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

Abdel Abdellaoui[*], Oana Borcan[†], Pierre Chiappori[‡] & David Hugh-Jones[§]

2022-03-31

Abstract

If social status and genetic variants are both assets in marriage markets, then the two will become associated in spouse pairs, and will be passed on to subsequent generations together. This process provides a new explanation for the surprising persistence of inequality across generations, and for observed genetic differences across the distribution of 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

Over the long run, inequality is surprisingly persistent across generations (Clark and Cummins 2015; Solon 2018). Intergenerational mobility is correlated with cross-sectional inequality (Becker et al. 2018; Krueger 2012), which has risen dramatically in high-income countries, at the same time as intergenerational absolute mobility has declined (Western, Bloome, and Percheski 2008; Chetty et al. 2017).[1] Assortative mating in marriage markets can increase the variance of human capital and income across families (Breen and Salazar 2011; Greenwood et al. 2014), and

[*]Department of Psychiatry, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands. Email: a.abdellaoui@amsterdamumc.nl

[†]School of Economics, University of East Anglia, Norwich, UK. Email: O.Borcan@uea.ac.uk

[‡]Department of Economics, University of Columbia, New York. Email: pc2167@columbia.edu

[§]Corresponding author. School of Economics, University of East Anglia, Norwich, UK. Email: D.Hugh-Jones@uea.ac.uk

[1]Though relative mobility has been stable (Chetty et al. 2014). 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 to 43.3% during 1962-2013.

increasing returns to human capital may explain rising inequality (e.g. Kaplan and Rauh 2013; Becker et al. 2018). It follows that how families are formed, and how they transmit traits and assets to their offspring, are critical for understanding inequality. These processes have been studied from both socio-economic and genetic angles. While educational homogamy is well established, genetic assortative mating has been demonstrated only recently (Hugh-Jones et al. 2016; Robinson et al. 2017). 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, and as a result, genetics and inherited social status may become associated in subsequent generations.[3] For example, suppose that wealth and intelligence are both positive assets in a potential 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; Solon 2018). At the same time, this advantage is not an exogenous fact of biology, but endogenous to the social structure. Indeed, under SGAM, environmental shocks to an individual's social status may 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 genetic variants. We show that social-genetic assortative mating in one generation increases the correlation between SES and genetic variants 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. It is already known that humans mate assortatively on PSEA, which

[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).

makes it a likely candidate for detecting SGAM (Hugh-Jones et al. 2016; Robinson et al. 2017). We depart from the assumption 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 status (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 as mentioned above, they may take partners based on genomic similarity. 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 Heckman, Pinto, and Savelyev (2013), 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 also includes measures of own socio-economic status. In the latter model, socio-economic status can be interpreted, under certain assumptions, as a mediator of the effect of birth order on spouse genetics. We also include controls to balance covariates across individuals with different birth orders in different cohorts.

We find that later-born children have spouses with significantly lower polygenic scores for educational attainment in the reduced-form regressions. When we include university attendance as a measure of socio-economic status, birth order is no longer significant, while university attendance increases the spouse's genetic endowment at 0.1% significance. A similar pattern holds when we proxy socio-economic status with a measure of income, although the sample is reduced. Thus, SES appears to mediate the effect of birth order on spouse genetics. 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 and intergenerational mobility. Clark and Cummins (2015) show using a database of surnames that long-run intergenerational persistence of wealth is higher than simple parent-child correlations would predict. In particular, grandparents' wealth predicts grandchildren's wealth even after controlling for parents' wealth. Clark (2021) argues that the data can be explained by an underlying process where unobserved genetic variation determines wealth. We show below that SGAM could also generate these patterns. The mechanism again is unobserved genetic variation, but the interpretation is slightly different, since we view genetic endowments not an exogenous source of variation, but as an asset effectively "traded" in marriage markets in exchange for wealth and social status.

SGAM also affects cross-sectional inequality, like other forms of assortative mating (Fernández and Rogerson 2001). The standard mechanism is that when couples assort with respect to some characteristic, the resulting households will have more variance in that characteristic than if couples match randomly. This may then carry over into higher inter-individual inequality in the next generation (Fernandez, Guner, and Knowles 2005; Eika, Mogstad, and Zafar 2019). In particular, a 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. From twin studies, the heritability of occupational class and educational attainment, i.e. the proportion of variance

[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).

explained by genetic differences between individuals, is around 50% (Tambs et al. 1989). Genome-wide Complex Trait Analysis shows that the family socio-economic status of 2-year-old children 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). Thus, an important component of inequality is the association between SES and genetic variation – the "genes-SES gradient".

The standard explanation for the genes-SES gradient is social meritocracy. Parents with higher ability reap higher market returns, and they 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. Even when social status is entirely ascribed, it may still become associated with certain genetic variants, so long as their associated phenotypes (and not only status) are prized assets in marriage markets. Since meritocracy is historically rare, while assortative mating is universal, this suggests that genes-SES gradients are likely to be historically widespread.

SGAM may 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; 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, high 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.

Lastly, 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). These papers – like the vast majority of the behavior genetics literature (see e.g. Plomin, DeFries, and McClearn 2008) – take genetic endowments as exogenous and examine how they affect individual outcomes, perhaps in interaction with the environment. We take a different approach by putting genetics on the "left hand side". Thus,

our paper challenges the assumption, in economics and beyond, that genetic endowment is exogenous to economic characteristics. 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 associated 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 not 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 aspects of social status ("high earning capacity," "professional status") as well as traits that are partly under genetic influence ("intelligent," "tall," "kind," "physically attractive"). Despite this, we have found almost no previous work in genetics or economics that analyses SGAM or its consequences.[6]

Model

People in the marriage market have two characteristics: $x = (x_1, x_2)$, drawn from a normal distribution

$$\mathcal{N}\left(\begin{array}{ccc} 0 & , & s^2 \quad \sigma \\ 0 & & \sigma \quad S^2 \end{array}\right).$$

We interpret x_1 as a genetic measure, and x_2 as a measure of SES or of social status more generally. The correlation between x_1 and x_2 is then

$$Corr = \frac{\sigma}{sS}.$$

People's attractiveness is given by

$$i\left(x\right) = ax_1 + \left(1 - a\right)x_2$$

[6]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.

where $a \in [0,1]$ is a parameter reflecting the relative importance of genetics to SES in the marriage market.[7] If $a = 0$, marriage markets are highly inegalitarian, such that only SES matters. If $a = 1$, marriage markets are egalitarian and only genetics matter. We expect realistic societies to fall between these extremes. Attractiveness i is distributed $N(0, \sigma_I^2)$, where

$$\sigma_I^2 = a^2 s^2 + (1-a)^2 S^2 + 2a(1-a)\sigma.$$

People form matches with transferable utility, where the surplus for a match between x and y is $S(i(x), i(y))$ such that $\partial^2 S / \partial i \partial j > 0$, i.e. S is supermodular. As a result there is positive assortative mating on attractiveness: x matches with y only if $i(x_1, x_2) = j(y_1, y_2)$. We describe this as social-genetic assortative mating (SGAM).

We also consider random matching (RM) as a benchmark to compare against SGAM. Under RM, the distribution of couples' characteristics is normal with mean 0 and covariance matrix

$$\mathbb{C}\begin{pmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} s^2 & \sigma & 0 & 0 \\ \sigma & S^2 & 0 & 0 \\ 0 & 0 & s^2 & \sigma \\ 0 & 0 & \sigma & S^2 \end{pmatrix}.$$

Couples under SGAM

Proposition 1. *Under SGAM, the distribution of couples' characteristics is normal, with mean 0 and covariance matrix*

$$\mathbb{C}\begin{pmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} s^2 & \sigma & A^2 & AC \\ \sigma & S^2 & AC & C^2 \\ A^2 & AC & s^2 & \sigma \\ AC & C^2 & \sigma & S^2 \end{pmatrix}$$

where:

$$A = \frac{as^2 + (1-a)\sigma}{\sqrt{a^2 s^2 + (1-a)^2 S^2 + 2a(1-a)\sigma}} = \frac{as^2 + (1-a)\sigma}{\sigma_I};$$

$$C = \frac{a\sigma + (1-a)S^2}{\sqrt{a^2 s^2 + (1-a)^2 S^2 + 2a(1-a)\sigma}} = \frac{a\sigma + (1-a)S^2}{\sigma_I}.$$

In particular, the covariance between x_2 and y_1, AC, is positive if either $\sigma > 0$ or $\sigma = 0$ and $0 < a < 1$.

[7] Note that since the variance of the shocks to x_1 and x_2 (see below) has been normalized to 1, a also reflects this variance. That is, a large variance of SES shocks (compared to genetic shocks) translates into a being large.

Proof. See Appendix. □

We consider the distribution of couples' SES. Under RM this has mean 0 and variance $2S^2$. Under SGAM, the variance is:

$$V(x_2 + y_2) = 2S^2 + 2C^2 > 2S^2$$

Note, however, that

$$V(x_2 + y_2) \leq 4S^2$$

which would be reached if $a = 0$, i.e. if people only matched on SES. Thus, SGAM increases cross-sectional inequality but less so than pure matching on SES.

