Trading social status for genetics in marriage markets: evidence from Great Britain and Norway

Abdel Abdellaoui[*], Oana Borcan[†], Pierre-André Chiappori[‡], David Hugh-Jones[§], Fartein Ask Torvik[¶] & Eivind Ystrøm[□]

2024-06-18

Abstract

Under social-genetic assortative mating (SGAM), socio-economic status (SES) and genetically inherited traits are both assets in marriage markets, become associated in spouse pairs, and are passed together to future generations. This gives a new explanation for persistent intergenerational inequality and the "genes-SES gradient" – observed genetic differences between high- and low-SES people. We model SGAM and test for it in two large surveys from Great Britain and Norway. Spouses of earlier-born siblings have genetics predicting more education. This effect is mediated by individuals' own education and income. Under SGAM, shocks to SES are reflected in the DNA of subsequent generations, and the distribution of genetic variants in society is endogenous to economic institutions.

Keywords: Assortative mating, MoBa, UK Biobank.

Introduction

How families are formed, and transmit traits and assets to their offspring, is crucial for understanding inequality and social structure. Assortative mating in marriage markets can increase inequality

[*]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. Email: davidhughjones@gmail.com

[¶]Norwegian Institute of Public Health, Oslo. Email: f.a.torvik@psykologi.uio.no

[□] Norwegian Institute of Public Health, Oslo. Email: eivind.ystrom@psykologi.uio.no

between families (Breen and Salazar 2011; Greenwood et al. 2014) and contribute to its persistence across generations, which is surprisingly high (Clark and Cummins 2015; Solon 2018). Wealthy families pass on advantages to their children through both genetic inheritance and environmental influence (Rimfeld et al. 2018; Björklund, Lindahl, and Plug 2006; Sacerdote 2011).

This paper examines a plausible aspect of marriage markets: both social status and genetics contribute to a person's attractiveness, and as a result, they may become associated in subsequent generations.[1] For example, suppose that wealth, intelligence and health are advantages in a potential spouse. Then wealthy people are more likely to marry intelligent or healthy people, and their children will inherit both wealth, and genetic variants associated with intelligence or health. We call this mechanism social-genetic assortative mating (SGAM). SGAM may be an important channel for the transmission of inequality. It creates a genetic advantage for privileged families, which may help to explain the long-run persistence of inequality. At the same time, this advantage is not a fact of biology, but is endogenous to the social structure. Indeed, under SGAM, environmental shocks to a person's social status may be reflected in the genetics of his or her children.

Below, we first write down a theory 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 *genes-SES gradient* – the systematic genetic differences between high- and low-SES people (Belsky et al. 2018; Rimfeld et al. 2018; Björklund, Lindahl, and Plug 2006). The dominant existing explanation for the gradient is meritocratic social mobility: if a genetic variant predicts success in the labour market, then it will become associated with high SES and will be inherited in high-SES families. While under meritocracy, genes causes SES, under SGAM causality goes both ways, from genes to SES and vice versa. Also, the size of the genes-SES gradient depends on economic institutions. Under institutions which increase intergenerational mobility, like high inheritance tax rates, the genes-SES gradient becomes weaker. On the other hand, an increase in meritocracy can make the gradient stronger.

Next, using data on spouse pairs from two large genetically-informed surveys in Great Britain and Norway, we test the hypothesis that a person's higher social status attracts spouses with genetic variants predicting greater educational attainment. Our genetic measure, the polygenic score for educational attainment (PSEA), derives from large-scale genome-wide association studies (Lee et al. 2018; Okbay et al. 2022). PSEA reflects a bundle of polygenic effects on underlying traits,

[1] *Social status* refers to characteristics that an individual possesses in virtue of their social position. For example, my wealth is a fact about me that holds in virtue of my relationship to certain social institutions (bank deposits, title deeds et cetera). Other examples include caste, class, income, and educational qualifications. *Socio-economic status* (SES) is a specific type of social status which exists in economically stratified societies, covering variables like educational attainment, occupational class, income and wealth (e.g. White 1982).

including intelligence, personality, and physical and mental health (Demange et al. 2021). PSEA predicts, and causes, educational attainment itself, as well as intelligence and labour market outcomes. It is already known that humans mate assortatively on PSEA (Hugh-Jones et al. 2016; Robinson et al. 2017; Torvik et al. 2022), which makes it a likely candidate for detecting SGAM.

The endogeneity of socio-economic status is the main challenge in identifying the effect of SES on the spouse's genetic endowment. For instance, people with high educational qualifications tend to also have high PSEA, and as mentioned above, they may take partners based on genetic similarity. Indeed, recent studies show strong assortative mating on PSEA, much more than we would expect if spouses matched only on observed measures of educational attainment (Okbay et al. 2022). To isolate the causal link from own SES to partner's genes, we use a shock to SES which is independent of own genetics. Specifically, we use a person's *birth order*. Earlier-born children receive higher parental investment and have better life outcomes, including measures of SES such as educational attainment and occupational status (Black, Devereux, and Salvanes 2011; Booth and Kee 2009; Lindahl 2008). At the same time, the facts of biology, in particular the so-called "lottery of meiosis", guarantee that siblings' birth order is independent of their genetic endowments.[2] Because birth order could affect partner choice through both SES and non-SES mechanisms, we run a mediation analysis similar to Heckman, Pinto, and Savelyev (2013), decomposing the treatment effect into effects of measured and unmeasured mediating variables. Specifically, we estimate a reduced-form model with spouse polygenic scores for educational attainment (PSEA) as the dependent variable, and own birth order as the main independent variable. We then add in to the model measures of own socio-economic status, including university attendance and income. Under certain assumptions, these variables can be interpreted as mediating the effect of birth order on spouse genetics.

In both Great Britain and Norway, later-born children have spouses with significantly lower PSEA in the reduced-form regressions. When we add mediators, including university attendance and/or income, the effect of birth order shrinks substantially, becoming insignificant in Great Britain, while the SES mediators significantly increase the spouse's PSEA. The results are robust to the inclusion of several controls, including non-SES mediators, and a rich set of own genetic traits. Thus, SES appears to mediate the effect of birth order on spouse genetics. The effects of individual mediators differ between the two countries. While university attendance explains more than a third of the effects of birth order in both Britain and Norway, income explains about 10% of the effects in Britain but has little or no independent effect in Norway. Although our main focus is on testing the basic mechanism of SGAM, this is suggestive evidence that in a more egalitarian society, some forms of SES are less important to the marriage market, with long-run implications for the genes-SES gradient.

[2] Although Muslimova et al. (2020) find that PSEA and birth order *interact* to produce human capital.

Both economists and geneticists study assortative mating. The economics literature has typically focused on educational similarities (e.g. Pencavel 1998; Chiappori, Salanié, and Weiss 2017) 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). Some papers have studied substitution between different traits.[3] For instance, Chiappori, Oreffice, and Quintana-Domeque (2012) showed that individuals trade off BMI for partners' income or education.

In genetics, Halsey (1958) showed that social mobility combined with assortative mating might increase the association between genetics and social class. Cloninger, Rice, and Reich (1979) model genetic and cultural transmission, where assortative mating is based directly on phenotype and culture is transmitted from parents. Assortative mating, modeled simply as a correlation coefficient, leads culture and genetics to be associated in offspring. Heath and Eaves (1985), following earlier papers (Rao, Morton, and Yee 1976; Rao, Morton, and Cloninger 1979), introduce "social homogamy", i.e. assortative mating by social background. Otto, Christiansen, and Feldman (1995) extend assortative mating to include both phenotypic and social homogamy.

More recently, interest in these topics has been revived by empirical findings from genomics. "Direct" effects of individual genetic variants, estimated by within-family studies, are different from "indirect" effects, i.e. associations found in the whole sample, and direct effects of polygenic scores can be smaller than population-wide associations (Howe et al. 2022; A. Young et al. 2022). Also, parental alleles which are *not* transmitted to the child correlate with child outcomes (Kong et al. 2018). Both these phenomena could be explained by confounding from gene-environment correlation, or by assortative mating (A. S. Young 2023; Nivard et al. 2024). Lastly, correlations between spouses' polygenic scores for education are higher than can be explained by assortative mating on measured phenotypic education alone (Okbay et al. 2022; Robinson et al. 2017; Torvik et al. 2022). To address this, several recent papers papers have estimated structural models of assortative mating in family data (Eaves et al. 1999; Torvik et al. 2022; Collado, Ortuño-Ortín, and Stuhler 2023; Rustichini et al. 2023). Because both cultural and genetic inheritance proceed from parents to children, it can be hard to differentiate them. For example, Collado, Ortuño-Ortín, and Stuhler (2023) derive extremely low estimates for heritability of education, within a model in which all genetic similarity between spouses is driven by matching either on the measured phenotype, or on a shared cultural factor; whereas Torvik et al. (2022) estimate partner correlation between "true" polygenic scores for education of 0.37, and heritability above 50%, in a model where environment is shared

[3] 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. Chiappori, Ciscato, and Guerriero (2021) analyse matching on multiple characteristics and show that a three-dimensional matching model fits their data.

between siblings but not across generations.[4] We get round this problem by cleanly identifying separate environmental and genetic contributions to assortative mating: environmental contributions using birth order, genetic contributions by comparing polygenic scores within siblings.

SGAM has consequences for long-run inequality. Long-run estimates of intergenerational persistence of wealth and status are often surprisingly higher than would be predicted from parent-child correlations (Clark and Cummins 2015; Barone and Mocetti 2021; Solon 2018), and distant relatives in the same generation are also more similar than parent-child and spousal correlations would predict (Collado, Ortuño-Ortín, and Stuhler 2023). Clark (2023) argues that this can be explained by an underlying process where unobserved genetic variation determines wealth. This requires a high degree of assortative mating. Our model shows that genetics may itself be a mediator for the transmission of SES, via "trading" in marriage markets. We also show how different social and economic institutions can affect that process. When SES is highly transmissible across generations, this increases the long-run association between SES and genetics. If so, institutional reforms that increase *intergenerational mobility*, like mass education or inheritance taxation, may affect not only economic but genetic inequality. Conversely, an increase in *economic meritocracy* increases the long-run association between SES and genetics,[5] posing the problem raised by M. Young (1958) and more recently Markovits (2019): meritocracy may be self-limiting or even self-undermining.

SGAM can also explain a large body of evidence for cross-sectional associations between genetics and social status. For example: from twin studies, the heritability of occupational class and educational attainment, i.e. the proportion of variance explained by genetic differences between individuals, is around 50% (Tambs et al. 1989). Genome-wide Complex Trait Analysis (GCTA) shows that the family socio-economic status of 2-year-old children can be predicted from their genes (Trzaskowski et al. 2014). Children born into higher-income families have more genetic variants predicting educational attainment (Belsky et al. 2018). Adoption studies show that both post-birth environment and pre-birth conditions (genetics and prenatal environment) contribute to the transmission of wealth and human capital (e.g. Björklund, Lindahl, and Plug 2006). There is also a genes-SES gradient in genetic predictors of health. DNA-derived scores predicting several health outcomes are associated with regional economic deprivation (Abdellaoui et al. 2019). The correlation between education and health may be mediated by shared genetic causes (Amin, Behrman, and Kohler 2015; Boardman, Domingue, and Daw 2015). Family SES correlates with several health-related polygenic scores (Selzam et al. 2019), and genetic variants associated with SES may explain the genetic correlations between many mental health outcomes (Marees et al.

[4] As Okbay et al. (2022) put it: "Because the parameters of a general biometric model cannot be separately identified from a small number of phenotypic correlations among different types of relatives, researchers typically have to assume that some of the parameters equal zero in order to estimate other parameters."

[5] See Proposition 4 below.

2021).

SGAM shows how marriage markets can lead high SES to be associated with different genetic variants, i.e. it can explain genes-SES gradients. The standard explanation for these gradients is returns to human capital in labour markets, also known as meritocratic mobility. Higher-ability parents 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).[6] 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). By contrast, SGAM does not require meritocracy. Even when social status is entirely ascribed, it can still become associated with certain genetic variants, so long as their associated phenotypes 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.

Lastly, we contribute to a literature in economics that examines the relationship between genetic and economic variables. Benjamin et al. (2011) and Benjamin et al. (2024) are reviews. Several recent papers use polygenic scores, in particular polygenic scores for educational attainment (e.g. Barth, Papageorge, and Thom 2020; Papageorge and Thom 2020; Ronda et al. 2020). Barban et al. (2021) use PSEA as an instrument for education in a marital matching model. These papers, like much of the behavior genetics literature, 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 of the estimating equation. Assortative mating and cultural inheritance are social processes, so we think there are good prospects for social scientists to contribute to understanding how genetic variants get distributed in society – what geneticists call "stratification" and "dynastic effects".

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." 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

[6] Belsky et al. (2018) offer three reasons for the association between education-linked genetic variants and SES, but do not consider SGAM.

that are partly under genetic influence ("intelligent," "tall," "kind," "physically attractive"). Despite this, to our knowledge, ours is the first work to model the socio-economic consequences of assortative mating between SES and genetics. In particular, no previous work has modelled how institutional variables moderate gradients between genes and SES, or how SGAM affects economic inequality.

Model

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

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

We interpret x_1 as a genetic measure, for example of genetic variants predicting height, physical attractiveness, health or intelligence. x_2 is a measure of socio-economic status, such as income or wealth, or social status more generally (we sometimes use "wealth" as a shorthand). The correlation between x_1 and x_2 is

$$Corr = \frac{\sigma}{sS} < 1.$$

People's attractiveness is given by

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

where $a \in [0,1]$ is a parameter reflecting the relative importance of genetics to wealth in the marriage market.[7] If $a = 0$, marriage markets are highly inegalitarian, such that only SES matters. If $a = 1$, marriage markets are economically egalitarian and only genetics matter. We expect realistic societies to fall between these extremes, with $0 < a < 1$. Then, both genes and SES matter to attractiveness, and as a result, social-genetic assortative mating (SGAM) takes place.[8]

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

[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.

[8] This model does not assume that people match *directly* on genetics, which most observers agree would be unlikely. Instead we assume that genetics may contribute to an attractive phenotype which is matched on.

assortative mating on attractiveness: x matches with y only if they are at the same quantile of attractiveness, i.e. if $i(x_1, x_2) = i(y_1, y_2)$. Within attractiveness quantiles, matching is random. This is the SGAM mechanism.

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

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

Our first proposition shows that if SGAM is taking place, i.e. if $0 < a < 1$, then there is a positive correlation between one partner's wealth and the other partner's genetics.

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

$$\mathbb{C}\begin{pmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} s^2 & \sigma & A^2 & AC \\ \sigma & S^2 & AC & C^2 \\ A^2 & AC & s^2 & \sigma \\ AC & C^2 & \sigma & S^2 \end{pmatrix} \tag{1}$$

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 x_1 and x_2 are already correlated ($\sigma > 0$) or if they are uncorrelated ($\sigma = 0$) and the attractiveness parameter a is strictly between 0 and 1.

Proof. See Appendix. □

We consider the distribution of couples' wealth. Under random matching this has mean 0 and

variance $2S^2$. Under SGAM, the variance is:

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

with strict inequality if $a < 1$ or $\sigma > 0$. The variance is decreasing in a and equals $4S^2$ if $a = 0$. Thus, SGAM increases cross-sectional inequality, but less so than pure matching on wealth.

