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

David Hugh-Jones, Oana Borcan & Abdel Abdellaoui

2021-03-04

Abstract

Some genetic variants are systematically associated with socio-economic status. We provide a new explanation for this, based on marriage markets: if social status and 'good genes' are both attractive in a partner, then they will become associated in the next generation. We test our theory on couples in UK Biobank, a large genetically-informed survey. We compare spouses of individuals with different birth order, which is known to affect socio-economic status and which is exogenous to own genetic endowments (among siblings). Spouses of earlier-born siblings have higher polygenic scores for educational attainment. The effect of birth order appears to be mediated by own educational attainment. Thus, environmental shocks to social status are reflected in the DNA of subsequent generations. This mechanism reveals a new aspect of the inheritance of inequality in contemporary and historical human societies.

Introduction

Inequality persists over generations, and this has important economic effects. Intergenerational mobility is correlated with (within-generation) inequality itself (Krueger 2012; Becker et al. 2018), and it may also affect economic efficiency by changing the payoffs to talent and hard work (Corak 2013). Intergenerational absolute mobility has fallen in America (Chetty et al. 2017), although relative mobility has been stable (Chetty et al. 2014).

Families play a role in this story, because wealth, human capital and other traits are transmitted from parents to children. In particular, assortative mating (when similar people marry each other) can increase social inequality in income, wealth, social status or human capital, compared to the case where people pair at random. It is often thought that the degree of assortative mating has increased in recent generations, and that this could help to explain increasing levels of social inequality (Fernandez, Guner, and Knowles 2005; Schwartz and Mare 2005; Greenwood et al. 2014; Eika, Mogstad, and Zafar 2019).

Another strand of research relates inequality to human genetics. Twin studies show that the heritability of occupational class and educational attainment, i.e. the proportion of variance explained by genetic differences between individuals, is around 50% (Tambs et al. 1989). Genome-wide Complex Trait Analysis (GCTA) shows that 2-year-old children's family socio-economic status (SES) can be predicted from their genes (Trzaskowski et al. 2014). Polygenic scores for educational attainment predict occupational class (Rimfeld et al. 2018). The current leading explanation for this *SES-genetics gradient* is that in meritocratic societies, individuals with genes associated with valuable skills and personality traits experience upward mobility. Their children will then inherit both their genes (via biology) and their increased social status (via social institutions).

This paper links assortative mating to genetics in order to give a new explanation for the SES-genetics gradient. The mechanism is that both social status[1] and genetics contribute to a person's attractiveness in marriage markets. For example, suppose that wealth and intelligence are both attractive qualities in a spouse. Then wealthy people will be more likely to marry intelligent people. Again, children of these marriages will then inherit both wealth, and genetic variants associated with intelligence. 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).

Under this mechanism, shocks to an individual's social status may alter the identity of his or her spouse, and so be reflected in the genetics of his or her children. Thus, in contrast to a large body of research that treats genetic data as an independent variable and traces its effects on social and economic outcomes, we make the genetics endogenous.[2] Our mechanism gives a new explanation for the link between social status and genetics. This explanation is likely to apply to a wider range of societies than the alternative "meritocratic mobility" explanation. That is because, while widespread social mobility is relatively rare and historically recent, assortative mating is a human universal (Buss 1989).

Below, after reviewing the literature, we develop a simple model of SGAM to show how parents' social status predicts children's genetics, and how this varies with a society's marriage market institutions. Surprisingly, there is a non-linear relationship between "egalitarianism" in marriage markets, 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 the other hand, in a wholly "inegalitarian" marriage market, such as a caste society where

[1] *Social status* refers to characteristics that an individual possesses 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 et cetera.). Other examples include caste, class, income, and educational qualifications. *Socio-economic status* (SES) is a specific type of social status which exists in economically stratified societies, and which refers to a combination of educational attainment, occupational class, income and wealth (e.g. White 1982).

[2] Almost all research in behavioural genetics fits this pattern (see e.g. Plomin, DeFries, and McClearn 2008). For a review of relevant research in economics, see Benjamin et al. (2011). Some more recent economics papers use polygenic scores (Barth, Papageorge, and Thom 2020; Papageorge and Thom 2020; Ronda et al. 2020; Sias, Starks, and Turtle 2020).

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, potential spouses trade social status for "good genes," and the two become associated.

We then test for SGAM using contemporary data from the UK. Our dependent variable is the polygenic score for educational attainment (PSEA) of a person's spouse. PSEA is a measure, derived solely from an individual's DNA, predicting the number of years that person will spend in education (Lee et al. 2018). To detect SGAM, and avoid the potential confound of pure genetic assortative mating, 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 (Black, Devereux, and Salvanes 2005; Booth and Kee 2009). We show that people born earlier within a sibling group marry spouses with different genetics, specifically, higher PSEA. We present evidence that this effect is mediated by measures of socio-economic status, including educational attainment and labour market income.

Related literature

There are several ways assortative mating can affect social inequality (Fernández and Rogerson 2001). The simplest is that if couples assort with respect to some characteristic, the resulting households will have more variance in that characteristic than if couples match randomly. For instance, if rich people only marry each other, then the wealth distribution of the resulting households will be more unequal than if they sometimes marry poor people. If the characteristic is inherited from parents to children, then this will carry over into higher inter-*individual* inequality in the next generation.

Assortative mating on one dimension can also affect inequality in other dimensions. For example, if parents invest in their children's human capital, and if households are credit-constrained, then assortative mating on income can increase inequality in children's human capital (Fernandez, Guner, and Knowles 2005). The mechanism is purely environmental, not genetic: some children are raised in a higher-quality environment (e.g. with more books and educational toys), and variance in these environments is increased by assortative mating.

Empirically, it is often thought that assortative mating has increased since World War II in developed societies, and that this has helped to increase inequality over time. Several papers test these claims against US and international data (Fernandez, Guner, and Knowles 2005; Schwartz and Mare 2005; Greenwood et al. 2014; Eika, Mogstad, and Zafar 2019). Overall, there is better evidence for a rise in assortative mating until the 1980s than afterwards, and the case that it has contributed to household inequality is not proven.

Differences in human capital across socio-economic groups also have a genetic component [Tambs et al. (1989);

@Trzaskowski_2014; Rimfeld et al. (2018)]. While most human geneticists treat genetic endowments as an independent variable, there is also research into factors that shape the distribution of genetic variants. The best-known explanation for the association between SES and genetics is that in meritocratic societies, "good 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). In a wholly non-meritocratic society, where social status was ascribed rather than earned, this mechanism could not take effect. 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).

Meritocratic mobility and SGAM are not competing explanations: both may operate in any specific setting. Both require that SES and genetics are inherited, which holds to varying degrees in all human societies. While mobility requires a degree of social openness, so that "good genes" can causally affect a person's social status, SGAM does not need this: 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, SGAM does require assortative mating (on a combination of SES and genetics), while meritocratic mobility does not. As a result, SGAM is likely to apply to a historically wider range of societies. Whilst a degree of meritocracy exists in modern capitalist economies, opportunities to earn status have been far more limited in most societies throughout history (Smelser and Lipset 1966). On the other hand, assortative mating is likely to be a cultural universal, and SGAM has probably existed in all societies with inherited status, i.e. all societies other than egalitarian hunter-gatherers (Fried 1967). Thus, SGAM predicts that genetic differences across social status are likely to be historically ancient.