XXX is the above true only if $\sigma = 0$?

Children

All couples have the same number of children. Assume that a child's characteristics are given by:

$$\begin{aligned} x_1' &= \frac{\tau}{2}(x_1 + y_1) + \varepsilon \\ x_2' &= \frac{\theta}{2}(x_2 + y_2) + \eta \end{aligned} \tag{1}$$

where x and y are the child's parents, and ε and η are normal random shocks with mean 0 and variance 1.

Parameter $\tau \approx 1$ reflects genetic inheritance. Under standard biological assumptions $\tau = 1$ and characteristics show no regression to the mean. In our model this leads the variance of x_1 to grow without limit over generations. In reality, we expect $\tau < 1$ because very extreme characteristics are selected against, a process known as stabilizing selection.

Parameter $\theta \in [0, 1]$ reflects inheritance of SES. Unlike τ it may vary between societies. θ is high when there is high intergenerational transmission of SES. If we interpret x_2 narrowly as wealth, $(1 - \theta)$ can be thought of as the rate of inheritance taxation.

For the time being, we assume that a person's genetic endowment has no impact on their SES. Technically, thus, x_2' does not directly depend on x_1'. In a meritocratic society we would expect this to be violated. We show that even absent meritocracy, correlations between x_1' and x_2' can arise.

We can now calculate the covariance matrix for $x' = (x_1', x_2')$ under SGAM as:

$$
\begin{aligned}
\mathbb{C} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} \frac{\tau}{2} & 0 & \frac{\tau}{2} & 0 \\ 0 & \frac{\theta}{2} & 0 & \frac{\theta}{2} \end{pmatrix} \begin{pmatrix} s^2 & \sigma & A^2 & AC \\ \sigma & S^2 & AC & C^2 \\ A^2 & AC & s^2 & \sigma \\ AC & C^2 & \sigma & S^2 \end{pmatrix} \begin{pmatrix} \frac{1}{2}\tau & 0 \\ 0 & \frac{1}{2}\theta \\ \frac{1}{2}\tau & 0 \\ 0 & \frac{1}{2}\theta \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{2}A^2\tau^2 + \frac{1}{2}s^2\tau^2 + 1 & \frac{1}{2}\theta\sigma\tau + \frac{1}{2}AC\theta\tau \\ \frac{1}{2}\theta\sigma\tau + \frac{1}{2}AC\theta\tau & \frac{1}{2}C^2\theta^2 + \frac{1}{2}S^2\theta^2 + 1 \end{pmatrix}
\end{aligned} \tag{2}
$$

We now explore two issues. First, under SGAM, genetic characteristics are no longer exogenous; because of assortative matching, they are (partly) socially determined. In particular, even if genetics and SES are uncorrelated among parents, the expected genetic endowment of the child is positively related to parental SES. Second, as a result, in the long run a correlation appears between traits; that is, high SES people have genes which are attractive in marriage markets.

Regarding point 1, we compute the expected genetic characteristic of the child, conditional on parental SES:

$$\mathbb{E}\left[\frac{\tau}{2}(x_1 + y_1) + \varepsilon \mid x_2 = v, y_2 = w\right]$$

Given the symmetry of the model, this conditional expectation only depends on the parents' total wealth, i.e. $v + w$.

Claim 1. *Under random matching, the expected genetic endowment of the children is proportional to the parents' SES and to the covariance between SES and genetics for the parents. In particular, if $\sigma = 0$ (i.e. genetics and SES are uncorrelated for the parents), then the expected genetic endowment of the children does not depend on parental SES.*

Claim 2. *Under SGAM, if $\sigma = 0$ (i.e. genetics and SES are uncorrelated for the parents), then the expected genetic endowment of the children is linearly increasing in parental SES. The relationship increases with the ratio of genetic variance to wealth variance, is zero for $a = 0$ or $a = 1$, and is highest for intermediate values of a.*

Next, we study the correlation between children's traits 1 and 2 as a function of σ, the covariance of parents' traits. We first consider the general case, then concentrate on $\sigma = 0$, i.e. when traits are initially uncorrelated.

Claim 3. *Under random matching, the correlation between characteristics is smaller for children than for parents. In particular, if genetics and wealth are uncorrelated for the parents, then they are uncorrelated for the children.*

Claim 4. *Under SGAM, if genetics and wealth are uncorrelated for the parents, then they are positively correlated for the children so long as $0 < a < 1$. The correlation is increasing in θ.*

Whether characteristics are more or less correlated for children than for parents depends on whether the initial correlation between parents' characteristics is larger or smaller than the asymptotic one, derived later on.

These results show that SGAM can lead to a genes-SES gradient, i.e. a positive correlation between genes and SES. Also, the strength of the genes-SES correlation is affected by economic institutions, as captured in θ. When θ is high, the genes-SES correlation is high too. As we would expect, these results carry over to the asymptotic distribution.

Asymptotics

Proposition 2. *Under RM, the dynamics converges to a stationary distribution that is normal with mean zero and covariance matrix*

$$C = \begin{pmatrix} \frac{2}{2-\tau^2} & 0 \\ 0 & \frac{2}{2-\theta^2} \end{pmatrix}$$

In particular, the traits are asymptotically uncorrelated and children's expected genetic endowment is independent of parents' SES.

Proposition 3. *Under SGAM, for $\theta < 1$ and $\tau < 1$, the dynamics converge to a stationary distribution that is normal with mean zero and covariance matrix*

$$C = \begin{pmatrix} \bar{s}^2 & \bar{\sigma} \\ \bar{\sigma} & \bar{S}^2 \end{pmatrix}$$

For $\theta = 1$, the dynamics diverge and S^2 goes to $+\infty$; for $\tau = 1$, the dynamics diverges and s^2 goes to $+\infty$.

XXX what can we say about $\bar{\sigma}$ e.g. if $0 < a < 1$. Is it positive? Increasing in θ? Increasing then decreasing in a?

Conditional expectations of children's genetics given parents' SES under SGAM are calculated using the same formula as before, plugging in moments of the asymptotic distribution:

$$\begin{aligned} \mathbb{E}\left[\frac{\tau}{2}(x_1 + y_1) + \varepsilon \mid x_2 = v, y_2 = w\right] &= \frac{1}{2}\tau \frac{\bar{\sigma} + \bar{A}\bar{C}}{\bar{C}^2 + \bar{S}^2}(v + w) \quad (3) \\ &= \mu \frac{\theta \bar{S}^2}{2\bar{S}^2 - 2} \quad (4) \end{aligned}$$

Numerical application

In what follows, we impose

$$\tau = 0.95.$$

The three components of the asymptotic covariance matrix (s^2, S^2, σ) are given in the following graph:

(XXX omitted)

Lastly, the asymptotic correlation between characteristics is:

$$Corr = \frac{\sigma}{sS} = \frac{\mu}{\sqrt{\lambda}} = \frac{\phi(a,\theta,\tau)}{\sqrt{\psi(a,\theta,\tau)}}$$

(XXX graph omitted)

and the conditional expectation of genetics given wealth is proportional to wealth, coefficient:

(XXX graph omitted)

Note that both $\bar{S}^2$ and $\bar{\sigma}$, as well as the correlation between characteristics and the conditional expectation of genetics given wealth, are increasing in θ, i.e. decreasing in the tax rate. Higher taxation reduces the asymptotic variance of wealth (not surprisingly), but also the correlation between genetic and wealth.

Extensions

We consider robustness in two dimensions. First, the weight of x_1 and x_2 might differ for men and women, for example if SES is more important for men's attractiveness. Suppose that men's and women's attractiveness are given by

$$i(x) = ax_1 + (1-a)x_2;$$
$$j(y) = by_1 + (1-b)y_2$$

respectively. Then Claims 2 and 4 continue to hold.

XXX put proof in appendix

Second, we consider non-normal distributions of x_1 and x_2, and non-linear attractiveness functions. Suppose

$$i(x) = f(ax_1, (1-a)x_2) \tag{5}$$

with f strictly increasing in both its arguments. Our sole condition on x is that a positive measure of the population has attractiveness $i(x) = i$ where the distribution of $(x_1, x_2)|i$ is non-degenerate, i.e. not everybody with attractiveness i is both genetically and socially identical.

Proposition 4. *Let attractiveness be given by* (5)*. If the correlation between x_1 and x_2 is 0, then the correlation among*

children between x_1' and x_2' is non-negative, and strictly positive if $0 < a < 1$.

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 a: 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 meanings of both social status, and "good genes" in the marriage market, are likely to vary across societies. Social status could encompass variables like social class or caste; ethnic identity in "ranked" ethnic systems; or in modern societies, SES. Regarding genetics, standards of physical attractiveness, and other characteristics which make someone a "good match", vary across societies and over time.

Recent empirical work shows high persistence of SES over time, in particular at the top. XXX insert stuff about grandchild-control here.

