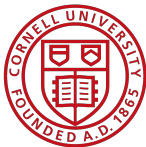


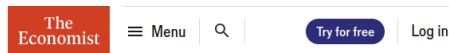
The Direct and Indirect Effects of Genetics and Education

Senan Hogan-Hennessy
Economics Department, Cornell University
seh325@cornell.edu



Cornell, Labor Work in Progress Seminar
10 April 2026

Introduction



Leaders | Modification revolution

Science has made a new genetic revolution possible

Now let it flourish



Nature vs Nurture

1. Genetics measure nature (initial endowment)
2. Find genes associated with outcomes
3. Deeper understanding of social outcomes, from genetics.

Introduction

Nature(?) vs Nurture

1. Genetics measure one component of nature (at most. . .)
2. Genes associated with outcomes are not the entire story
3. Genetics gives new findings, with caveats.

INTERDISCIPLINARY RESEARCH

Economics and genetics meet in uneasy union

Use of population-genetic data to predict economic success sparks war of words.

BY EWEN CALLAWAY

The invalid assumption that correlation implies cause is probably among the two or three most serious and common errors of human reasoning." Evolutionary biologist Stephen Jay Gould was referring to purported links between genetics and an individual's intelligence when he made this familiar complaint in his 1981 book *The Mismeasure of Man*.

Fast-forward three decades, and leading geneticists and anthropologists are levelling a similar charge at economics researchers who claim that a country's genetic diversity can predict the success of its economy. To critics, the economists' paper seems to suggest that a country's poverty could be the result of its citizens' genetic make-up, and the paper is attracting charges of genetic determinism, and even racism. But the economists say that they have been misunderstood, and are merely using

genetics as a proxy for other factors that can drive an economy, such as history and culture. The debate holds cautionary lessons for a nascent field that blends genetics with economics, sometimes called *genoconomics*. The work could have real-world pay-offs, such as helping policy-makers "reduce barriers to the flows of ideas and innovations across populations", says Enrico Spolaore, an economist at Tufts University near Boston, Massachusetts, who has also used global genetic-diversity data in his research. But the economists at the forefront of this field clearly need to be prepared for harsh scrutiny of their techniques and conclusions. At the centre of the storm is a 107-page paper by Oded Galor of Brown University in Providence, Rhode Island, and Quamrul Ashraf of Williams College in Williamstown, Massachusetts. It has been peer-reviewed by economists and biologists, and will soon appear in *American Economic Review*, one of the most prestigious economics journals.

The paper argues that there are strong links between estimates of genetic diversity for 145 countries and per capita income, and



Introduction

Genes associated with education causally improve labour market outcomes, but the **mechanism** is unclear

This paper:

1. Estimate the causal effect of the Education PGI on education and income using quasi-random genetic variation from parents in UK Biobank (replicating Carvalho 2025)
 2. Decompose the labour market genetic effect into an **education channel** + **direct genetic effect** using *design-based* causal mediation framework
 3. Sensitivity analysis for different values for returns to an extra education year, transparently showing identification uncertainty.
-

Preview: Education PGI causal effects transmit via **education channel**
→ **Direct genetic effect** closes to zero at plausible education returns values.

Introduction

How do genes associated with education affect labour outcomes?

Introduction

1. Genetics & Education

Data: UK Biobank

2. Genetic Effects

Mendelian Independent Assortment

3. Causal Mediation Analysis

Sensitivity analysis

Conclusion

Introduction — Research Ethics

Genetic differences are real, but they are not fate.

- ▶ A higher Ed PGI reflects a *statistical tendency*, not a destiny — social policies and institutions can modify or eliminate the consequences of genetic variation
- ▶ These results provide no support for “natural” social hierarchies or genetic justifications for inequality
- ▶ Genetic associations estimated within one ancestry group **cannot** be meaningfully compared across groups; any such inference would be both invalid and ethically deficient

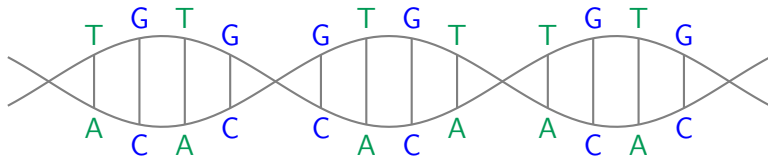
The appropriate interpretation:

- ▶ Both social *and* genetic factors shape socioeconomic outcomes
- ▶ This should inspire support for policies promoting **equality of opportunity** — not acceptance of inequality as inevitable or “natural.”

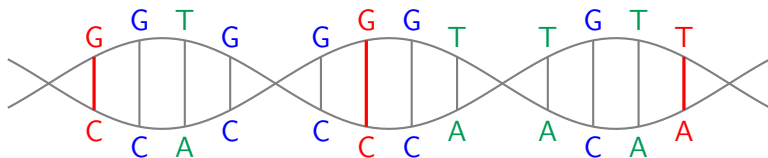
Genetics: a Primer

- ▶ The human genome is a 3 billion-long sequence of base pairs
- ▶ $\approx 0.1\%$ these pairs differ between two average people

Person 1's DNA:



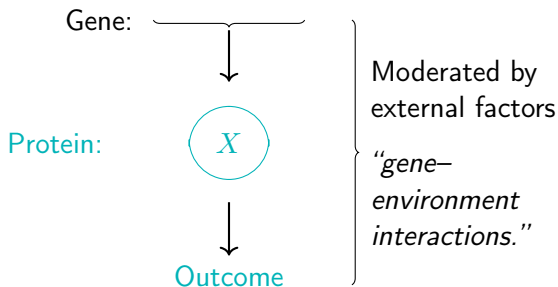
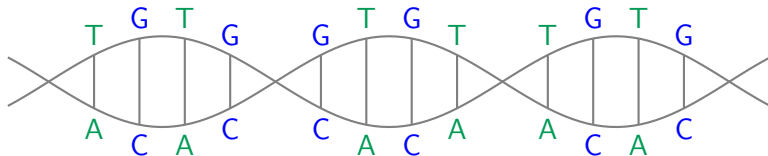
Person 2's DNA:



Genetics: a Primer

- ▶ The human genome is a 3 billion-long sequence of base pairs
- ▶ $\approx 0.1\%$ these pairs differ between two average people

