209–22. https://doi.org/10.1007/bf01065905.

Trzaskowski, Maciej, Nicole Harlaar, Rosalind Arden, Eva Krapohl, Kaili Rimfeld, Andrew McMillan, Philip S. Dale, and Robert Plomin. 2014. "Genetic Influence on Family Socioeconomic Status and Childrens Intelligence." *Intelligence* 42 (January): 83–88. https://doi.org/10.1016/j.intell.2013.11.002.

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

White, Karl R. 1982. "The Relation Between Socioeconomic Status and Academic Achievement." *Psychological Bulletin* 91 (3): 461.