The converse also holds: regressions of children's characteristics on their genetics alone risk overestimating the effect of genetics, by confounding it with the effects of correlated socio-economic status. Recent work in genetics has shown just this. Polygenic scores for educational attainment have smaller effects in between-sibling regressions, where between-family variation in SES is partialled out and where genes are guaranteed to be randomly allocated, than in regressions which pool the whole sample (Howe et al. 2021). Parents' genetic variants which are *not* passed on to children predict children's characteristics, via environmental effects (Kong et al. 2018).

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. Also, SGAM is increased by θ. This implies that policy has long-run effects on the social structure: reducing θ not only increases intergenerational mobility, but reduces the correlation of genes with SES, and hence the unfairness of what has been called the "genetic lottery" (Harden 2021).

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. (XXX needs model it with gamma)

Data and methods

To test for the existence of SGAM, 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;[8]
- both reported the same homeownership/renting status, length of time at the address, and number of children;
- attended the same UK Biobank assessment center on the same day;
- both reported living with their spouse ("husband, wife or partner");
- consisted of one male and one female.

We also eliminate all pairs where either spouse appeared more than once in the data. This leaves a total of 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. So, to validate the accuracy of our pairs, 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

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

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).[9]

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

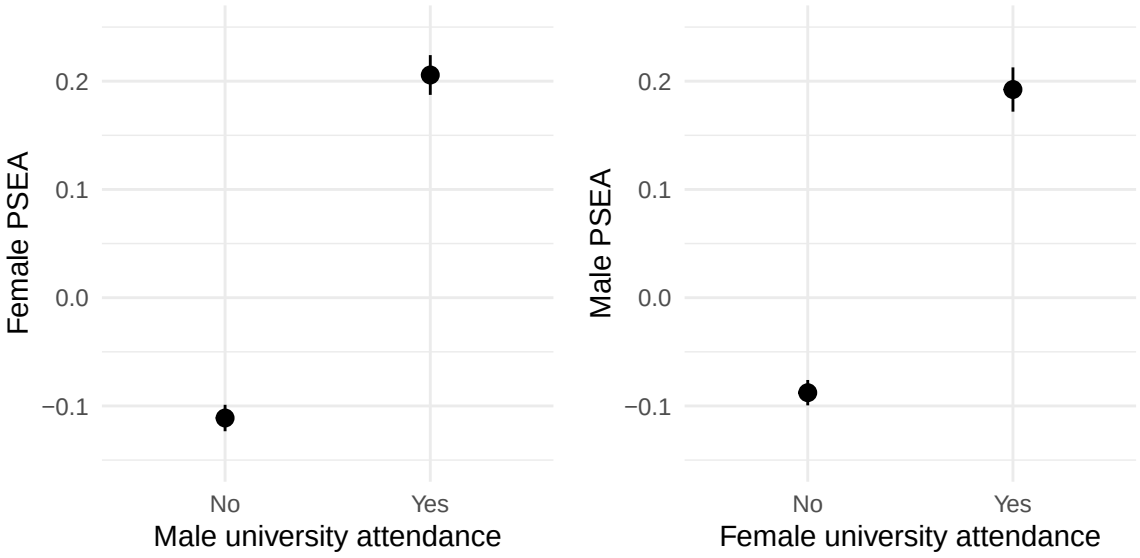
Despite these points, PSEA has non-trivial estimated effects on educational attainment. PSEA correlates with measures of education, including university attendance and years of full-time education; within-siblings regressions, where PSEA is randomly assigned by the "lottery of meiosis", confirm this 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 1 and 2 illustrate the core idea of SGAM within our pair data. The X axis shows a measure of one partner's socio-economic status: university attendance (Figure 1) or income (Figure 2). The Y axis plots the other partner's mean 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.

These figures do not prove that SGAM is taking place. Since an individual's own PSEA correlates with both their educational attainment, and their income, both figures could be a result of genetic assortative mating (GAM) alone (Hugh-Jones et al. 2016). To demonstrate SGAM, we need a source of social status which is exogenous to genetics. Also, the link between social status and spouse genetics is likely to be noisy, for three reasons: first, polygenic scores contain a large amount of error, as discussed above; second, causal mechanisms behind variation in social status are

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

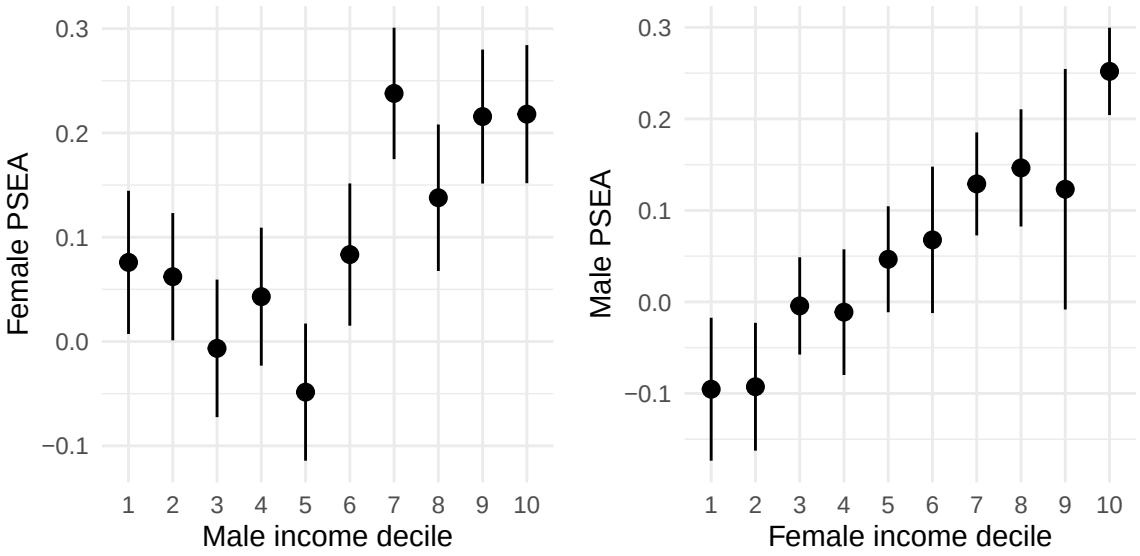
[10] PSEA was computed by summing the alleles across ~1.3 million genetic variants weighted by their effect sizes as estimated in genome-wide association studies (GWASs) that excluded UK Biobank. PSEA was then residualized on the first 100 principal components of the SNP array data. Further details can be found in Abdellaoui et al. (2019).



(a) Female PSEA by male educational attainment

(b) Male PSEA by female educational attainment

Figure 1: Spouse PSEA against own university attendance. Lines show 95% confidence intervals



(a) Female EA3 by male income

(b) Male EA3 by female income

Figure 2: Spouse PSEA against own estimated income in first job. Lines show 95% confidence intervals

likely to be noisy; third, to paraphrase Shakespeare (1595), the spouse matching process is highly 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. Despite this language, 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 be related 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. respondents' total number of siblings including themselves; month of birth; age at interview; respondents' own PSEA; and 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 size of 2-6.

To test whether birth order effects are mediated by SES, we use two measures: income, and university attendance. Income is a direct measure of SES. University attendance is a predictor of income over the whole life course, and a form of SES itself. The UK Biobank data only contains a direct measure of current household income, which is inappropriate for our purposes because it includes income from both spouses. Instead, we estimate income in the respondent's first job, from the job's Standard Occupational Classification (SOC) code. We match this with information on average earnings by SOC from the Annual Survey of Hours and Earnings 2007 (National Statistics 2007). This data is only available for a subset of respondents.

Decomposing the birth order effect on spouse genetics

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 attended university, but not which university. Birth order likely affects both measured and unmeasured aspects of SES. So, an instrumental variables approach would be likely to fall foul of the exclusion restriction.

[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. Isungset et al. (2021) also find that birth order differences in education are not genetic.

Instead, we conduct a mediation analysis, following the strategy of Heckman, Pinto, and Savelyev (2013). We first confirm statistically that birth order affects our measures of respondents' SES (income and education). Then, we regress spouse's PSEA on birth order, with and without controlling for SES. Under the assumption that birth order is exogenous to own genetics, these regressions identify the effect of birth order, plus other environmental variables that correlate with it, on own social status and spouse's genetics. Also, if the estimated effect of birth order on spouse's PSEA changes when SES is controlled for, that is evidence that SES mediates the effect of birth order.

Linearizing our model so that $A(g, s) = (1 - k)g + ks$ and applying (**??**) 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 measures of genetic endowment, such as PSEA, are noisy and incomplete, it is not enough to include them as controls in the regression. Instead, we use birth order as a source of variation in s_i which is orthogonal to g_i.

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

Assume B is a multivalued variable indicating birth order. Let Y_b be the counterfactual outcome (spouse PSEA) for the first-born, second-born etc. Given b, spouse PSEA is assumed to be independent across observations conditional on some predetermined controls which are assumed not to be affected 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. 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. We are mainly interested in estimating the effect of SES on spouse PSEA.