Another difference is in the chain of causality. In meritocratic mobility, 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. A shock to one individual's status causes their spouse to have different genetic variants; again, both genetics and status are then inherited. As a result, shocks to social status may be encoded in the genetics of subsequent generations.

Another factor affecting the distribution of genetic endowments is genetic assortative mating (GAM), the phenomenon that people with similar genes marry each other. Just as assortative mating can increase inequality in income or wealth, GAM can increase genetic inequality, i.e. the population variance of genetic endowments. Recent research has confirmed the long-standing conjecture that GAM takes place in contemporary human populations (Hugh-Jones et al. 2016; Robinson et al. 2017; Howe et al. 2019). Geneticists have also developed the concept of cross-trait assortative mating (Sundet et al. 2005; Beauchamp et al. 2010), which 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.

To sum up, there are two parallel research traditions on the inheritance of inequality within families. Each has explored the effects of assortative mating on different kinds of inequality, and each has developed theories in which different dimensions of inequality interact. It seems natural to bring the two literatures together. Here, we do that, extending the idea of cross-trait assortative mating to encompass both socially inherited status, and biologically inherited genetic variants. This leads to a new explanation of the link between socio-economic status and genetics.

The underlying observations behind SGAM are hardly new. That status and physical attractiveness tend to assort in marriage markets is a commonplace, as well as a perennial theme of literature. In the Iliad, powerful leaders fight over the beautiful slave-girl Bryseis. In Jane Austen's novels, wealth, attractiveness and "virtue" all make a good match. Marx (1844) wrote "the effect of ugliness, its repelling power, is destroyed by money." And Donald Trump claimed: "part of the beauty of me is that I am very rich." There is also a rich literature on mate preference from evolutionary psychology (Buss 1989; Buss and Barnes 1986; Buss and Schmitt 2019). Attractive mate characteristics include elements of social status ("high earning capacity," "professional status") as well as traits that are partly under genetic influence ("intelligent," "tall," "kind," "physically attractive").

Despite this, surprisingly few papers have analysed the consequences of SGAM, either within genetics or economics. In economics, several empirical papers examine cross-trait matching in marriage markets. Chiappori, Oreffice, and Quintana-Domeque (2012) show empirically that body mass index and SES (specifically, earnings and education) are both valued in a match. Oreffice and Quintana-Domeque (2010) show that height and BMI are associated with spouse earnings. Dupuy and Galichon (2014) find spouse matching on multiple independent dimensions, including education, height, BMI and personality; see also Chiappori, Oreffice, and Quintana-Domeque (2018). However, these papers do not use genetic data, and do not attempt to disentangle genetic and environmental contributions to attractiveness. They also do not theorize the consequences of SGAM. Within genetics, 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.

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 $0 \leq k \leq 1$. Our sole condition on G and S is that a positive measure of the population has attractiveness A where the distribution of $(G, S)|A$ is non-degenerate, i.e. not everybody with attractiveness A is both genetically and socially identical. We assume that people marry at random within their attractiveness isoquant. One microfoundation is that A measures the utility a person provides as a spouse, and the marriage market works via a Gale-Shapley (1962) matching mechanism. Or, A could represent inputs into a household production function where male and female inputs are complementary (Becker 1973). Other interpretations are possible: for instance, different values of A could represent segmented marriage markets, like different schools or workplaces.

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 pure *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:

[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.

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

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

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

This is a strong simplifying assumption; we relax it below. For real world examples approximated by the model, 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* $corr(G_p, S_p) > 0$.

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 marriage markets.

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

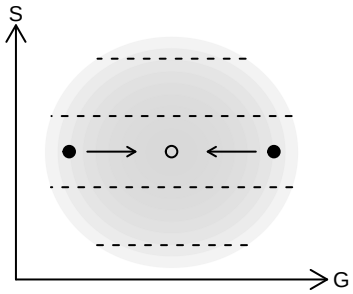
$$\begin{aligned} g_{c(i)} &= \bar{g}_i + \varepsilon_i^G \\ s_{c(i)} &= \bar{s}_i + \varepsilon_i^S \end{aligned}$$

where

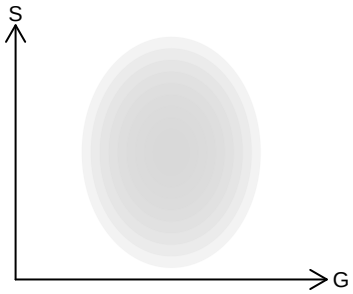
$$\bar{g}_i = \frac{g_i + g_{p(i)}}{2}; \bar{s}_i = \frac{s_i + s_{p(i)}}{2};$$

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

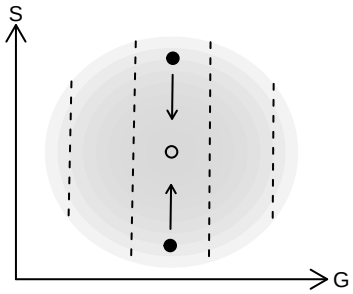
Proposition 2. *1. If* σ_G^2 *and* σ_S^2 *are small enough and* $corr(G_p, S_p) \geq 0$*, then* $corr(G_c, S_c) > corr(G_p, S_p)$ *for* $k \in (0, 1)$.



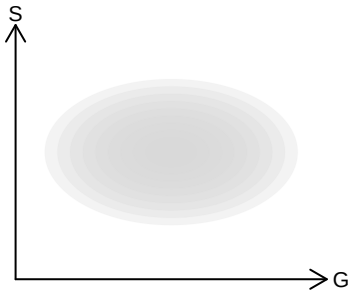
(a) Caste society ($k = 1$): parents



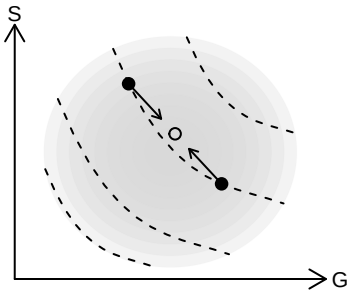
(b) Caste society: children



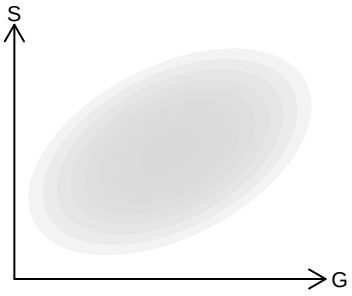
(c) Egalitarian society ($k = 0$): parents



(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.

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

$$A(g_i, s_i) = (1 - k)g_i + ks_i \tag{3}$$

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.

Recent empirical work shows high persistence of SES over time, in particular at the top. One suggested reason for this is that unmeasured family characteristics persist along with measured wealth (Clark and Cummins 2015). Our model captures this idea. For simplicity, assume the original model with no noise in children's values of g or s (condition ((2))), and let condition ((3)) hold. Consider a regression of children's social status on parent's social status:

$$s_{c(i)} = \alpha + \beta_S s_i + \varepsilon_{c(i)}.$$

The value of β_S will be less than one, since parents match on downward sloping isoattractiveness curves: within each curve, relatively wealthy individuals match with less wealthy individuals on average. Now consider a regression of child on parent "attractiveness" A:

$$A_{c(i)} = \alpha + \beta_A A_i + \varepsilon_{c(i)}.$$

Since children are on the same attractiveness curve as both their parents, $\beta_A = 1$. Thus, regressions on measured components of social status will underestimate true persistence over time, embodied in genetic variation. Indeed, grandparents' social status will independently predict grandchildren's social status, even after controlling for parents' social status, because of the unmeasured pathway via parents' genetics.