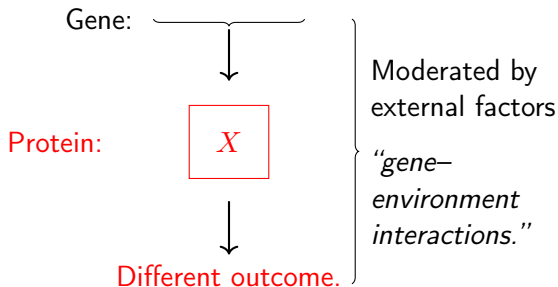
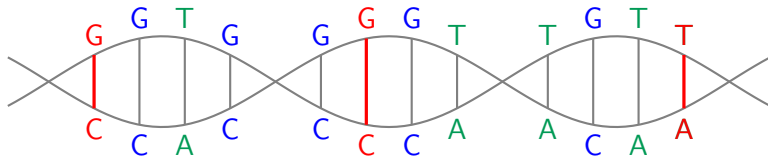
Person 1's DNA:



Genetics: a Primer

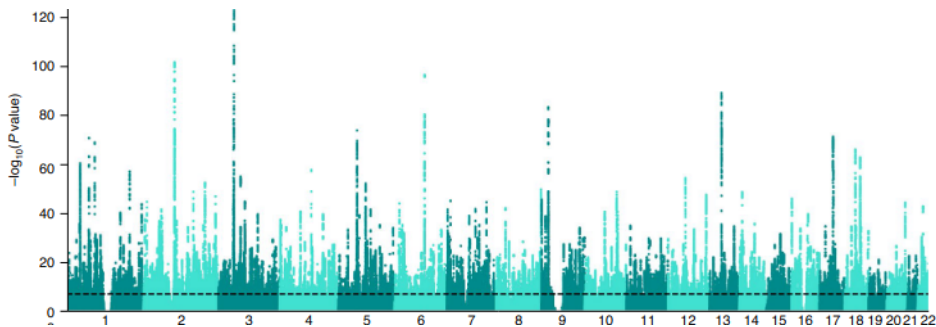
- ▶ The human genome is a 3 billion-long sequence of base pairs
- ▶ $\approx 0.1\%$ these **pairs differ** between two average people

Person 2's DNA:



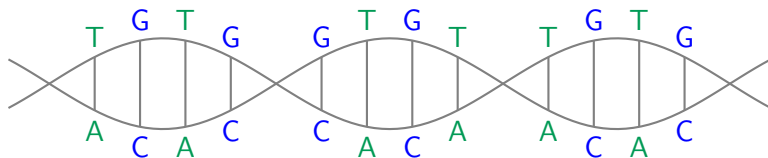
Education PolyGenic Index (Ed PGI)

“We conduct a genome-wide association study (GWAS) of educational attainment (EA) in a sample of 3 million individuals and **identify 3,952 approximately uncorrelated genome-wide-significant single-nucleotide polymorphisms (SNPs)**. A genome-wide [...] polygenic index (PGI), explains 12–16% of EA variance and contributes to risk prediction for ten diseases.”
Abstract — Okbay et al. (2022, Nature Genetics)

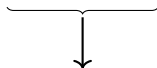


Education PolyGenic Index (Ed PGI)

Person 1's DNA:



3,952 Different Genes:



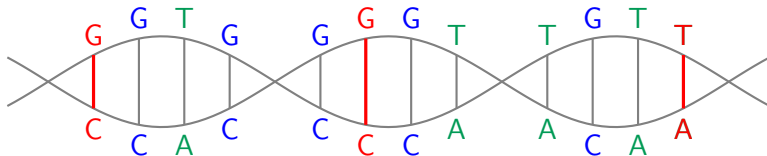
3,952 Different Proteins:



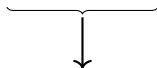
Education Years

Education PolyGenic Index (Ed PGI)

Person 2's DNA:



3,952 Different Genes:



3,952 Different Proteins:

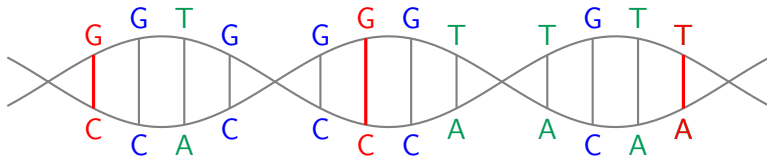


Education Years

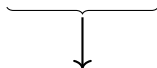
- ▶ In practice, we do not know exact location of relevant gene regions
- ▶ Nor the associated protein + how it works.

Education PolyGenic Index (Ed PGI)

Person 2's DNA:



3,952 Different Genes:



3,952 Different Proteins:

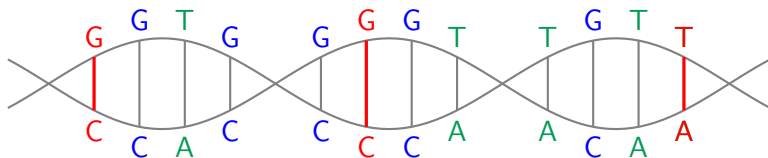


Education Years

- ▶ In practice, we do not know exact location of relevant gene regions
- ▶ Nor the associated protein + how it works.

Education PolyGenic Index (Ed PGI)

Person 2's DNA:



↓
Education Years

Calculate Polygenic Index (PGI):

1. Identify mutation regions (3,952 for Ed PGI)
2. Calculate weights for each
3. Assign a 1-dim genetic measure, summarising + dimension-reducing this topic.

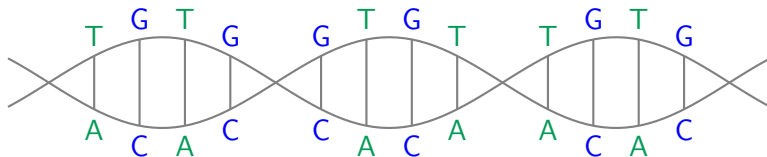
- ▶ In practice, we do not know exact location of relevant gene regions
- ▶ Nor the associated protein + how it works.

Education PolyGenic Index (Ed PGI)

Education Polygenic Index (Ed PGI):

A weighted sum of single letter gene changes predicting years of education.