Our linear model is:

$$Y_b = \kappa_b + \sum_{j \in J_m} \alpha_b^j \theta_b^j + \sum_{j \in J_u} \alpha_b^j \theta_b^j + \mathbf{X}' \boldsymbol{\beta}_\mathbf{b} + \tilde{\varepsilon}_b = \tau_b + \sum_{j \in J_m} \alpha_b^j \theta_b^j + \mathbf{X}' \boldsymbol{\beta}_\mathbf{b} + \varepsilon_b \tag{6}$$

where $\tilde{\varepsilon}_b$ is a mean-zero residual assumed independent of θ_b and $\mathbf{X}$; $\tau_b = \kappa_b + \sum_{j \in J_u} \alpha_b^j E(\theta_b^j)$; and $\varepsilon_b = \tilde{\varepsilon}_b + \sum_{j \in J_u} (\theta_b^j - E(\theta_b^j))$. 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.[12] We assume differences in unmeasured investments due to b are independent of $\mathbf{X}$.

We use a linear model for each observed mediator variable:

$$\theta_b^j = \mu_{0,j} + \mathbf{X'}\boldsymbol{\mu}_{\mathbf{1,j}} + \mu_{2,j} \cdot b + \eta_j, j \in J_m \tag{7}$$

where η_j is a mean-zero residual. We also assume the treatment-specific intercepts are linear in b:

$$\tau_b = \tau_0 + \tau b. \tag{8}$$

With the simplifying assumptions above and substituting (7)- (8) into (6) we obtain:

$$Y_b = \tau_0 + \tau b + \sum_{j \in J_m} \alpha^j (\mu_{0,j} + \mathbf{X'}\boldsymbol{\mu}_{\mathbf{1,j}} + \mu_{2,j} \cdot b + \eta_j) + \mathbf{X}'\boldsymbol{\beta} + \varepsilon_b \tag{9}$$

Using equation (9), we can decompose the average treatment effect of B (birth order b' instead of b) into the effect of measured mediators θ^j and unmeasured mediators on the outcome:

$$E(Y_b' - Y_b) = \tau(b' - b) + \sum_{j \in J_m} \alpha^j E(\theta_{b'}^j - \theta_b^j) = \underbrace{\tau(b' - b)}_{\text{Effect of unmeasured mediators}} + \underbrace{\sum_{j \in J_m} \alpha^j \mu_{2,j}(b' - b)}_{\text{Effect of measured mediators}} \tag{10}$$

We are primarily interested in estimating the effect of SES on spouse PSEA, amongst the measured mediators, and furthermore we would like to measure the relative importance of SES compared to other factors in predicting spouse PSEA.

We therefore estimate:

$$Y = \tau_0 + \tau B + \sum_{j \in J_m} \alpha^j \theta_b^j + \mathbf{X}'\boldsymbol{\beta} + \varepsilon \tag{11}$$

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, there are two identifying assumptions that need to hold for unbiased OLS estimates: (a) the measured

[12] Under the assumption that measured and unmeasured mediators are uncorrelated, we can test these assumptions by running an OLS regression of an extended model (11) where we interact the measured mediators and controls with the treatment B, and test the significance of the coefficients on the interaction terms ($\boldsymbol{\alpha}_\mathbf{b} = 0$ and $\boldsymbol{\beta}_\mathbf{b} = 0$). See Heckman, Pinto, and Savelyev (2013) and Fagereng, Mogstad, and Rønning (2021) for details and different applications.

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. Second, (b) the measured investments should be uncorrelated with other shocks $\tilde{\varepsilon}_b$.

By running a least square regression of (11), we can estimate τ and α^j. If assumption (a) 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 mediators. We can estimate these effects in a second step, from OLS regressions based on equation (7) for each measured mediator (in particular university attendance and income) on X and B. The part of the birth order effect that is due to university attendance (or income) on spouse PSEA will be the coefficient of university (income) in the regression of spouse PSEA in equation (11) multiplied by the coefficient of birth order from equation (7).

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 mediators that could be affected by birth order: fluid IQ and height. We control for respondent's own PSEA and their parents' age at birth (see below). Table 1 shows that birth order significantly predicts all these variables. Effect sizes are quite substantial: a single extra elder sibling reduces the chance of attending university by about 7.9 percentage points, income by about £1,088, fluid IQ by about 0.27 points on a 13 point test, and height by about 0.71 centimeters.

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 and insignificant. 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 parents' 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.[13] Including parents' age means we can separate the effect of parental age from birth order. This reduces the N by a lot, since only respondents with a live parent reported the necessary data. However, the effect of birth order jumps in size and becomes significant at the 5 per cent level. Meanwhile, parents' age has a positive effect. This suggests that estimates in columns 1-2 mixed two opposite-signed effects: having older parents versus being later in birth order.

[13] We often only have data only for one parent. We use this, or take the mean if we have both. There are also potential genetic effects from parental age, 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 mother's age. We

Table 1: Regressions of variables on birth order

	University	Income	Fluid IQ	Height
Birth order	−0.0789 ***	−1.0882 *	−0.2718 ***	−0.7051 ***
	(0.0067)	(0.4264)	(0.0304)	(0.1353)
PSEA	0.0890 ***	1.5211 ***	0.3172 ***	0.1921 *
	(0.0046)	(0.3296)	(0.0200)	(0.0919)
Parents' age at birth	0.0164 ***	0.2631 ***	0.0588 ***	0.1515 ***
	(0.0012)	(0.0721)	(0.0053)	(0.0241)
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes
N	10243	3419	10243	10243
R2	0.074	0.026	0.058	0.017

Estimates from OLS regressions with the mediators (university attendance, income, fluid IQ and height) as dependent variables, and own birth order as the main independent variable. PSEA is the polygenic score for educational attainment, which is normalized with mean 0 and standard deviation 1. We include parents' age at birth (the mean of parents' ages) and further controls to ensure the balance of covariates across birth order. All data is from the UK Biobank for a sample of UK individuals born between 1935 and 1970. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. 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' parents' 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. 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 in first job. Column 4 includes both.

When we add university attendance (column 2), the effect of birth order drops and becomes insignificant, while the

observe very similar results if we control only for father's age at respondent's birth.

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

Table 2: Regressions of spouse PSEA on birth order

	(1)	(2)	(3)
Birth order	−0.0099	−0.0079	−0.0312 *
	(0.0074)	(0.0074)	(0.0146)
Own PSEA		0.0651 ***	0.0583 ***
		(0.0065)	(0.0100)
Parents' age at birth			0.0113 ***
			(0.0026)
Family size dummies	Yes	Yes	Yes
Birth month dummies	No	Yes	Yes
Birth year dummies	No	Yes	Yes
N	23904	23861	10229
R2	0.003	0.011	0.013

Estimates from OLS regressions with spouse PSEA as dependent variable, and own birth order as the main independent variable. PSEA is the polygenic score for educational attainment, which is normalized with mean 0 and standard deviation 1. We include parents' age at birth (the mean of parents' ages), and further controls in column 3 to ensure the balance of covariates across birth order. All data is from the UK Biobank for a sample of UK individuals born between 1935 and 1970. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors: robust.

coefficient for university is positive and highly significant. Fluid IQ and height are also positive and significant. Controlling for income alone (column 3), again the effect of birth order shrinks and becomes insignificant, while income has a positive and highly significant effect. Lastly, the same pattern holds when we control for both university and income (column 4).

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.

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. Birth order loses significance in these subsets. However, the pattern of coefficient sizes is the same as in the main regression: the coefficient of birth order is about -0.3, and adding university attendance reduces the absolute size of the birth order effect. Columns 5 and 6 use only couples with children. Here, birth order is significant in the base specification, and again, university attendance still seems to mediate the birth order effect.

Spouse matching is strictly a single process in which neither side is an "independent variable". In Appendix Table 12, we estimate a matching model in the style of Chiappori, Oreffice, and Quintana-Domeque (2012), estimating the effects of birth order (and PSEA) on both spouse's PSEA and spouse's birth order, and testing for proportionality of coefficients across the dependent variables. For women only we reject proportionality of coefficients, suggesting that there may be some assortative matching separately on each dimension.

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

	(1)	(2)	(3)	(4)
Birth order	−0.0312 *	−0.0065	−0.0128	−0.0059
	(0.0146)	(0.0146)	(0.0270)	(0.0269)
University		0.2276 ***		0.1596 ***
		(0.0224)		(0.0377)
Income			0.0037 ***	0.0030 **
			(0.0011)	(0.0011)
Own PSEA	0.0583 ***	0.0320 **	0.0299	0.0189
	(0.0100)	(0.0101)	(0.0184)	(0.0185)
Parents' age at birth	0.0113 ***	0.0061 *	0.0100 *	0.0086 +
	(0.0026)	(0.0026)	(0.0047)	(0.0047)
Fluid IQ		0.0174 ***	0.0198 *	0.0104
		(0.0052)	(0.0093)	(0.0097)
Height		0.0028 **	0.0047 *	0.0043 *
		(0.0011)	(0.0020)	(0.0019)
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes
N	10229	10229	3414	3414
R2	0.013	0.029	0.027	0.032
logLik	−14327.102	−14242.367	−4826.981	−4817.738
AIC	28754.204	28590.734	9759.962	9743.475

Estimates from OLS regressions with spouse PSEA as dependent variable, and own Birth Order and mediators (university attendance and income) as the main independent variables. Columns 2−4 correspond to model 7. PSEA is the polygenic score for educational attainment, which is normalized with mean 0 and standard deviation 1. We include parents' age at birth (the mean of parent's ages) and further controls to ensure the balance of covariates across birth order. All data is from the UK Biobank for a sample of UK individuals born between 1935 and 1970. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors: robust.