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 meritocratic mobility 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;[4]
- both reported the same homeownership/renting status, length of time at the address, and number of children;
- attended the same UK Biobank assessment centre on the same day;
- both reported living with their spouse ("husband, wife or partner");
- consisted of one male and one female.

We also eliminate all pairs where either spouse appeared more than once in the data. This leaves a total of 35682 pairs. Some of these 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 use genetic relationships. Some respondents in the Biobank sample have a child who is also in the sample, as inferred from genetic data. Among our spouse pairs, 511 have a genetic child of at least one partner in the sample. For 441 of these, the child is the genetic child of both partners. If this subsample is representative, then at least 86% 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. As a point of comparison, 11% of families with dependent children included a stepchild in England and Wales in 2011 (National Statistics 2014).

It is still possible that some pairs in our data may not be actual spouses. In the appendix, to sign any possible bias in our estimates resulting from this, we use a dataset of known "fake pairs." We show that estimated coefficients of interest are closer to zero among these fake pairs than among our "real pairs." Because of this, any fake pairs remaining in our data are likely to bias our coefficients towards zero.

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), re-estimated excluding UK Biobank participants, to calculate PSEA.[5] We normalize the score to have mean 0 and variance

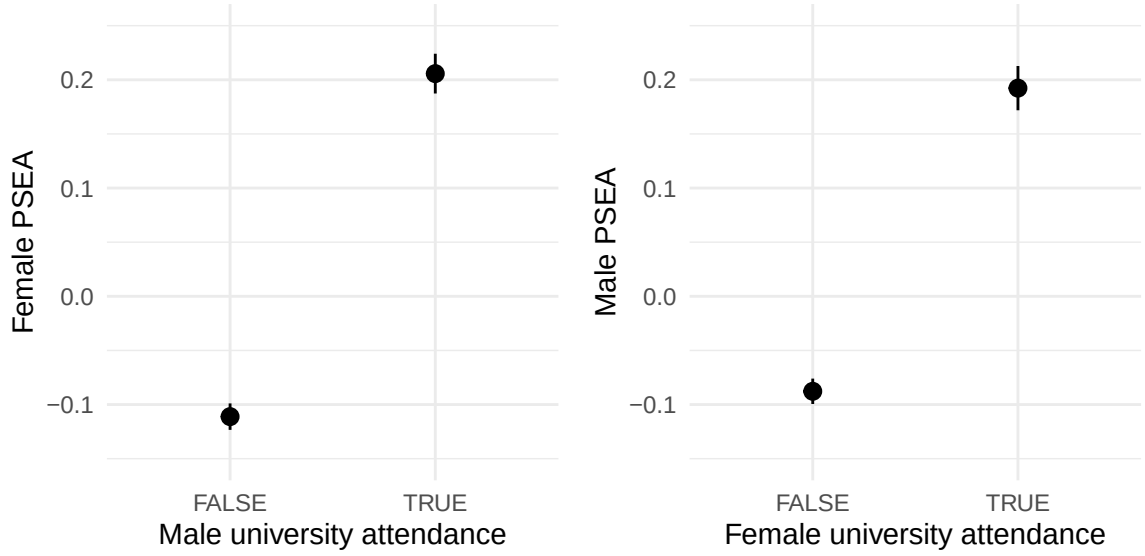
[4] Typically a UK postcode contains about 15 properties.

[5] PSEA was computed by summing the alleles across ~1.3 million genetic variants weighted by their effect sizes as estimated in genome-wide association studies (GWASs) that excluded UK Biobank. PSEA was then residualized on the first 100 principal components of the SNP array data. Further details can be found in @abdellaoui2019genetic.

1.

PSEA is known to correlate with measures of education, including university attendance and years of full-time education; from within-siblings regressions, where PSEA is randomly assigned by the "lottery of meiosis," we know that this correlation is at least partly causal. We recheck these facts within the UK Biobank sample. In a simple linear regression (N = 408,524) of university attendance on PSEA, a one-standard-deviation increase in PSEA was associated with a 9.23% increase in the probability of university attendance ($p < 2e\text{-}16$). In a within-siblings regression among genetic full siblings (N = 36,748), the increase was 4.46% ($p < 2e\text{-}16$). This suggests that about half of the raw correlation of PSEA with university attendance is down to confounds like good environments or parental nurture, while the remainder is causal.

Figures 2 and 3 illustrate the possible consequences of SGAM. The X axis shows a measure of one partner's socioeconomic 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.

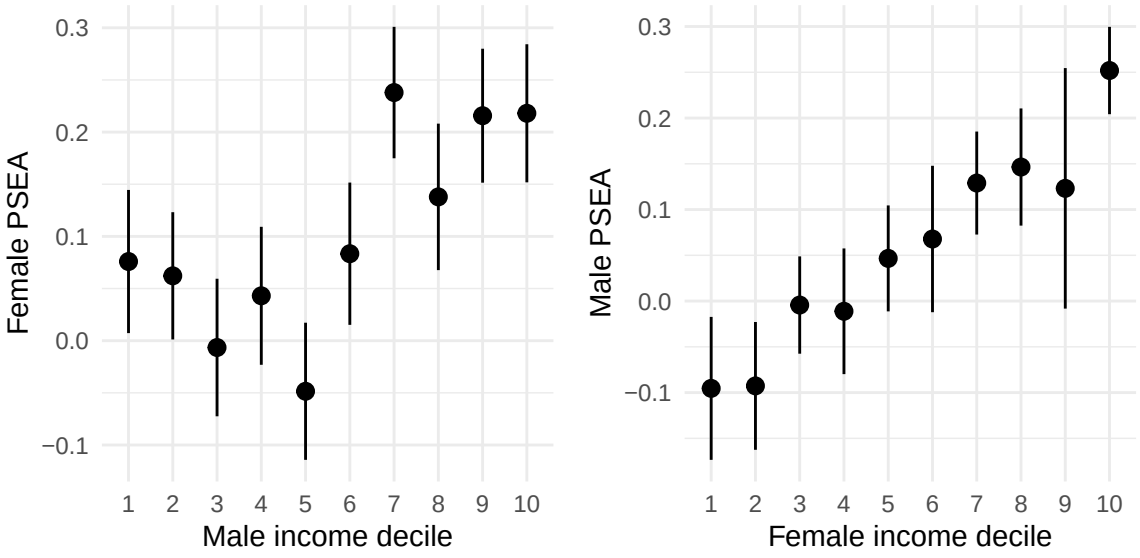


(a) Female PSEA by male educational attainment

(b) Male PSEA by female educational attainment

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

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,



(a) Female EA3 by male income

(b) Male EA3 by female income

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

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.[6] 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). For most regressions, we use only respondents with between 1 and 7 siblings, i.e. with a family size of 2-8. 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

[6] 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.

entire life course.

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.

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.