Person 1's DNA:



Weights: w_1 0 0 0 0 w_2 0 0 0 0 0 w_3

Person 1's Ed PGI:

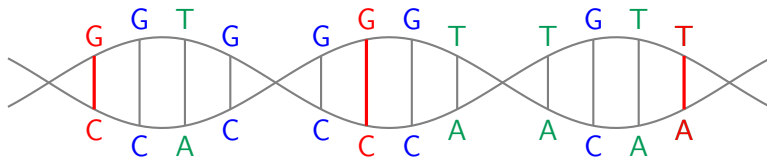
$$Z_i = \sum_{j=1}^J w_j \mathbb{1} \{ \text{person 1 has one-letter change at } j \} = 0$$

Education PolyGenic Index (Ed PGI)

Education Polygenic Index (Ed PGI):

A weighted sum of single letter gene changes predicting years of education.

Person 2's DNA:



Weights: w_1 0 0 0 0 w_2 0 0 0 0 0 w_3

Person 2's Ed PGI:

$$Z_i = \sum_{j=1}^J w_j \mathbb{1} \{ \text{person 2 has one-letter change at } j \} = w_1 + w_2 + w_3$$

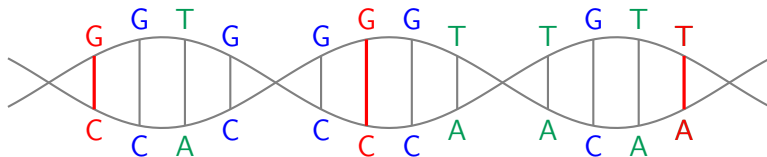
Note: Humans have two copies of each chromosome, so indicator is 0, 1, 2 corresponding to how many copies one has.

Education PolyGenic Index (Ed PGI)

Education Polygenic Index (Ed PGI):

A weighted sum of single letter gene changes predicting years of education.

Person 2's DNA:



Weights: w_1 0 0 0 0 w_2 0 0 0 0 0 w_3

Person 2's Ed PGI:

$$Z_i = \sum_{j=1}^J w_j \mathbb{1} \{ \text{person 2 has one-letter change at } j \} = w_1 + w_2 + w_3$$

Note: Humans have two copies of each chromosome, so indicator is 0, 1, 2 corresponding to how many copies one has.

Education PolyGenic Index (Ed PGI)

Previous economics + science research shows Ed PGI is strongly associated with multiple economics-relevant outcomes:

Retirement wealth	(Barth et al., 2020)
Life-cycle income	(Barth et al., 2022)
Longevity	(Marioni et al., 2016)
Primary school performance	(Aagaard-Houmark et al., 2022)
Parental-child care, nurture	(Aagaard-Houmark et al., 2025)
	(Muslimova et al. 2025)
...	
Labour market outcomes (causal)	(Carvalho, 2025)

Open question: What are the mechanisms behind these relationships?

- ▶ Is **education the only way** it is related to outcomes?
- ▶ Is there a **direct genetic relationship**, absent returns to education?

Data: UK Biobank

This is **the** dataset that genetics researchers have used for genetics research.

The first “Biobank”

- ▶ Tissue samples (e.g., DNA) from 500,000 Britons
- ▶ Large-scale population study started in 2004, to investigate biological + genetic basis of low-frequency diseases
- ▶ Includes surveys for social outcomes, including education, occupation, etc.



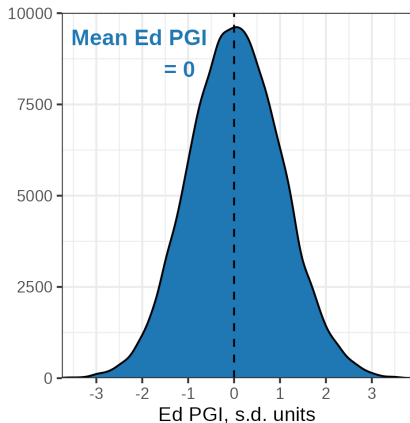
UK Biobank has $N \approx 25,000$ subsample of people with genetic data from siblings included.¹

¹Limited to race = white, for whom Ed PGI is defined.

Ed PGI and Education

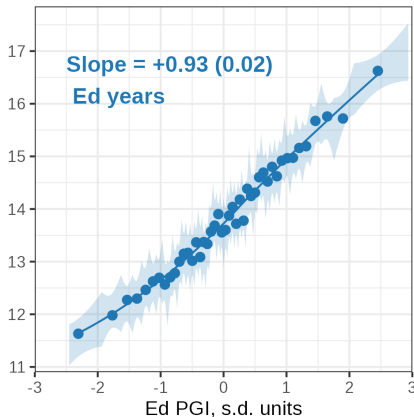
(a) Ed PGI.

Observations



(b) Ed PGI + Education Years, Binscatter.

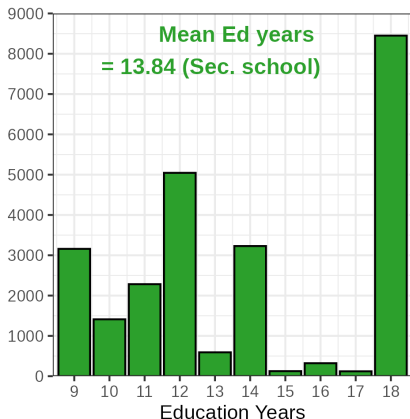
Education Years



Education and Labour Market Outcomes

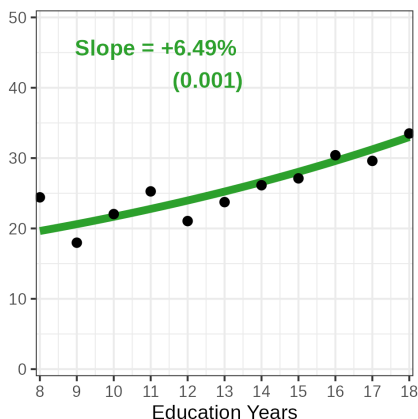
(a) Education Years.

Observations, N



(b) Education Years + Income.

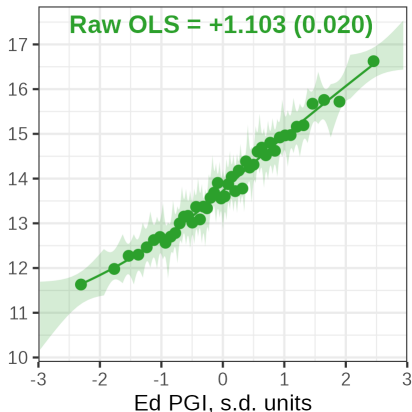
Annual Income, £ thousands



Associations with the Ed PGI

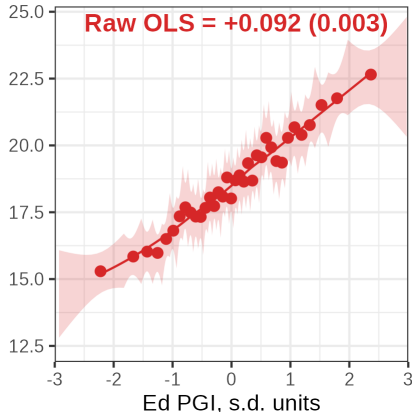
(a) Ed PGI $\rightarrow$ Education Years.

