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

David Hugh-Jones, Oana Borcan & Abdel Abdellaoui

2021-04-23

Abstract

If social status and genetics are both attractive in marriage markets, then they will become associated in subsequent generations. This process provides a new explanation for the surprising persistence of inequality across generations, and for observed genetic differences across socio-economic status. We model Social-Genetic Assortative Mating (SGAM) and test for its existence in 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 more genetic variants that predict educational attainment. We provide evidence that this effect is mediated by individuals' own educational attainment and income. Thus, environmental shocks to socio-economic status are reflected in the DNA of subsequent generations. SGAM reveals a new aspect of the inheritance of inequality in contemporary and historical societies.

Introduction

Inequality persists over generations, and this has important economic effects. Intergenerational mobility is correlated with cross-sectional inequality (Becker et al. 2018; Krueger 2012), which has risen dramatically in high-income countries (Western, Bloome, and Percheski 2008).[1] At the same time intergenerational absolute mobility has fallen (Chetty et al. 2017), although relative mobility has been stable (Chetty et al. 2014). Over the long run, inequality is surprisingly persistent across generations (Clark and Cummins 2015; Solon 2018). To explain inequality and its persistence, social scientists have analysed assortative mating (homogamy) in the marriage market (e.g. Breen and Salazar 2011; Greenwood et al. 2014) and increasing returns to human capital (e.g. Kaplan and Rauh 2013; Becker et al. 2018). It follows

[1] In the United Kingdom the Gini coefficient has increased from 26% to 34.6% between 1977 and 2020. The United States has seen a 10 percentage point rise up to 43.3% during 1962-2013.

that how families are formed, and how they transmit traits and assets to their offspring, are critical for understanding inequality. These processes are traditionally separated into socio-economic and genetic. While educational homogamy is well established, genetic assortative mating has been demonstrated only recently (Hugh-Jones et al. 2016). Similarly, wealthy families pass on advantages to their children through both genetic inheritance and environmental influence (Rimfeld et al. 2018; Björklund, Lindahl, and Plug 2006).[2]

This paper examines a plausible, but under-analysed, aspect of the spouse matching process: that both social status and genetics contribute to a person's attractiveness in marriage markets.[3] As a result, genetics and inherited social status may become associated in subsequent generations. For example, suppose that wealth and intelligence are both attractive qualities in a spouse. Then wealthy people are more likely to marry intelligent people, and their children will inherit both wealth, and genetic variants associated with intelligence. We call this mechanism Social-Genetic Assortative Mating (SGAM).

SGAM may be an important channel for the transmission of inequality. It leads to a hidden dimension of advantage for privileged families – hidden because most social science datasets do not include genetic information. This dimension may help to explain the surprising long-run persistence of inequality (Clark and Cummins 2015). At the same time, this advantage is not an exogenous fact of biology, but endogenous to the social structure: under SGAM, shocks to an individual's social status can be reflected in the genetics of his or her children.

Below, we first outline a theoretical framework where attractiveness in the marriage market is a function of both socio-economic status (SES) and genes. We show that Social-Genetic Assortative Mating in one generation leads to increased correlation between social and genetic traits in the offspring generation . This result provides a new explanation of the association between genetics and socio-economic status (Belsky et al. 2018; Rimfeld et al. 2018; Björklund, Lindahl, and Plug 2006). While existing explanations have focused on meritocratic social mobility (genes cause SES), under SGAM causality goes both ways, from genes to SES and vice versa.

Next, using novel data on matched spouses born between 1935 and 1970 from the UK Biobank, we empirically test the hypothesis that an individual's higher social status attracts spouses with higher genetic potential for educational attainment. Our genetic measure, the Polygenic Score for Educational Attainment (PSEA), derives from large-scale genome-wide association studies (Lee et al. 2018) and is causally related to educational attainment itself, as well as to intelligence and labour market outcomes. Humans also mate assortatively on PSEA, suggesting that it is relevant in marriage markets, presumably via its associated phenotypes (Hugh-Jones et al. 2016). We depart from the assumption

[2] See Sacerdote (2011) for a review of the behavioural genetics and economics literatures on the nature vs nurture debate; for a broader review of the studies on intergenerational transmission of income see Black and Devereux (2010).

[3] *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).

that both socio-economic status and genetic traits can enhance the attractiveness of potential spouses, and are substitutable. This assumption has received support in recent economic studies of marriage markets, which suggest that people trade off physical characteristics for higher earnings (e.g. Chiappori, Oreffice, and Quintana-Domeque 2012) or matching social class (Banerjee et al. 2013).

The endogeneity of social status is the main challenge in identifying the causal effect of social status on the spouse's genetic endowment. For instance, individuals with high education qualifications tend to also have high educational attainment genes, and they may take partners based on genomic similarity (see e.g. Robinson et al. 2017). In order to isolate the causal link from own socio-economic status to partner genes, we use the "accident of birth" as a source of exogenous variation in socio-economics status. Specifically, we use the birth order of individuals in the sample as a "treatment" which affects their partner choice through a range of mechanisms, of which the most salient one is own socio-economic status. It is well documented that earlier-born children enjoy higher parental investment and have better life outcomes, including measures of socio-economic status such as educational attainment and occupational status (Black, Devereux, and Salvanes 2011; Booth and Kee 2009; Lindahl 2008). At the same time, birth order is independent of siblings' genetic endowments, a fact guaranteed by the biological mechanism involved (the "lottery of meiosis").

While birth order is plausibly independent of siblings' genes, it cannot be used as a valid instrument for socio-economic status, because it may affect partner choice through alternative mechanisms. Hence, we steer away from a two-stage procedure and instead rely on a mediation analysis similar to J. Heckman, Pinto, and Savelyev (2013) (henceforth HPS), who decompose the average treatment effect into effects of measured and unmeasured consequences of treatment. Specifically, we estimate a reduced-form model with spouse genes associated with educational attainment as the dependent variable, and own birth order as the main independent variable. We then estimate a model which further includes own socio-economic status. In the latter model, socio-economic status can be interpreted as the measurable component of the effect of birth order that impacts spouse genetics choice, under a set of assumptions. We also include controls to balance covariates across individuals with different birth orders in different cohorts and we control for own genetic traits for education attainment.

We find that a higher own birth order significantly lowers the spouse's genetic variants associated with educational attainment in the reduced-form regressions, at 0.1% significance level. When we include university attendance as a proxy for socio-economic status, birth order is no longer significant, while university attendance increases the spouse's genetic endowment, again at 0.1% significance. A similar pattern holds when we proxy socio-economic status with a measure of income, although the sample is reduced and the coefficients are imprecisely estimated. The results are robust to the inclusion of several own phenotype traits and a rich set of own genetic traits.