Decomposing 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:

$$Y = \delta + \alpha \cdot BO + \beta \cdot \mathbf{X} + \varepsilon, \tag{4}$$

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:

$$Y = BO \cdot Y_1 + (1 - BO) \cdot Y_0 \tag{5}$$

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 (5) 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:

$$Y_d = \kappa_d + \alpha_d \cdot \theta_d + \beta_d \cdot \mathbf{X} + \tilde{\varepsilon_d}, \tag{6}$$

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.

$$Y_d = \tau_d + \sum_{j \in J_m} \alpha_d^j \cdot \theta_d^j + \beta \cdot \mathbf{X} + \varepsilon_d, \tag{7}$$

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 (7) into (5) we obtain:

$$Y = \tau_0 + \tau \cdot BO + \sum_{j \in J_m} \alpha^j \cdot \theta^j + \beta \cdot \mathbf{X} + \varepsilon, \tag{8}$$

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 (8) by OLS will generate unbiased estimates of $\alpha^j, j \in J_m$ 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 (8), 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 = 6.93e-08).

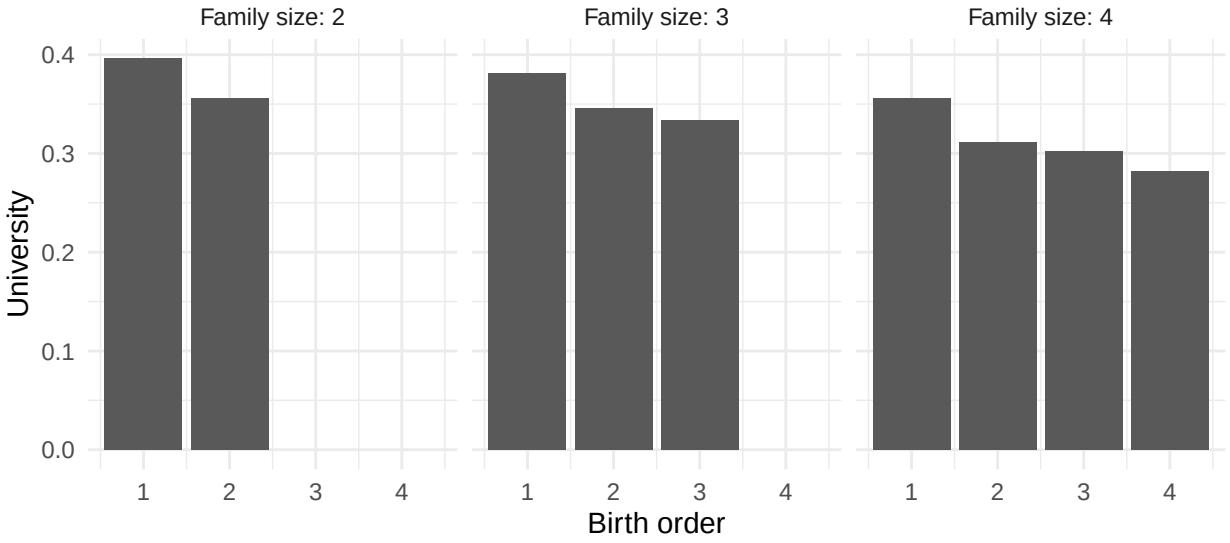


Figure 4: University attendance by birth order and family size

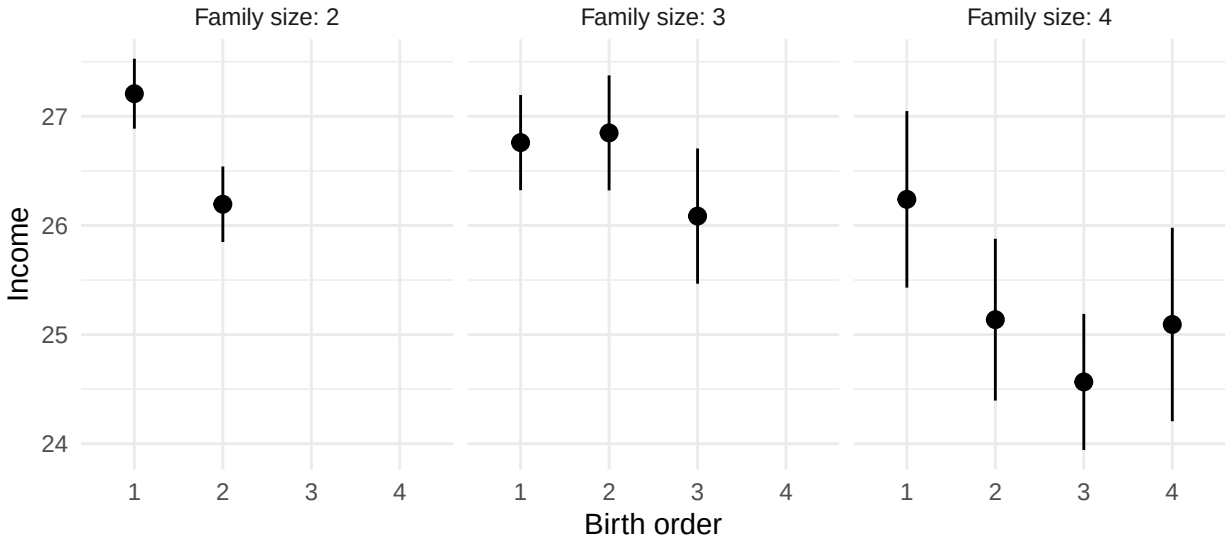


Figure 5: Median income of first job by birth order and family size

Next we run regressions of spouse PSEA on birth order, within our dataset of spouse pairs. Table 1 reports the results.

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 mother'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.[7] Including mother's age means we can separate the effect of mother's age from birth order. This reduces the N by a lot, since only respondents with live mothers reported the necessary data. However, the effect of birth order jumps in size. Meanwhile, mother'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.[8]

Table 1: Regressions of spouse PSEA on birth order

	(1)	(2)	(3)
Birth order	-0.0105	-0.0085	-0.0412 *
	(0.0061)	(0.0061)	(0.0186)
Own PSEA		0.0661 ***	0.0604 ***
		(0.0082)	(0.0118)
Mother's age at birth			0.0108 ***
			(0.0027)
Family size dummies	Yes	Yes	Yes
Birth month dummies	No	Yes	Yes
Birth year dummies	No	Yes	Yes
N	25439	25395	9175
R2	0.003	0.011	0.014

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors clustered by assessment centre. Respondents with 1-7 siblings included.

[7] There are also potential genetic effects, though recent research has rejected these in favour of "social" explanations (Kristensen and Bjerkedal 2007; Black, Devereux, and Salvanes 2011). 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 father's age at respondent's birth.

[8] 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.

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.[9]

Table 2 shows the results. Column 1 shows the effect of birth order, using the same specification as column 3 of the previous table. (We exclude respondents without data on income or university attendance, to make comparison easier across the columns: this reduces the N.) 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 3 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 university as a mediator. Columns 3 and 4 repeat the exercise for female respondents, using their birth order to predict male spouses' PSEA. Effect sizes are smaller for female respondents. This might be because social status is less advantageous to females in the marriage markets, or because females place less value on PSEA-related phenotypes in a spouse. Columns 5 and 6 use only couples with children. University attendance still seems to mediate the birth order effect.