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

	Model 2	**Model 3**	**Model 4**
University	57.6%	-	40.4%
Income	-	13.0%	10.5%
Fluid IQ	15.2%	17.3%	9.0%
Height	6.3%	10.5%	9.7%

Percentage of the effects of birth order in Table 3, columns 2 to 4, that are explained by mediators (university attendance, income in first job, fluid IQ and height).

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

	Male respondents	Male respondents	Female respondents	Female respondents	With children	With children
Birth order	−0.030	−0.002	−0.031	−0.011	−0.034 *	−0.009
	(0.022)	(0.022)	(0.019)	(0.019)	(0.015)	(0.015)
University		0.283 ***		0.177 ***		0.227 ***
		(0.033)		(0.031)		(0.024)
Own PSEA	0.059 ***	0.025 +	0.059 ***	0.037 **	0.063 ***	0.035 **
	(0.015)	(0.015)	(0.014)	(0.014)	(0.011)	(0.011)
Parents' age at birth	0.013 **	0.006	0.010 **	0.006 +	0.012 ***	0.007 *
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Fluid IQ		0.020 **		0.015 *		0.023 ***
		(0.008)		(0.007)		(0.006)
Height		0.004 +		0.005 *		0.002 +
		(0.002)		(0.002)		(0.001)
Family size dummies	Yes	Yes	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	4684	4684	5545	5545	9148	9148
R2	0.018	0.041	0.018	0.028	0.015	0.032

Estimates from OLS regressions corresponding to columns 1 and 2 in Table 3, separately for males, females and respondents with children. Spouse PSEA is the dependent variable, and own birth order and university attendance are the main independent variables. PSEA is the polygenic score for educational attainment, which is normalized with mean 0 and standard deviation 1. We include parents' age at birth (the mean of parent's ages) and further controls to ensure the balance of covariates across birth order. All data is from the UK Biobank for a sample of UK individuals born between 1935 and 1970. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors: robust.

Robustness

Although all children of the same parents have the same polygenic scores in expectation, it might 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 choose family size in a way related to genetics. For example, suppose that when 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.[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 effect sizes of less than 0.02 per standard deviation. 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 marginally significant positive correlation of birth order with own PSEA (effect size 0.0276, $p = 0.06$). 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.

The appendix reports other robustness checks, including a specification with separate dummies for each value of birth order, and replacing university attendance with age of leaving full-time education.

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. We interpret this as evidence of social-genetic assortative mating.

[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. For full details of score construction, see Abdellaoui et al. (2019).

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 testing the theory, rather than on estimating effect sizes. This is by necessity: even the latest measures 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 greater

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

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. As we know, marriage markets can play an important role in explaining inequality and intergenerational mobility (Fernández and Rogerson 2001): SGAM shows that this is also true for genetic inequality.

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

Proofs

Proof of Proposition 1. By a change of variable, rewrite:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to \begin{pmatrix} x_1 \\ u \end{pmatrix} \text{ where } u = \frac{ax_1 + (1-a)\,x_2}{\sqrt{a^2 s^2 + (1-a)^2 S^2 + 2a\,(1-a)\,\sigma}} = \frac{ax_1 + (1-a)\,x_2}{\sigma_I}$$

is the attractiveness rescaled to $\mathcal{N}(0,1)$. Thus,

$$\begin{pmatrix} x_1 \\ u \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ a/\sigma_I & (1-a)/\sigma_I \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Note that the means are still zero, but the covariance of (x_1, u) is:

$$\begin{aligned} C\begin{pmatrix} x_1 \\ u \end{pmatrix} &= \begin{pmatrix} 1 & 0 \\ a/\sigma_I & (1-a)/\sigma_I \end{pmatrix} \begin{pmatrix} s^2 & \sigma \\ \sigma & S^2 \end{pmatrix} \begin{pmatrix} 1 & a/\sigma_I \\ 0 & (1-a)/\sigma_I \end{pmatrix} \\ &= \begin{pmatrix} s^2 & \left(as^2 + \sigma(1-a)\right)/\sigma_I \\ \left(as^2 + \sigma(1-a)\right)/\sigma_I & 1 \end{pmatrix} \end{aligned}$$

Under SGAM, individual $\begin{pmatrix} x_1 \\ u \end{pmatrix}$ is matched with $\begin{pmatrix} y_1 \\ v \end{pmatrix}$ such that $u = v = t$.

The distribution of t is $\mathcal{N}(0,1)$. Therefore the vector $\begin{pmatrix} x_1 \\ y_1 \\ t \end{pmatrix}$ is normally distributed, with mean 0, and covariance

$$\Sigma = \begin{pmatrix} s^2 & d^2 & d \\ d^2 & s^2 & d \\ d & d & 1 \end{pmatrix}$$

where

$$d = \frac{as^2 + (1-a)\,\sigma}{\sqrt{S^2\,(a-1)^2 + a^2 s^2 - 2a\sigma\,(a-1)}} = \frac{as^2 + (1-a)\,\sigma}{\sigma_I}$$

Finally, we are interested in

$$\begin{pmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{a}{1-a} & 0 & \frac{\sqrt{a^2s^2+(1-a)^2S^2+2a(1-a)\sigma}}{1-a} \\ 0 & 1 & 0 \\ 0 & -\frac{a}{1-a} & \frac{\sqrt{a^2s^2+(1-a)^2S^2+2a(1-a)\sigma}}{1-a} \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ t \end{pmatrix}$$

therefore again the means are 0 and

$$\begin{aligned} \mathbb{C}\begin{pmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{pmatrix} &= \begin{pmatrix} 1 & 0 & 0 \\ -\frac{a}{1-a} & 0 & \frac{\sqrt{a^2s^2+(1-a)^2S^2+2a(1-a)\sigma}}{1-a} \\ 0 & 1 & 0 \\ 0 & -\frac{a}{1-a} & \frac{\sqrt{a^2s^2+(1-a)^2S^2+2a(1-a)\sigma}}{1-a} \end{pmatrix} \Sigma \begin{pmatrix} 1 & 0 & 0 \\ -\frac{a}{1-a} & 0 & \frac{\sqrt{a^2s^2+(1-a)^2S^2+2a(1-a)\sigma}}{1-a} \\ 0 & 1 & 0 \\ 0 & -\frac{a}{1-a} & \frac{\sqrt{a^2s^2+(1-a)^2S^2+2a(1-a)\sigma}}{1-a} \end{pmatrix}^T \\ &= \begin{pmatrix} s^2 & \sigma & A^2 & AC \\ \sigma & S^2 & AC & C^2 \\ A^2 & AC & s^2 & \sigma \\ AC & C^2 & \sigma & S^2 \end{pmatrix} \end{aligned}$$

where:

$$\begin{aligned} A &= \frac{as^2 + (1-a)\sigma}{\sqrt{a^2s^2 + (1-a)^2 S^2 + 2a(1-a)\sigma}} \text{ and} \\ C &= \frac{a\sigma + (1-a)S^2}{\sqrt{a^2s^2 + (1-a)^2 S^2 + 2a(1-a)\sigma}}. \end{aligned}$$

□

Proof of Claim 1. Under RM, the joint distribution of $\left(\frac{\tau}{2}(x_1 + y_1) + \varepsilon, x_2, y_2\right)$ is normal with mean $(0,0,0)$ and covariance

$$C = \begin{pmatrix} \frac{\tau^2}{2}(s^2+\sigma) + 1 & \frac{\tau}{2}\sigma & \frac{\tau}{2}\sigma \\ \frac{\tau}{2}\sigma & S^2 & 0 \\ \frac{\tau}{2}\sigma & 0 & S^2 \end{pmatrix}$$

Therefore

$$\begin{aligned}\mathbb{E}\left[\frac{\tau}{2}(x_1+y_1)+\varepsilon \mid x_2=v, y_2=w\right] &= \begin{pmatrix} \frac{\tau}{2}\sigma & \frac{\tau}{2}\sigma \end{pmatrix}\begin{pmatrix} S^2 & 0 \\ 0 & S^2 \end{pmatrix}^{-1}\begin{pmatrix} v \\ w \end{pmatrix} \\ &= \frac{\sigma\tau}{2S^2}(v+w)\end{aligned}$$