Our paper contributes to several literatures. Firstly, we highlight a novel mechanism of assortative mating. The eco-

nomics literature on matching in marriage markets has typically focused on educational similarities (e.g. Pencavel 1998) or social class or caste (e.g. Abramitzky, Delavande, and Vasconcelos 2011; Banerjee et al. 2013), but also sorting based on age, physical traits and ethnicity (Hitsch, Hortaçsu, and Ariely 2010). Matching decisions on the marriage market have also been shown to follow multiple criteria, with some degree of substitutability between them.[4] For instance, Chiappori, Oreffice, and Quintana-Domeque (2012) showed that individuals trade off BMI for partners' income or education and that the marginal rate of substitution between these characteristics is different for males and females. The genetics literature has focused on genetic assortative mating (GAM), the phenomenon that people with similar genes marry each other. Recent research has confirmed the long-standing conjecture that GAM takes place in contemporary human populations (Howe et al. 2019; Hugh-Jones et al. 2016; Robinson et al. 2017). Geneticists have also developed the concept of cross-trait assortative mating (Beauchamp et al. 2010; Sundet et al. 2005), which refers to people with (genes for) e.g. height marrying people with (genes for) e.g. intelligence. As a result, the two types of variation become associated. In this paper we bring the two literatures together, extending the idea of cross-trait assortative mating to encompass both socially inherited status, and biologically inherited genetic variants. Our results confirm that individuals with higher social status are more likely to attract a spouse with higher innate cognitive ability.

Secondly, our findings have implications for understanding the sources of economic inequality. There are several ways assortative mating can affect 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. Research has shown that assortative mating is associated with higher inequality within generations, although there is no conclusive evidence that it underpins the inequality trend over time (Eika, Mogstad, and Zafar 2019).

A more likely driver of the rise in inequality is the increase in market returns to human capital (e.g. Kaplan and Rauh 2013; Eika, Mogstad, and Zafar 2019). In this context, the distribution of human capital is a key contributor to inequality, in addition to inherited wealth.[5] Of course, one part of human capital is acquired and the other is genetic.

[4] 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.

[5] See Adermon, Lindahl, and Waldenström (2018), Mulder et al. (2009).

From twin studies, 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 can be predicted from their genes (Trzaskowski et al. 2014). Studies comparing parent-child income and education associations between adoptees and non-adoptees show that both post-birth environment and pre-birth conditions (genetics and to a lesser extent prenatal environment) contribute to the transmission of wealth and human capital (e.g. Björklund, Lindahl, and Plug 2006). Several channels shape the distribution of heritable ability. Parents with higher ability reap higher market returns, and more successful parents may then pass both higher socio-economic status 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 social institutions (Branigan, McCallum, and Freese 2013; Heath et al. 1985); in a society where social status was ascribed rather than earned, it could not take effect. Indeed, after the fall of communism in Estonia, the heritability of SES increased, presumably because post-communist society allowed higher returns to talent (Rimfeld et al. 2018). Social-genetic assortative mating provides a complementary explanation for the association between genes and SES, one which does not require social meritocracy.

SGAM may also increase social inequality overall, if there are complementarities between genetic and environmental components of human capital. For example, higher-ability parents may make more productive investments in children's human capital (Cunha and Heckman 2007; Cunha, Heckman, and Schennach 2010; J. J. Heckman and Mosso 2014; Kong et al. 2018). Becker et al. (2018) demonstrated how inequality can arise in a model of intergenerational transmission of human capital where high-income and human capital parents are able to invest more in their children's human capital than low-income parents. Thus, by bringing good genes and enriched environments together, SGAM may increase inequality in the next generation.

Thirdly, we contribute to a literature in economics that examines the relationship between genetic and economic variables. Benjamin et al. (2011) is an early review. Several more recent papers use polygenic scores, in particular polygenic scores for educational attainment (Barth, Papageorge, and Thom 2020; Papageorge and Thom 2020; Ronda et al. 2020).

Finally, our paper challenges the widespread assumption, in economics and beyond, that genetic endowment is exogenous to economic characteristics.[6] While this may be tenable in within-generation studies, it ceases to hold in intergenerational models. Social-genetic assortative mating is a causal mechanism going from socio-economic status to genetic traits. Furthermore, our model shows that the strength of this mechanism varies with the structure of the society's marriage market. When both genes and status are both relevant in marriage markets, then they become asso-

[6] A small literature in economics examines the relationship between genetic and economic variables. Benjamin et al. (2011) is an early review. Several recent papers use polygenic scores, in particular polygenic scores for educational attainment (Barth, Papageorge, and Thom 2020; Papageorge and Thom 2020; Ronda et al. 2020).

ciated in the next generation. When social status is irrelevant and only good genes matter – or vice versa – then genes and status do not become associated. Social structure matters for the redistribution of genes, just as for other forms of capital.

The observations behind SGAM are hardly new. That status and physical attractiveness assort in marriage markets is a commonplace and 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." The literature on mate preference from evolutionary psychology (Buss and Barnes 1986; Buss 1989; Buss and Schmitt 2019) confirms that 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"). It is therefore surprising that papers in genetics and economics have not formalised the mechanism and consequences of SGAM.[7]

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 $Var(G) = Var(S) = 1$.[8] 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,

[7] 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.

[8] 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.

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:

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

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

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

where

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

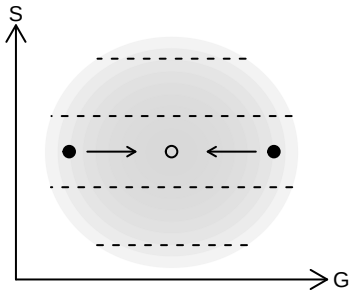
ε^G has mean 0 and variance σ_G^2; and ε^S has mean 0 and variance σ_S^2.

Proposition 2.

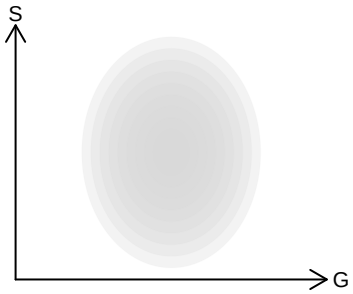
1. *If* σ_G^2 *and* σ_S^2 *are small enough and* $corr(G_p, S_p) \geq 0$*, then* $corr(G_c, S_c) > corr(G_p, S_p)$ *for* $k \in (0, 1)$.
2. *If* ε^G *and* ε^S *are uncorrelated with each other and with* $\bar{G}$ *and* $\bar{S}$*; and if* G_p *and* S_p *are uncorrelated, then* $corr(G_c, S_c) \geq 0$*,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

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

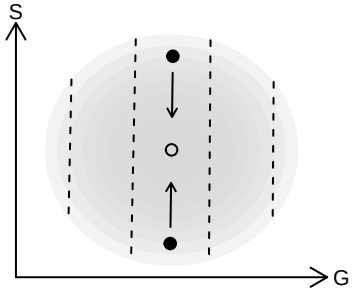
It is worth considering what kind of social arrangements would *violate* these conditions. For example, suppose that parents' combined wealth is inherited by the child with the lowest value of $g_{c(i)}$. This creates a negative correlation between $s_{c(i)}$ and $g_{c(i)}$.