Lastly, we examine whether the level of assortative mating has increased over time within our sample. This has been an important question in the literature, since it is might be a cause of increasing inequality. Table 4 splits our sample by year of birth at 1954, and estimates columns 2-3 of Table 2 for each subsample. The broad pattern of coefficients is the same, although the lower N makes all coefficients insignificant. Interestingly, the effect of birth order on spouse

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

PSEA is about 40% larger for the later-born group, We do not place too much emphasis on this result, as it could be driven by characteristics of the sample.

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

	(1)	(2)	(3)	(4)	(5)
Birth order	-0.0412 *	-0.0256	-0.0132	-0.0024	0.0044
	(0.0186)	(0.0179)	(0.0175)	(0.0378)	(0.0373)
University			0.2287 ***		0.1508 ***
			(0.0305)		(0.0378)
Income				0.0030 **	0.0024 *
				(0.0009)	(0.0009)
Own PSEA	0.0604 ***	0.0491 ***	0.0340 *	0.0386 +	0.0279
	(0.0118)	(0.0120)	(0.0126)	(0.0211)	(0.0210)
Mother's age at birth	0.0108 ***	0.0076 *	0.0050 +	0.0068	0.0054
	(0.0027)	(0.0028)	(0.0028)	(0.0050)	(0.0051)
Fluid IQ		0.0341 ***	0.0184 *	0.0243 +	0.0154
		(0.0055)	(0.0065)	(0.0121)	(0.0126)
Height		0.0029 *	0.0022	0.0054 +	0.0050 +
		(0.0014)	(0.0014)	(0.0028)	(0.0028)
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	9175	8977	8975	2982	2981
R2	0.014	0.020	0.030	0.033	0.037
logLik	-12846.738	-12538.313	-12489.041	-4207.137	-4198.973
AIC	25797.477	25184.626	25088.081	8524.274	8509.946

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors clustered by assessment centre.

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

	Male respondents	Male respondents	Female respondents	Female respondents	With children	With children
Birth order	-0.035	-0.023	-0.019	-0.007	-0.027	-0.015
	(0.023)	(0.022)	(0.020)	(0.020)	(0.023)	(0.022)
University		0.257 ***		0.198 ***		0.226 ***
		(0.041)		(0.033)		(0.033)
Mother's age at birth	0.008 *	0.005	0.007 +	0.005	0.008 *	0.006
	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)
Own PSEA	0.047 ***	0.029 *	0.051 **	0.038 *	0.052 ***	0.036 *
	(0.011)	(0.012)	(0.014)	(0.014)	(0.013)	(0.014)
Fluid IQ	0.044 ***	0.026 *	0.025 **	0.011	0.040 ***	0.024 **
	(0.009)	(0.011)	(0.008)	(0.007)	(0.006)	(0.007)
Height	0.004	0.004	0.005 *	0.005 *	0.002	0.001
	(0.004)	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)
N	4108	4107	4869	4868	8033	8031
R2	0.029	0.042	0.023	0.030	0.023	0.033
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

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors clustered by assessment centre.

Table 4: Regressions of spouse PSEA on birth order: early and late born

	Born 1934-54	Born 1934-54	Born 1955-70	Born 1955-70
Birth order	-0.020	-0.007	-0.028	-0.015
	(0.020)	(0.020)	(0.028)	(0.028)
University		0.240 ***		0.221 ***
		(0.044)		(0.037)
Mother's age at birth	0.005	0.002	0.009 +	0.007
	(0.005)	(0.004)	(0.005)	(0.005)
Own PSEA	0.060 **	0.044 *	0.038 +	0.024
	(0.017)	(0.016)	(0.021)	(0.022)
Fluid IQ	0.033 **	0.016	0.036 ***	0.021 **
	(0.009)	(0.010)	(0.005)	(0.006)
Height	0.004	0.003	0.002	0.002
	(0.002)	(0.002)	(0.002)	(0.002)
N	4408	4406	4569	4569
R2	0.019	0.030	0.023	0.033
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors clustered by assessment centre.

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.[10] Second, there could be selection 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 has 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 the latter two problems, we run balance tests on 33 different polygenic scores, standardized to have mean zero and unit variance.[11] We regress each score on own birth order, controlling for family size. No scores were significant at $p < 0.10/33$. Four scores were significant at $p < 0.10$, all with coefficients less than 0.02. Table 7 in the appendix reruns our regressions controlling for these scores. Results are almost unchanged. To test whether polygenic scores might vary across birth orders within a 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 $p < 0.001$. However, among families of size 3, there is a significant positive correlation of birth order with own PSEA (effect size 0.0299, $p = 0.039$) . Table 9 in the appendix therefore reruns our regressions with families of size 3 excluded. Results are substantially unchanged.

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 makes us more confident that birth order is indeed exogenous to genetics.

Conclusion

Our empirical analysis shows that in a modern Western democracy, earlier-born children had spouses with higher PSEA. We also provide evidence that these effects are mediated by social status: income and education.

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

[11] 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.

Inequality gets transmitted within families by multiple mechanisms, from bequests to investment in children's human capital. Our analysis shows that genetics are another mechanism. Socio-economic status in one generation translates into the next generation's genetic endowments, via marriage markets. This may be one reason why inequality between families is surprisingly persistent in the long-run. Because our measures of PSEA are so noisy, it is hard to make confident estimates of the effect size of the genetic channel. This will be an important research topic as polygenic scores become more accurate.

Our explanation for the genes-SES gradient applies to a wider range of societies than the leading alternative explanation. If this mechanism has operated, then we would expect the gradient to be widespread or universal among all stratified societies. In fact, people in many societies have believed 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").[12] This belief has been explained by the psychological 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 any case, our prediction may be directly testable, for example by using ancient DNA samples.

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 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.

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

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

different societies.

Marriage markets are likely to respond to other features of the economy. For instance, a society's level of redistribution could affect the relative weight of social status in the attractiveness function. If so, then part of the effect of large-scale policy changes may take place over the very long run, by altering marriage markets and thus affecting the genes-status gradient. Intuitively one would expect that a more "egalitarian" society and marriage market would lead to a weaker genes-status gradient. But in our model, the covariance between genes G and status S is non-monotonic in the marriage market parameter k, so this intuition may not always hold true. A natural extension of our theory would be to endogenize the marriage market to investigate further. In any case, the potential for economic changes to have such long-run effects suggests the potential for hysteresis, with initial policy differences cumulating over time via their effect on genetic inequality.

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 individual 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." 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.

Appendix

Proof of Proposition 1

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.

To prove the second part, write

$$corr(G_j, S_j) = \frac{cov(G_j, S_j)}{\sqrt{var(G_j)var(S_j)}} \text{ for both generations } j \in \{p, c\}. \tag{9}$$

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.