In particular, if $\sigma = 0$, this expectation is equal to 0. □

Proof of Claim 2. From (2), the joint distribution of $\left(\frac{\tau}{2}(x_1+y_1)+\varepsilon, x_2, y_2\right)$ is normal with mean $(0,0,0)$ and covariance

$$\Sigma = \begin{pmatrix} \frac{1}{2}\tau^2(A^2+s^2)+1 & \frac{\tau}{2}(\sigma+AC) & \frac{\tau}{2}(\sigma+AC) \\ \frac{\tau}{2}(\sigma+AC) & S^2 & C^2 \\ \frac{\tau}{2}(\sigma+AC) & C^2 & S^2 \end{pmatrix}$$

Therefore

$$\begin{aligned}\mathbb{E}\left[\frac{\tau}{2}(x_1+y_1)+\varepsilon \mid x_2=v, y_2=w\right] &= \begin{pmatrix} \frac{\tau}{2}(\sigma+AC) & \frac{\tau}{2}(\sigma+AC) \end{pmatrix}\begin{pmatrix} S^2 & C^2 \\ C^2 & S^2 \end{pmatrix}^{-1}\begin{pmatrix} v \\ w \end{pmatrix} \\ &= \frac{1}{2}\tau\frac{\sigma+AC}{C^2+S^2}(v+w) \qquad (12)\end{aligned}$$

In particular, if $\sigma = 0$, we have

$$\begin{aligned}A &= \frac{as^2}{\sqrt{a^2s^2+(1-a)^2S^2}} \text{ and} \\ C &= \frac{(1-a)S^2}{\sqrt{a^2s^2+(1-a)^2S^2}},\end{aligned}$$

and (12) becomes

$$\begin{aligned}\mathbb{E}\left[\frac{\tau}{2}(x_1+y_1)+\varepsilon \mid x_2=v, y_2=w\right] &= \frac{1}{2}\tau\frac{a(1-a)s^2}{a^2s^2+2(1-a)^2S^2}(v+w) \\ &= \frac{1}{2}\tau\frac{a(1-a)\lambda}{a^2\lambda+2(1-a)^2}(v+w)\end{aligned}$$

where $\lambda = s^2/S^2$ is the ratio of genetic variance to wealth variance. The coefficient $\frac{a(1-a)\lambda}{a^2\lambda+2(1-a)^2}$ is increasing, then decreasing in a and is 0 for $a=0$ or $a=1$. □

Proof of Claim 3. Under RM, the correlation matrix for children's characteristics is:

$$\begin{aligned}\mathbb{C} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} \frac{\tau}{2} & 0 & \frac{\tau}{2} & 0 \\ 0 & \frac{\theta}{2} & 0 & \frac{\theta}{2} \end{pmatrix} \begin{pmatrix} s^2 & \sigma & 0 & 0 \\ \sigma & S^2 & 0 & 0 \\ 0 & 0 & s^2 & \sigma \\ 0 & 0 & \sigma & S^2 \end{pmatrix} \begin{pmatrix} \frac{1}{2}\tau & 0 \\ 0 & \frac{1}{2}\theta \\ \frac{1}{2}\tau & 0 \\ 0 & \frac{1}{2}\theta \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{2}s^2\tau^2 + 1 & \frac{1}{2}\theta\sigma\tau \\ \frac{1}{2}\theta\sigma\tau & \frac{1}{2}S^2\theta^2 + 1 \end{pmatrix}\end{aligned}$$

so that the correlation between characteristics for children is:

$$Corr\left(x_1', x_2'\right) = \frac{\frac{1}{2}\theta\sigma\tau}{\sqrt{\frac{1}{2}\tau^2 s^2 + 1}\sqrt{\frac{1}{2}\theta^2 S^2 + 1}}$$

Note that $\sigma = 0$ gives a zero correlation for children as well. Also, because $\theta < 1$ and $\tau < 1$, the correlation is less than the parents' correlation of σ/sS.

□

Proof of Claim 4. Again applying (2), under SGAM, the correlation between children's traits is:

$$Corr\left(x_1', x_2'\right) = \frac{\frac{1}{2}\theta\tau\left(\sigma + AC\right)}{\sqrt{\frac{1}{2}\tau^2\left(A^2 + s^2\right) + 1}\sqrt{\frac{1}{2}\theta^2\left(C^2 + S^2\right) + 1}}$$

This is positive if $\sigma = 0$ so long as $AC > 0$ i.e. $0 < a < 1$, and linear in θ.

□

Proof of Proposition 2. The fixed point condition on the covariance matrix is

$$\begin{pmatrix} s^2 & \sigma \\ \sigma & S^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2}s^2\tau^2 + 1 & \frac{1}{2}\theta\sigma\tau \\ \frac{1}{2}\theta\sigma\tau & \frac{1}{2}S^2\theta^2 + 1 \end{pmatrix}$$

which gives

$$s^2 = \frac{2}{2 - \tau^2}, S^2 = \frac{2}{2 - \theta^2}, \sigma = 0.$$

The asymptotic conditional expectation of children's genetics given parental SES is:

$$\mathbb{E}\left[\frac{\tau}{2}\left(x_1+y_1\right)+\varepsilon \mid x_2=v, y_2=w\right]=0$$

since the traits x_1, x_2, y_1, y_2 are uncorrelated.

□

Proof of Proposition 3. Start by characterizing the invariant distribution. This must satisfy:

$$\begin{pmatrix} \bar{s}^2 & \bar{\sigma} \\ \bar{\sigma} & \bar{S}^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\tau^2\left(\bar{A}^2+\bar{s}^2\right)+1 & \frac{1}{2}\theta\tau\left(\bar{\sigma}+\bar{A}\bar{C}\right) \\ \frac{1}{2}\theta\tau\left(\bar{\sigma}+\bar{A}\bar{C}\right) & \frac{1}{2}\theta^2\left(\bar{C}^2+\bar{S}^2\right)+1 \end{pmatrix}$$

where

$$\begin{aligned} \bar{A} &= \frac{a\bar{s}^2+(1-a)\,\bar{\sigma}}{\sqrt{a^2\bar{s}^2+(1-a)^2\,\bar{S}^2+2a\,(1-a)\,\bar{\sigma}}} \text{ and} \\ \bar{C} &= \frac{a\bar{\sigma}+(1-a)\,\bar{S}^2}{\sqrt{a^2\bar{s}^2+(1-a)^2\,\bar{S}^2+2a\,(1-a)\,\bar{\sigma}}}, \end{aligned}$$

Therefore

$$\begin{aligned} \bar{s}^2 &= \frac{1}{2}\tau^2\left(\frac{\left(a\bar{s}^2+(1-a)\,\bar{\sigma}\right)^2}{a^2\bar{s}^2+(1-a)^2\,\bar{S}^2+2a\,(1-a)\,\bar{\sigma}}+\bar{s}^2\right)+1 \\ \bar{S}^2 &= \frac{1}{2}\theta^2\left(\frac{\left(a\bar{\sigma}+(1-a)\,\bar{S}^2\right)^2}{a^2\bar{s}^2+(1-a)^2\,\bar{S}^2+2a\,(1-a)\,\bar{\sigma}}+\bar{S}^2\right)+1 \\ \left(1-\frac{1}{2}\theta\tau\right)\bar{\sigma} &= \frac{1}{2}\theta\tau\frac{\left(a\bar{\sigma}+(1-a)\,\bar{S}^2\right)\left(a\bar{s}^2+(1-a)\,\bar{\sigma}\right)}{a^2\bar{s}^2+(1-a)^2\,\bar{S}^2+2a\,(1-a)\,\bar{\sigma}} \end{aligned} \tag{13}$$

Define

$$\lambda=\bar{s}^2/\bar{S}^2 \text{ and } \mu=\bar{\sigma}/\bar{S}^2$$

The last equation gives

$$\left(1-\frac{1}{2}\theta\tau\right)\mu=\frac{1}{2}\theta\tau\frac{\left(a\mu+(1-a)\right)\left(a\lambda+(1-a)\,\mu\right)}{a^2\lambda+(1-a)^2+2a\,(1-a)\,\mu}$$

which is linear in λ; therefore

$$\lambda=\frac{\mu\left(\frac{1}{2}\theta\tau-1\right)\left((a-1)^2-2a\mu\,(a-1)\right)-\frac{1}{2}\theta\tau\mu\,(a-1)\,(-a+a\mu+1)}{a^2\mu\left(1-\frac{1}{2}\theta\tau\right)-\frac{1}{2}a\theta\tau\,(1-a+a\mu)} \tag{14}$$

The first two give:

$$\begin{aligned}&1-\frac{1}{2}\theta^2\left(\frac{\left(a\mu+(1-a)\right)^2}{a^2\bar{s}^2+(1-a)^2\bar{S}^2+2a(1-a)\bar{\sigma}}+1\right)\\=&\lambda-\frac{1}{2}\tau^2\left(\frac{\left(a\lambda+(1-a)\mu\right)^2}{a^2\bar{s}^2+(1-a)^2\bar{S}^2+2a(1-a)\bar{\sigma}}+\lambda\right)\end{aligned}$$

where λ is given by (14). This is equivalent to $F(\mu)=0$, where

$$\begin{aligned}F(\mu)=&\lambda-\frac{1}{2}\tau^2\left(\frac{\left(a\lambda+(1-a)\mu\right)^2}{a^2\lambda+(1-a)^2+2a(1-a)\mu}+\lambda\right)-\\&\left(1-\frac{1}{2}\theta^2\left(\frac{\left(a\mu+(1-a)\right)^2}{a^2\lambda+(1-a)^2+2a(1-a)\mu}+1\right)\right)\end{aligned}$$