Children

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

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

where x and y are the child's parents, and ε and η are independent 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 (Schmalhausen 1949; Sanjak et al. 2018).

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. Thus, θ captures social and economic institutions that affect this intergenerational transmission, from taxation and public education to hereditary nobility. If we interpret x_2 narrowly as wealth, $1 - \theta$ can be thought of as the rate of inheritance tax.

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 adult SES to partly depend on genetics. We show that even absent meritocracy, correlations between x_1' and x_2' can arise. In an extension below, we relax this assumption and allow meritocracy.

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

$$
\begin{aligned}
\mathbb{C} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} \frac{\tau}{2} & 0 & \frac{\tau}{2} & 0 \\ 0 & \frac{\theta}{2} & 0 & \frac{\theta}{2} \end{pmatrix} \begin{pmatrix} s^2 & \sigma & A^2 & AC \\ \sigma & S^2 & AC & C^2 \\ A^2 & AC & s^2 & \sigma \\ AC & C^2 & \sigma & S^2 \end{pmatrix} \begin{pmatrix} \frac{1}{2}\tau & 0 \\ 0 & \frac{1}{2}\theta \\ \frac{1}{2}\tau & 0 \\ 0 & \frac{1}{2}\theta \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{2}A^2\tau^2 + \frac{1}{2}s^2\tau^2 + 1 & \frac{1}{2}\theta\sigma\tau + \frac{1}{2}AC\theta\tau \\ \frac{1}{2}\theta\sigma\tau + \frac{1}{2}AC\theta\tau & \frac{1}{2}C^2\theta^2 + \frac{1}{2}S^2\theta^2 + 1 \end{pmatrix} \qquad (3)
\end{aligned}
$$

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 inherit genes that are attractive in marriage markets.

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

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

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 SES variance, is zero for $a = 0$ or $a = 1$, and is highest for intermediate values of a.*

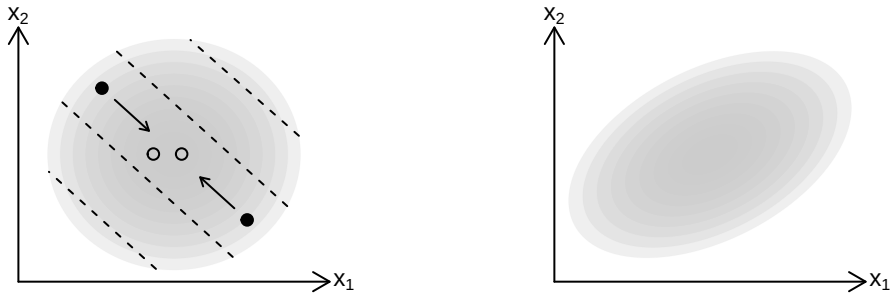
Next, we study the correlation between children's traits 1 and 2 as a function of σ, the covariance of parents' traits. We first consider the general case, then concentrate on $\sigma = 0$, i.e. when traits are initially uncorrelated.

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

Claim 4. *Under SGAM, if genetics and SES 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 below.

Figure 1 shows the intuition behind the model. Parents match on downward-sloping attractiveness isoquants given by $ax_1 + (1-a)x_2 = u$. Their children are in between them on both dimensions. This compresses the distribution along the attractiveness isoquants, which leads to a positive correlation between genetics and SES. The correlation between x_1' and x_2' is 0 when $a = 0$ or $a = 1$, because then spouses don't trade off SES for genes. It is highest for intermediate values of a.



Parents Children

Figure 1: Theory. The shaded area is the population distribution. Parents (solid circles) match along attractiveness isoquants (dotted lines). Children (hollow circles) are between them. As a result, the children's distribution is squeezed along attractiveness isoquants, and x_1 and x_2 become associated.

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.

We now consider the asymptotic distribution of x_1 and x_2 when the matching process is repeated over many generations. As we would expect, our main results continue to hold.

Proposition 2. *Under random matching, the dynamics converges to a stationary distribution that*

is normal with mean zero and covariance matrix

$$\mathbb{C}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{2}{2-\tau^2} & 0 \\ 0 & \frac{2}{2-\theta^2} \end{pmatrix}$$

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

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*

$$\mathbb{C}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \bar{s}^2 & \bar{\sigma} \\ \bar{\sigma} & \bar{S}^2 \end{pmatrix}$$

Moreover, the asymptotic correlation between characteristics, $corr = \bar{\sigma}/\bar{s}\bar{S}$, is non-negative, positive for $0 < a < 1$, increasing in θ and increasing then decreasing in a. The coefficient of parents' wealth on children's genetics is also positive for $0 < a < 1$.

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

Figure 2 plots the asymptotic correlation between x_1 and x_2, for $\tau = 0.95$. It is maximized for intermediate levels of a. 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 (unsurprisingly), but also the correlation between genetics and wealth.

Extensions

We consider three extensions. First, the relative attractiveness of genes and SES might differ for men and women. Our basic result extends to this setup.

Claim 5. *Suppose that men's and women's attractiveness is given by*

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

respectively, with $0 \leq a \leq 1, 0 \leq b \leq 1$. Then if $\sigma = 0$, children's characteristics x_1' and x_2' will be positively correlated unless $a = b = 0$ or $a = b = 1$. The correlation is increasing in θ.

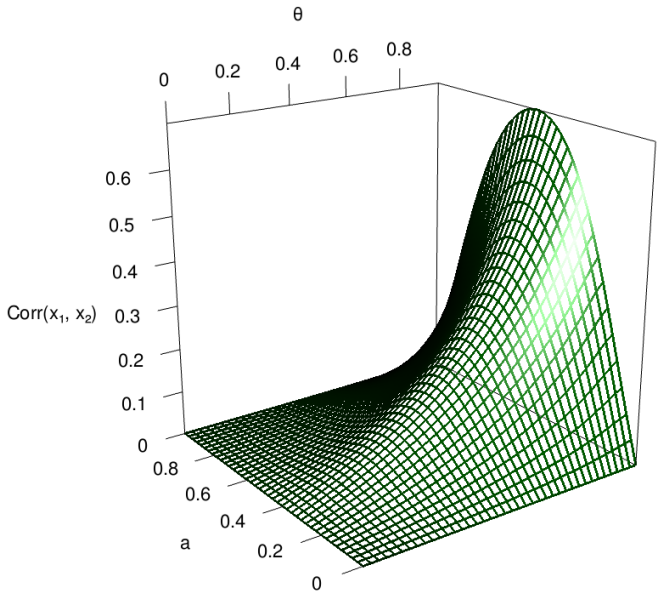


Figure 2: Long-run correlation between genetics x_1 and SES x_2, by weight of genetics in spouse matching (a) and strength of inheritance of SES (θ). τ = 0.95.

Interestingly, the x_1-x_2 correlation is highest when a and b are most different from each other. So gender differences in what counts as attractive make the effects of SGAM stronger. Intuitively, if one sex only assorts on SES while the other sex only assorts on genetics, this induces a very reliable correlation between genes and SES in couples, since (e.g.) every high-SES male is matched for sure with a high-genetics female.

Second, in modern meritocracies, people's adult SES depends not just on their parents' social status and on chance, but also on their own effort and skills, which might be related to their genetics. So, let

$$\begin{aligned} x_1' &= \tau \frac{x_1 + y_1}{2} + \varepsilon \\ x_2' &= \gamma x_1' + \theta \frac{x_2 + y_2}{2} + \eta \end{aligned} \tag{4}$$

where $\gamma > 0$ represents the effect of own genetics on own SES. The basic result continues to hold, and also, the degree of meritocracy γ increases the correlation between genes and SES; a highly meritocratic society may in the long run lead to a steep genes-SES gradient.

Proposition 4. *Under SGAM and equation* (4)*, if genetics and SES are uncorrelated for the parents, then they are positively correlated for the children so long as* $0 < a < 1$ *or* $\gamma > 0$*. The correlation is increasing in* γ*. Also, so long as* $\gamma > 0$ *and either* $0 < a < 1$ *or* $\sigma > 0$*, the coefficient of parents' wealth on children's wealth exceeds* θ*.*

Surprisingly, in this case, the children's genes-SES correlation is not always increasing in θ. The reason is that when γ is high, a higher θ decreases the proportion of x_2' that comes via γ from own genetics, and increases the proportion that comes from parents' SES, which may be less strongly correlated with own genetics. However, computing the long-run asymptotic correlation along the lines of Proposition 3 shows that it is increasing in θ for all but values of θ very close to 1.

Of interest to economists, Figure 3 shows how meritocracy, wealth transmission and assortative mating interact to produce economic inequality as measured by the equilibrium standard deviation of wealth x_2. In a 'status-based' society with persistent wealth and no meritocracy (solid line, θ = 0.9, γ = 0), inequality is highest for assortative mating on wealth alone ($a = 0$). In a society with less persistent wealth and more meritocracy (dashed line, θ = 0.3, γ = 0.3), inequality is highest when people assort on genetics ($a = 1$). Assortative matching on genetics can also change the rankings of the two societies with respect to inequality.

Third, we consider non-normal distributions of x_1 and x_2, non-normal shocks ε and η, and non-

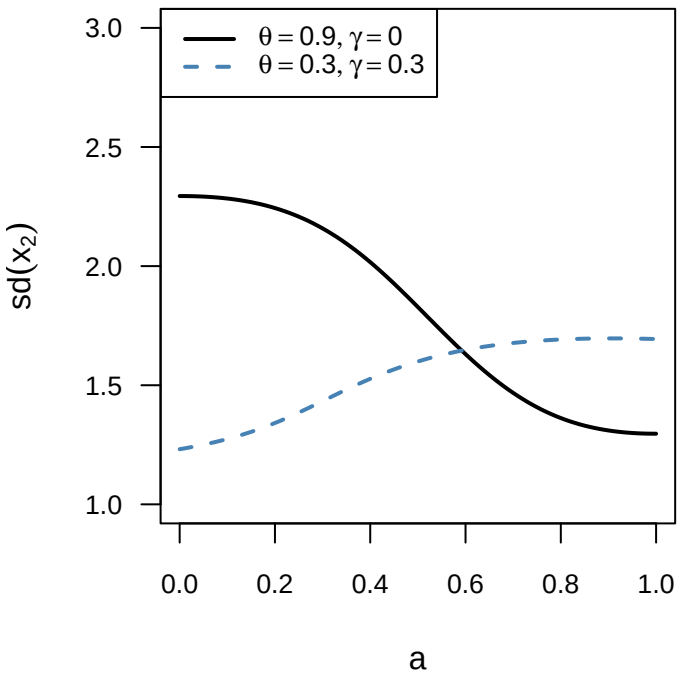


Figure 3: Meritocracy, assortative mating and economic inequality. Curves show equilibrium standard deviation of wealth x_2 for different parameter combinations.

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 the distribution of x is that not everybody with attractiveness i is both genetically and socially identical. In particular, this allows for discrete distributions, like some kinds of social status for x_2, and monogenic or oligogenic attractiveness phenotypes for x_1.

Proposition 5. *Let attractiveness be given by* (5)*. Let* (x_1, x_2) *have any distribution such that a positive measure of the population has* $i(x) = i$ *where the conditional distribution of* (x_1, x_2) *given* $i(x) = i$ *is non-degenerate. Let* η *and* ε *be mean 0 and independent of* x *and each other. If genetics and SES are uncorrelated for the parents, then the correlation among children is non-negative, and strictly positive if* $0 < a < 1$.

Other extensions are possible. We assumed that all couples have the same number of children. If fertility increased with x_1 or x_2, we would expect this to reduce the variance of traits in the children's generation and possibly also their covariance. Here, matching preferences, summarized by the a parameter, are exogenous; it would be natural to model a as an equilibrium outcome.

For example, if parents care about their children's wealth, a might decrease in θ and increase in γ. Indeed, below we see suggestive evidence that income is valued differently in Norwegian and British marriage markets.

Discussion

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, including wealth, income and occupation. Regarding genetics, standards of physical attractiveness, and other genetically-influenced characteristics which make someone a "good match", vary across societies and over time. The central prediction of the model is that whatever those characteristics, in the long run they will become correlated with SES.

Recent empirical work shows high persistence of SES over time, in particular at the top. Clark (2023) argues that this could be explained by unobserved genetic variation. Proposition 4 shows that if genes affect own wealth directly, under assortative mating, the regression coefficient of parents' wealth on own wealth exceeds the "direct" coefficient θ, because parents' wealth correlates with parents' genetics and via that with own wealth. Thus, regressions of wealth on wealth may include the effect of unobserved genetic variation. This may be a confound due to pre-existing gene-SES correlation (if $\sigma > 0$). But under SGAM ($0 < a < 1$) it can also be a mediating variable, since changes in someone's wealth may indeed affect the identity of their spouse, hence the genetics of their offspring, and from that their offspring's adult wealth.

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 this. Polygenic scores for educational attainment have smaller effects in between-sibling regressions, where between-family variation in SES is partialled out and where genetic variants are guaranteed to be randomly allocated, than in regressions which pool the whole sample (Howe et al. 2022). Parents' genetic variants which are *not* passed on to children predict children's characteristics, partly due to social stratification in the geneticists' sense of non-random mating (Kong et al. 2018; A. S. Young 2023).

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 ($a = 0$). Also, SGAM is increased by the institutional variable θ, which captures intergenerational persistence of SES. This implies that policy has long-run effects on biosocial structure: reducing θ not only increases intergenerational

mobility, but reduces the correlation of genes with SES, and hence the unfairness of what Harden (2021) calls the "genetic lottery". Conversely, reforms that increase meritocracy (γ, Proposition 4) may strengthen the genes-SES gradient.

Data and methods

The central insight in our model is that higher SES and good genes assort in the marriage market. We wish to test this directly, i.e. to test whether $0 < a < 1$ in the attractiveness equation

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

where x_2 is social status and x_1 is genetic endowment. Consider the effect of a change in x_2 holding x_1 constant. If $a = 1$ then this will not change $i(x)$ and therefore will not change the expected characteristics of the spouse. So, if we regress spouse's x_1 on own x_2, and reject the null of no effect, we can reject $a = 1$.[9]

We use data from two sources: Great Britain and Norway. This allows us to check our basic result in two different societies, and also to make (tentative) comparisons between them. Our Great Britain data comes from the UK Biobank, a study of about 500,000 individuals born between 1935 and 1970 (Bycroft et al. 2018). 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;[10] 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"); and 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.[11]

Our Norway data comes from the Norway Mother, Father and Child Cohort Study (MoBa), a population-based study of pregnant women and their partners and children (Magnus et al. 2016;

[9] Conceivably, if $a = 0$ but there is a pre-existing correlation between x_1 and x_2 in the population, then an increase in own x_2 will increase spouse's expected x_2 and therefore spouse's expected x_1, even though the latter does not enter the attractiveness equation. We can separately test the null that $a = 0$ by regressing spouse's x_2 on own x_1, holding own x_2 constant. Existing work has already linked own genetics to spouse's SES, e.g. education, so we focus on the other direction and treat this direction as a robustness check below.