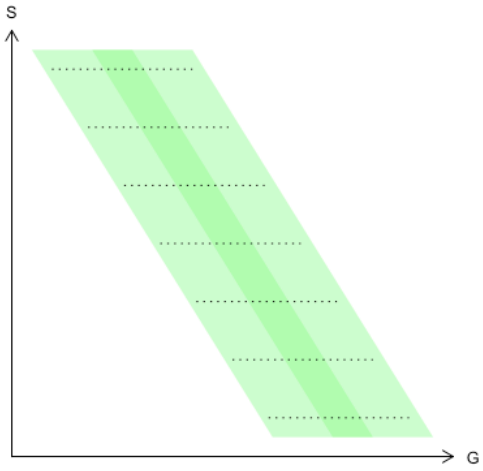


Figure 6: Correlation counterexample

Proof of Proposition 2

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

$$\begin{aligned} cov(G_c, S_c) &= cov(\bar{G} + \varepsilon^G, \bar{S} + \varepsilon^S) \\ &= cov(\bar{G}, \bar{S}) + cov(\varepsilon^G, \bar{S}) + cov(\bar{G}, \varepsilon^S) + cov(\varepsilon^G, \varepsilon^S). \end{aligned} \tag{10}$$

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 (9)

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 (10) 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 reruns our central regressions, dropping the control for mother's age at birth. Results show the same pattern as in the main text: the negative coefficient for birth order shrinks when controls are added, and then changes sign when university attendance is added as a potential mediator. However, the birth order effect is smaller overall, and is never significant.

Table 6 reruns our central regressions but includes a separate coefficient for each position in the birth order (with firstborn as the baseline). The basic pattern of our main result holds: birth order coefficients are negative and increase going down the birth order; in column 2, adding controls reduces their size; in column 3, adding university attendance reduces their size and significance still further. Results are slightly messier when we add income, probably because of the low N.

We also ran a specification with separate birth order dummies within each family size. Figure 7 shows 95% confidence intervals for the birth order coefficients, from the column 2 specification including height and IQ controls but no mediators. Most birth order coefficients are negative (compared to the baseline for the firstborn) and become more negative as birth order increases. However, there is some suggestion of a positive "last born" effect. This could happen because last born children receive more attention after elder children have left home.

Table 7 reruns our regressions controlling for 3 polygenic scores (scores for BMI, conscientiousness and neuroticism). Results are very close to those in the main text.

Table 8 reruns relevant columns of Table 2 using age of leaving full-time education as a measure of educational SES, instead of the university attendance dummy. Results are similar to those in the main text: controlling for age of leaving full-time education shrinks the effect of birth order and makes it insignificant.

Table 9 reruns Table 2 excluding families of size 3. Results are very similar to those in the main text.

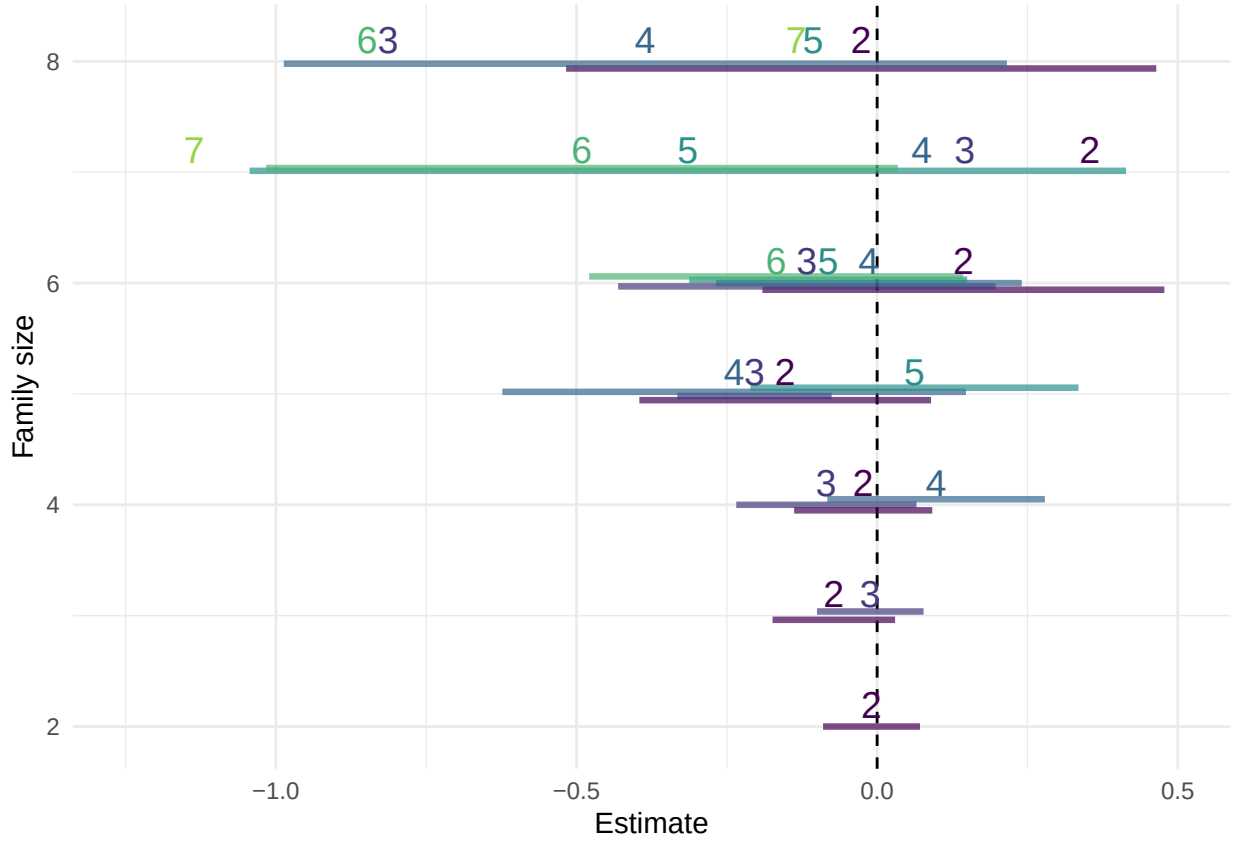


Figure 7: Regressions of spouse PSEA: birth order dummies within different family sizes. Labels show birth order. Lines are 95 per cent confidence intervals clustered by assessment centre. The dummy for birth order 8 is not shown.

Table 5: Regressions of spouse PSEA, without controls for mother's age at respondent's birth

	(1)	(2)	(3)	(4)	(5)
Birth order	-0.0058	-0.0032	0.0022	-0.0021	0.0026
	(0.0144)	(0.0148)	(0.0140)	(0.0148)	(0.0141)
University			0.2216 ***		0.2090 ***
			(0.0259)		(0.0260)
Income				0.0032 ***	0.0020 **
				(0.0007)	(0.0006)
Own PSEA	0.0579 **	0.0521 **	0.0363 *	0.0486 **	0.0350 *
	(0.0165)	(0.0160)	(0.0150)	(0.0165)	(0.0153)
Fluid IQ		0.0197 ***	0.0056	0.0172 ***	0.0048
		(0.0040)	(0.0036)	(0.0038)	(0.0036)
Height		0.0052 ***	0.0043 **	0.0041 **	0.0036 *
		(0.0013)	(0.0013)	(0.0013)	(0.0013)
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	8017	7868	7868	7868	7868
R2	0.011	0.015	0.026	0.017	0.026
logLik	-11292.019	-11062.255	-11019.983	-11052.559	-11016.175
AIC	22688.039	22232.511	22149.966	22215.118	22144.351

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Standard errors clustered by assessment centre.