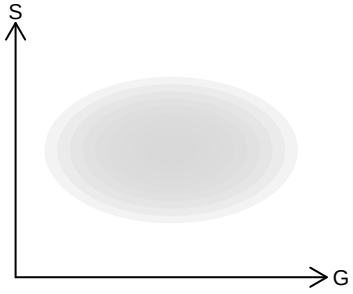
(a) Caste society ($k = 1$): parents



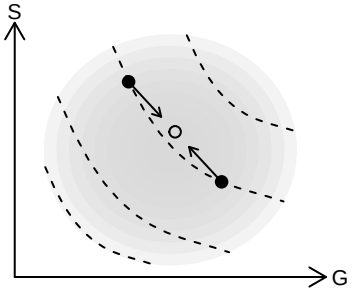
(b) Caste society: children



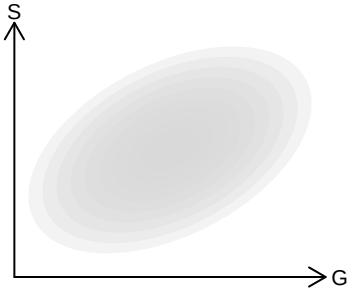
(c) Egalitarian society ($k = 0$): parents



(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.

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;[9]
- 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.

[9] A typical UK postcode contains about 15 properties.

We also eliminate all pairs where either spouse appeared more than once in the data. This leaves a total of 35,682 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 candidate "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 for two reasons. First, 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). Second, PSEA predicts a set of important socioeconomic variables, including not only education but also social and geographic mobility, IQ, future income and wealth (Belsky et al. 2016; Barth, Papageorge, and Thom 2020; Papageorge and Thom 2020).[^discuss]

[^discuss] See Papageorge and Thom (2020) for a detailed discussion of polygenic scores, aimed at economists.

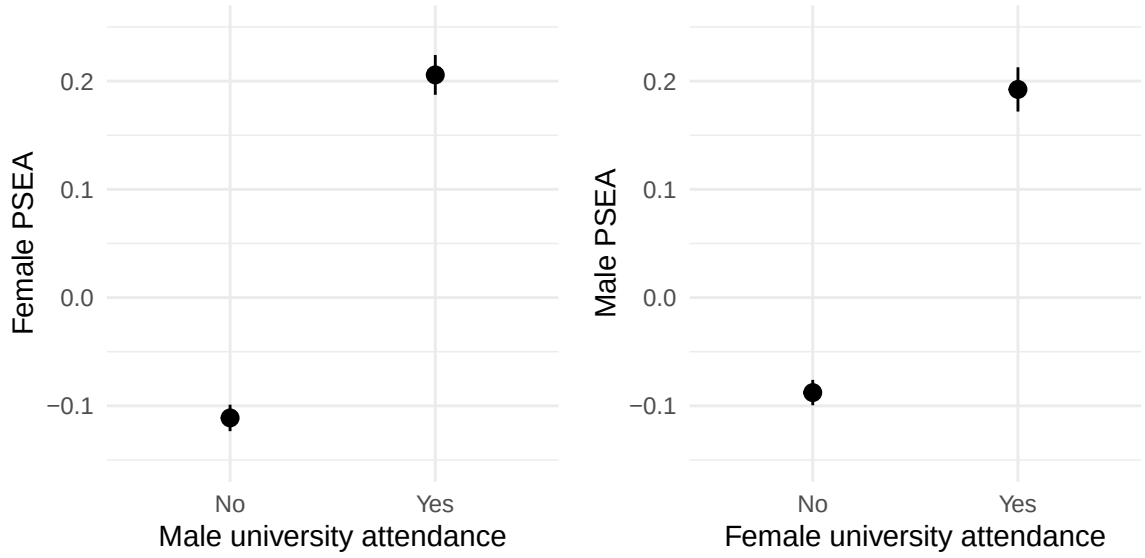
We calculate PSEA using per-SNP summary statistics from (Lee et al. 2018), re-estimated excluding UK Biobank participants.[10] We normalize the score to have mean 0 and variance 1. Because polygenic scores are created from estimates of many presumably tiny effects, they contain a large amount of noise relative to the true best estimator that could be derived from genetic data. For instance, PSEA explains only 11–13% of variance in educational attainment [out of sample; Lee et al. (2018)], whereas the true proportion explained by genetic variation - the heritability - is estimated from twin studies to be about 40% (Branigan, McCallum, and Freese 2013). In addition, polygenic scores are no more guaranteed to be causal than any other independent variable. For example, social stratification by descent

[10] 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 Abdellaoui et al. (2019).

may lead genes to be associated with educational attainment even while playing no causal role (Selzam et al. 2019).

Despite these points, PSEA has non-trivial estimated effects on educational attainment. PSEA correlates with measures of education, including university attendance and years of full-time education; within-siblings regressions, where PSEA is randomly assigned by the "lottery of meiosis," confirm this correlation is at least partly causal (Lee et al. 2018). 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 9.2 percentage point increase in the probability of university attendance ($p < 2 \times 10^{-16}$). In a within-siblings regression among genetic full siblings (N = 36,748), the increase was 4.5 ($p < 2 \times 10^{-16}$). 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. Nevertheless, the causal effect remains substantial: for a rough comparison, the (ITT) effect on college attendance of the Moving To Opportunity experiment in the US was 2.5 percentage points (Chetty, Hendren, and Katz 2016).

Figures 2 and 3 illustrate the core idea 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.

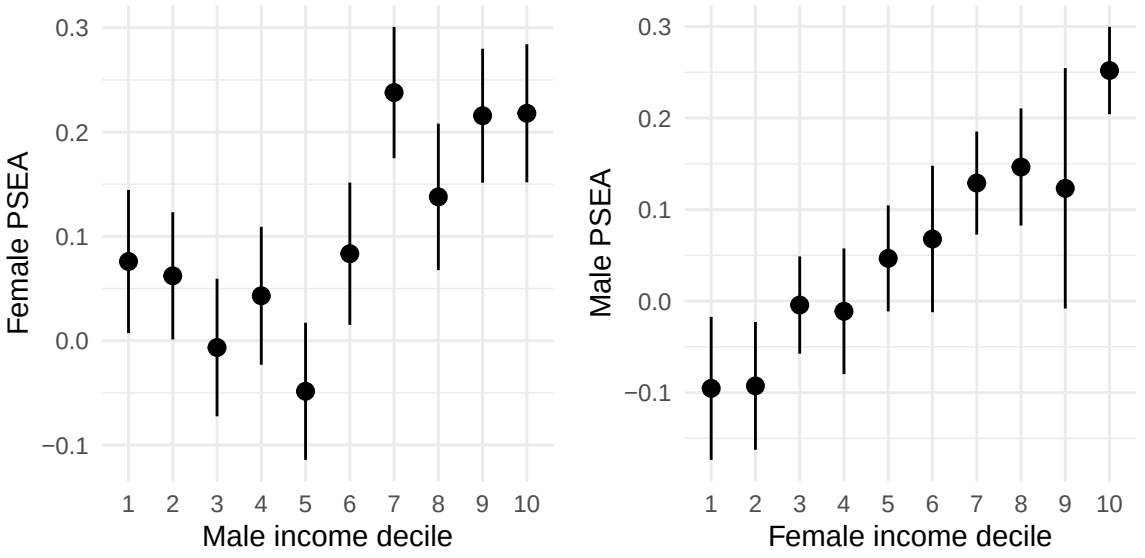


(a) Female PSEA by male educational attainment

(b) Male PSEA by female educational attainment

Figure 2: Spouse PSEA against own university attendance. Lines show 95% confidence intervals

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.