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

[11] In the appendix, we test the validity of our matching process by counting the proportion of pairs who had a shared genetic child, in a subsample of the data. We also check whether any misidentified pairs might have biased our results, by constructing a dataset of "known fake pairs".

Paltiel et al. 2014). Participants were recruited from all over Norway from 1999-2008. 41% of women consented to participation. In this paper, we use about 100,000 genotyped individuals and about 45,000 genotyped spouse pairs. The Norway data has some advantages over UK Biobank, including higher participation, larger sample size, and spouse pairs which are known rather than inferred. On the other hand it is missing some variables, including IQ measures and self-reported health.

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; Torvik et al. 2022). Second, PSEA predicts a set of important socioeconomic variables, including not only education but also social and geographic mobility, IQ, future income and wealth (Belsky et al. 2016; Barth, Papageorge, and Thom 2020; Papageorge and Thom 2020).[12]

PSEA in the UK was calculated using per-SNP summary statistics from Lee et al. (2018), re-estimated excluding UK Biobank participants; in Norway, using statistics from Okbay et al. (2022). The score was normalized to have mean 0 and variance 1. Because polygenic scores are created from estimates of many small 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 (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). Also, polygenic scores are no more guaranteed to be causal than any other independent variable. For example, social stratification by ancestry may lead genes to be associated with educational attainment even if they play 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. Effect sizes are smaller but still non-trivial in within-siblings regressions (Lee et al. 2018), where they can be interpreted as causal, since genetic variation across siblings is guaranteed to be random by the biological mechanism involved – the "lottery of meiosis" (see below). 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

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

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 percentage points ($p < 2 \times 10^{-16}$). This suggests that about half of the raw correlation of PSEA with university attendance is down to environmental confounds like parental nurture, while the remainder is causal (cf. Lee et al. 2018). Still, 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).

We use two measures of socio-economic status: 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 in itself. The MoBa data includes both university attendance and income. UK Biobank includes university attendance, but only has a direct measure of current household income, which is inappropriate for our purposes because it includes income from both spouses and is measured after marriage. Instead, we estimate income in the respondent's first job, by matching the job's Standard Occupational Classification (SOC) code with average earnings by SOC from National Statistics (2007). Job codes are only available for a subset of respondents. We convert income to a z score among each group of respondents with the same gender and year of birth.

Figure 4 illustrates the core idea of SGAM within the UK Biobank data. The X axis shows a measure of one partner's socio-economic status: university attendance or income. 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 in their first job. Since DNA is inherited, these people's children will also have higher PSEA.[13]

These plots 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). Indeed, recent studies show much higher levels of GAM than could be explained by matching on the observed education phenotype alone (Okbay et al. 2022). So, 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 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

[13] Figure 5 in the appendix shows the same plot for the MoBa sample.

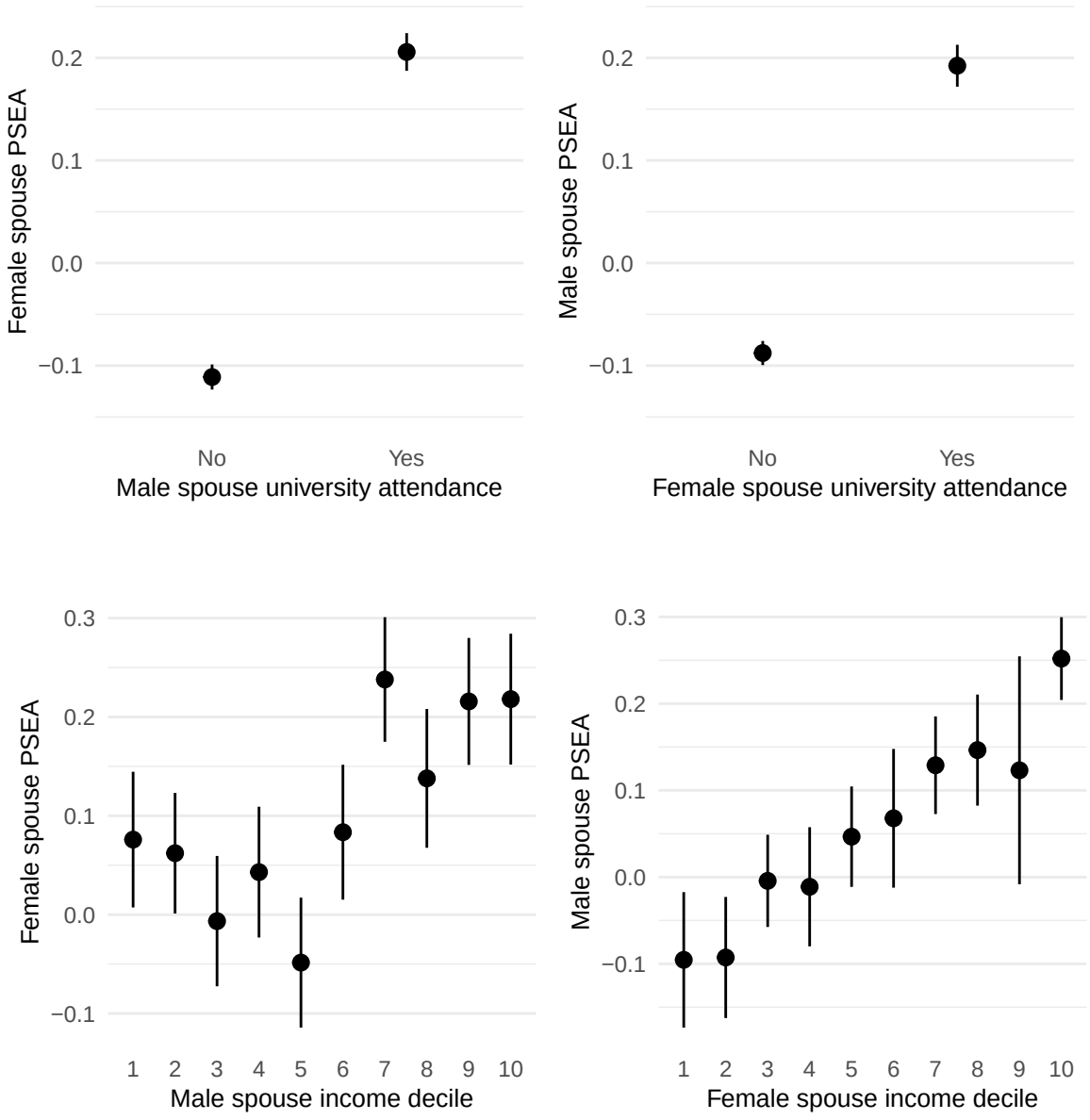


Figure 4: Spouse polygenic score for educational attainment (PSEA) against own university attendance and own income in first job (Great Britain). Lines show 95% confidence intervals. PSEA is normalized to have mean 0 and variance 1. Income is estimated from the respondent's first job, as the average income of the SOC job code.

(Lindahl 2008; Booth and Kee 2009; Black, Devereux, and Salvanes 2011).[14] On the other hand, all full siblings have the same *ex ante* expected genetic endowment from their parents, irrespective of their birth order. This is guaranteed by the biological mechanism of meiosis, which ensures that any gene is transmitted from either the mother or the father to the child, with independent 50% probability (Mendel 1865; Lawlor et al. 2008). For example, siblings' expected polygenic score is equal to the mean of their parents' polygenic scores.[15] We can therefore use birth order as a "shock" to social status. "Shock" is in quotes because 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 correlate with household SES at the time of birth. We only claim that birth order is exogenous to genetic variation.

Although birth order within a given family is independent of genetics, birth order in the whole sample might be correlated with other environmental and genetic factors. In particular, birth order naturally correlates with family size, i.e. the total number of siblings born, because e.g. third-born children must be in a family of at least three siblings. It also correlates with parental age, which may affect the income and maturity of the parents. To control for family size, we use dummies for each possible value of family size. In most regressions, we use only respondents with between 1 and 5 siblings, i.e. with a family size of 2-6: beyond that, estimation would get noisy because of small cell sizes. To control for parental age, we use father's and/or mother's age at the respondent's birth. This is calculated from the relevant parent's current age. In the UK Biobank, this data is only available if the respondent's parent was still alive. For other controls we use respondent's month of birth, year of birth, and own PSEA. Our claim is that controlling for these variables, birth order is independent of genetic variation.

Decomposing the birth order effect on spouse genetics

Birth order offers a way to test the central assumption in our theoretical model – that SES is traded for genes in the marriage markets – by providing a shock to SES which is exogenous to own genet-

[14] Earlier work was ambiguous on the effects of birth order (e.g. Hauser and Sewell 1985; Hanushek 1992). However, this work often used unrepresentative samples and/or did not control for family size or parental age. More recent work improves on this and shows clear birth order effects. Kantarevic and Mechoulan (2006) show that parental age is an important confound for birth order. Black, Devereux, and Salvanes (2005) show substantial birth order effects in the whole Norwegian population, even in a family fixed-effects specification, and after controlling for mother's age. Booth and Kee (2009) examine UK families, controlling for family size and for parental age at birth, and show significant and substantial birth order effects on education. Some studies (e.g. Black, Devereux, and Salvanes 2005; De Haan 2010) use twin births or the gender mix of children as instruments for family size. They too typically find that birth order has large negative effects.

[15] 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.

ics. 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 fall foul of the exclusion restriction.

Instead, we conduct a mediation analysis, following the strategy of Heckman, Pinto, and Savelyev (2013). We first confirm 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. Most importantly, 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.

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 m_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 m_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 m_b into a set J_m of measured mediators, including university attendance and income in first job, and a set J_u of mediators that we cannot measure.

Our linear model is:

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

where $\tilde{\varepsilon}_b$ is a mean-zero residual assumed independent of m_b and $\mathbf{X}$; $\tau_b = \kappa_b + \sum_{j \in J_u} \alpha_b^j E(m_b^j)$; and $\varepsilon_b = \tilde{\varepsilon}_b + \sum_{j \in J_u} (m_b^j - E(m_b^j))$. We simplify by assuming that $\beta_b = \beta$ and $\alpha_b = \alpha$ for all b, i.e. that the effects of $\mathbf{X}$ and m_B don't differ by birth order.[16] We assume differences in unmeasured

[16] Under the assumption that measured and unmeasured mediators are uncorrelated, we can test these assumptions

investments due to b are independent of $\mathbf{X}$.

We use a linear model for each observed mediator variable:

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

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) and (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 a change from birth order b to b' into the effect of measured mediators m^j and unmeasured mediators on the outcome:

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

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:

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

Estimating the above by OLS will generate unbiased estimates of α^j if m^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 investments (specifically SES) should

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. When we run the model with interactions, only one interaction is significant after Bonferroni correction at $p < 0.05/34$: the interaction of income in first job with the dummy for birth order 6. So overall, the uninteracted model seems a good enough approximation.

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 $\tilde{\varepsilon}_b$. With respect to assumption (a), in our regressions we control for a set of potential alternative mediators available in the data: height, BMI, and (in UK Biobank only) fluid IQ and self-reported general health. With respect to assumption (b), we use further controls, such as parental age, year of birth, and own PSEA, to reduce unobserved variation in the error term.

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 from OLS regressions based on equation (7) for each measured mediator (in particular university attendance and income) on $\mathbf{X}$ 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 on university/income from equation (7). We now apply this framework to each of our two samples.

Results: Great Britain

We first regress our measures of socio-economic status, university attendance and income from first job, on birth order in the UK Biobank spouse pairs. We also do the same for four non-SES mediators that could be affected by birth order: fluid IQ, height, body mass index (BMI) and a measure of self-reported health. We control for respondent's own PSEA and their parents' age at birth (see below). Table 1 shows that birth order significantly predicts all the mediators. Effects are quite substantial: on average, one extra elder sibling reduces the chance of attending university by about 7.9 percentage points, income by about 0.077 standard deviations, fluid IQ by about 0.27 points on a 13 point test, height by about 0.7 centimeters, and self reported health by 0.043 points on a 4-point scale; and increases BMI by 0.19.

Next we run regressions of spouse PSEA on birth order. Table 2 reports the results. Column 1 reports results controlling 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 reports results controlling for 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 seasonal effects. The effect size of birth order is not much changed.

Column 3 reports results controlling for 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 may help later children.[17] Kantarevic and Mechoulan (2006) show that mother's age at childbirth indeed mechanically offsets the negative effect of birth order. Including parents' age means we can separate the effect of parental age from birth order.[18] 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.

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.[19]

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 controls for our first measure of socio-economic status: university attendance. We also include potential non-SES mediators, which are affected by birth order and might affect spouse matching: fluid IQ, height, BMI and self-reported health. Column 3 adds our second measure of socio-economic status, income in first job. Column 4 includes both.

When we add university attendance and other mediators (column 2), the effect of birth order drops and becomes insignificant at 5%, while the coefficient for university is positive and highly significant. Fluid IQ, height and BMI are also positive and significant, while self-reported health has the right sign but is insignificant. Controlling for income instead of university attendance (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

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

[18] A possible critique is that later siblings by definition have older parents, so the parental age control is inappropriate. This may be true if we are estimating overall inequality between siblings. But here we are interested in using birth order as a shock, so it's reasonable to separate out its effects from those of parental age. The resulting estimate gives the counterfactual impact of having one more elder sibling, holding the age of one's parents constant.

[19] Table 11 in the appendix reports results without controlling for parents' age.

spouse PSEA, divided by the coefficient of birth order on spouse PSEA estimated from column 1, i.e. without mediators. Education explains about 38-55 percent of the effect, much more than all the other mediators. Income, fluid IQ, height and BMI all explain between 6 and 17 percent of the effect, depending on the specification.

These results provide evidence that birth order affects spouse PSEA via education and income, with education being especially important. The effect size of birth order is small (a few percent of a standard deviation of PSEA), but what matters is the effect size of education and income. The effect of education in particular is quite large as measured here, and since it also appears to mediate the purely environmental shock of birth order, it cannot just be due to an unobserved correlation with own genetics.

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. Claim 5 showed that these differences would strengthen the effects of SGAM. To test for this, we separately regress female spouses' PSEA on male birth order, and male 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 present results using birth order of male respondents to predict female spouses' PSEA. Column 1 shows the regression of birth order plus controls; in column 2, we add university attendance and non-SES mediators (here, we exclude first job income 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. The effect of birth order is imprecisely estimated in these subsets due to the lower sample size. However, the pattern of coefficient sizes is the same as in the main regression: the coefficient of birth order is about -0.3 (and very similar between the sexes), and adding university attendance reduces the absolute size of the birth order effect. Columns 5 and 6 show results from regressions on the subsample of couples with children. Here, birth order is significant in the base specification, and again, university attendance still seems to mediate the birth order effect.

