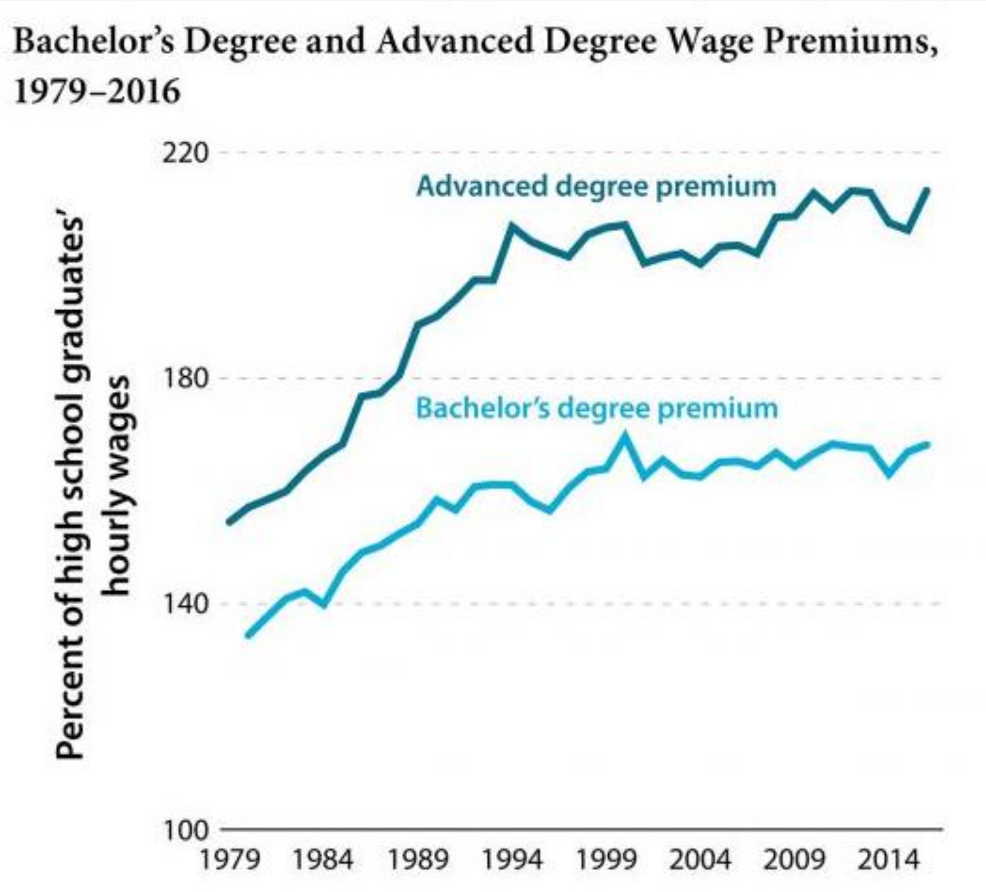
GENETIC ADVANTAGE AND EQUALITY OF OPPORTUNITY IN EDUCATION:

TWO DEFINITIONS AND AN EMPIRICAL ILLUSTRATION

Rita Dias Pereira

MOTIVATION

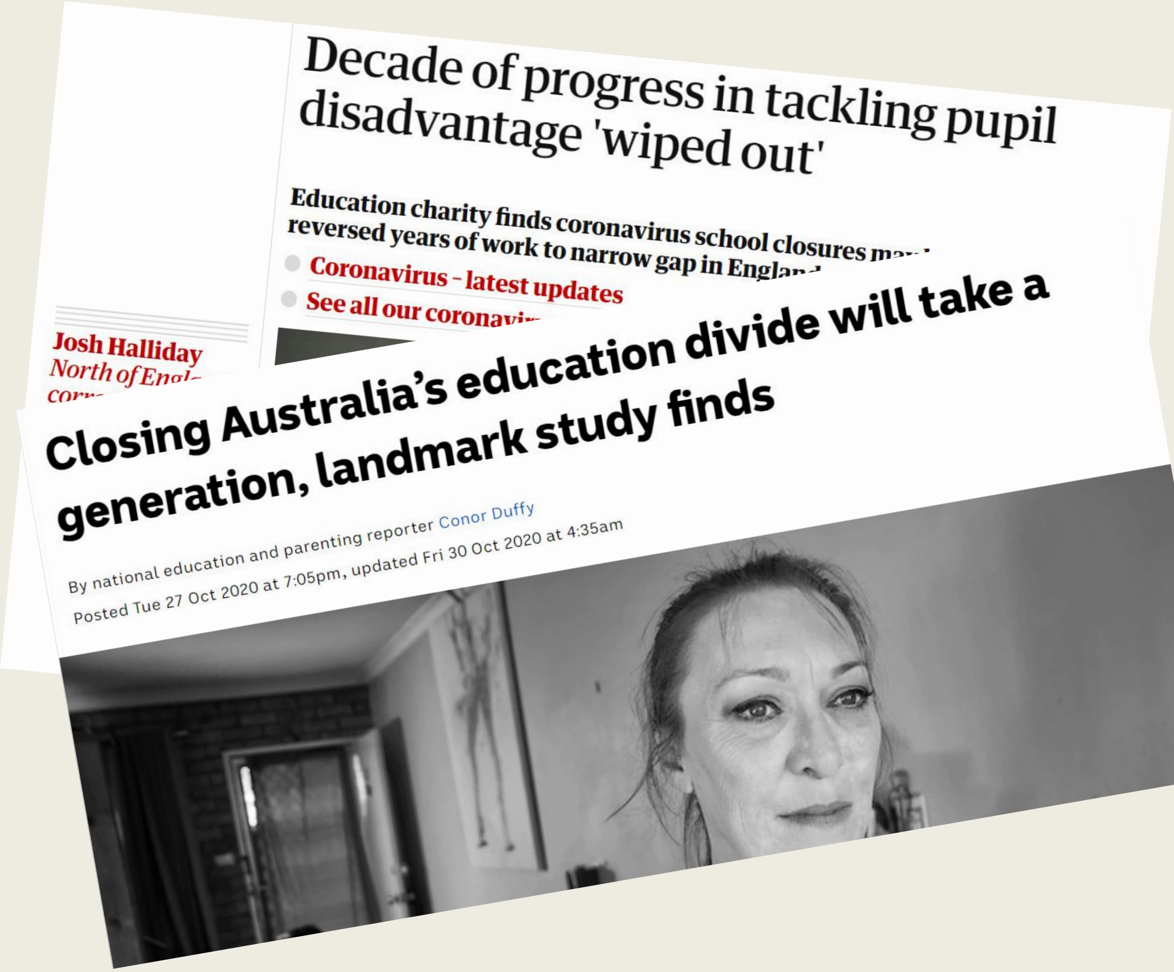
- Large education earnings premium in developed countries
- As of 2019, earnings premium of a bachelor or above (OECD, 2019):
  - U.S. 194%
  - EU 163%
  - OECD 158%
- Well documented health-education gradient
- Fairness in education is considered important