Education Years



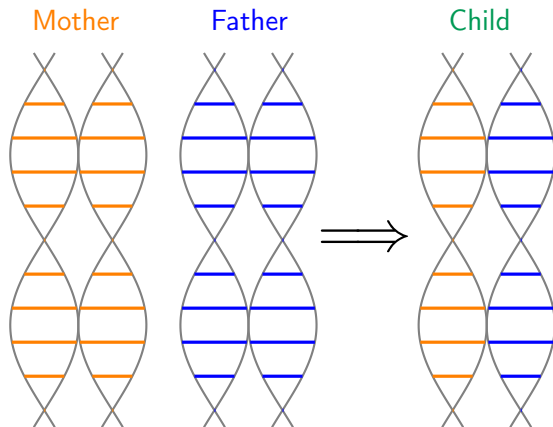
(b) Ed PGI $\rightarrow$ Hourly Wages.

Occupation Hourly Wages, £



Mendelian Independent Assortment

Question: Is **Ed PGI** randomly assigned?



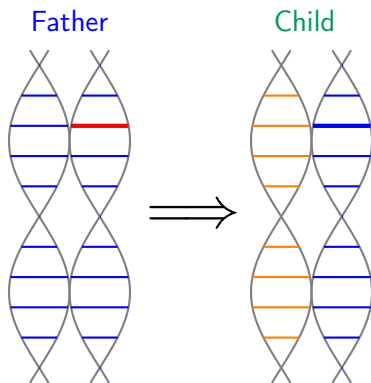
Answer:

No, inherited from parents.

$$\begin{aligned}\mathbb{E} \left[Z_i^{\text{Child}} \mid Z_{m(i)}^{\text{Mother}}, Z_{f(i)}^{\text{Father}} \right] \\ = \frac{1}{2} \left(Z_{m(i)}^{\text{Mother}} + Z_{f(i)}^{\text{Father}} \right)\end{aligned}$$

Mendelian Independent Assortment

Difference in Ed PGI from parents' arises from random inheritance + mixing, **Mendelian Independent Assortment**.



$$\Pr(\text{Child inherits} \mid \text{Father one copy}) = \frac{1}{2}$$

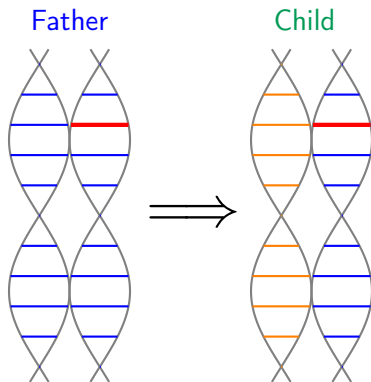
I.e., inheritance probability depends only on parents' genetics.

Propagate this reasoning across all 3,952 single-letter changes in the Ed PGI, and both parents:

$$Z_i^{\text{Child}} - \mathbb{E} \left[Z_i^{\text{Child}} \mid Z_{f(i)}^{\text{Father}}, Z_{m(i)}^{\text{Mother}} \right] \longrightarrow \text{Independently assigned.}$$

Mendelian Independent Assortment

Difference in Ed PGI from parents' arises from random inheritance + mixing, **Mendelian Independent Assortment**.



$$\Pr(\text{Child inherits} \mid \text{Father one copy}) = \frac{1}{2}$$

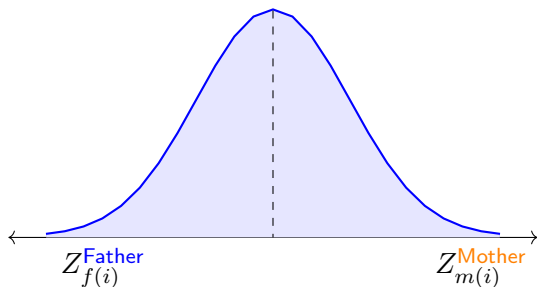
I.e., inheritance probability depends only on parents' genetics.

Propagate this reasoning across all **3,952 single-letter gene changes** in the Ed PGI, and both parents:

$$Z_i^{\text{Child}} - \mathbb{E} \left[Z_i^{\text{Child}} \mid Z_{f(i)}^{\text{Father}}, Z_{m(i)}^{\text{Mother}} \right] \rightarrow \text{Independently assigned.}$$

Ed PGI Causal Effects: Identification.

Difference in Ed PGI from parents' arises from random inheritance + mixing, **Mendelian Independent Assortment**.

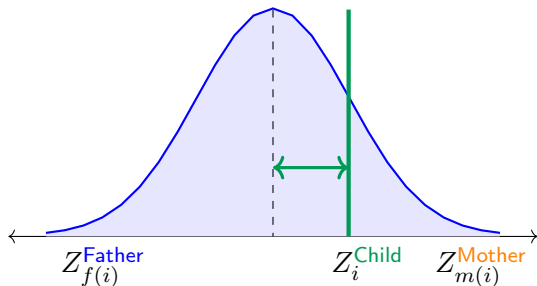


Propagate this reasoning across all **3,952 single-letter changes** in the Ed PGI, and both parents:

$$Z_i^{\text{Child}} - \mathbb{E} \left[Z_i^{\text{Child}} \mid Z_{f(i)}^{\text{Father}}, Z_{m(i)}^{\text{Mother}} \right] \longrightarrow \text{Independently assigned.}$$

Ed PGI Causal Effects: Identification

Difference in Ed PGI from parents' arises from random inheritance + mixing, **Mendelian Independent Assortment**.