Table 1: Regressions of mediators on birth order (Great Britain)

	University	Income	Fluid IQ	Height	BMI	Health
Birth order	−0.0790 ***	−0.0771 **	−0.2733 ***	−0.7012 ***	0.1907 **	−0.0430 ***
	(0.0068)	(0.0283)	(0.0306)	(0.1346)	(0.0661)	(0.0104)
PSEA	0.0889 ***	0.1050 ***	0.3180 ***	0.1970 *	−0.4281 ***	0.0533 ***
	(0.0046)	(0.0194)	(0.0199)	(0.0915)	(0.0460)	(0.0069)
Parents' age at birth	0.0163 ***	0.0174 ***	0.0588 ***	0.1514 ***	−0.0989 ***	0.0110 ***
	(0.0012)	(0.0044)	(0.0053)	(0.0240)	(0.0118)	(0.0018)
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	10220	3412	10220	10220	10220	10220
R^2	0.074	0.027	0.058	0.017	0.023	0.018

Estimates from OLS regressions with the mediators (university attendance, income, fluid IQ, height, BMI, self-reported health) 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 clustered by spouse pair in parentheses.

Table 2: Regressions of spouse PSEA on birth order (Great Britain)

	(1)	(2)	(3)
Birth order	−0.0091	−0.0075	−0.0314 *
	(0.0074)	(0.0074)	(0.0145)
Own PSEA		0.0650 ***	0.0573 ***
		(0.0086)	(0.0120)
Parents' age at birth			0.0116 ***
			(0.0026)
Family size dummies	Yes	Yes	Yes
Birth month dummies	No	Yes	Yes
Birth year dummies	No	Yes	Yes
N	23840	23797	10206
R^2	0.003	0.010	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 own PSEA, parents' age at birth (the mean of parents' ages), and further controls (family size, birth year, and birth month dummies) in columns 2−3 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 clustered by spouse pair in parentheses.

Table 3: Regressions of spouse PSEA on birth order and mediators (Great Britain)

	(1)	(2)	(3)	(4)
Birth order	−0.0314 *	−0.0045	−0.0092	−0.0034
	(0.0145)	(0.0145)	(0.0270)	(0.0270)
University		0.2179 ***		0.1498 ***
		(0.0225)		(0.0378)
Income			0.0621 ***	0.0488 **
			(0.0178)	(0.0181)
Fluid IQ		0.0172 **	0.0194 *	0.0109
		(0.0052)	(0.0093)	(0.0096)
Height		0.0029 **	0.0058 **	0.0052 **
		(0.0011)	(0.0019)	(0.0019)
BMI		−0.0109 ***	−0.0110 **	−0.0106 **
		(0.0022)	(0.0040)	(0.0040)
Self-reported health		0.0181	0.0147	0.0082
		(0.0152)	(0.0274)	(0.0273)
Own PSEA	0.0573 ***	0.0263 *	0.0212	0.0117
	(0.0120)	(0.0121)	(0.0202)	(0.0203)
Parents' age at birth	0.0116 ***	0.0053 *	0.0089 +	0.0077
	(0.0026)	(0.0026)	(0.0047)	(0.0047)
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes
N	10206	10206	3407	3407
R^2	0.013	0.032	0.030	0.034

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 (11). PSEA is the polygenic score for educational attainment, which is normalized with mean 0 and standard deviation 1. We include own PSEA, mean of parents' ages at birth, potential non-SES mediators (fluid IQ, height, BMI, self-reported health) and further controls (family size, birth year, and birth month dummies) 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 clustered by spouse pair in parentheses.

Table 4: Percent of birth order effects accounted for by mediators (Great Britain)

	Model 2 (%)	**Model 3 (%)**	**Model 4 (%)**
University	54.9		37.7
Income		15.3	12.0
Fluid IQ	15.0	16.9	9.5
Height	6.6	13.0	11.7
BMI	6.6	6.7	6.4
Self-reported health	2.5	2.0	1.1

Percentage of the effects of birth order in Table 3, columns 2 to 4, explained by by each mediating variable.

Table 5: Regressions of spouse PSEA on birth order: subsets (Great Britain)

	Male respondents	Male respondents	Female respondents	Female respondents	With children	With children
Birth order	−0.030	−0.001	−0.031	−0.009	−0.035 *	−0.007
	(0.022)	(0.022)	(0.019)	(0.019)	(0.015)	(0.015)
University		0.272 ***		0.169 ***		0.217 ***
		(0.033)		(0.031)		(0.024)
Fluid IQ		0.019 *		0.015 *		0.022 ***
		(0.008)		(0.007)		(0.006)
Height		0.004		0.004 +		0.002 *
		(0.002)		(0.002)		(0.001)
BMI		−0.008 *		−0.012 ***		−0.011 ***
		(0.004)		(0.003)		(0.002)
Self-reported health		0.028		0.010		0.022
		(0.022)		(0.021)		(0.016)
Own PSEA	0.059 ***	0.022	0.057 ***	0.030 *	0.062 ***	0.029 *
	(0.015)	(0.015)	(0.014)	(0.014)	(0.013)	(0.013)
Parents' age at birth	0.013 **	0.005	0.011 **	0.005	0.013 ***	0.006 *
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
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	4675	4675	5531	5531	9127	9127
R^2	0.017	0.043	0.017	0.031	0.015	0.035

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 own PSEA, parents' age at birth (the mean of parent's ages) and further controls (family size, birth year, and birth month dummies) 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 clustered by spouse pair in parentheses.

Results: Norway

Now we turn to the results from MoBa in Norway. Some of the variables are different from those for UK Biobank. Spouse PSEA is calculated using summary statistics from Okbay et al. (2022), aka "EA4", rather than "EA3".[20] Income is from all sources, reported at age 30, converted to a z score among respondents with the same gender, year of birth and year of reported income. In particular, some low-income individuals may be in continuing education or in relationships already.[21] Data on IQ and self-reported health is unavailable. The sample is also younger than UK Biobank, spouse pairs are given rather than constructed, and all couples have at least one child.

Tables 6 and 7 are the equivalent of Tables 3 and 4 for respondents in the MoBa dataset. (Equivalents to Tables 1 and 2 are in the appendix.) The broad pattern of results is similar to Britain. The larger N gives higher statistical significance. Effects of all the variables are in the expected direction, except that the coefficient on income changes sign when university is also included. In particular, the point estimate of the total effect of birth order is about twice as high as in Britain, and the estimated effect of university attendance is also higher. The effect of own PSEA is also about twice as high. Note that these differences could be driven by the PSEA score containing less noise in the Norwegian sample.

Adding university attendance again substantially and significantly reduces the effect of birth order, though here, birth order remains independently significant and substantively large even controlling for university. In Norway, however, it is much less clear that income is an important mediator on its own. Adding income barely changes the effect of birth order (column 3). Table 7 computes the percentages of the effects that are mediated by our variables. The percentage effects of education, when controlling for income, are about the same in both countries. But the effect of income in the Norwegian sample is much smaller and very close to zero.

Table 8 runs our regressions separately for males and females.[22] As in Britain, coefficients look very similar across the sexes. The effects of birth order are highly significant in either sex, and adding the mediators significantly reduces them. Interestingly, the effect of BMI is about 50% larger for women, as in Britain, and here the difference is significant. On the other hand, in Norway, there is no difference between the male and female coefficients on university attendance.

Overall, the Norway results show two things. First, they clearly confirm that birth order affects

[20] The R^2 on own university attendance is 0.0814 for EA4 in MoBa, compared to 0.04 for EA3 in UK Biobank.

[21] We also tried income at age 25. This gave odd results, with a significant negative beta on spouse PSEA. A possible reason is that in Norway many potential high earners are still in higher education at age 25. Income at age 25 correlated only at 0.1 with income at age 30, and was negatively correlated with educational attainment. Overall, we think income at 30 is more informative as a measure of SES.

[22] We don't run separate regressions for families with children, since MoBa only includes families with children by design.

spouse PSEA, with education a key mediator. Second, they suggest that the effects of SES in marriage markets vary between the two countries. Education has a similar effect on spouse PSEA in Norway, but income has a much smaller effect. Of course this is a loose comparison, since even conditioning on our controls, samples and measures in the two countries are different. Still, it is interesting that in Norway, a more egalitarian country than the UK, income seems to have a smaller effect on spouse PSEA, while own PSEA, a genetic characteristic, seems to have a larger effect. This suggests that genetic and SES contributions to attractiveness may vary between countries, and perhaps be endogenous to economic institutions.

Table 6: Regressions of spouse PSEA (Norway)

	(1)	(2)	(3)	(4)
Birth order	−0.0728 ***	−0.0454 ***	−0.0689 ***	−0.0454 ***
	(0.0034)	(0.0043)	(0.0041)	(0.0043)
University		0.3092 ***		0.3109 ***
		(0.0051)		(0.0058)
Income			0.0201 *	−0.0088
			(0.0047)	(0.0054)
Height		0.0034 ***	0.0062 ***	0.0034 ***
		(0.0001)	(0.0002)	(0.0001)
BMI		−0.0204 ***	−0.0239 ***	−0.0205 ***
		(0.0012)	(0.0011)	(0.0012)
Own PSEA	0.1166 ***	0.0696 ***	0.1071 ***	0.0698 ***
	(0.0078)	(0.0059)	(0.0075)	(0.0059)
Parents' age at birth	0.0140 ***	0.0091 ***	0.0133 ***	0.0091 ***
	(0.0004)	(0.0006)	(0.0005)	(0.0006)
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes
N	75055	75055	75055	75055
R^2	0.031	0.061	0.041	0.061

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 (11). PSEA is the polygenic score for educational attainment, which is normalized with mean 0 and standard deviation 1. We include own PSEA, mean of parents' ages at birth, potential non-SES mediators (height and BMI) and further controls (family size, birth year, and birth month dummies) to ensure the balance of covariates across birth order. All data is from the MoBa dataset for a sample of spouse pairs with a child between 1999 and 2008. *** p < 0.001; ** p < 0.01; * p < 0.05; + p < 0.1. Standard errors clustered by spouse pair in parentheses.

Table 7: Percent of birth order effects accounted for by mediators (Norway)

	Model 2 (%)	**Model 3 (%)**	**Model 4 (%)**
University	35.0		35.2
Income		0.7	−0.3
Height	2.4	4.4	2.4
BMI	0.3	0.3	0.3

Percentage of the effects of birth order in Table 6, columns 2 to 4, explained by by each mediating variable.

Table 8: Regressions of spouse PSEA: subsets (Norway)

	Male respondents	Male respondents	Female respondents	Female respondents
Birth order	−0.074 ***	−0.047 ***	−0.071 ***	−0.041 **
	(0.004)	(0.005)	(0.005)	(0.005)
University		0.320 ***		0.318 ***
		(0.012)		(0.006)
Income		−0.023 *		0.005
		(0.007)		(0.004)
Height		0.008 ***		0.006 ***
		(0.001)		(0.000)
BMI		−0.016 ***		−0.023 ***
		(0.001)		(0.001)
Own PSEA	0.121 ***	0.071 ***	0.111 ***	0.067 ***
	(0.009)	(0.005)	(0.007)	(0.006)
Parents' age at birth	0.013 ***	0.008 **	0.015 ***	0.010 **
	(0.001)	(0.001)	(0.002)	(0.002)
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes
N	37495	37495	37560	37560
R^2	0.032	0.065	0.032	0.062

Estimates from OLS regressions corresponding to columns 1 and 2 in Table 6, separately for males and females. 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 own PSEA, parents' age at birth (the mean of parent's ages) and further controls (family size, birth year, and birth month dummies) to ensure the balance of covariates across birth order. All data is from the MoBa dataset for a sample of spouse pairs with a child between 1999 and 2008. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors clustered by spouse pair in parentheses.

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 in the UK Biobank sample.[23] We regress each score on own birth order, controlling for family size. No scores were significant at $p < 0.10/33$. Four scores were significant at $p < 0.10$, all with effect sizes of less than 0.02 per standard deviation. Table 15 in the appendix reports 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 0.0277, $p = 0.06$). Table 18 in the appendix therefore reports 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.

Another concern is that our chosen SES mediators might not be exogenous. We have already seen that birth order affects intelligence, height, BMI and health. So there might be other unobserved variables which mediate the effect of birth order on spousal PSEA, and which correlate with education or income, but which do not themselves capture SES. If so, that would threaten our claim that

[23] 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). We also ran similar tests on MoBa data, and found no significant associations between birth order and polygenic scores.

education and income are important mediators. However, the effects of education on spouse PSEA in Table 3, and of birth order on education in Table 1, are both large and highly significant. In other literature on spouse matching, education is a common, robust and significant predictor. For these reasons, we think that our results are unlikely to be driven wholly by other, unobserved mediators. In the appendix we run Oster (2019) style robustness checks where we formally ask how strong selection on unobservables would need to be to reduce the effect of our key mediators to zero.

A final concern is that polygenic scores, including PSEA, contain noise from correlated environments. (That is, effect sizes of individual SNPs may be confounded in the underlying statistical analyses used to create polygenic scores.) It is conceivable that birth order could only affect the noise component of spouse PSEA, rather than the component which is truly causal for education. We think this is unlikely for several reasons. First, our polygenic scores were calculated using per-SNP summary statistics estimated on non-UK populations. So they will only include noise that correlates with social environment insofar as non-causal correlations of SNPs with social status are the same across different countries. Second, we have residualized PSEA on 100 principal components of the genetic data, a standard technique in genetics to avoid confounding causal effects with population stratification. Third, true causal effects of individual SNPs are highly correlated (r = 0.74) with their "population effects" including noise (A. Young et al. 2022). Lastly, it is hard to imagine an assortative mating process by which people match spouses who have genetic variants that correlate with educational attainment, but not genetic variants that cause it.

Our main specification is linear in birth order. Tables 12 and 13 in the appendix run specifications with separate dummies for each birth order. The pattern that birth order coefficients shrink after controlling for SES mediators holds robustly across all birth orders in both Britain and Norway.

UK Biobank is not a representative sample of the population. Table 14 in the appendix weights cases to match the Biobank's sampling frame. Results are similar to those in the main text. Although this is still not representative of the population as a whole, it provides some assurance that our results are not driven by volunteering bias.

As noted earlier, our results could conceivably be explained by spouses *only* mating on SES ($a = 0$ in our model), but with a pre-existing correlation between SES and PSEA in the population. That said, existing work strongly suggests that own PSEA affects spouse's education (Robinson et al. 2017; Torvik et al. 2022). To confirm this we use a sample of siblings from the MoBa data[24], and regress spouse's university attendance on own PSEA, including sibling group fixed effects. Again, this uses the lottery of meiosis to guarantee that between-sibling differences in PSEA are exogenous to environmental characteristics including SES. Table 19 in the appendix shows that spouses of higher-PSEA siblings were substantially more likely to have attended university, allowing us to

[24]The UK Biobank sample has too few siblings with spouses for this analysis to be informative.

rule out $a = 0$. Interestingly, they did not have higher income, which again suggests that income is not an important form of SES in Norway marriage markets.

The appendix reports other robustness checks, including replacing university attendance with age of leaving full-time education. Overall, while significance sometimes varies, the pattern of results is remarkably consistent. Birth order is always negatively associated with spouse PSEA, and this effect is always reduced in magnitude after adding education as a mediator. Effect sizes are also consistent, with the exception that they are smaller if we do not control for parental age (just as in Table 2).

Conclusion

Our empirical analysis shows that in Great Britain and Norway, two contemporary developed countries, earlier-born siblings 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 (SGAM).

Advantage is transmitted across generations by many mechanisms. Rich 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 2023). 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. As our model shows, genetic variation is not an exogenous input into the social system, but an endogenous outcome – not a confound for wealth, but a mediator.