(a) Female EA3 by male income

(b) Male EA3 by female income

Figure 3: Spouse PSEA against own estimated income in first job. Lines show 95% confidence intervals

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, as discussed above; 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.[11] 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 5 siblings, i.e. with a family

[11] 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.

size of 2-6. 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.

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 (J. 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 SGAM 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 J. 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.

Assume B is a variable indicating birth order. The observed outcome (spouse PSEA) is

$$Y = \sum_{b \in B} Y_b \delta_b \tag{4}$$

where Y_1, Y_2 etc. are counterfactual outcomes for the first-born, second-born etc. and δ_b is a dummy variable which is 1 if birth order equals b. Given b, spouse PSEA is assumed to be independent across observations conditional on some predetermined controls which are assumed to be unaffected by b.

Let θ_b be a set of mediators, i.e. proximate outcomes determined by b, which account (at least in part) for the b treatment effect on spouse PSEA. We can think of θ_b as all the effects on attractiveness, such as increments to SES, health, cognitive and non-cognitive skills, that individuals receive due to their birth rank.

Our linear model is

$$Y_b = \kappa_b + \alpha_b \theta_b + \beta_b \cdot \mathbf{X} + \tilde{\varepsilon}_b \tag{5}$$

where $\tilde{\varepsilon}_b$ is a mean-zero residual assumed independent of θ_b and $\mathbf{X}$. We simplify by assuming that $\beta_b = \beta$ and $\alpha_b = \alpha$ for all b, i.e. that the effects of $\mathbf{X}$ and θ_b don't differ by birth order.

We can split the mediators in θ_b into a set J_m of measured mediators, including measures of SES, and a set J_u of mediators that we cannot measure. Write

$$Y_b = \tau_b + \sum_{j \in J_m} \alpha^j \theta_b^j + \beta \cdot \mathbf{X} + \varepsilon_b \tag{6}$$

where $\tau_b = \kappa_b + \sum_{j \in J_u} \alpha^j E(\theta_b^j)$ and $\varepsilon_b = \tilde{\varepsilon}_b + \sum_{j \in J_u} (\theta_b^j - E(\theta_b^j))$. We assume differences in unmeasured investments due to b are independent of $\mathbf{X}$.

With the assumptions above and substituting (6) into (4) we obtain:

$$Y = \tau_1 + \sum_{b \in B \setminus \{1\}} \delta_b (\tau_b - \tau_1) + \sum_{j \in J_m} \alpha^j \theta^j + \beta \cdot \mathbf{X} + \varepsilon \tag{7}$$

where $\theta^j = \sum_{b \in B} \theta_b^j \delta_b$ and $\varepsilon = \sum_{b \in B} \varepsilon_b \delta_b$; $\tau_b - \tau_1$ measures the contribution of unmeasured variables to the average treatment effect of birth order b, relative to birth order 1; and $\theta^j, j \in J_m$ denote the observed mediators that we can measure.

Estimating the above by OLS will generate unbiased estimates of α^j if θ^j is measured without error and is uncorrelated with the error term ε. Since ε contains both individual disturbances and differences in unmeasured investments due to birth order, the identifying assumptions that need to hold for unbiased OLS estimates are (a) the measured investments (specifically SES) should be independent of unmeasured investments generated by birth order. Failing this, the estimates of α^j will be conflated with the effects of unmeasured investments. (b) The measured investments should be uncorrelated with other shocks $\tilde{\varepsilon}_b$. The overall treatment effect (relative to birth order $b = 1$) can then be decomposed as follows:

$$E(Y_b - Y_1) = \tau_b - \tau_1 + \sum_{j \in J_m} \alpha^j E(\theta_b^j - \theta_1^j) \tag{8}$$

where $\tau_b - \tau_1$ is the unmeasured component of the treatment effect, and $\sum_{j \in J_m} \alpha^j E(\theta_b^j - \theta_1^j)$ is the treatment effect due to measured investments.

By running a least square regression of (8), we can estimate $\tau_b - \tau_1$. If assumption (a) above holds, the part of the birth order treatment effect on spouse PSEA that is due to measured mediators, including SES, can be constructed using the estimated α^j and the effects of birth order on measured investments. Specifically, the part of the birth order effect that is due to university education (or income) on spouse PSEA will be the coefficient of university (income) in the regression of spouse PSEA in equation (7), multiplied by the coefficient of birth order from a model where the dependent variable is an indicator for university (income).

Results

We first regress our measures of socio-economic status, university attendance and income from first job, on birth order in our sample. We also do the same for two non-SES controls that could be affected by birth order: fluid IQ and height. We control for respondent's own PSEA and mother's age at birth (see below). Table 1 shows that birth order strongly predicts both university attendance and logged income. Effect sizes are quite substantial: a single extra elder sibling reduces the chance of attending university by about 8 percentage points, log income by about 0.05, fluid IQ by about 0.3 points on a 13 point test, and height by about three-quarters of a centimetre.

Table 1: Regressions of socio-economic status variables on birth order

	University	Income	Fluid IQ	Height
Birth order	−0.0842 ***	−1.3010 **	−0.2993 ***	−0.7681 ***
	(0.0072)	(0.4619)	(0.0333)	(0.1438)
PSEA	0.0894 ***	1.4594 ***	0.3213 ***	0.1839 +
	(0.0049)	(0.3481)	(0.0217)	(0.0991)
Mother's age at birth	0.0167 ***	0.2790 ***	0.0625 ***	0.1519 ***
	(0.0013)	(0.0779)	(0.0057)	(0.0252)
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes
N	8980	2970	8803	8967
R2	0.076	0.028	0.062	0.017

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors: robust.

Next we run regressions of spouse PSEA on birth order, within our dataset of spouse pairs. Table 2 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.[12] 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.[13]

[12] 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.

[13] Note that parental age would not be a good independent variable for testing genetic encoding of social advantage, since it is likely to correlate

Table 2: Regressions of spouse PSEA on birth order

	(1)	(2)	(3)
Birth order	−0.0084	−0.0066	−0.0325 *
	(0.0072)	(0.0073)	(0.0158)
Own PSEA		0.0655 ***	0.0604 ***
		(0.0064)	(0.0107)
Mother's age at birth			0.0102 ***
			(0.0028)
Family size dummies	Yes	Yes	Yes
Birth month dummies	No	Yes	Yes
Birth year dummies	No	Yes	Yes
N	24624	24580	8970
R2	0.003	0.010	0.013

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors: robust.

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.[14]

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 comparison easier across the columns: this reduces the N.) The remaining columns add potential mediators of birth order effects. Column 2 includes our first measure of socio-economic status: university attendance. As controls, we also include fluid IQ and height, both of which could be affected by birth order and affect spouse matching. Column 3 adds our second measure of socio-economic status, income. Column 4 includes both.

When we add university attendance (column 2), the effect of birth order drops and becomes insignificant, while the coefficient for university is positive and highly significant. Fluid IQ and height are also positive and significant. Controlling for income alone (column 3), birth order again becomes insignificant and the coefficient changes sign. Lastly, when we control for both university and income (column 4), again birth order's effect size is positive.

with parents' genetics.