This equation is quadratic in μ; thus it has two closed form solutions, of which one, denoted $\phi(a,\theta,\tau)$, is positive. Then

$$\begin{aligned}\lambda=&\psi(a,\theta,\tau)\\=&\frac{\phi(a,\theta,\tau)\left(\frac{1}{2}\theta\tau-1\right)\left((a-1)^2-2a\phi(a,\theta,\tau)(a-1)\right)-\frac{1}{2}\theta\tau\phi(a,\theta,\tau)(a-1)\left(-a+a\phi(a,\theta,\tau)+1\right)}{a^2\phi(a,\theta,\tau)\left(1-\frac{1}{2}\theta\tau\right)-\frac{1}{2}a\theta\tau\left(1-a+a\phi(a,\theta,\tau)\right)}\end{aligned}$$

Finally, the second equation in (13) gives

$$1=\frac{1}{2}\theta^2\left(\frac{\left(a\phi(a,\theta,\tau)+(1-a)\right)^2}{a^2\psi(a,\theta,\tau)+(1-a)^2+2a(1-a)\phi(a,\theta,\tau)}+1\right)+\frac{1}{\bar{S}^2}$$

which gives $\bar{S}^2$, then $\bar{s}^2$ and $\bar{\sigma}$ follow. □

Proof of Proposition 4. The correlation is signed by the covariance. Write C for the set of couples in the parents' generation with typical member $c=(x,y)$. Without loss of generality let $x_1\geq y_1$. Then, since the iso-attractiveness curves defined by f are downward-sloping, $x_2\leq y_2$.

Since $Ex=Ey=0$, the covariance among the parents' generation is

$$\int_C (x_1x_2+y_1y_2)/2\,dc$$

Write

$$x_1' = \tau x_1^* + \varepsilon \qquad \text{where } x_1^* = (x_1 + y_1)/2$$
$$x_2' = \theta x_2^* + \eta \qquad \text{where } x_2^* = (x_2 + y_2)/2$$

and write the children's covariance as

$$Cov(x_1', x_2') = Cov(\tau x_1^*, \theta x_2^*) + Cov(\tau x_1^*, \eta) + Cov(\varepsilon, \theta x_2^*) + Cov(\varepsilon, \eta).$$

By independence of the shocks, the last 3 terms are zero. So we need to show that

$$Cov(\tau x_1^*, \theta x_2^*) = \tau^2 \theta^2 Cov(x_1^*, x_2^*) > 0$$

Write

$$Cov(x_1^*, x_2^*) = \int_C x_1^* x_2^* \, dc$$

using that $Ex_1^* = Ex_2^* = 0$.

Take a typical parent, and write

$$x_1 x_2 = (x_1^* - \Delta_1)(x_2^* - \Delta_2)$$
$$y_1 y_2 = (x_1^* + \Delta_1)(x_2^* + \Delta_2)$$

where

$$\Delta_1 = (x_1 - y_1)/2; \ \Delta_2 = (x_2 - y_2)/2.$$

By assumption $\Delta_1 \geq 0$ and $\Delta_2 \leq 0$. Furthermore, if $a \in (0, 1)$, then for a set of positive measure, $\Delta_1 > 0$ and $\Delta_2 < 0$, by our assumption that not all matching couples are identical.

Taking the average of the parents gives

$$(x_1 x_2 + y_1 y_2)/2 = x_1^* x_2^* + \Delta_1 \Delta_2$$

and if $a \in (0, 1)$, this is strictly less than $x_1^* x_2^*$ for a set of positive measure. Plugging this into the integral gives

$$Cov(x_1, x_2) \leq Cov(x_1^*, x_2^*) = \int_C x_1^* x_2^* \, dc$$

with strict inequality if $a \in (0, 1)$. Since the parental covariance was 0 by assumption, this completes the proof.

□

Robustness checks

Table 6 reruns our central regressions, dropping the control for parents' 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 only; 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 3 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 = 0.24 for males and p = 0.022 for females. So for females there is some evidence against this linear model.

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

	(1)	(2)	(3)	(4)
Birth order	−0.0079	0.0026	−0.0030	0.0018
	(0.0074)	(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	0.0651 ***	0.0366 ***	0.0471 ***	0.0335 **
	(0.0065)	(0.0066)	(0.0120)	(0.0120)
Fluid IQ		0.0165 ***	0.0173 **	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	23861	23861	7673	7673
R2	0.011	0.028	0.018	0.027
logLik	−33521.263	−33311.976	−10771.954	−10737.107
AIC	67144.525	66731.953	21649.909	21582.215

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

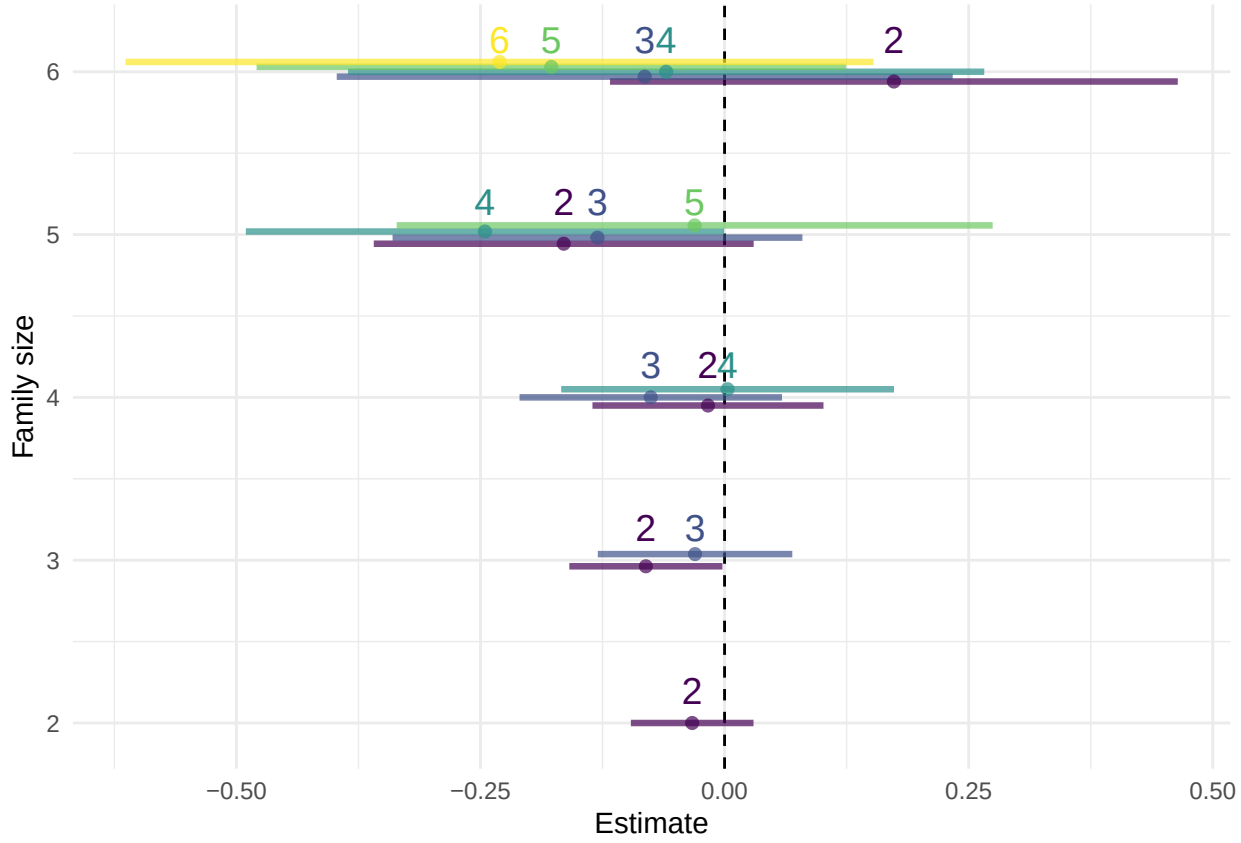


Figure 3: Regressions of spouse PSEA: birth order dummies within different family sizes. Labels show birth order. Lines are 95 per cent confidence intervals.