SGAM also provides a new explanation for the genes-SES gradient – the observed association of genes with SES – which is an important cause of educational and occupational inequality, and perhaps also of health inequalities. The leading alternative explanation is meritocratic social mobility. Whilst meritocracy exists in modern capitalist economies, it has 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 genes-SES gradients should exist in all stratified societies. In fact, people in many societies have believed that innate traits do vary

by social status.[25] In future, it may be possible to directly test for genetic differences across social status in ancient DNA samples.

Under SGAM, the association between SES and genetic variation depends on economic and social institutions. Institutions that make wealth more persistent across generations also increase the correlation between SES and genetics. If so, then institutional differences may have long-run effects over generations by altering the genes-SES gradient. There could be hysteresis, with initial social differences cumulating over time via their effect on genetic inequality. On the other hand, while lowering the intergenerational transmission of wealth may eventually flatten the genes-SES gradient, increases in the level of meritocracy paradoxically make it steeper, suggesting a deep conflict between meritocracy and egalitarianism (M. Young 1958; Markovits 2019). Lastly, the structure of marriage markets also affects the gradient. Our empirical analysis suggests that in relatively egalitarian Norway, income plays a less important role in assortative mating than it does in the UK. However, this is a loose comparison, and we see careful tests of comparative statics across different contexts as an important challenge for future work.

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 reflects the fact that our individual genetic endowment is fixed at birth, affects our body and brain through proximate biological mechanisms, and cannot be changed by our social environment. 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 and bear children within social institutions. A person's genetic inheritance is a social and historical fact about them, not just a fact of nature. As Marx (1844) wrote, "History is the true natural history of man".

The theory of evolution suggests that two motivations are likely to be central for fitness-maximizing organisms: acquiring material resources so as to survive and raise offspring, and pursuing reproductive partners who themselves have high fitness value. Arguably, these two motives structure nearly all of human society. On this view, the genetics-SES trade in marriage markets is the most basic trade there is. Genetic endowments can be thought of 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.

[25] The appendix has a selection of relevant historical quotations.

Acknowledgements

The Norwegian Mother, Father and Child Cohort Study (MoBa) is a population-based pregnancy cohort study conducted by the Norwegian Institute of Public Health. Participants were recruited from all over Norway from 1999-2008. The women consented to participation in 41% of the pregnancies. The cohort includes approximately 114,500 children, 95,200 mothers and 75,200 fathers. The current study is based on version 12 of the quality-assured data files released for research in 2019. The establishment of MoBa and initial data collection was based on a license from the Norwegian Data Protection Agency and approval from The Regional Committees for Medical and Health Research Ethics. The MoBa cohort is currently regulated by the Norwegian Health Registry Act. The current study was approved by The Regional Committees for Medical and Health Research Ethics.

We thank the Norwegian Institute of Public Health (NIPH) for generating high-quality genomic data. This research is part of the HARVEST collaboration, supported by the Research Council of Norway (#229624). We also thank the NORMENT Centre for providing genotype data, funded by the Research Council of Norway (#223273), South East Norway Health Authorities and Stiftelsen Kristian Gerhard Jebsen. We further thank the Center for Diabetes Research, the University of Bergen for providing genotype data and performing quality control and imputation of the data funded by the ERC AdG project SELECTionPREDISPOSED, Stiftelsen Kristian Gerhard Jebsen, Trond Mohn Foundation, the Research Council of Norway, the Novo Nordisk Foundation, the University of Bergen, and the Western Norway Health Authorities.

This study was conducted using UK Biobank resources under application number 40310.

AA is supported by the Foundation Volksbond Rotterdam and by ZonMw grant 849200011 from The Netherlands Organisation for Health Research and Development.

Code to reproduce is available at https://github.com/hughjonesd/trading-genetics.

Appendix: for online publication

Contents

Proofs 43

More empirical results 57

Testing the UK Biobank spouse pair matching 72

Quotations on natural inequality 74

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:

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

where

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

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

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

Finally, we are interested in

$$\begin{pmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{a}{1-a} & 0 & \frac{\sigma_I}{1-a} \\ 0 & 1 & 0 \\ 0 & -\frac{a}{1-a} & \frac{\sigma_I}{1-a} \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ t \end{pmatrix}$$

therefore again the means are 0 and

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

where:

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

□

Lemma 1. *$C^2 \leq S^2$, with strict inequality if $a > 0$ and $\sigma < 1$.*

Proof. Write

$$\begin{aligned} C^2 &= \frac{a^2\sigma^2 + (1-a)^2S^4 + 2a(1-a)\sigma S^2}{a^2s^2 + (1-a)^2S^2 + 2a(1-a)\sigma} \\ &\leq \frac{a^2s^2S^2 + (1-a)^2S^4 + 2a(1-a)\sigma S^2}{a^2s^2 + (1-a)^2S^2 + 2a(1-a)\sigma}, \text{since } \sigma/sS = Corr(x_1, x_2) \leq 1 \\ &= S^2 \end{aligned}$$

and observe that the inequality is strict if $a > 0$ and $\sigma < sS$. □

Proof of Claim 1. Under random matching, 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

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

Using the matrix formula for the conditional mean of normal variables,

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

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

Proof of Claim 2. From (3), the joint distribution of $\left(\frac{\tau}{2}\left(x_1+y_1\right)+\varepsilon, x_2, y_2\right)$ is normal with mean $(0,0,0)$ and covariance

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

Therefore

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

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

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

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 random matching, the covariance 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:

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

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 (3), under SGAM, the correlation between children's traits is:

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

This is positive if $\sigma = 0$ so long as $AC > 0$ i.e. $0 < a < 1$. To show it is increasing in θ, strip out constant terms and take the derivative of

$$\frac{\theta}{\sqrt{\frac{1}{2}\theta^2(C^2+S^2)+1}}$$

The derivative is signed by

$$\begin{aligned}
&\left(\frac{1}{2}\theta^2(C^2+S^2)+1\right)^{0.5} - \frac{1}{2}\theta^2(C^2+S^2)\left(\frac{1}{2}\theta^2(C^2+S^2)+1\right)^{-0.5} \\
&> \left(\frac{1}{2}\theta^2(C^2+S^2)+1\right)^{0.5} - \left(\frac{1}{2}\theta^2(C^2+S^2)+1\right)\left(\frac{1}{2}\theta^2(C^2+S^2)+1\right)^{-0.5} \\
&= 0.
\end{aligned}$$

□

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:

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

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:

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

where

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

Note that if the distribution converges, s^2 and S^2 must be above 1. Also, for A and C to have a real-valued solution, it must be that $a^2s^2 + (1-a)^2S^2 + 2a(1-a)\sigma > 0$; using this,

$$AC = \frac{a^2s^2\sigma + (1-a)^2S^2\sigma + a(1-a)(\sigma^2 + S^2s^2)}{a^2s^2 + (1-a)^2S^2 + 2a(1-a)\sigma} > \sigma, \text{ by } \sigma^2 < S^2s^2.$$

Since $AC > \sigma$, $\frac{\theta\tau}{2}(\sigma + AC) > \theta\tau\sigma$. If $\sigma < 0$ then $\sigma = \frac{\theta\tau}{2}(\sigma + AC) > \theta\tau\sigma > \sigma$, a contradiction. Thus $\sigma \geq 0$. Also if $\sigma = 0$ then

$$\sigma = \frac{\theta\tau}{2}\left(\frac{a(1-a)(S^2s^2)}{a^2s^2 + (1-a)^2S^2}\right)$$

which implies $a = 0$ or $a = 1$. This proves that σ is non-negative, and positive if $a \in (0,1)$, so long as the distribution converges.

From the invariant distribution, first:

$$\sigma\left(1 - \frac{1}{2}\theta\tau\right) = \frac{\theta\tau}{2}\frac{\left(as^2 + (1-a)\,\sigma\right)\left(a\sigma + (1-a)\,S^2\right)}{a^2s^2 + (1-a)^2\,S^2 + 2a\,(1-a)\,\sigma}$$

or

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

where

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

Solving for λ gives

$$\lambda = \frac{1-a}{a}\mu\frac{(-2a + 4a\mu - 2\theta\tau + 2a\theta\tau - 3a\theta\tau\mu + 2)}{(1-a)\,\theta\tau - 2a\mu\,(1-\theta\tau)} \tag{13}$$

Then

$$s^2\left(1 - \frac{1}{2}\tau^2\right) = \frac{1}{2}A^2\tau^2 + 1$$
$$S^2\left(1 - \frac{1}{2}\theta^2\right) = \frac{1}{2}C^2\theta^2 + 1$$

give

$$s^2\left(1-\frac{1}{2}\tau^2\right)-\frac{1}{2}A^2\tau^2=S^2\left(1-\frac{1}{2}\theta^2\right)-\frac{1}{2}C^2\theta^2, \text{ therefore}$$
$$\lambda\left(1-\frac{1}{2}\tau^2\right)-\frac{1}{2}\frac{A^2}{S^2}\tau^2=\left(1-\frac{1}{2}\theta^2\right)-\frac{1}{2}\frac{C^2}{S^2}\theta^2$$

Here

$$\frac{A^2}{S^2}=\frac{\left(a\lambda+(1-a)\,\mu\right)^2}{D} \text{ and}$$
$$\frac{C^2}{S^2}=\frac{\left(a\mu+(1-a)\right)^2}{D},$$
$$D=a^2\lambda+(1-a)^2+2a\,(1-a)\,\mu$$

which give a quadratic equation in μ:

$$\lambda\left(1-\frac{1}{2}\tau^2\right)D-\frac{1}{2}\left(a\lambda+(1-a)\,\mu\right)^2\tau^2-\left(1-\frac{1}{2}\theta^2\right)D+\frac{1}{2}\left(a\mu+(1-a)\right)^2\theta^2=0$$

Plugging in λ given by (13), this can be rewritten to

$$F(\mu)=\frac{(1-a+a\mu)^2}{a}\frac{N(\mu)}{D(\mu)}=0$$

where

$$N(\mu)=X\mu^2+Y\mu+Z, \text{ with } X,Y,Z \text{ polynomials in } a,\theta,\tau$$

and

$$D(\mu)=\left(\theta\tau\,(1-a)-2a\mu\,(1-\theta\tau)\right)^2$$

One can check that the discriminant is always positive. Therefore this has two solutions (not shown), of which only one is acceptable (it goes to the exact solution when the coefficient of μ^2 goes to 0). Writing

$$\mu=\phi_1(a,\theta,\tau)$$

for this solution:

$$\lambda=\psi(a,\theta,\tau)=\frac{1-a}{a}\phi_1(a,\theta,\tau)\frac{\left(-2a+4a\phi_1(a,\theta,\tau)-2\theta\tau+2a\theta\tau-3a\theta\tau\phi_1(a,\theta,\tau)+2\right)}{(1-a)\,\theta\tau-2a\phi_1(a,\theta,\tau)\,(1-\theta\tau)}.$$

Finally

$$S^2 = \frac{1}{1 - \frac{1}{2}\theta^2 - \frac{1}{2}\frac{C^2}{S^2}\theta^2} \text{ where}$$
$$\frac{C^2}{S^2} = \frac{\left(a\phi_1(a,\theta,\tau) + (1-a)\right)^2}{a^2\psi(a,\theta,\tau) + (1-a)^2 + 2a(1-a)\phi_1(a,\theta,\tau)}$$

and

$$s^2 = \lambda S^2 = \frac{\psi(a,\theta,\tau)}{1 - \frac{1}{2}\theta^2 - \frac{1}{2}\frac{C^2}{S^2}\theta^2};$$
$$\sigma = \mu S^2 = \frac{\phi_1(a,\theta,\tau)}{1 - \frac{1}{2}\theta^2 - \frac{1}{2}\frac{C^2}{S^2}\theta^2}.$$

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

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

□

Proof of Claim 5. Since men and women have different distributions of attractiveness, we have to match them by quantiles of their respective distributions. Men's and women's attractiveness are distributed

$$N(0, \sigma_I^2) \text{ where } \sigma_I = \sqrt{a^2 s^2 + (1-a)^2 S^2 + 2a(1-a)\sigma};$$
$$N(0, \sigma_J^2) \text{ where } \sigma_J = \sqrt{b^2 s^2 + (1-b)^2 S^2 + 2b(1-b)\sigma}.$$

Thus, men with normalized attractiveness $i(x)/\sigma_I$ match women with normalized attractiveness $j(y)/\sigma_J$.

Change variables so that

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to \begin{pmatrix} x_1 \\ u \end{pmatrix} \text{ where } u = \frac{ax_1 + (1-a)x_2}{\sigma_I};$$
$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \to \begin{pmatrix} y_1 \\ v \end{pmatrix} \text{ where } v = \frac{by_1 + (1-b)y_2}{\sigma_J}.$$

Thus

$$\begin{pmatrix} x_1 \\ u \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ a/\sigma_I & (1-a)/\sigma_I \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix};$$
$$\begin{pmatrix} y_1 \\ v \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ b/\sigma_J & (1-b)/\sigma_J \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

and their respective covariance matrices are

$$\begin{aligned} \mathbb{C}\begin{pmatrix} x_1 \\ u \end{pmatrix} &= \begin{pmatrix} 1 & 0 \\ a/\sigma_I & (1-a)/\sigma_I \end{pmatrix} \begin{pmatrix} s^2 & \sigma \\ \sigma & S^2 \end{pmatrix} \begin{pmatrix} 1 & a/\sigma_I \\ 0 & (1-a)/\sigma_I \end{pmatrix} \\ &= \begin{pmatrix} s^2 & A \\ A & 1 \end{pmatrix}, \text{ where } A = \frac{as^2 + (1-a)\sigma}{\sigma_I}; \end{aligned}$$

similarly

$$\mathbb{C}\begin{pmatrix} y_1 \\ v \end{pmatrix} = \begin{pmatrix} s^2 & B \\ B & 1 \end{pmatrix}, \text{ where } B = \frac{bs^2 + (1-b)\sigma}{\sigma_J}.$$

Under SGAM, couples have characteristics $\begin{pmatrix} x_1 \\ t \\ y_1 \\ t \end{pmatrix}$, where $\begin{pmatrix} x_1 \\ y_1 \\ t \end{pmatrix}$ is trivariate normal with mean 0 and covariance matrix

$$\Sigma = \begin{pmatrix} s^2 & AB & A \\ AB & s^2 & B \\ A & B & 1 \end{pmatrix}$$

Lastly, we calculate the covariance matrix of couples' original characteristics. Since

$$\begin{pmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{-a}{1-a} & 0 & \frac{\sigma_I}{1-a} \\ 0 & 1 & 0 \\ 0 & \frac{-b}{1-b} & \frac{\sigma_J}{1-b} \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ t \end{pmatrix}$$

we have that the mean is again 0 and the covariance matrix is

$$\mathbb{C}\begin{pmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{-a}{1-a} & 0 & \frac{\sigma_I}{1-a} \\ 0 & 1 & 0 \\ 0 & \frac{-b}{1-b} & \frac{\sigma_J}{1-b} \end{pmatrix} \Sigma \begin{pmatrix} 1 & \frac{-a}{1-a} & 0 & 0 \\ 0 & 0 & 1 & \frac{-b}{1-b} \\ 0 & \frac{\sigma_I}{1-a} & 0 & \frac{\sigma_J}{1-b} \end{pmatrix}$$
$$= \begin{pmatrix} s^2 & \sigma & AB & AD \\ \sigma & S^2 & BC & CD \\ AB & BC & s^2 & \sigma \\ AD & CD & \sigma & S^2 \end{pmatrix}$$

where

$$C = \frac{a\sigma + (1-a)S^2}{\sigma_I}; D = \frac{b\sigma + (1-b)S^2}{\sigma_J}.$$