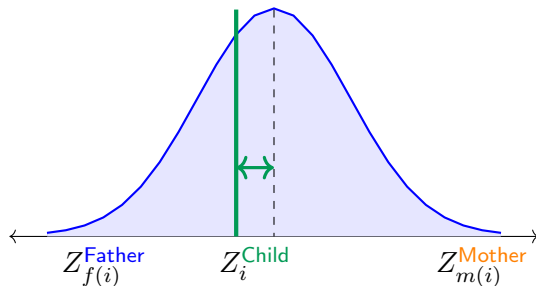


Propagate this reasoning across all **3,952 single-letter changes** in the Ed PGI, and both parents:

$$Z_i^{\text{Child}} - \mathbb{E} \left[Z_i^{\text{Child}} \mid Z_{f(i)}^{\text{Father}}, Z_{m(i)}^{\text{Mother}} \right] \longrightarrow \text{Independently assigned.}$$

Ed PGI Causal Effects: Identification

Difference in Ed PGI from parents' arises from random inheritance + mixing, **Mendelian Independent Assortment**.

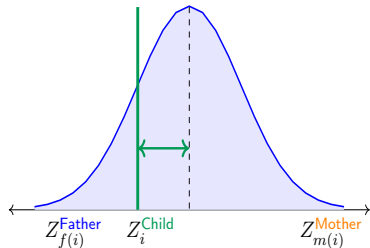


Propagate this reasoning across all **3,952 single-letter changes** in the Ed PGI, and both parents:

$$Z_i^{\text{Child}} - \mathbb{E} \left[Z_i^{\text{Child}} \mid Z_{f(i)}^{\text{Father}}, Z_{m(i)}^{\text{Mother}} \right] \longrightarrow \text{Independently assigned.}$$

Ed PGI Causal Effects: Identification

Difference in Ed PGI from parents' arises from random inheritance + mixing, **Mendelian Independent Assortment**. $\implies$ Use the difference from parents as **plausibly random genetic variation**.



Remark on Estimation:

Control for imputed mean parental Ed PGI from siblings, because few parents observed in UK Biobank, losing statistical efficiency.

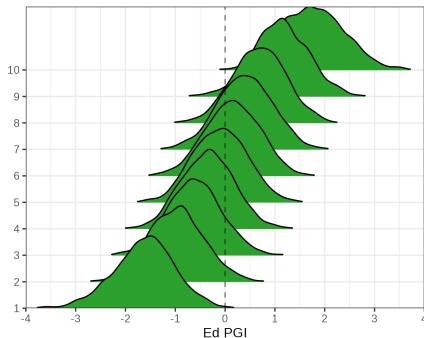
$$Z_i^{\text{Child}} - \mathbb{E} \left[Z_i^{\text{Child}} \mid Z_{f(i)}^{\text{Father}}, Z_{m(i)}^{\text{Mother}} \right] \longrightarrow \text{Independently assigned.}$$

Ed PGI Causal Effects: Estimates

Figure: Distribution of the Ed PGI, Relative to Imputed Parental Decile.

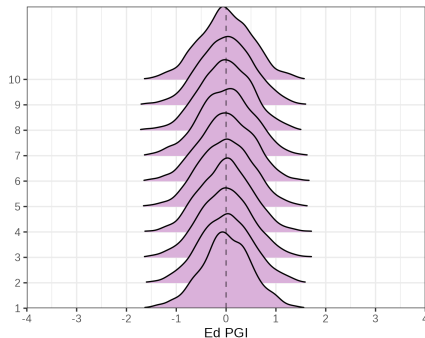
(a) Ed PGI.

Parental Quantile, 1st through 10th



(b) Ed PGI minus imputed parental mean.

Parental Quantile, 1st through 10th

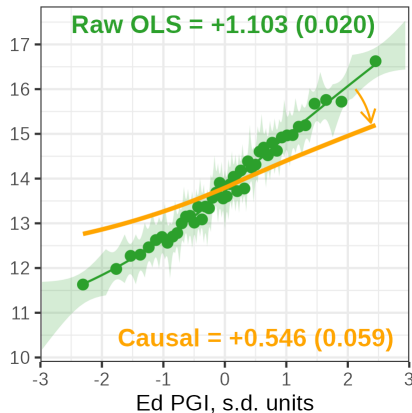


$$Z_i^{\text{Child}} - \mathbb{E} \left[Z_i^{\text{Child}} \mid Z_{f(i)}^{\text{Father}}, Z_{m(i)}^{\text{Mother}} \right] \longrightarrow \text{Independently assigned.}$$

Causal Effects of the Ed PGI

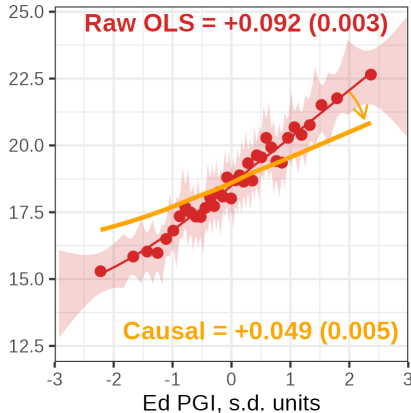
(a) Ed PGI $\rightarrow$ Education Years.

Education Years



(b) Ed PGI $\rightarrow$ Hourly Wages.

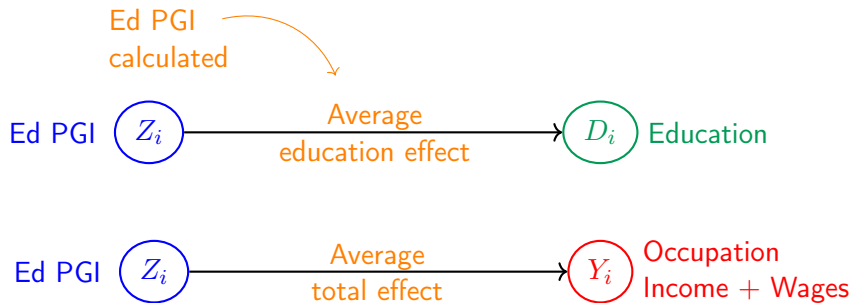
Occupation Hourly Wages, £



Replicates main estimates in Carvalho (2025).