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

	(1)	(2)	(3)	(4)
Birth order 2	−0.0492 *	−0.0204	−0.0466	−0.0445
	(0.0232)	(0.0231)	(0.0411)	(0.0410)
Birth order 3	−0.0560	−0.0049	−0.0275	−0.0066
	(0.0375)	(0.0374)	(0.0672)	(0.0672)
Birth order 4	−0.0739	−0.0006	−0.0150	0.0028
	(0.0653)	(0.0650)	(0.1270)	(0.1267)
Birth order 5	−0.0773	0.0060	0.0897	0.1094
	(0.1189)	(0.1182)	(0.2296)	(0.2291)
Birth order 6	−0.2726	−0.2014	0.1936	0.2392
	(0.2370)	(0.2352)	(0.5972)	(0.5958)
University		0.2279 ***		0.1611 ***
		(0.0220)		(0.0374)
Income			0.0037 ***	0.0030 **
			(0.0010)	(0.0010)
Own PSEA	0.0583 ***	0.0320 **	0.0300 +	0.0189
	(0.0099)	(0.0100)	(0.0179)	(0.0180)
Parents' age at birth	0.0113 ***	0.0061 *	0.0102 *	0.0087 +
	(0.0026)	(0.0026)	(0.0047)	(0.0047)
Wald p-value, birth order	0.2553	0.8302	0.8294	0.7945
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	10229	10229	3414	3414
R2	0.013	0.029	0.028	0.033
logLik	−14326.354	−14241.674	−4826.210	−4816.814
AIC	28762.709	28599.348	9768.420	9751.628

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

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

	Males	Males	Females	Females	With children	With children
Birth order 2	−0.0261	0.0037	−0.0685 *	−0.0441	−0.0507 *	−0.0217
	(0.0356)	(0.0353)	(0.0309)	(0.0311)	(0.0246)	(0.0245)
Birth order 3	−0.0224	0.0348	−0.0826	−0.0408	−0.0626	−0.0088
	(0.0578)	(0.0570)	(0.0508)	(0.0508)	(0.0406)	(0.0403)
Birth order 4	−0.0991	−0.0064	−0.0493	0.0055	−0.0804	−0.0062
	(0.1001)	(0.0997)	(0.0850)	(0.0850)	(0.0685)	(0.0684)
Birth order 5	−0.2141	−0.1399	0.0280	0.1015	−0.1325	−0.0484
	(0.1457)	(0.1467)	(0.1468)	(0.1471)	(0.1070)	(0.1073)
Birth order 6	−0.6084 *	−0.5003 *	−0.0473	−0.0071	−0.1863	−0.0980
	(0.2580)	(0.2243)	(0.2246)	(0.2407)	(0.1891)	(0.1904)
University		0.2837 ***		0.1763 ***		0.2271 ***
		(0.0330)		(0.0306)		(0.0238)
Own PSEA	0.0591 ***	0.0252 +	0.0587 ***	0.0372 **	0.0631 ***	0.0346 **
	(0.0148)	(0.0150)	(0.0135)	(0.0137)	(0.0106)	(0.0107)
Parents' age at birth	0.0124 **	0.0055	0.0103 **	0.0064 +	0.0123 ***	0.0070 *
	(0.0040)	(0.0040)	(0.0034)	(0.0035)	(0.0028)	(0.0028)
Wald p-value, birth order	0.2262	0.2517	0.2962	0.6595	0.3239	0.9538
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	4684	4684	5545	5545	9148	9148
R2	0.018	0.042	0.018	0.028	0.015	0.032
logLik	−6597.708	−6540.149	−7698.634	−7670.271	−12799.184	−12718.639
AIC	13303.416	13194.298	15503.267	15452.543	25706.369	25551.278

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

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

	(1)	(2)	(3)	(4)
Birth order	−0.0311	−0.0065	−0.0132	−0.0063
	(0.0181)	(0.0177)	(0.0311)	(0.0307)
University		0.2276 ***		0.1590 ***
		(0.0259)		(0.0258)
Income			0.0037 ***	0.0030 ***
			(0.0008)	(0.0007)
Own PSEA	0.0529 ***	0.0286 *	0.0258	0.0155
	(0.0111)	(0.0113)	(0.0240)	(0.0237)
Parents' age at birth	0.0111 ***	0.0060 *	0.0102 *	0.0087 *
	(0.0028)	(0.0028)	(0.0040)	(0.0041)
Fluid IQ		0.0171 *	0.0197 +	0.0104
		(0.0065)	(0.0112)	(0.0118)
Height		0.0027 *	0.0046 *	0.0042 *
		(0.0011)	(0.0018)	(0.0018)
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	10229	10229	3414	3414
R2	0.014	0.030	0.028	0.033

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. 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	−0.0312 *	0.0005	0.0018
	(0.0146)	(0.0147)	(0.0270)
Age left full-time educ.		0.0497 ***	0.0417 ***
		(0.0044)	(0.0078)
Income			0.0028 *
			(0.0011)
Parents' age at birth	0.0113 ***	0.0049 +	0.0071
	(0.0026)	(0.0026)	(0.0047)
Own PSEA	0.0583 ***	0.0305 **	0.0198
	(0.0100)	(0.0101)	(0.0185)
Fluid IQ		0.0143 **	0.0068
		(0.0053)	(0.0097)
Height		0.0028 **	0.0042 *
		(0.0011)	(0.0019)
Family size dummies	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes
N	10229	10179	3407
R2	0.013	0.032	0.035
logLik	−14327.102	−14159.366	−4803.939
AIC	28754.204	28424.732	9715.878

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

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

	(1)	(2)	(3)	(4)
Birth order	−0.0356 *	−0.0133	−0.0230	−0.0202
	(0.0170)	(0.0170)	(0.0329)	(0.0329)
University		0.2155 ***		0.1678 ***
		(0.0270)		(0.0462)
Income			0.0018	0.0011
			(0.0015)	(0.0015)
Parents' age at birth	0.0121 ***	0.0073 *	0.0084	0.0071
	(0.0031)	(0.0032)	(0.0056)	(0.0056)
Own PSEA	0.0538 ***	0.0277 *	0.0210	0.0074
	(0.0121)	(0.0122)	(0.0230)	(0.0232)
Fluid IQ		0.0201 **	0.0111	0.0021
		(0.0062)	(0.0113)	(0.0116)
Height		0.0024 +	0.0044 +	0.0037
		(0.0013)	(0.0024)	(0.0023)
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes
N	6979	6979	2292	2292
R2	0.016	0.031	0.031	0.036
logLik	−9750.380	−9696.075	−3241.813	−3234.948
AIC	19598.761	19496.149	6585.625	6573.895

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

Table 12: Seemingly Unrelated Regressions on spouse characteristics

	Males		Females	
	Spouse PSEA	Spouse birth order	Spouse PSEA	Spouse birth order
Birth order	−0.040 +	0.072 **	−0.029	0.079 **
	(0.023)	(0.028)	(0.020)	(0.025)
PSEA	0.062 ***	−0.040 *	0.064 ***	−0.015
	(0.015)	(0.019)	(0.014)	(0.017)
Controls (parents' 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	4463	4463	5283	5283

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

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 center, 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, and significant when controlling for parent's age, they are always absolutely smaller than the corresponding coefficient in the main text. This 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	−0.0064	−0.0051	−0.0283 *
	(0.0080)	(0.0080)	(0.0144)
Own PSEA		0.0511 ***	0.0511 ***
		(0.0068)	(0.0099)
Parents' age at birth			0.0099 ***
			(0.0025)
Family size dummies	Yes	Yes	Yes
Birth month dummies	No	Yes	Yes
Birth year dummies	No	Yes	Yes
N	21604	21562	10423
R2	0.001	0.007	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, 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.

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 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.

Fagereng, Andreas, Magne Mogstad, and Marte Rønning. 2021. "Why Do Wealthy Parents Have Wealthy Children?" *Journal of Political Economy* 129 (3): 703–56.

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.

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.

Harden, Kathryn Paige. 2021. *The Genetic Lottery: Why DNA Matters for Social Equality*. Princeton University Press.

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, 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.

Howe, Laurence J, Michel G Nivard, Tim T Morris, Ailin F Hansen, Humaira Rasheed, Yoonsu Cho, Geetha Chittoor,

et al. 2021. "Within-Sibship GWAS Improve Estimates of Direct Genetic Effects." *bioRxiv*. https://doi.org/10.1101/2021.03.05.433935.

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.

Isungset, Martin Arstad, Jeremy Freese, Ole Andreassen, and Torkild Hovde Lyngstad. 2021. "Birth Order Differences in Education Are Environmental in Origin." *bioRxiv*.

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. 2007. "Annual Survey of Hours and Earnings." https://webarchive.nationalarchives.gov.uk/20160108022706/http://www.ons.gov.uk/ons/rel/ashe/annual-survey-of-hours-and-earnings/2007-results/index.html.

———. 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.

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

Rimfeld, Kaili, Eva Krapohl, Maciej Trzaskowski, Jonathan R. I. Coleman, Saskia Selzam, Philip S. Dale, Tonu Esko, Andres Metspalu, and Robert Plomin. 2018. "Genetic Influence on Social Outcomes During and After the Soviet Era in Estonia." *Nature Human Behaviour* 2 (4): 269–75. https://doi.org/10.1038/s41562-018-0332-5.

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

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

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.

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 Children's 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.