From the above and (2) we can calculate the covariance matrix of children's characteristics as

$$\mathbb{C}\begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} \frac{\tau}{2} & 0 & \frac{\tau}{2} & 0 \\ 0 & \frac{\theta}{2} & 0 & \frac{\theta}{2} \end{pmatrix} \begin{pmatrix} s^2 & \sigma & AB & AD \\ \sigma & S^2 & BC & CD \\ AB & BC & s^2 & \sigma \\ AD & CD & \sigma & S^2 \end{pmatrix} \begin{pmatrix} \frac{\tau}{2} & 0 \\ 0 & \frac{\theta}{2} \\ \frac{\tau}{2} & 0 \\ 0 & \frac{\theta}{2} \end{pmatrix}$$
$$= \begin{pmatrix} \frac{\tau^2}{2}(s^2 + AB) + 1 & \frac{\tau\theta}{4}(2\sigma + AD + BC) \\ \frac{\tau\theta}{4}(2\sigma + AD + BC) & \frac{\theta^2}{2}(S^2 + CD) + 1 \end{pmatrix}.$$

Thus x_1' and x_2' will be positively correlated if $2\sigma + AD + BC > 0$. This is always positive if $\sigma > 0$; if $\sigma = 0$ it reduces to

$$\frac{(a + b - 2ab)s^2S^2}{\sigma_I\sigma_J}$$

which is positive unless $a = b = 0$ or $a = b = 1$. The correlation is

$$\frac{\frac{\tau\theta}{4}(2\sigma + AD + BC)}{\sqrt{\frac{\tau^2}{2}(s^2 + AB) + 1}\sqrt{\frac{\theta^2}{2}(S^2 + CD) + 1}}$$

and taking the derivative shows it is increasing in θ, as in the proof for Claim 4. □

Proof of Proposition 4. Write

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

Since

$$\begin{pmatrix} \tau\frac{x_1+y_1}{2} \\ \gamma\tau\frac{x_1+y_1}{2}+\theta\frac{x_2+y_2}{2} \end{pmatrix} = \begin{pmatrix} \frac{\tau}{2} & 0 & \frac{\tau}{2} & 0 \\ \frac{\gamma\tau}{2} & \frac{\theta}{2} & \frac{\gamma\tau}{2} & \frac{\theta}{2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{pmatrix}$$

we can use (1) to derive the covariance matrix for children:

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

The first claim in the proof follows from the covariance:

$$\frac{\gamma\tau^2}{2}(s^2+A^2)+\frac{\tau\theta}{2}(\sigma+AC)+\gamma$$

This is increasing in γ, and positive if any of $\sigma>0, \gamma>0$, or $AC>0$ (which holds if $0<a<1$ when $\sigma=0$).

The correlation $Cov(x_1', x_2')/\sqrt{Var(x_1')Var(x_2')}$ is proportional to

$$\frac{\gamma p+q}{\sqrt{\gamma^2 p+\gamma 2q+r}}$$

where

$$p = \tau^2(s^2 + A^2) + 2;$$
$$q = \tau\theta(\sigma + AC);$$
$$r = \theta^2(S^2 + C^2) + 2.$$

The derivative of this with respect to γ is signed by $pr - q^2$, which equals

$$\begin{aligned}&[\tau^2(s^2 + A^2) + 2][\theta^2(S^2 + C^2) + 2] - [\tau\theta(\sigma + AC)]^2\\ =&\tau^2\theta^2(s^2S^2 + A^2S^2 + s^2C^2 - 2\sigma AC - \sigma^2) + 2[\theta^2(S^2 + C^2) + \tau^2(s^2 + A^2)] + 4\end{aligned}$$

The last two terms are positive. In the first term, $\sigma^2 < s^2S^2$, and

$$\begin{aligned}0 &< (AS - Cs)^2\\ &= A^2S^2 + C^2s^2 - 2ACSs\\ &< A^2S^2 + C^2s^2 - 2\sigma AC \text{ , again using } \sigma < sS.\end{aligned}$$

Hence the whole sum is positive.

Now we can calculate

$$\mathbb{E}[x_2'|x_2 + y_2]$$

using

$$\begin{pmatrix} \gamma\tau\frac{x_1+y_1}{2} + \theta\frac{x_2+y_2}{2} \\ x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} \frac{\gamma\tau}{2}0 & \frac{\theta}{2}1 & \frac{\gamma\tau}{2}0 & \frac{\theta}{2}0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{pmatrix}$$

to give

$$\begin{aligned}\mathbb{C}\begin{pmatrix} x_2' \\ x_2 \\ y_2 \end{pmatrix} &= \begin{pmatrix} \frac{\gamma\tau}{2} & \frac{\theta}{2} & \frac{\gamma\tau}{2} & \frac{\theta}{2} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} s^2 & \sigma & A^2 & AC \\ \sigma & S^2 & AC & C^2 \\ A^2 & AC & s^2 & \sigma \\ AC & C^2 & \sigma & S^2 \end{pmatrix} \begin{pmatrix} \frac{\gamma\tau}{2} & 0 & 0 \\ \frac{\theta}{2} & 1 & 0 \\ \frac{\gamma\tau}{2} & 0 & 0 \\ \frac{\theta}{2} & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1+\gamma^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\\ &= \begin{pmatrix} \ldots & \frac{\gamma\tau}{2}(\sigma + AC) + \frac{\theta}{2}(S^2 + C^2) & \frac{\gamma\tau}{2}(\sigma + AC) + \frac{\theta}{2}(S^2 + C^2) \\ \frac{\gamma\tau}{2}(\sigma + AC) + \frac{\theta}{2}(S^2 + C^2) & S^2 & C^2 \\ \frac{\gamma\tau}{2}(\sigma + AC) + \frac{\theta}{2}(S^2 + C^2) & C^2 & S^2 \end{pmatrix}.\end{aligned}$$

Next

$$\mathbb{E}[x_2'|x_2,y_2] = \begin{pmatrix} \frac{\gamma\tau}{2}(\sigma+AC)+\frac{\theta}{2}(S^2+C^2) & \frac{\gamma\tau}{2}(\sigma+AC)+\frac{\theta}{2}(S^2+C^2) \end{pmatrix} \begin{pmatrix} S^2 & C^2 \\ C^2 & S^2 \end{pmatrix}^{-1} \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$$
$$= \left(\gamma\tau\frac{\sigma+AC}{S^2+C^2}+\theta\right)\frac{x_2+y_2}{2}$$

So long as $0 < a < 1$ or $\sigma > 0$, the coefficient on parents' wealth is thus higher than θ.

□

Proof of Proposition 5. Without loss of generality let $Ex_1 = Ex_2 = 0$. The correlation is signed by the covariance. Write K for the set of couples in the parents' generation with typical member $k = (x, y)$. Without loss of generality let $x_1 \geq y_1$. Then, since the iso-attractiveness curves defined by f are downward-sloping, $x_2 \leq y_2$. (If $a = 1$ then $x_1 = y_1$; pick x so that $x_2 \leq y_2$.) Also, for $a \in (0, 1)$, if $x_1 > y_1$ then $x_2 < y_2$.

Since $Ex_1 = Ex_2 = 0$, the covariance among the parents' generation is

$$\int_K (x_1x_2 + y_1y_2)/2 \, dk$$

Write

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

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

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

Write

$$Cov(x_1^*, x_2^*) = \int_K x_1^* x_2^* \, dk$$

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

$$Cov(x_1, x_2) \leq Cov(x_1^*, x_2^*) = \int_K x_1^* x_2^*\, dk$$

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

□

More empirical results

Figure 5 redoes Figure 4 for the MoBa data, plotting measures of individual SES against spouse PSEA. While university attendance gives results similar to the UK, the relationship between income decile and spouse PSEA is interestingly nonlinear in both sexes. Some low-income individuals may be out of the labour market, either in continuing education or as a stay-at-home spouse.

Table 9 shows regressions of birth order on mediators for the MoBa data. Birth order predicts all the mediators significantly and with large effect sizes, except for BMI.

Table 10 regresses birth order on spouse PSEA for the MoBa data, starting with a simple bivariate regression and adding controls as in Table 2. Here, the negative effect of birth order is highly significant even before we control for parental age. As in Britain, the effect is greatly increased when we control for parental age.

Table 11 reruns our central regressions in the UK, 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.

Tables 12 and 13 rerun our central regressions estimating a separate coefficient for each position in the birth order (with firstborn as the baseline). The basic pattern of our main result is remarkably robust, in both Great Britain and Norway: birth order coefficients are generally negative, and adding mediators always causes them to increase towards zero or to change sign.

We also ran a specification with separate birth order dummies within each family size. Figure 6 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 14 re-estimates Table 3 using weights from Alten et al. (2022). These weight the UK Biobank sample to match its sampling frame of 40-69 year olds living close to 22 assessment centres. Although these weights probably bring the sample closer to the UK population, the sampling frame is still not representative of that population: for instance, urban areas are oversampled. Results are similar to those in the main text, although birth order coefficients are absolutely larger in all the specifications.

Table 15 reruns our regressions controlling for several polygenic scores. Results are very close to those in the main text.

Table 16 and 17 rerun our regressions using age of leaving full-time education as a measure of edu-

cational 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 substantially.

Table 18 reruns Table 3 excluding families of size 3. Results are very similar to those in the main text.

Table 19 regresses spouse university attendance and income z-score on own PSEA on a subset of sibling within the MoBa dataset. This tests the null hypothesis $a = 0$ in our model. If $a = 0$, own polygenic score has no effect on spouse matching. To ensure that PSEA is exogenously varied, we again rely on the lottery of meiosis by including sibling fixed effects. Analogously to our main results in MoBa, own PSEA significantly increases the probability of spouse university attendance, by about 5% per standard deviation; but its effect on spouse income is insignificant and tightly bounded around zero. Overall, these results confirm that PSEA affects the set of potential spouses.

Lastly, we run an Oster (2019) style robustness analysis on the effect of education and income on spouse PSEA. This is a formalization of the informal idea that if a coefficient does not change much when a known set of controls is added, it may be robust to other, unobserved controls. The method works by comparing a "short" regression, without alternative controls like BMI and height, to a "medium" regression with those controls. An important input is the maximum R^2 of independent variables on the dependent variable. Here, we use our knowledge about the noise in measured PSEA. We take the ratio of the R^2 of measured PSEA on education to the maximum R^2 of all genetic variables on education, a.k.a. the heritability. We assume that this ratio s gives the proportion of "signal" in measured PSEA, with the rest being noise (from sampling error in the construction of the polygenic score). We then make an assumption about the maximum R^2 of any own variable on spouse's true PSEA - i.e., about how random the spouse matching process is. We take this to be 50%. Multiplying s by 50% gives the maximum possible R^2 of own variables on spouse's measured PSEA. The output of the analysis is a value δ^* representing the degree of selection on unobservables relative to observables (with respect to the treatment variable) that would be necessary to eliminate the effect of the independent variable. A δ^* of about 1 is considered a reasonable threshold for robustness, if we assume that measured control variables ought to be at least as important as unmeasured ones.

In UK Biobank, we calculate the maximum R^2 as 0.056.[26] For university attendance, δ^* is 0.941, and for income in first job δ^* is 2.03. In MoBa, we calculate the maximum R^2 as 0.131.[27] For university attendance, δ^* is 0.99, and for income at age 30, δ^* is 0.275. These results confirm that the results on university attendance are relatively robust, except for results on income in MoBa.

[26] In other regressions of spouse traits on PSEA, R^2 are indeed below this value (Hugh-Jones et al. 2016; Okbay et al. 2022).

[27] The MoBa number is larger because the R^2 of EA4 on education is higher than for EA3.

Of course, this technique depends crucially on the input assumptions, especially about noise in measured PSEA, the true heritability, and the maximum possible R^2 in spouse matching.

Table 9: Regressions of mediators on birth order (Norway)

	University	Income	Height	BMI
Birth order	−0.0825 ***	−0.0241 **	−0.5132 **	0.0103
	(0.0071)	(0.0046)	(0.0849)	(0.0498)
Own PSEA	0.1313 ***	0.0571 ***	0.3678 **	−0.2544 ***
	(0.0039)	(0.0029)	(0.0440)	(0.0280)
Parents' age at birth	0.0145 ***	0.0037 ***	0.0720 ***	−0.0080
	(0.0004)	(0.0003)	(0.0081)	(0.0041)
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes
N	75055	75055	75055	75055
R2	0.118	0.027	0.025	0.006

Estimates from OLS regressions with the mediators (university attendance, income, height, BMI) 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 and further controls to ensure the balance of covariates across birth order. All data is from the MoBa dataset for a sample of spouse pairs with a child between 1999 and 2008. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors clustered by spouse pair in parentheses.