Direct and Indirect Effects

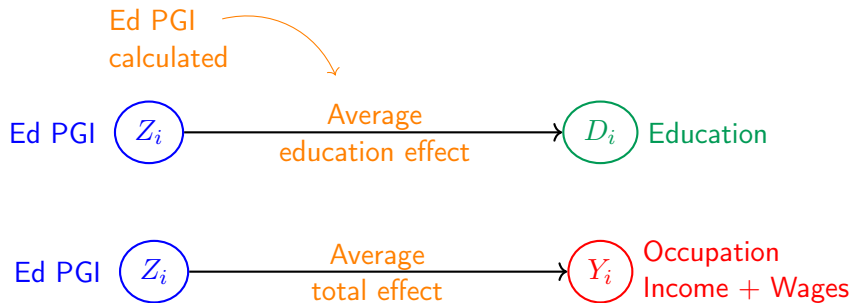
Ed PGI causally affects education years and labour market outcomes.
→ Not surprising, Ed PGI computed from data **only** genes + education.



Usual approach: Observe education years → labour market outcomes, and conclude with suggestive evidence that education “is a mechanism.”

Direct and Indirect Effects

Ed PGI causally affects education years and labour market outcomes.
→ Not surprising, Ed PGI computed from data only genes + education.



Usual approach: Observe education years → labour market outcomes, and conclude with **suggestive evidence** that education “is a mechanism.”

Direct and Indirect Effects

Ed PGI causally affects education years and labour market outcomes.

⇒ Consider the mechanisms:

Direct Genetic Channel

- ▶ Creativity
- ▶ Soft skills

Returns to Education

- ▶ Classroom learning
- ▶ Hard skills
- ▶ Occupational qualifications

A direct effect implies Ed PGI genes have direct labour market benefits.

Motivated by literature noting that Ed PGI improves/correlated with

- ▶ fluid intelligence test score (Carvalho 2025)
- ▶ Income, after controlling for education (correlationally, Papageorge Thom 2020, Bryson et al. 2025)
- ▶ Financial literacy at retirement (Barth et al. 2020).

Direct and Indirect Effects

Ed PGI causally affects education years and labour market outcomes.

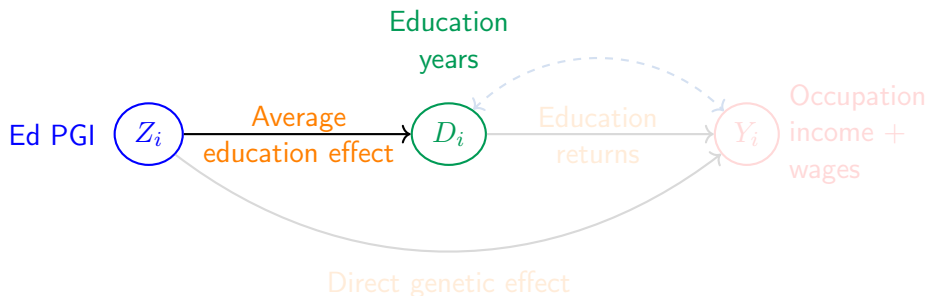
Figure: Model for Causal Effect of the Ed PGI.



Direct and Indirect Effects

Ed PGI causally affects education years and labour market outcomes.

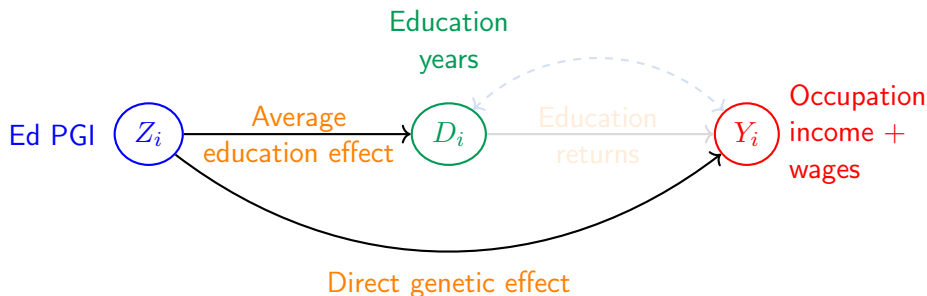
Figure: Model for Causal Mediation of the Ed PGI.



Direct and Indirect Effects

Ed PGI causally affects education years and labour market outcomes.

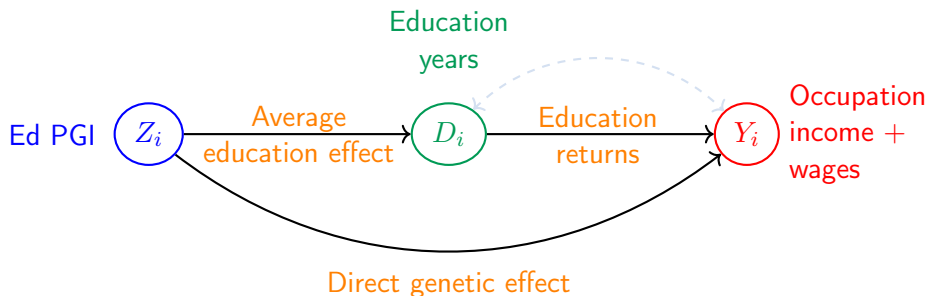
Figure: Model for Causal Mediation of the Ed PGI.



Direct and Indirect Effects

Ed PGI causally affects **education years** and **labour market outcomes**.

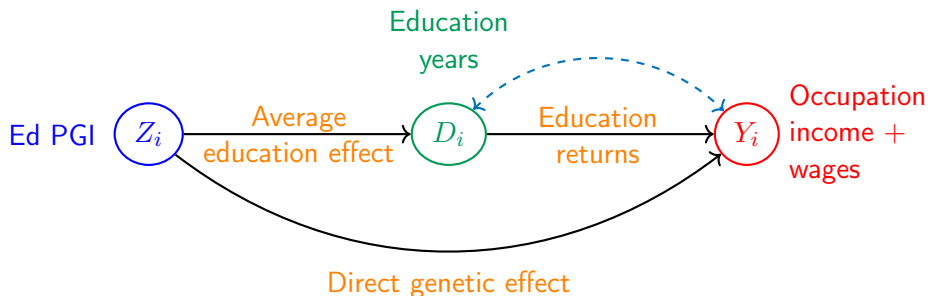
Figure: Model for Causal Mediation of the **Ed PGI**.



Direct and Indirect Effects

Ed PGI causally affects **education years** and **labour market outcomes**.