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

Under the assumptions discussed above, we can estimate the proportion of the birth order effect that is mediated by these variables. Table 4 reports this for each model in columns 2-4. Each estimate is the coefficient of birth order on the mediator, times the coefficient of the mediator on spouse PSEA, divided by the coefficient of birth order on spouse PSEA estimated from column 1, i.e. without mediators. 90% confidence intervals are estimated by 199 bootstraps.

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 5 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; column 2 adds university attendance as a mediator (here, we focus on university attendance alone so as to keep our N large). Columns 3 and 4 repeat the exercise for female respondents, using their birth order to predict male spouses' PSEA. In these subsets, the coefficient on birth order is insignificant. However, the pattern of coefficient sizes is the same as in the main regression: university attendance is highly significant, and adding it reduces the absolute size of the birth order effect. 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. Here, birth order is significant in the base specification; university attendance still seems to mediate the birth order effect.

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

	(1)	(2)	(3)	(4)
Birth order	−0.0325 *	−0.0040	0.0094	0.0158
	(0.0158)	(0.0160)	(0.0296)	(0.0296)
University		0.2282 ***		0.1417 ***
		(0.0242)		(0.0408)
Income			0.0034 **	0.0027 *
			(0.0012)	(0.0012)
Own PSEA	0.0604 ***	0.0338 **	0.0365 +	0.0263
	(0.0107)	(0.0110)	(0.0203)	(0.0206)
Mother's age at birth	0.0102 ***	0.0045	0.0067	0.0054
	(0.0028)	(0.0028)	(0.0050)	(0.0050)
Fluid IQ		0.0179 **	0.0228 *	0.0145
		(0.0057)	(0.0101)	(0.0105)
Height		0.0024 *	0.0050 *	0.0047 *
		(0.0012)	(0.0021)	(0.0021)
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes
N	8970	8778	2920	2919
R2	0.013	0.029	0.032	0.036
logLik	−12561.372	−12215.967	−4118.480	−4111.272
AIC	25222.744	24537.934	8342.961	8330.543

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors: robust.

Table 4: Percent of birth order effects accounted for by mediators, models 2-4

	Model 2	**Model 3**	**Model 4**
University	64.0% (15.6%, 218.8%)	-	219.9% (-258.4%, 223.7%)
Income	-	78.0% (-82.7%, 97.8%)	65.7% (-73.8%, 78.6%)
Fluid IQ	17.9% (0.1%, 69.9%)	121.0% (-149.5%, 202.9%)	80.1% (-114.2%, 104.3%)
Height	6.0% (-0.8%, 23.1%)	68.8% (-82.2%, 81.5%)	66.5% (-74.6%, 72.8%)

Bootstrap 90% confidence intervals in brackets

Table 5: Regressions of spouse PSEA on birth order: subsets

	Male respondents	Male respondents	Female respondents	Female respondents	With children	With children
Birth order	−0.039	−0.008	−0.028	−0.004	−0.036 *	−0.006
	(0.024)	(0.024)	(0.021)	(0.021)	(0.017)	(0.017)
University		0.258 ***		0.198 ***		0.225 ***
		(0.036)		(0.033)		(0.026)
Own PSEA	0.065 ***	0.031 +	0.057 ***	0.036 *	0.064 ***	0.036 **
	(0.016)	(0.016)	(0.015)	(0.015)	(0.011)	(0.012)
Mother's age at birth	0.011 **	0.004	0.009 **	0.005	0.011 ***	0.005
	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
Fluid IQ		0.024 **		0.012		0.024 ***
		(0.008)		(0.008)		(0.006)
Height		0.003		0.005 *		0.002
		(0.003)		(0.002)		(0.001)
N	4093	4011	4877	4767	8023	7846
R2	0.018	0.040	0.017	0.030	0.015	0.032
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: robust.

Robustness

Although all children of the same parents have the same polygenic scores in expectation, it could still be possible that genetics correlates with birth order within the sample. This could happen in three ways. First, siblings with high birth order will typically come from larger families than those with low birth order, and parents of different-sized families are likely to differ systematically on many dimensions, including genetics. We controlled for this by including a full set of family size dummies in the regression. Second, there could be 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.[15] We regress each score on own birth order, controlling for family size. No scores were significant at $p < 0.10/33$. Five scores were significant at $p < 0.10$, all with coefficients less than 0.02. Table 9 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 495 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 11 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 contemporary developed society, earlier-born children had spouses with higher PSEA. We also provide evidence that these effects are mediated by socio-economic status, specifically income and education.

[15] 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.

There are multiple mechanisms by which advantage is transmitted across generations. High-income parents may invest more in their children's human capital, transfer wealth via gifts and bequests, model valuable skills, or provide them with advantageous social networks. They may also pass on causally relevant genetic variants. This channel has been proposed as a reason for the surprising persistence of inequality over generations (Clark and Cummins 2015; Clark 2021). One problem with this theory is that in the absence of assortative mating, genetic variation regresses swiftly to the mean, with coefficient $r = 0.5$ per generation. Thus to explain long-run persistence, the genetic theory seems to require very high levels of genetic assortative mating. SGAM may help to solve this puzzle. Persistence will be increased if, in addition to genetic assortative mating, high SES itself attracts "good genes." At the same time, SGAM changes the interpretation of genetics. Genetic variation is not an exogenous input into the social system, but an endogenous outcome of the structure of the marriage market, as our theoretical model shows.

SGAM also provides a new explanation for the observed association of genes with SES. Unlike meritocratic social mobility, the leading alternative explanation, SGAM may apply to a historically wide 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 (Buss 1989). Thus, SGAM predicts that genetic differences across social status should exist in all stratified societies. In fact, people in many societies have believed that innate traits do vary by social status. The ancient Greeks described the social elite as *καλοί κἀγαθοί* ("fine and good"), while the Roman nobility were the *optimates* ("best").[16] This belief has been explained by the 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 may be held for similar reasons. In future, it may be possible to directly test for genetic differences across social status in ancient DNA samples.

Our empirical analysis has focused on significance rather than effect size. This is by necessity: even the latest estimates of PSEA contain substantial noise compared to the "true" polygenic score. As polygenic scores become more accurate, we will be better able to gauge the true effect size of SGAM on both the inheritance of inequality and the genes-SES gradient. We expect this to be an increasingly important research topic.

In our model, the association between social status and genetic variation depends on the society's marriage market. The association is weaker when marriage markets are very socially egalitarian, with spouse matching driven only by genetics, or very inegalitarian, with matching driven only by status. Thus, there is a non-monotonic relationship between "meritocracy" in marriage markets, and the gene-environment correlation ("rGE") between genes and SES. This logic is different from the standard mechanism, in which more meritocratic societies give genetic variants a

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

greater role in causing labour market success, so that the relationship between social meritocracy and rGE is monotonic. Comparing rGE across societies or over time is beyond our scope here, but we see it as a good area for future theory and empirics. Marriage markets may play an important role in explaining how inequality and intergenerational mobility vary across 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 social changes may take place over the very long run, by altering marriage markets and thus affecting the genes-SES gradient and intergenerational mobility. Intuitively one would expect that a more "egalitarian" society and marriage market would lead to a weaker genes-SES gradient. But in our model, the covariance between genes G and status S is non-monotonic in the weight of status in marriage markets 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 social 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 "environment" (e.g. Chakravarti and Little 2003; Plomin 2019). This idea expresses the fact that our genetic endowment is fixed at birth and cannot be influenced by our social environment (though genes *always* 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 4. 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.