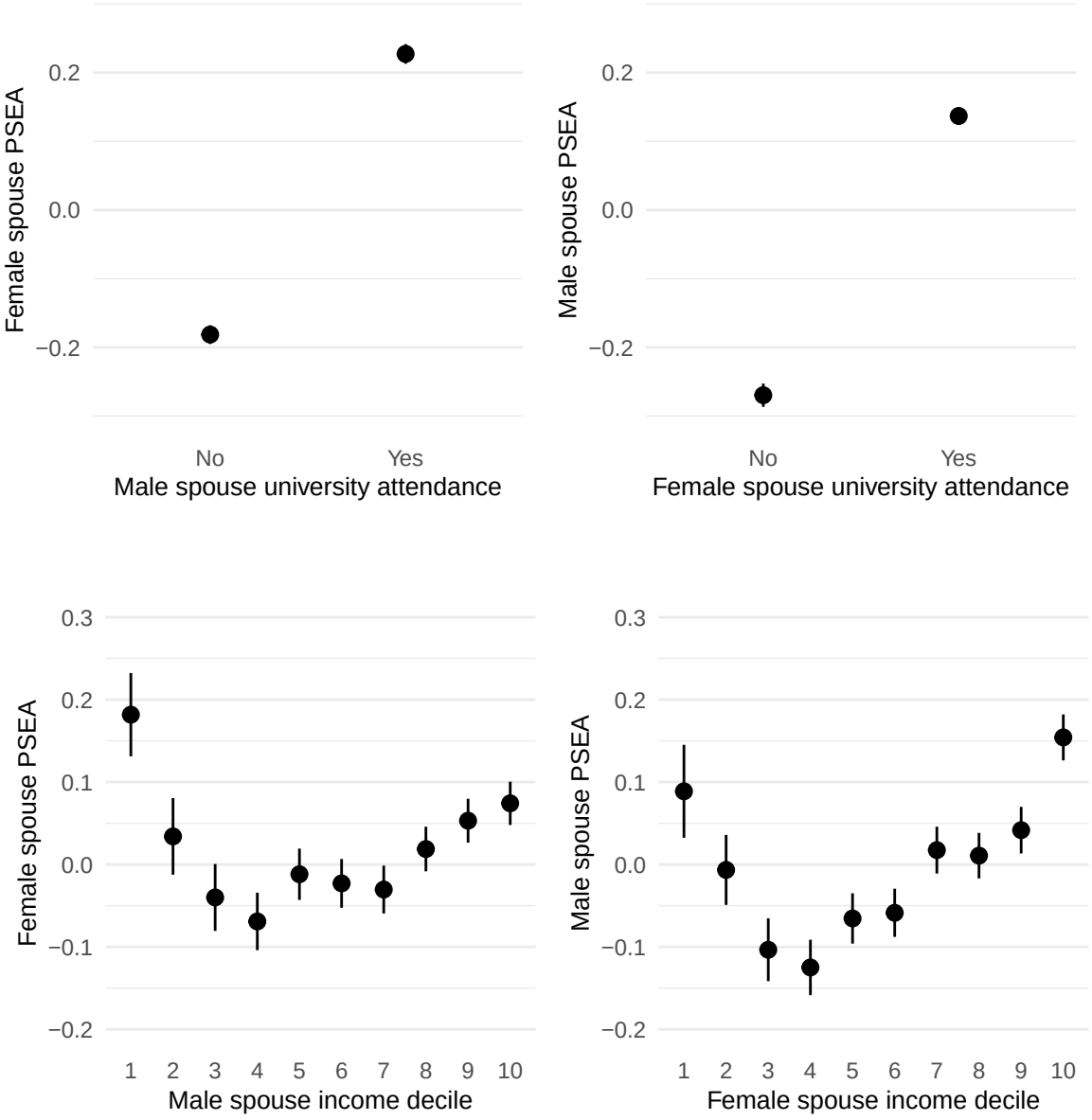


Figure 5: Spouse polygenic score for educational attainment (PSEA) against own university attendance and own income at age 30 (Norway). Lines show 95% confidence intervals. PSEA is normalized to have mean 0 and variance 1.

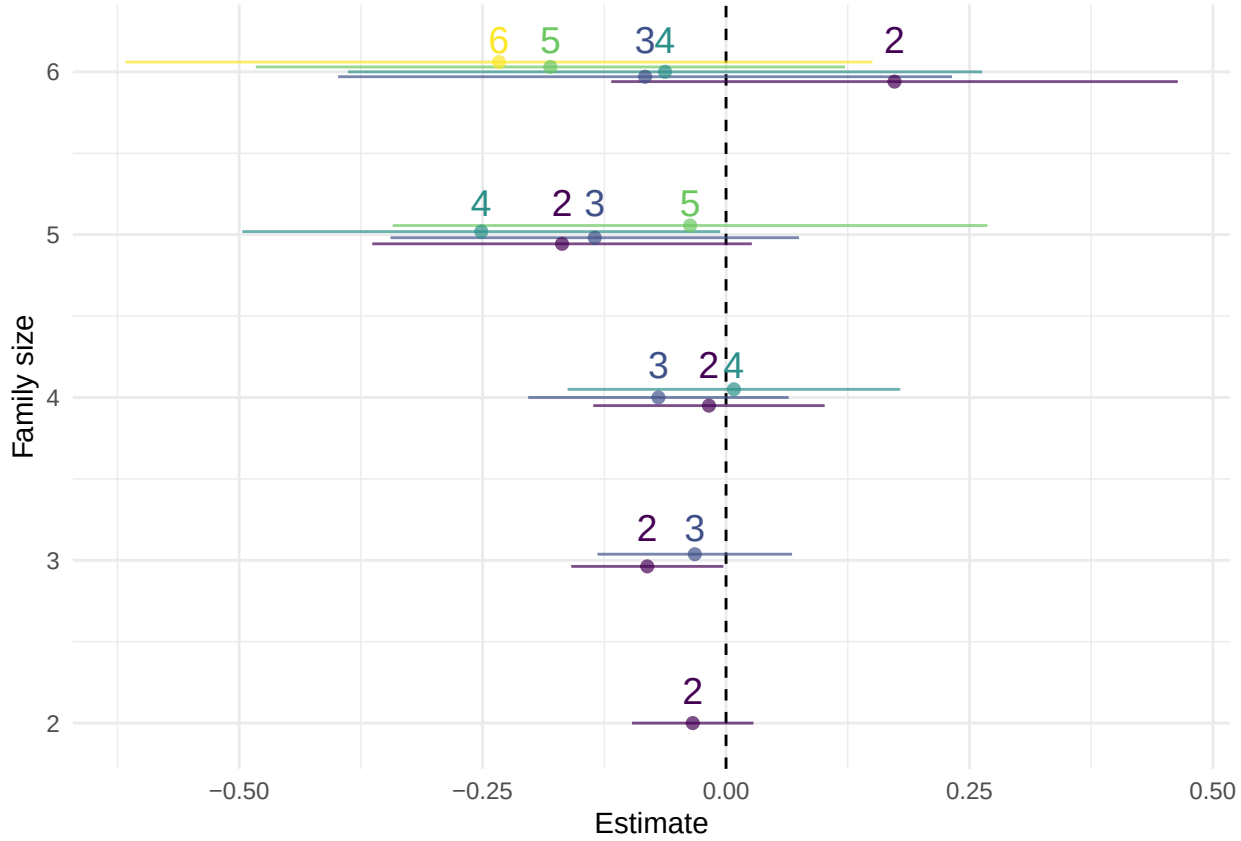


Figure 6: Regressions of spouse PSEA: birth order dummies within different family sizes (Great Britain). Labels show birth order. Lines are 95 per cent confidence intervals. The omitted category is birth order 1.

Table 10: Regressions of spouse PSEA on birth order (Norway)

	(1)	(2)	(3)
Birth order	−0.0136 *	−0.0189 ***	−0.0728 ***
	(0.0037)	(0.0037)	(0.0062)
Own PSEA		0.1248 ***	0.1166 ***
		(0.0030)	(0.0032)
Parents' age at birth			0.0140 ***
			(0.0011)
Family size dummies	Yes	Yes	Yes
Birth month dummies	No	Yes	Yes
Birth year dummies	No	Yes	Yes
N	75055	75055	75055
R2	0.000	0.028	0.031

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 own PSEA, parents' age at birth (the mean of parents' ages), and further controls (family size, birth year, and birth month dummies) in columns 2−3 to ensure the balance of covariates across birth order. All data is from the MoBa dataset for a sample of spouse pairs with a child between 1999 and 2008. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors clustered by spouse pair in parentheses.

Table 11: Regressions of spouse PSEA, without controls for parents' age at respondent's birth (Great Britain)

	(1)	(2)	(3)	(4)
Birth order	−0.0075	0.0027	−0.0020	0.0021
	(0.0074)	(0.0074)	(0.0137)	(0.0136)
University		0.2386 ***		0.1913 ***
		(0.0149)		(0.0251)
Income			0.0609 ***	0.0407 **
			(0.0121)	(0.0124)
Fluid IQ		0.0156 ***	0.0159 **	0.0051
		(0.0034)	(0.0060)	(0.0062)
Height		0.0019 **	0.0047 ***	0.0039 **
		(0.0007)	(0.0012)	(0.0012)
BMI		−0.0115 ***	−0.0146 ***	−0.0137 ***
		(0.0015)	(0.0027)	(0.0027)
Self-reported health		0.0184 +	0.0129	0.0047
		(0.0097)	(0.0181)	(0.0180)
Own PSEA	0.0650 ***	0.0316 ***	0.0378 **	0.0266 +
	(0.0086)	(0.0087)	(0.0142)	(0.0142)
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes
N	23797	23797	7658	7658
R2	0.010	0.031	0.022	0.030
logLik	−33426.686	−33179.643	−10732.036	−10702.746
AIC	66953.372	66469.287	21572.072	21515.492

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors clustered by spouse pair in parentheses.

Table 12: Regressions of spouse PSEA, separate birth order dummies (Great Britain)

	(1)	(2)	(3)	(4)
Birth order 2	−0.0500 *	−0.0204	−0.0460	−0.0441
	(0.0232)	(0.0231)	(0.0411)	(0.0410)
Birth order 3	−0.0557	0.0013	−0.0133	0.0045
	(0.0376)	(0.0374)	(0.0673)	(0.0673)
Birth order 4	−0.0736	0.0099	−0.0117	0.0026
	(0.0655)	(0.0651)	(0.1270)	(0.1268)
Birth order 5	−0.0801	0.0022	0.1032	0.1195
	(0.1190)	(0.1181)	(0.2293)	(0.2289)
Birth order 6	−0.2746	−0.1997	0.1713	0.2125
	(0.2371)	(0.2349)	(0.5965)	(0.5953)
University		0.2182 ***		0.1513 ***
		(0.0221)		(0.0378)
Income			0.0622 ***	0.0487 **
			(0.0167)	(0.0170)
Own PSEA	0.0574 ***	0.0263 **	0.0213	0.0117
	(0.0099)	(0.0101)	(0.0180)	(0.0181)
Parents' age at birth	0.0116 ***	0.0052 *	0.0090 +	0.0078
	(0.0026)	(0.0026)	(0.0047)	(0.0047)
Wald p-value, birth order	0.2452	0.7917	0.8266	0.7815
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes
Other mediators (IQ, height, BMI, s.-r. health)	No	Yes	Yes	Yes
N	10206	10206	3407	3407
R2	0.013	0.032	0.030	0.035
logLik	−14296.681	−14196.868	−4809.676	−4801.539
AIC	28703.362	28513.735	9739.351	9725.078

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors clustered by spouse pair in parentheses. Grey background: coefficients are higher than column 1.

Table 13: Regressions of spouse PSEA, separate birth order dummies (Norway)

	(1)	(2)	(3)	(4)
Birth order 2	−0.0645 **	−0.0378 *	−0.0635 **	−0.0379 *
	(0.0128)	(0.0112)	(0.0122)	(0.0113)
Birth order 3	−0.1482 ***	−0.0886 ***	−0.1395 ***	−0.0889 ***
	(0.0102)	(0.0096)	(0.0094)	(0.0096)
Birth order 4	−0.2317 ***	−0.1559 ***	−0.2201 ***	−0.1559 ***
	(0.0107)	(0.0117)	(0.0115)	(0.0117)
Birth order 5	−0.2879 **	−0.1853 **	−0.2694 **	−0.1852 **
	(0.0375)	(0.0398)	(0.0387)	(0.0399)
Birth order 6	−0.3288 ***	−0.1906 ***	−0.2917 ***	−0.1902 ***
	(0.0143)	(0.0170)	(0.0153)	(0.0171)
University		0.3093 ***		0.3110 ***
		(0.0051)		(0.0059)
Income			0.0201 *	−0.0088
			(0.0047)	(0.0054)
Height		0.0034 ***	0.0062 ***	0.0034 ***
		(0.0001)	(0.0002)	(0.0001)
BMI		−0.0204 ***	−0.0239 ***	−0.0205 ***
		(0.0012)	(0.0012)	(0.0012)
Own PSEA	0.1165 ***	0.0695 ***	0.1071 ***	0.0698 ***
	(0.0077)	(0.0058)	(0.0074)	(0.0059)
Parents' age at birth	0.0141 ***	0.0092 ***	0.0134 ***	0.0092 ***
	(0.000)	(0.001)	(0.000)	(0.001)
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes
N	75055	75055	75055	75055
R2	0.031	0.061	0.041	0.061

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors clustered by spouse pair in parentheses.
Grey background: coefficients are higher than column 1.

Table 14: Regressions of spouse PSEA, weighted to match UK Biobank sampling frame (Great Britain)

	(1)	(2)	(3)	(4)
Birth order	−0.0382 *	−0.0176	−0.0262	−0.0240
	(0.0190)	(0.0190)	(0.0347)	(0.0347)
University		0.1932 ***		0.0865 +
		(0.0278)		(0.0471)
Income			0.0538 *	0.0463 *
			(0.0215)	(0.0213)
Fluid IQ		0.0169 *	0.0135	0.0082
		(0.0069)	(0.0117)	(0.0122)
Height		0.0038 **	0.0051 *	0.0051 *
		(0.0014)	(0.0023)	(0.0023)
BMI		−0.0110 ***	−0.0100 *	−0.0099 *
		(0.0028)	(0.0049)	(0.0049)
Self-reported health		0.0169	0.0144	0.0095
		(0.0201)	(0.0322)	(0.0321)
Own PSEA	0.0558 ***	0.0242	0.0141	0.0086
	(0.0157)	(0.0161)	(0.0261)	(0.0263)
Parents' age at birth	0.0112 ***	0.0064 +	0.0085	0.0080
	(0.0034)	(0.0034)	(0.0058)	(0.0059)
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes
N	10005	10005	3356	3356
R2	0.018	0.035	0.040	0.042
logLik	−14040.496	−13942.942	−4761.416	−4754.938
AIC	28180.992	27995.884	9632.832	9621.875

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors clustered by spouse pair in parentheses.

Table 15: Regressions of spouse PSEA with controls for polygenic scores (Great Britain)

	(1)	(2)	(3)	(4)
Birth order	−0.0313	−0.0047	−0.0100	−0.0042
	(0.0183)	(0.0177)	(0.0315)	(0.0310)
University		0.2178 ***		0.1493 ***
		(0.0245)		(0.0230)
Income			0.0620 ***	0.0488 **
			(0.0134)	(0.0130)
Fluid IQ		0.0168 *	0.0191	0.0107
		(0.0066)	(0.0116)	(0.0120)
Height		0.0029 *	0.0058 **	0.0052 **
		(0.0011)	(0.0018)	(0.0018)
BMI		−0.0109 ***	−0.0108 *	−0.0104 *
		(0.0023)	(0.0040)	(0.0039)
Self-reported health		0.0173	0.0140	0.0076
		(0.0199)	(0.0340)	(0.0339)
Own PSEA	0.0519 ***	0.0231 +	0.0172	0.0083
	(0.0111)	(0.0115)	(0.0247)	(0.0244)
Parents' age at birth	0.0114 ***	0.0052 +	0.0091 *	0.0079 +
	(0.0028)	(0.0028)	(0.0041)	(0.0041)
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	10206	10206	3407	3407
R2	0.013	0.032	0.030	0.035

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors clustered by spouse pair in parentheses.
Polygenic scores: alzheimer's, cognitive ability, neuroticism, substance use.

Table 16: Regressions of spouse PSEA using age of leaving full-time education (Great Britain)

	(1)	(2)	(3)
Birth order	−0.0314 *	0.0022	0.0041
	(0.0145)	(0.0146)	(0.0270)
Age left full-time educ.		0.0475 ***	0.0394 ***
		(0.0044)	(0.0078)
Income			0.0439 *
			(0.0182)
Fluid IQ		0.0144 **	0.0076
		(0.0053)	(0.0097)
Height		0.0029 **	0.0051 **
		(0.0011)	(0.0019)
BMI		−0.0105 ***	−0.0103 **
		(0.0022)	(0.0040)
Self-reported health		0.0148	0.0094
		(0.0152)	(0.0272)
Own PSEA	0.0573 ***	0.0252 *	0.0128
	(0.0120)	(0.0121)	(0.0203)
Parents' age at birth	0.0116 ***	0.0041	0.0063
	(0.0026)	(0.0026)	(0.0047)
Family size dummies	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes
N	10206	10156	3400
R2	0.013	0.035	0.037
logLik	−14297.465	−14116.670	−4789.163
AIC	28694.930	28343.341	9690.326

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors clustered by spouse pair in parentheses.