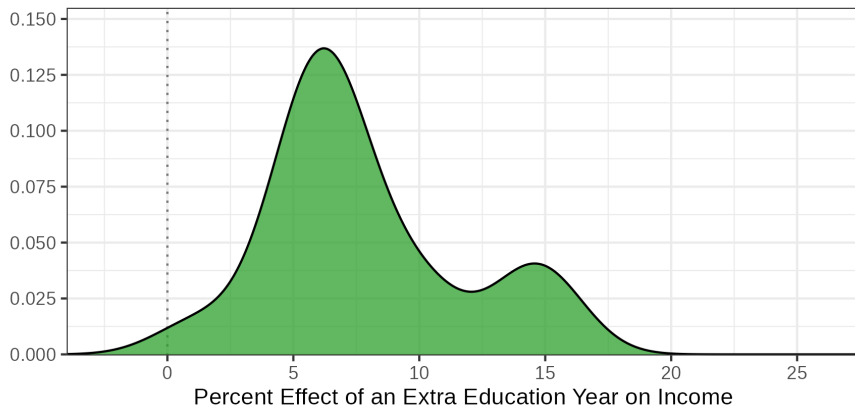
Figure: Model for Causal Mediation of the Ed PGI.



Returns to Education

The only missing piece for a valid Causal Mediation analysis are causal estimates for returns to education (and statistical uncertainty).

Density of British IV Estimates

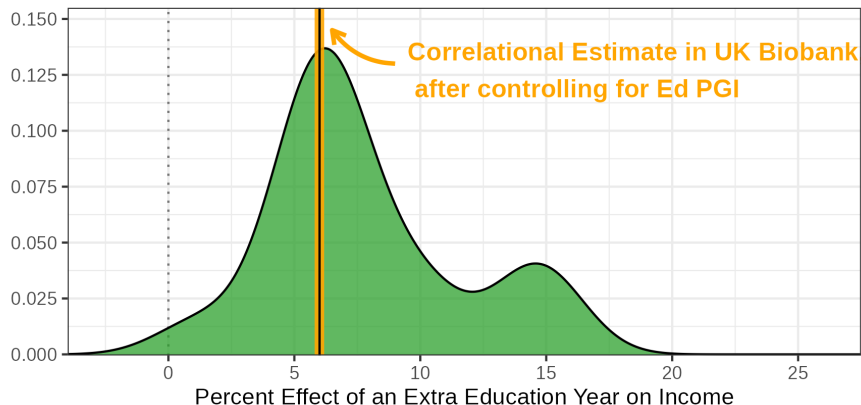


Note: Distribution from 17 IV estimates for Britain (Patrinos Psacharopoulos 2026).

Returns to Education

The only missing piece for a valid Causal Mediation analysis are causal estimates for returns to education (and statistical uncertainty).

Density of British IV Estimates

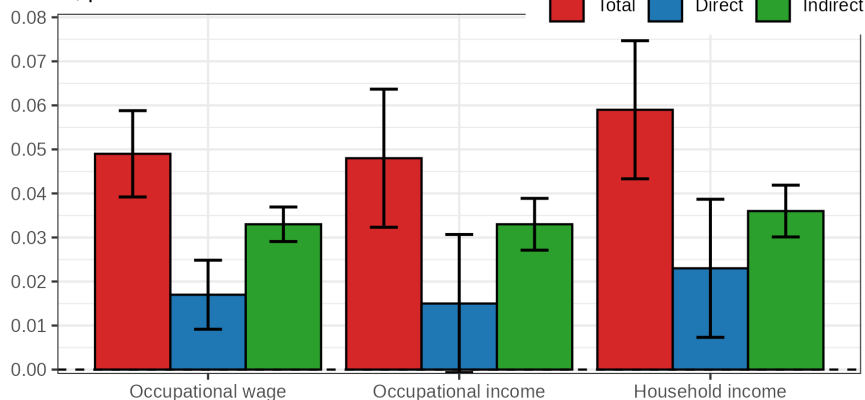


Note: Distribution from 17 IV estimates for Britain (Patrinos Psacharopoulos 2026).

Direct and Indirect Effects — Estimates

Using Correlation return estimate of $\approx 6\%$ (0.1 SE):

Estimate, percent Ed PGI effect on income outcomes



Direct and Indirect Effects — Estimates

Using Correlation return estimate of $\approx 6\%$ (0.1 SE):

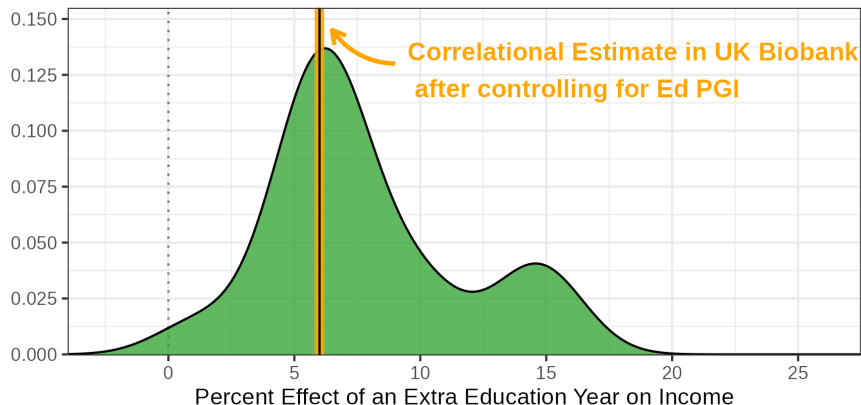
Estimate, Proportion mediated through education



Direct and Indirect Effects — Sensitivity Analysis

The only missing piece for a valid Causal Mediation analysis are causal estimates for returns to education (and statistical uncertainty).

Density of British IV Estimates

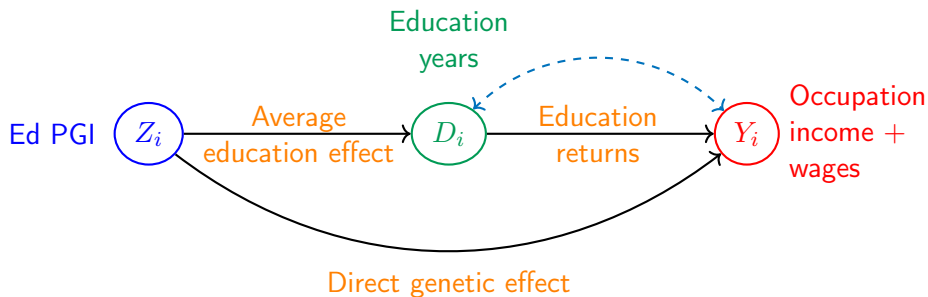


Note: Distribution from 17 IV estimates for Britain (Patrinos Psacharopoulos 2026).