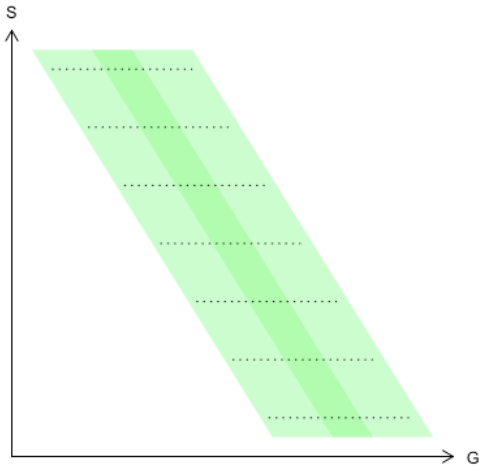


Figure 4: Correlation counterexample

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^2_G and σ^2_S 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^2_G 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 6 reruns our central regressions, dropping the control for mother's age at birth. Results show the same pattern as in the main text: the coefficient for birth order is negative, but changes sign when university attendance is added as a potential mediator. However, the birth order effect is smaller overall, and is never significant. We also ran regressions using father's age instead of mother's age; results are similar to those in the main text.

Table 7 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 generally negative; adding mediators causes them to increase towards zero or to change sign. Birth order effects appear largest for birth order 2-3. However, effects for later birth orders are also imprecisely estimated (since fewer respondents come from large families).

Notably when we add income, dummies for birth order 5 and 6 become large and positive. This could be (for instance) because being the last born has advantages after effects on SES have been netted out. Table 8 runs the same exercise for different subsets: male respondents, female respondents, and couples with children. The basic pattern that birth order coefficients shrink after adding mediators is quite robust. Note however that here, the estimates of effects for birth order 2-3 are larger for females.

We also ran a specification with separate birth order dummies within each family size. Figure 5 shows 95% confidence intervals for the birth order coefficients, from the column 2 specification including height and IQ controls but no mediators. Not surprisingly, coefficients are imprecisely estimated. But most birth order coefficients are negative compared to the baseline for firstborns.

Table 9 reruns our regressions controlling for several polygenic scores. Results are very close to those in the main text.

Table 10 reruns relevant columns of Table 3 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 11 reruns Table 3 excluding families of size 3. Results are very similar to those in the main text.

Chiappori, Oreffice, and Quintana-Domeque (2012) write down a matching model in which a person's attractiveness is summarized by a single index. The linear version of the model can be tested by regressing one partner's characteristics on each of the other partner's characteristics in turn (using Seemingly Unrelated Regressions), and checking that the coefficients have the same proportions across each regression. We do this for birth order and PSEA. Table 12 shows the results. We exclude year of birth dummies, since they cause the estimation procedure to fail.

We also check the proportionality of coefficients. We run a Wald test that

$$\frac{\beta_{PSEA}^{PSEA}}{\beta_{BO}^{PSEA}} = \frac{\beta_{PSEA}^{BO}}{\beta_{BO}^{BO}}$$

where e.g. β_{BO}^{PSEA} is the coefficient of birth order in the regression targeting PSEA. P values are p = 0.37 for males and p = 0.042 for females. So for females there is some evidence against this linear model.

Table 6: Regressions of spouse PSEA, without controls for mother's age at respondent's birth

	(1)	(2)	(3)	(4)
Birth order	−0.0066	0.0026	−0.0029	0.0018
	(0.0073)	(0.0074)	(0.0137)	(0.0137)
University		0.2504 ***		0.2064 ***
		(0.0148)		(0.0250)
Income			0.0033 ***	0.0022 **
			(0.0008)	(0.0008)
Own PSEA	0.0655 ***	0.0366 ***	0.0471 ***	0.0335 **
	(0.0064)	(0.0066)	(0.0120)	(0.0120)
Fluid IQ		0.0165 ***	0.0174 **	0.0051
		(0.0034)	(0.0060)	(0.0062)
Height		0.0019 **	0.0037 **	0.0032 *
		(0.0007)	(0.0013)	(0.0013)
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes
N	24580	23861	7678	7673
R2	0.010	0.028	0.018	0.027
logLik	−34546.361	−33311.976	−10777.375	−10737.107
AIC	69194.721	66731.953	21660.751	21582.215