Table 17: Regressions of spouse PSEA using age of leaving full-time education (Norway)

	(1)	(2)	(3)
Birth order	−0.0728 ***	−0.0382 ***	−0.0382 ***
	(0.0034)	(0.0041)	(0.0042)
Age left fulltime educ.		0.0692 ***	0.0703 ***
		(0.0007)	(0.0008)
Income			−0.0225 *
			(0.0056)
Height		0.0034 ***	0.0034 ***
		(0.0001)	(0.0001)
BMI		−0.0192 ***	−0.0193 ***
		(0.0012)	(0.0012)
Own PSEA	0.1166 ***	0.0570 ***	0.0574 ***
	(0.0078)	(0.0061)	(0.0060)
Parents' age at birth	0.0140 ***	0.0078 ***	0.0078 ***
	(0.000)	(0.001)	(0.001)
Family size dummies	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes
N	75055	75055	75055
R2	0.031	0.067	0.067
logLik	−105282.932	−103844.384	−103835.829
AIC	210699.865	207828.768	207813.657

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors clustered by spouse pair in parentheses.

Table 18: Regressions of spouse PSEA, excluding family size 3 (Great Britain)

	(1)	(2)	(3)	(4)
Birth order	−0.0360 *	−0.0126	−0.0226	−0.0203
	(0.0171)	(0.0170)	(0.0330)	(0.0330)
University		0.2056 ***		0.1615 ***
		(0.0272)		(0.0466)
Income			0.0300	0.0163
			(0.0240)	(0.0242)
Fluid IQ		0.0198 **	0.0112	0.0030
		(0.0062)	(0.0112)	(0.0116)
Height		0.0025 *	0.0049 *	0.0041 +
		(0.0013)	(0.0023)	(0.0023)
BMI		−0.0124 ***	−0.0128 **	−0.0125 *
		(0.0026)	(0.0049)	(0.0049)
Self-reported health		0.0121	0.0071	−0.0012
		(0.0182)	(0.0337)	(0.0337)
Own PSEA	0.0524 ***	0.0210	0.0123	−0.0000
	(0.0136)	(0.0137)	(0.0245)	(0.0247)
Parents' age at birth	0.0125 ***	0.0065 *	0.0075	0.0064
	(0.0032)	(0.0032)	(0.0056)	(0.0056)
Family size dummies	Yes	Yes	Yes	Yes
Birth month dummies	Yes	Yes	Yes	Yes
Birth year dummies	Yes	Yes	Yes	Yes
N	6959	6959	2286	2286
R2	0.016	0.034	0.033	0.039
logLik	−9723.678	−9656.955	−3227.648	−3221.403
AIC	19545.356	19421.909	6561.296	6550.807

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors clustered by spouse pair in parentheses.

Table 19: Within-siblings regressions of spouse university attendance and income on own PSEA (Norway)

	University	Income
Own PSEA	0.0477 ***	−0.0007
	(0.0098)	(0.0174)
Sibling group dummies	Yes	Yes
Birth month dummies	Yes	Yes
Birth year dummies	Yes	Yes
N	9729	9755
Adj. R^2	0.182	0.083

Estimates from within-sibling-group regressions. PSEA is the polygenic score for educational attainment, which is normalized with mean 0 and standard deviation 1. Sibling group dummies are included to ensure exogeneity of PSEA. *** p < 0.001; ** p < 0.01; * p < 0.05; + p < 0.1. Standard errors in parentheses.

Testing the UK Biobank spouse pair matching

Some of our spouse pairs in UK Biobank could be false positives, i.e. people who are not each others' spouse but simply live in the same postcode. To validate the accuracy of our pairs, we use genetic relationships. Some respondents in the UK Biobank sample have a child (inferred from genetic data) who is also in the sample. Among our spouse pairs, 463 have a genetic child of at least one partner in the sample. For 425 of these, at least one child is the genetic child of both partners. If this subsample is representative, then about 92% of the pairs who have had a child, have had a child together. This is a lower bound estimate, because some of the remaining couples may have had a genetically shared child who is not in the UK Biobank sample. 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. These pairs might 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 786 genetic children of the fake pairs, only 32 were children of both parents. Thus, the vast majority of fake pairs do not appear to be spouses. Table 20 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 20: Regressions of PSEA on birth order: fake pairs (Great Britain)

	(1)	(2)	(3)
Birth order	−0.0074	−0.0061	−0.0273 +
	(0.0080)	(0.0080)	(0.0144)
Own PSEA		0.0510 ***	0.0514 ***
		(0.0089)	(0.0116)
Parents' age at birth			0.0096 ***
			(0.0025)
Family size dummies	Yes	Yes	Yes
Birth month dummies	No	Yes	Yes
Birth year dummies	No	Yes	Yes
N	21550	21508	10400
R2	0.001	0.007	0.011

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$. Standard errors clustered by spouse pair in parentheses.

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*

Abilities come from innate talents, which differ in their capacities and take on different responsibilities in government.

– Liu Shao *Study of human abilities*

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.

Alten, Sjoerd van, Benjamin W Domingue, Titus J Galama, and Andries T Marees. 2022. "Reweighting the UK Biobank to Reflect Its Underlying Sampling Population Substantially Reduces Pervasive Selection Bias Due to Volunteering." *medRxiv*. https://doi.org/10.1101/2022.05.16.22275048.

Amin, Vikesh, Jere R Behrman, and Hans-Peter Kohler. 2015. "Schooling Has Smaller or Insignificant Effects on Adult Health in the US Than Suggested by Cross-Sectional Associations: New Estimates Using Relatively Large Samples of Identical Twins." *Social Science & Medicine* 127: 181–89.

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.

Barban, Nicola, Elisabetta De Cao, Sonia Oreffice, and Climent Quintana-Domeque. 2021. "The Effect of Education on Spousal Education: A Genetic Approach." *Labour Economics* 71: 102023.

Barone, Guglielmo, and Sauro Mocetti. 2021. "Intergenerational Mobility in the Very Long Run: Florence 1427-2011." *Review of Economic Studies* 88 (4): 1863–91.

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

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 Gudhnason, et al. 2011. "The Promises and Pitfalls of Genoeconomics."

Benjamin, Daniel J, David Cesarini, Patrick Turley, and Alexander Strudwick Young. 2024. "Social-Science Genomics: Progress, Challenges, and Future Directions."

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

———. 2011. "Older and Wiser? Birth Order and IQ of Young Men." *CESifo Economic Studies* 57 (1): 103–20.

Boardman, Jason D, Benjamin W Domingue, and Jonathan Daw. 2015. "What Can Genes Tell Us about the Relationship Between Education and Health?" *Social Science & Medicine* 127: 171–80.

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.

Bycroft, Clare, Colin Freeman, Desislava Petkova, Gavin Band, Lloyd T Elliott, Kevin Sharp, Allan Motyer, et al. 2018. "The UK Biobank Resource with Deep Phenotyping and Genomic Data." *Nature* 562 (7726): 203–9.

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

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.

Chiappori, Pierre-André, Edoardo Ciscato, and Carla Guerriero. 2021. "Analyzing Matching Pat-

terns in Marriage: Theory and Application to Italian Data."

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.

Chiappori, Pierre-André, Bernard Salanié, and Yoram Weiss. 2017. "Partner Choice, Investment in Children, and the Marital College Premium." *American Economic Review* 107 (8): 2109–67.

Clark, Gregory. 2023. "The Inheritance of Social Status: England, 1600 to 2022." *Proceedings of the National Academy of Sciences* 120 (27): e2300926120.

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

Cloninger, C Robert, John Rice, and Theodore Reich. 1979. "Multifactorial Inheritance with Cultural Transmission and Assortative Mating. II. A General Model of Combined Polygenic and Cultural Inheritance." *American Journal of Human Genetics* 31 (2): 176.

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

Collado, M Dolores, Ignacio Ortuño-Ortín, and Jan Stuhler. 2023. "Estimating Intergenerational and Assortative Processes in Extended Family Data." *The Review of Economic Studies* 90 (3): 1195–1227.

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.

De Haan, Monique. 2010. "Birth Order, Family Size and Educational Attainment." *Economics of Education Review* 29 (4): 576–88.

Demange, Perline A, Margherita Malanchini, Travis T Mallard, Pietro Biroli, Simon R Cox, Andrew D Grotzinger, Elliot M Tucker-Drob, et al. 2021. "Investigating the Genetic Architecture of Noncognitive Skills Using GWAS-by-Subtraction." *Nature Genetics* 53 (1): 35–44.

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

Eaves, Lindon, Andrew C Heath, Nicholas Martin, Hermine Maes, Michael Neale, Kenneth Kendler, Katherine Kirk, and Linda Corey. 1999. "Comparing the Biological and Cultural Inheritance of Personality and Social Attitudes in the Virginia 30 000 Study of Twins and Their Relatives." *Twin Research and Human Genetics* 2 (2): 62–80.

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

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.

Hanushek, Eric A. 1992. "The Trade-Off Between Child Quantity and Quality." *Journal of Political Economy* 100 (1): 84–117.

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

Hauser, Robert M, and William H Sewell. 1985. "Birth Order and Educational Attainment in Full Sibships." *American Educational Research Journal* 22 (1): 1–23.

Heath, Andrew 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.

Heath, Andrew C, and Lindon J Eaves. 1985. "Resolving the Effects of Phenotype and Social Background on Mate Selection." *Behavior Genetics* 15 (1): 15–30.

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, Michel G Nivard, Tim T Morris, Ailin F Hansen, Humaira Rasheed, Yoonsu Cho, Geetha Chittoor, et al. 2022. "Within-Sibship Genome-Wide Association Analyses Decrease Bias in Estimates of Direct Genetic Effects." *Nature Genetics* 54 (5): 581–92.

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*.

Kantarevic, Jasmin, and Stéphane Mechoulan. 2006. "Birth Order, Educational Attainment, and Earnings an Investigation Using the PSID." *Journal of Human Resources* 41 (4): 755–77.

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.

Lawlor, Debbie A, Roger M Harbord, Jonathan AC Sterne, Nic Timpson, and George Davey Smith. 2008. "Mendelian Randomization: Using Genes as Instruments for Making Causal Inferences in Epidemiology." *Statistics in Medicine* 27 (8): 1133–63.

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.

Magnus, Per, Charlotte Birke, Kristine Vejrup, Anita Haugan, Elin Alsaker, Anne Kjersti Daltveit, Marte Handal, et al. 2016. "Cohort Profile Update: The Norwegian Mother and Child Cohort Study (MoBa)." *International Journal of Epidemiology* 45 (2): 382–88.

Marees, Andries T, Dirk JA Smit, Abdel Abdellaoui, Michel G Nivard, Wim Van Den Brink, Damiaan Denys, Titus J Galama, Karin JH Verweij, and Eske M Derks. 2021. "Genetic Correlates of Socio-Economic Status Influence the Pattern of Shared Heritability Across Mental Health Traits." *Nature Human Behaviour* 5 (8): 1065–73.

Markovits, Daniel. 2019. *The Meritocracy Trap*. Penguin UK.

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

Mendel, Gregor. 1865. "Experiments in Plant Hybridization."

Muslimova, Dilnoza, Hans Van Kippersluis, Cornelius A Rietveld, Stephanie Von Hinke, and Fleur Meddens. 2020. "Dynamic Complementarity in Skill Production: Evidence from Genetic Endowments and Birth Order."

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.

Nivard, Michel G, Daniel W Belsky, K Paige Harden, Tina Baier, Ole A Andreassen, Eivind Ystrøm, Elsje van Bergen, and Torkild H Lyngstad. 2024. "More Than Nature and Nurture, Indirect Genetic Effects on Children's Academic Achievement Are Consequences of Dynastic Social Processes." *Nature Human Behaviour*, 1–8.

Okbay, A, Y Wu, N Wang, H Jayashankar, M Bennett, SM Nehzati, J Sidorenko, et al. 2022. "Polygenic Prediction of Educational Attainment Within and Between Families from Genome-Wide Association Analyses in 3 Million Individuals." *Nature Genetics*.

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

Oster, Emily. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence."

Journal of Business & Economic Statistics 37 (2): 187–204.

Otto, Sarah P, Freddy B Christiansen, and Marcus W Feldman. 1995. "Genetic and Cultural Inheritance of Continuous Traits." *Morrison Institute for Population and Resource Studies Paper Number* 64.

Paltiel, Liv, Haugan Anita, Trine Skjerden, Kari Harbak, Siri Bækken, Stensrud Nina Kristin, Gun Peggy Knudsen, and Per Magnus. 2014. "The Biobank of the Norwegian Mother and Child Cohort Study–Present Status." *Norsk Epidemiologi* 24 (1-2).

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.

Rao, DC, NE Morton, and CR Cloninger. 1979. "Path Analysis Under Generalized Assortative Mating: I. Theory." *Genetics Research* 33 (2): 175–88.

Rao, DC, Newton E Morton, and S Yee. 1976. "Resolution of Cultural and Biological Inheritance by Path Analysis." *American Journal of Human Genetics* 28 (3): 228.

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).

Rustichini, Aldo, William G Iacono, James J Lee, and Matt McGue. 2023. "Educational Attainment and Intergenerational Mobility: A Polygenic Score Analysis." *Journal of Political Economy* 131 (10): 2724–79.

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.

Sanjak, Jaleal S, Julia Sidorenko, Matthew R Robinson, Kevin R Thornton, and Peter M Visscher. 2018. "Evidence of Directional and Stabilizing Selection in Contemporary Humans."

Proceedings of the National Academy of Sciences 115 (1): 151–56.

Schmalhausen, Ivan Ivanovich. 1949. "Factors of Evolution: The Theory of Stabilizing Selection."

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.

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.

Torvik, Fartein Ask, Espen Moen Eilertsen, Laurie J Hannigan, Rosa Cheesman, Laurence J Howe, Per Magnus, Ted Reichborn-Kjennerud, et al. 2022. "Modeling Assortative Mating and Genetic Similarities Between Partners, Siblings, and in-Laws." *Nature Communications* 13 (1): 1108.

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.

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

Young, Alexander Strudwick. 2023. "Estimation of Indirect Genetic Effects and Heritability Under Assortative Mating." *bioRxiv*, 2023–07.

Young, Alexander, Seyed Moeen Nehzati, Stefania Benonisdottir, Aysu Okbay, Hariharan Jayashankar, Chanwook Lee, David Cesarini, Daniel J Benjamin, Patrick Turley, and Augustine Kong. 2022. "Mendelian Imputation of Parental Genotypes Improves Estimates of Direct Genetic Effects." *Nature Genetics* 54 (6): 897–905.

Young, M. 1958. *The Rise of the Meritocracy*. Thames; Hudson.