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

	(1)	(2)	(3)	(4)	(5)
Birth order 2	-0.0560	-0.0345	-0.0213	-0.0416	-0.0403
	(0.0374)	(0.0377)	(0.0381)	(0.0477)	(0.0491)
Birth order 3	-0.0913 *	-0.0611	-0.0337	-0.0126	0.0074
	(0.0405)	(0.0394)	(0.0396)	(0.0689)	(0.0683)
Birth order 4	-0.0686	-0.0124	0.0302	0.0631	0.0850
	(0.0739)	(0.0732)	(0.0706)	(0.1533)	(0.1513)
Birth order 5	-0.0720	-0.0455	-0.0077	0.1850	0.2036
	(0.1075)	(0.1156)	(0.1162)	(0.2917)	(0.2906)
Birth order 6	-0.4608 **	-0.4166 **	-0.3730 **	-0.1257	-0.0947
	(0.1225)	(0.1241)	(0.1231)	(0.4132)	(0.4191)
Birth order 7	-0.5268	-0.3865	-0.3443	-1.7474 ***	-1.7018 ***
	(0.3170)	(0.3383)	(0.3435)	(0.2529)	(0.2451)
Birth order 8	2.5272 ***	2.5650 ***	2.6524 ***	2.6819 ***	2.7110 ***
	(0.1798)	(0.1888)	(0.2003)	(0.3976)	(0.3955)
University			0.2294 ***		0.1519 ***
			(0.0306)		(0.0376)
Income				0.0030 **	0.0023 *
				(0.0009)	(0.0009)
Own EA3	0.0607 ***	0.0496 ***	0.0344 *	0.0392	0.0285
	(0.0116)	(0.0117)	(0.0124)	(0.0210)	(0.0208)
Wald p-value, birth order	0.0000	0.0000	0.0000	0.0000	0.0000
Family size dummies	Yes	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes	Yes
Controls (IQ, height)	No	Yes	Yes	Yes	Yes
N	9175	8977	8975	2982	2981
R2	0.015	0.021	0.031	0.038	0.042
logLik	-12840.097	-12531.936	-12482.324	-4199.400	-4191.119
AIC	25796.194	25183.873	25086.649	8520.799	8506.238

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Standard errors clustered by assessment centre.

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

	(1)	(2)	(3)	(4)	(5)
Birth order	-0.0413 *	-0.0257	-0.0132	-0.0014	0.0053
	(0.0184)	(0.0178)	(0.0174)	(0.0383)	(0.0379)
University			0.2289 ***		0.1507 ***
			(0.0304)		(0.0377)
Income				0.0030 **	0.0023 *
				(0.0009)	(0.0009)
Mother's age at birth	0.0106 ***	0.0075 *	0.0049	0.0065	0.0051
	(0.0027)	(0.0027)	(0.0027)	(0.0050)	(0.0051)
Own PSEA	0.0589 ***	0.0477 ***	0.0326 *	0.0381	0.0274
	(0.0118)	(0.0120)	(0.0127)	(0.0217)	(0.0217)
Fluid IQ		0.0342 ***	0.0184 *	0.0245	0.0157
		(0.0055)	(0.0066)	(0.0122)	(0.0128)
Height		0.0029 *	0.0022	0.0054	0.0051
		(0.0014)	(0.0014)	(0.0027)	(0.0027)
Family size dummies	Yes	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes	Yes
Polygenic score controls	Yes	Yes	Yes	Yes	Yes
N	9175	8977	8975	2982	2981
R2	0.014	0.020	0.030	0.033	0.038

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Standard errors clustered by assessment centre.
Polygenic scores: age at menopause, caffeine, coronary artery disease, neuroticism.

Table 8: Regressions of spouse PSEA using age of leaving full-time education

	(1)	(2)	(3)
Birth order	-0.0256	-0.0072	0.0115
	(0.0179)	(0.0188)	(0.0366)
Age left full-time educ.		0.0500 ***	0.0424 ***
		(0.0056)	(0.0075)
Income			0.0021 *
			(0.0009)
Mother's age at birth	0.0076 *	0.0039	0.0041
	(0.0028)	(0.0029)	(0.0048)
Own PSEA	0.0491 ***	0.0319 *	0.0268
	(0.0120)	(0.0122)	(0.0219)
Fluid IQ	0.0341 ***	0.0152 *	0.0111
	(0.0055)	(0.0064)	(0.0118)
Height	0.0029 *	0.0022	0.0050 +
	(0.0014)	(0.0014)	(0.0027)
Family size dummies	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes
N	8977	8938	2976
R2	0.020	0.034	0.041
logLik	-12538.313	-12426.282	-4187.055
AIC	25184.626	24962.564	8486.111

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors clustered by assessment centre.

Table 9: Regressions of spouse PSEA, excluding family size 3

	(1)	(2)	(3)	(4)	(5)
Birth order	-0.0477 *	-0.0313	-0.0205	-0.0145	-0.0117
	(0.0223)	(0.0224)	(0.0215)	(0.0432)	(0.0424)
University			0.2235 ***		0.1541 ***
			(0.0270)		(0.0324)
Income				0.0010	0.0004
				(0.0013)	(0.0014)
Mother's age at birth	0.0123 **	0.0088 *	0.0063 +	0.0065	0.0053
	(0.0035)	(0.0035)	(0.0036)	(0.0044)	(0.0045)
Own PSEA	0.0581 ***	0.0464 ***	0.0303 *	0.0264	0.0135
	(0.0135)	(0.0120)	(0.0120)	(0.0240)	(0.0232)
Fluid IQ		0.0364 ***	0.0224 **	0.0169	0.0087
		(0.0064)	(0.0070)	(0.0167)	(0.0164)
Height		0.0020	0.0011	0.0046	0.0040
		(0.0014)	(0.0014)	(0.0028)	(0.0029)
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	6306	6165	6164	2019	2018
R2	0.018	0.024	0.033	0.037	0.042
logLik	-8804.901	-8589.110	-8557.018	-2846.458	-2840.438
AIC	17711.801	17284.221	17222.037	5798.916	5788.875

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors clustered by assessment centre.

Regressions with "fake pairs"

Our dataset of pairs could still contain pairs who live in the same postcode but are not spouses. These pairs might still show a relationship between one partner's phenotype and the other's genotype. For example, maybe early-born children grow up to live in richer postcodes, along with people who have higher PSEA scores (Abdellaoui et al. 2019). This could then bias the results. If the coefficient for "fake pairs" is absolutely larger (smaller) than for real pairs, then our results will be biased away from zero (towards zero).

To sign the bias, we create a dataset of "known fake pairs." These are opposite-sexed pairs who live in the same postcode, but do not share all the characteristics listed for the real pairs. Specifically, from the list of characteristics used to create our real pairs (same homeownership status, same length of time at address, same number of children, attended same assessment centre, attended on same day, husband reported living with spouse, wife reported living with spouse) the fake pairs ticked exactly 5 out of 7 boxes.

We again use genetic children to confirm that the fake pairs are "real fakes." Out of 817 genetic children of the fake pairs, only 33 were children of both parents. Thus, the vast majority of fake pairs do not appear to be spouses. Table 10 reruns the regressions of Table 1 using the fake pairs. Although the coefficients on birth order are always negative, they are never significant, and they are always absolutely smaller than the corresponding coefficient in the main text. This strongly suggests that any fake pairs remaining in our data will have the effect of biasing our results towards zero.