*** p $<$ 0.001; ** p $<$ 0.01; * p $<$ 0.05; + p $<$ 0.1. Standard errors: robust.

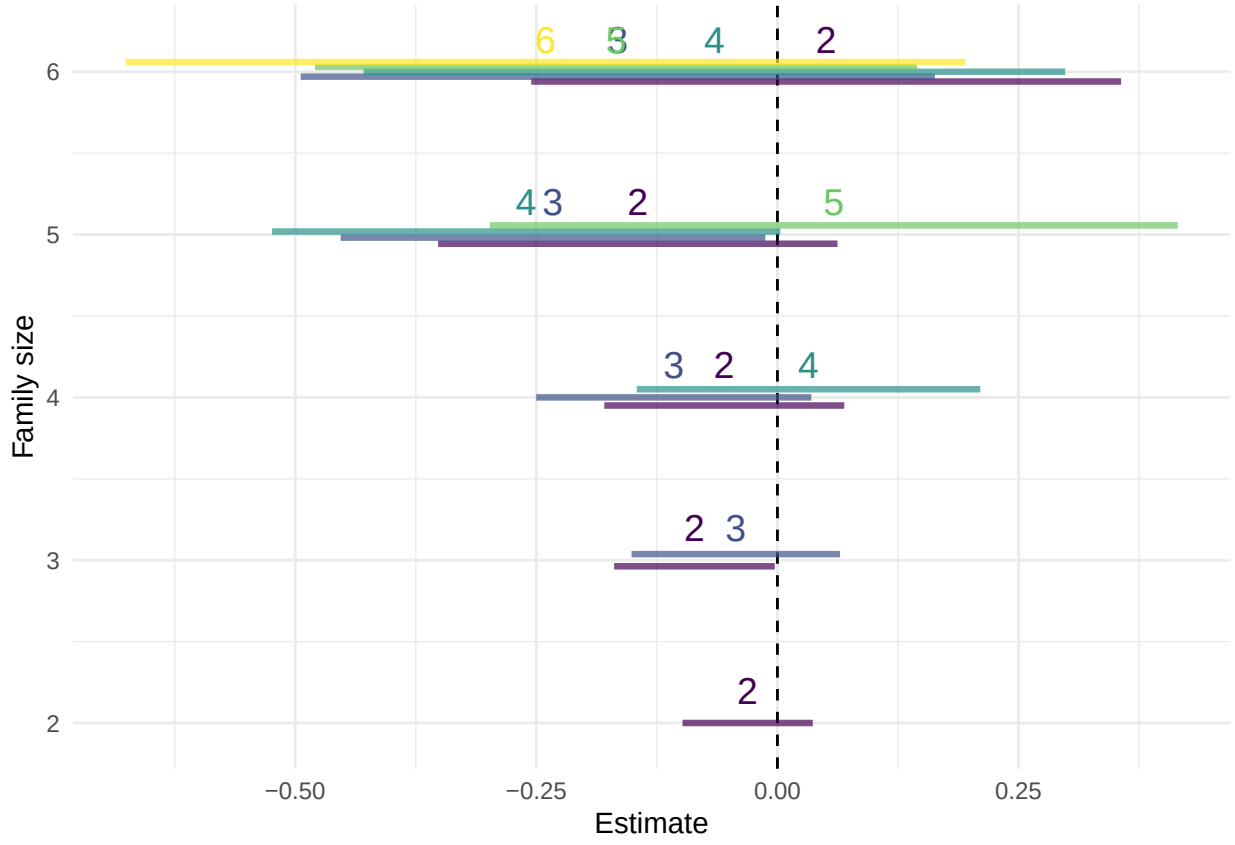


Figure 5: Regressions of spouse PSEA: birth order dummies within different family sizes. Labels show birth order. Lines are 95 per cent confidence intervals.

Table 7: Regressions of spouse PSEA, separate birth order coefficient dummies

	(1)	(2)	(3)	(4)
Birth order 2	−0.0573 *	−0.0239	−0.0389	−0.0379
	(0.0248)	(0.0250)	(0.0444)	(0.0443)
Birth order 3	−0.0841 *	−0.0264	0.0008	0.0197
	(0.0403)	(0.0406)	(0.0728)	(0.0729)
Birth order 4	−0.0451	0.0499	0.0529	0.0728
	(0.0697)	(0.0706)	(0.1338)	(0.1336)
Birth order 5	0.0015	0.0796	0.4368	0.4608 +
	(0.1300)	(0.1320)	(0.2783)	(0.2778)
Birth order 6	−0.2347	−0.1572	0.6275	0.6500
	(0.2658)	(0.2640)	(0.7288)	(0.7275)
University		0.2284 ***		0.1438 ***
		(0.0237)		(0.0405)
Income			0.0033 **	0.0027 *
			(0.0011)	(0.0011)
Own PSEA	0.0608 ***	0.0342 **	0.0372 +	0.0269
	(0.0106)	(0.0109)	(0.0195)	(0.0197)
Mother's age at birth	0.0104 ***	0.0046 +	0.0069	0.0056
	(0.0027)	(0.0028)	(0.0050)	(0.0050)
Wald p-value, birth order	0.1564	0.6395	0.0110	0.0205
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes
Controls (IQ, height)	No	Yes	Yes	Yes
N	8970	8778	2920	2919
R2	0.013	0.030	0.034	0.038
logLik	−12559.612	−12214.465	−4116.234	−4108.854
AIC	25229.224	24544.930	8348.469	8335.707

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors: robust.

Table 8: Regressions of spouse PSEA, separate birth order coefficient dummies: subsets

	Males	Males	Females	Females	With children	With children
Birth order 2	−0.0425	−0.0057	−0.0698 *	−0.0426	−0.0617 *	−0.0281
	(0.0379)	(0.0381)	(0.0332)	(0.0337)	(0.0261)	(0.0264)
Birth order 3	−0.0491	0.0041	−0.1089 *	−0.0539	−0.0862 *	−0.0265
	(0.0626)	(0.0622)	(0.0547)	(0.0554)	(0.0437)	(0.0439)
Birth order 4	−0.1006	0.0095	−0.0056	0.0753	−0.0499	0.0457
	(0.1089)	(0.1112)	(0.0903)	(0.0915)	(0.0736)	(0.0751)
Birth order 5	−0.2616	−0.1439	0.1700	0.2074	−0.0880	0.0023
	(0.1681)	(0.1738)	(0.1603)	(0.1639)	(0.1208)	(0.1227)
Birth order 6	−0.5222 +	−0.4184 +	−0.0445	0.0170	−0.0978	0.0003
	(0.2877)	(0.2518)	(0.2867)	(0.3123)	(0.2261)	(0.2269)
University		0.2584 ***		0.1969 ***		0.2256 ***
		(0.0358)		(0.0331)		(0.0257)
Own PSEA	0.0652 ***	0.0307 +	0.0580 ***	0.0365 *	0.0646 ***	0.0363 **
	(0.0158)	(0.0163)	(0.0147)	(0.0150)	(0.0113)	(0.0116)
Mother's age at birth	0.0110 *	0.0040	0.0097 **	0.0051	0.0108 ***	0.0050 +
	(0.0043)	(0.0044)	(0.0037)	(0.0037)	(0.0030)	(0.0030)
Wald p-value, birth order	0.3758	0.6456	0.0988	0.3147	0.2240	0.8207
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
Controls (IQ, height)	Yes	Yes	Yes	Yes	Yes	Yes
N	4093	4011	4877	4767	8023	7846
R2	0.019	0.040	0.019	0.031	0.016	0.033
logLik	−5752.671	−5597.747	−6780.108	−6589.496	−11209.236	−10892.137
AIC	11613.341	11309.494	13664.216	13288.992	22526.471	21898.275

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors: robust.

Table 9: Regressions of spouse PSEA with controls for polygenic scores

	(1)	(2)	(3)	(4)
Birth order	−0.0325 +	−0.0043	0.0080	0.0144
	(0.0169)	(0.0163)	(0.0357)	(0.0357)
University		0.2280 ***		0.1405 **
		(0.0303)		(0.0374)
Income			0.0033 **	0.0027 **
			(0.0009)	(0.0009)
Own PSEA	0.0528 **	0.0287 +	0.0307	0.0212
	(0.0140)	(0.0148)	(0.0240)	(0.0238)
Mother's age at birth	0.0099 **	0.0043	0.0070	0.0057
	(0.0026)	(0.0027)	(0.0052)	(0.0053)
Fluid IQ		0.0174 *	0.0228 +	0.0147
		(0.0070)	(0.0121)	(0.0129)
Height		0.0023 +	0.0050 +	0.0047 +
		(0.0013)	(0.0027)	(0.0027)
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes
Polygenic score controls	Yes	Yes	Yes	Yes
N	8970	8778	2920	2919
R2	0.014	0.030	0.034	0.038

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors: robust.
Polygenic scores: alzheimer's, caffeine, cognitive ability, neuroticism, substance use.

Table 10: Regressions of spouse PSEA using age of leaving full-time education

	(1)	(2)	(3)
Birth order	−0.0325 *	0.0016	0.0223
	(0.0158)	(0.0161)	(0.0296)
Age left full-time educ.		0.0496 ***	0.0403 ***
		(0.0047)	(0.0083)
Income			0.0025 *
			(0.0012)
Mother's age at birth	0.0102 ***	0.0034	0.0041
	(0.0028)	(0.0028)	(0.0050)
Own PSEA	0.0604 ***	0.0319 **	0.0253
	(0.0107)	(0.0110)	(0.0205)
Fluid IQ		0.0149 **	0.0104
		(0.0057)	(0.0105)
Height		0.0023 *	0.0047 *
		(0.0012)	(0.0021)
Family size dummies	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes
N	8970	8743	2914
R2	0.013	0.033	0.040
logLik	−12561.372	−12156.366	−4099.989
AIC	25222.744	24418.732	8307.978

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors: robust.