Direct and Indirect Effects — Sensitivity Analysis

The only missing piece for a valid Causal Mediation analysis are causal estimates for returns to education (and statistical uncertainty).

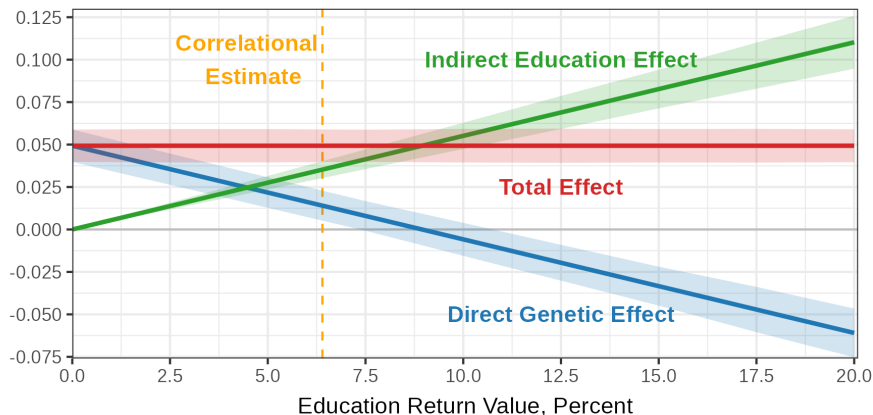
Figure: Model for Causal Mediation of the Ed PGI.



Direct and Indirect Effects — Sensitivity Analysis

Figure: Sensitivity Analysis, Effect of **Ed PGI** on **Log Occupational Wages** via **Education Years**.

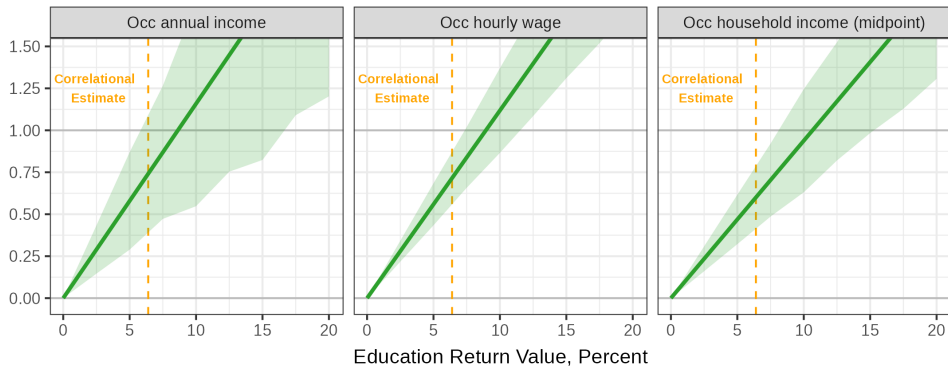
Direct and Indirect Effect Estimates



Direct and Indirect Effects — Sensitivity Analysis

Figure: Sensitivity Analysis, Proportion of **Ed PGI Effect** on **Income Outcomes** via **Education Years**.

Indirect Effect / Total Effect

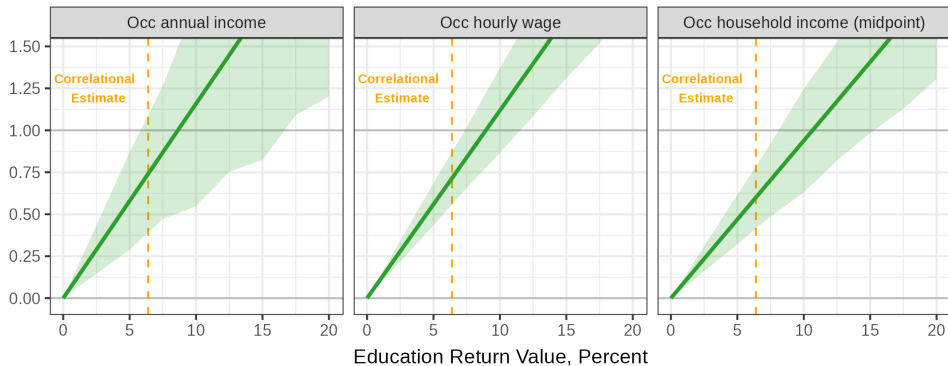


Direct genetic effect closes $\rightarrow 0$ at plausible education returns values.

Direct and Indirect Effects — Sensitivity Analysis

Figure: Sensitivity Analysis, Proportion of **Ed PGI Effect** on **Income Outcomes** via **Education Years**.

Indirect Effect / Total Effect



Direct genetic effect closes $\rightarrow 0$ at plausible education returns values.

Conclusion

Genes correlated with education impact income inequality primarily through education — implying this inequality gradient could be tackled via policies aimed at education access.

1. Ed PGI raises **education years** and **labour market outcomes** (replicating Carvalho 2025)
2. Economics literature has speculated (with suggestive evidence) on **direct genetic effects**, but lacked quantitative evidence
3. At plausible causal returns to education values, most of the total genetic income effect operates through the **education channel**, with a **direct genetic effect** that is small, and $\rightarrow 0$ under higher return values.

“The difference of natural talents in different men, is, in reality, much less than we are aware of [...]. The difference between the most dissimilar characters, between a philosopher and a common street porter, seems to arise **not so much from nature, as from habit, custom, and education.**”

— A Smith (1776), *The Wealth of Nations*, Volume 1 p. 16.

Conclusion

Genes correlated with education impact income inequality primarily through education — implying this inequality gradient could be tackled via policies aimed at education access.

1. Ed PGI raises **education years** and **labour market outcomes** (replicating Carvalho 2025)
2. Economics literature has speculated (with suggestive evidence) on **direct genetic effects**, but lacked quantitative evidence
3. At plausible causal returns to education values, most of the total genetic income effect operates through the **education channel**, with a **direct genetic effect** that is small, and $\rightarrow 0$ under higher return values.

“The difference of natural talents in different men, is, in reality, much less than we are aware of [...]. The difference between the most dissimilar characters, between a philosopher and a common street porter, seems to arise **not so much from nature, as from habit, custom, and education.”**

— A Smith (1776), *The Wealth of Nations*, Volume 1 p. 16.