Table 10: Regressions of PSEA on birth order: fake pairs

	(1)	(2)	(3)
Birth order	-0.0076	-0.0059	-0.0193 +
	(0.0051)	(0.0048)	(0.0095)
Own PSEA		0.0498 ***	0.0525 ***
		(0.0056)	(0.0110)
Mother's age at birth			0.0095 ***
			(0.0019)
Family size dummies	Yes	Yes	Yes
Birth month dummies	No	Yes	Yes
Birth year dummies	No	Yes	Yes
N	22885	22838	9372
R2	0.002	0.007	0.011

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors clustered by assessment centre. Respondents with 1-7 siblings included.

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

Abdellaoui, Abdel, David Hugh-Jones, Loïc Yengo, Kathryn E Kemper, Michel G Nivard, Laura Veul, Yan Holtz, et al. 2019. "Genetic Correlates of Social Stratification in Great Britain." *Nature Human Behaviour* 3 (12): 1332–42.

Barth, Daniel, Nicholas W Papageorge, and Kevin Thom. 2020. "Genetic Endowments and Wealth Inequality." *Journal of Political Economy* 128 (4): 1474–1522.

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.

Becker, Gary S. 1973. "A Theory of Marriage: Part i." *Journal of Political Economy* 81 (4): 813–46.

Becker, Gary S, Scott Duke Kominers, Kevin M Murphy, and Jörg L Spenkuch. 2018. "A Theory of Intergenerational Mobility." *Journal of Political Economy* 126 (S1): S7–25.

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): E7275–84.

Benjamin, Daniel J, David Cesarini, Christopher F Chabris, Edward L Glaeser, David I Laibson, Gene Age, Vilmundur Gunason, et al. 2011. "The Promises and Pitfalls of Genoeconomics."

Black, Sandra E, Paul J Devereux, and Kjell G Salvanes. 2005. "The More the Merrier? The Effect of Family Size and Birth Order on Children's Education." *The Quarterly Journal of Economics* 120 (2): 669–700.

———. 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. 1989. "Sex Differences in Human Mate Preferences: Evolutionary Hypotheses Tested in 37 Cultures." *Behavioral and Brain Sciences* 12 (1): 1–14.

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

———. 1986. "Preferences in Human Mate Selection." *Journal of Personality and Social Psychology* 50 (3): 559.

Buss, David M, and David P Schmitt. 2019. "Mate Preferences and Their Behavioral Manifestations." *Annual Review of Psychology* 70: 77–110.

Chakravarti, Aravinda, and Peter Little. 2003. "Nature, Nurture and Human Disease." *Nature* 421 (6921): 412–14.

Chetty, Raj, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang. 2017. "The Fading American Dream: Trends in Absolute Income Mobility Since 1940." *Science* 356 (6336): 398–406.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. 2014. "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility." *American Economic Review* 104 (5): 141–47.

Chiappori, Pierre-André, Sonia Oreffice, and Climent Quintana-Domeque. 2012. "Fatter Attraction: Anthropometric and Socioeconomic Matching on the Marriage Market." *Journal of Political Economy* 120 (4): 659–95.

———. 2018. "Bidimensional Matching with Heterogeneous Preferences: Education and Smoking in the Marriage Market." *Journal of the European Economic Association* 16 (1): 161–98.

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

Cochran, Gregory, and Henry Harpending. 2013. "Paternal Age and Genetic Load." *Human Biology* 85 (4): 515–28.

Corak, Miles. 2013. "Income Inequality, Equality of Opportunity, and Intergenerational Mobility." *Journal of Economic Perspectives* 27 (3): 79–102.

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.

Dupuy, Arnaud, and Alfred Galichon. 2014. "Personality Traits and the Marriage Market." *Journal of Political Economy* 122 (6): 1271–1319.

Eika, Lasse, Magne Mogstad, and Basit Zafar. 2019. "Educational Assortative Mating and Household Income Inequality." *Journal of Political Economy* 127 (6): 2795–835.

Fernandez, Raquel, Nezih Guner, and John Knowles. 2005. "Love and Money: A Theoretical and Empirical Analysis of Household Sorting and Inequality." *The Quarterly Journal of Economics* 120 (1): 273–344.

Fernández, Raquel, and Richard Rogerson. 2001. "Sorting and Long-Run Inequality." *The Quarterly Journal of Economics* 116 (4): 1305–41.

Fried, Morton Herbert. 1967. *The Evolution of Political Society: An Essay in Political Anthropology*. Vol. 7. Random House.

Furnham, Adrian. 1993. "Just World Beliefs in Twelve Societies." *The Journal of Social Psychology* 133 (3): 317–29.

Gale, David, and Lloyd S Shapley. 1962. "College Admissions and the Stability of Marriage." *The American Mathematical Monthly* 69 (1): 9–15.

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.

Howe, Laurence J, Daniel J Lawson, Neil M Davies, Beate St Pourcain, Sarah J Lewis, George Davey Smith, and Gibran Hemani. 2019. "Genetic Evidence for Assortative Mating on Alcohol Consumption in the UK Biobank." *Nature Communications* 10 (1): 1–10.

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): 1717–17.

Krueger, Alan. 2012. "The Rise and Consequences of Inequality." *Presentation Made to the Center for American Progress, January 12th*.

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.

National Statistics, Office for. 2014. "Stepfamilies in 2011." 2014. https://webarchive.nationalarchives.gov.uk/20160105222243/http://www.ons.gov.uk/ons/rel/family-demography/stepfamilies/2011/stepfamilies-rpt.html.

Oreffice, Sonia, and Climent Quintana-Domeque. 2010. "Anthropometry and Socioeconomics Among Couples: Evidence in the United States." *Economics & Human Biology* 8 (3): 373–84.

Papageorge, Nicholas W, and Kevin Thom. 2020. "Genes, Education, and Labor Market Outcomes: Evidence from the Health and Retirement Study." *Journal of the European Economic Association* 18 (3): 1351–99.

Plomin, Robert. 1994. *Genetics and Experience: The Interplay Between Nature and Nurture*. Sage Publications, Inc.

Plomin, Robert, John C DeFries, and Gerald E McClearn. 2008. *Behavioral Genetics*. Macmillan.

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.

Robinson, Matthew R, Aaron Kleinman, Mariaelisa Graff, Anna AE Vinkhuyzen, David Couper, Michael B Miller, Wouter J Peyrot, et al. 2017. "Genetic Evidence of Assortative Mating in Humans." *Nature Human Behaviour* 1 (1): 1–13.

Ronda, Victor, Esben Agerbo, Dorthe Bleses, Preben Bo Mortensen, Anders Børglum, David M Hougaard, Ole Mors, Merete Nordentoft, Thomas Werge, and Michael Rosholm. 2020. "Family Disadvantage, Gender and the Returns to Genetic Human Capital." Institute of Labor Economics (IZA).

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

———. 2005. "Trends in Educational Assortative Marriage from 1940 to 2003." *Demography* 42 (4): 621–46.

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

Sias, Richard, Laura Starks, and Harry J Turtle. 2020. "Molecular Genetics, Risk Aversion, Return Perceptions, and Stock Market Participation." National Bureau of Economic Research.

Smelser, Neil J, and Seymour Martin Lipset. 1966. *Social Structure and Mobility in Economic Development*. Transaction Publishers.

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.