Table 11: Regressions of spouse PSEA, excluding family size 3

	(1)	(2)	(3)	(4)
Birth order	−0.0363 *	−0.0085	0.0005	0.0026
	(0.0184)	(0.0187)	(0.0362)	(0.0361)
University		0.2223 ***		0.1410 **
		(0.0292)		(0.0502)
Income			0.0014	0.0008
			(0.0017)	(0.0018)
Mother's age at birth	0.0116 ***	0.0058 +	0.0066	0.0056
	(0.0033)	(0.0034)	(0.0060)	(0.0060)
Own PSEA	0.0582 ***	0.0301 *	0.0225	0.0104
	(0.0131)	(0.0134)	(0.0254)	(0.0258)
Fluid IQ		0.0217 **	0.0154	0.0080
		(0.0068)	(0.0123)	(0.0127)
Height		0.0013	0.0040	0.0034
		(0.0014)	(0.0025)	(0.0025)
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes
N	6101	5967	1957	1956
R2	0.017	0.033	0.037	0.041
logLik	−8519.039	−8283.627	−2756.450	−2751.373
AIC	17136.077	16671.255	5614.899	5606.745

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors: robust.

Table 12: Seemingly Unrelated Regressions on spouse characteristics

	Males		Females	
	PSEA	Birth order	PSEA	Birth order
Birth order	−0.051 *	0.068 *	−0.027	0.086 ***
	(0.024)	(0.030)	(0.021)	(0.026)
PSEA	0.067 ***	−0.041 *	0.060 ***	−0.027
	(0.016)	(0.020)	(0.015)	(0.018)
Controls (mother's age)	Yes	Yes	Yes	Yes
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Birth year dummies	No	No	No	No
N	3899	3899	4627	4627

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$.

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 13 reruns the regressions of Table 2 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 13: Regressions of PSEA on birth order: fake pairs

	(1)	(2)	(3)
Birth order	−0.0069	−0.0057	−0.0257 +
	(0.0079)	(0.0079)	(0.0156)
Own PSEA		0.0497 ***	0.0510 ***
		(0.0067)	(0.0105)
Mother's age at birth			0.0103 ***
			(0.0027)
Family size dummies	Yes	Yes	Yes
Birth month dummies	No	Yes	Yes
Birth year dummies	No	Yes	Yes
N	22195	22151	9175
R2	0.002	0.006	0.011

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors: robust.

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.

Abramitzky, Ran, Adeline Delavande, and Luis Vasconcelos. 2011. "Marrying up: The Role of Sex Ratio in Assortative Matching." *American Economic Journal: Applied Economics* 3 (3): 124–57.

Adermon, Adrian, Mikael Lindahl, and Daniel Waldenström. 2018. "Intergenerational Wealth Mobility and the Role of Inheritance: Evidence from Multiple Generations." *The Economic Journal* 128 (612): F482–513.

Banerjee, Abhijit, Esther Duflo, Maitreesh Ghatak, and Jeanne Lafortune. 2013. "Marry for What? Caste and Mate Selection in Modern India." *American Economic Journal: Microeconomics* 5 (2): 33–72.

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 Height-Intelligence 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.

Belsky, Daniel W, Terrie E Moffitt, David L Corcoran, Benjamin Domingue, HonaLee Harrington, Sean Hogan, Renate Houts, et al. 2016. "The Genetics of Success: How Single-Nucleotide Polymorphisms Associated with Educational Attainment Relate to Life-Course Development." *Psychological Science* 27 (7): 957–72.

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."

Björklund, Anders, Mikael Lindahl, and Erik Plug. 2006. "The Origins of Intergenerational Associations: Lessons from Swedish Adoption Data." *The Quarterly Journal of Economics* 121 (3): 999–1028.

Black, Sandra E, and Paul J Devereux. 2010. "Recent Developments in Intergenerational Mobility." National Bureau of Economic Research.

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.

Breen, Richard, and Leire Salazar. 2011. "Educational Assortative Mating and Earnings Inequality in the United States." *American Journal of Sociology* 117 (3): 808–43.

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, and Lawrence F Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *American Economic Review* 106 (4): 855–902.

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.

Clark, Gregory. 2021. "For Whom the Bell Curve Tolls: A Lineage of 400,000 English Individuals 1750-2020 Shows Genetics Determines Most Social Outcomes." Working Paper. http://faculty.econ.ucdavis.edu/faculty/gclark/ClarkGlasgow2021.pdf.

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.

Cunha, Flavio, and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2): 31–47.

Cunha, Flavio, James J Heckman, and Susanne M Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica* 78 (3): 883–931.

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.

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 J, and Stefano Mosso. 2014. "The Economics of Human Development and Social Mobility." *Annu. Rev. Econ.* 6 (1): 689–733.

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.

Hitsch, Gunter J, Ali Hortaçsu, and Dan Ariely. 2010. "Matching and Sorting in Online Dating." *American Economic Review* 100 (1): 130–63.

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.

Kaplan, Steven N, and Joshua Rauh. 2013. "It's the Market: The Broad-Based Rise in the Return to Top Talent." *Journal of Economic Perspectives* 27 (3): 35–56.

Kong, Augustine, Gudmar Thorleifsson, Michael L Frigge, Bjarni J Vilhjalmsson, Alexander I Young, Thorgeir E Thorgeirsson, Stefania Benonisdottir, et al. 2018. "The Nature of Nurture: Effects of Parental Genotypes." *Science* 359 (6374): 424–28.

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.

Mulder, Monique Borgerhoff, Samuel Bowles, Tom Hertz, Adrian Bell, Jan Beise, Greg Clark, Ila Fazzio, et al. 2009. "Intergenerational Wealth Transmission and the Dynamics of Inequality in Small-Scale Societies." *Science* 326 (5953): 682–88.

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.

Pencavel, John. 1998. "Assortative Mating by Schooling and the Work Behavior of Wives and Husbands." *The American Economic Review* 88 (2): 326–29.

Plomin, Robert. 2019. *Blueprint: How DNA Makes Us Who We Are*. Mit Press.

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).

Sacerdote, Bruce. 2011. "Nature and Nurture Effects on Children's Outcomes: What Have We Learned from Studies of Twins and Adoptees?" In *Handbook of Social Economics*, 1:1–30. Elsevier.

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.

Selzam, Saskia, Stuart J Ritchie, Jean-Baptiste Pingault, Chandra A Reynolds, Paul F O'Reilly, and Robert Plomin. 2019. "Comparing Within-and Between-Family Polygenic Score Prediction." *The American Journal of Human Genetics* 105 (2): 351–63.

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

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

Solon, Gary. 2018. "What Do We Know so Far about Multigenerational Mobility?" *The Economic Journal* 128 (612): F340–52.

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.

Western, Bruce, Deirdre Bloome, and Christine Percheski. 2008. "Inequality Among American Families with Children, 1975 to 2005." *American Sociological Review* 73 (6): 903–20.

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