Source: U.S. current population survey, the Hamilton project

MOTIVATION

- There is an ongoing debate about fairness in education
- This fairness usually has to do with pupils’ socio-economic background disadvantage
- But there is another important factor that leads to inequality in educational outcomes: Genetic inequality

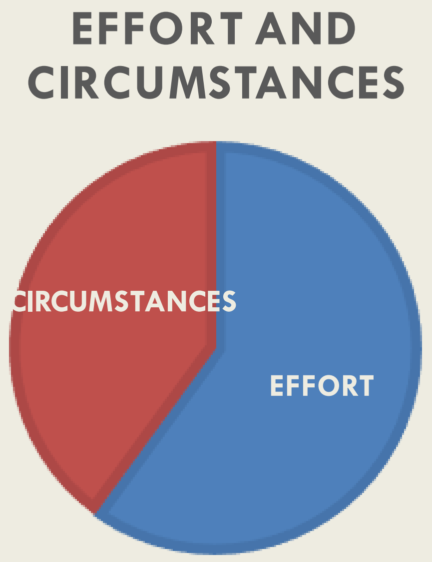


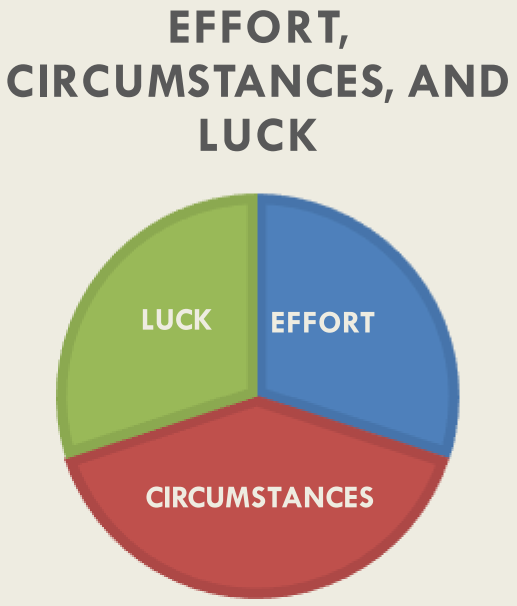
LITERATURE REVIEW

- Literature of social-science genetics quantifies genetic components (e.g. Lee et al., 2018), but without a measure of equality of opportunity
- Literature of Equality of Opportunity (EOp) tries to quantify EOp (e.g. Roemer & Trannoy, 2016), but without a measure of genetic components
- This work attempts to bridge this gap

ESTIMATING EOP WITHOUT A MEASURE OF GENETIC COMPONENTS

- Two main ways of decomposing outcome variance





DIFFERENT CLASSIFICATIONS OF GENETIC ADVANTAGE

Equality of Opportunity (EOp) literature:

- Some authors think that talents should be considered a fair (e.g., Rawls, 1971; Nozick, 1974), unfair (e.g. Roemer, 2004; Trannoy, 2019) or even partly fair and partly unfair (e.g. Lefranc et al., 2009; Lee and Seshadri, 2018) source of advantage.

Social-Science Genetics literature:

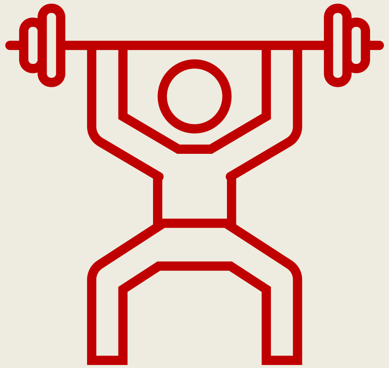
- Some authors support the view inequalities driven by genetic differences are fair (e.g. Rimfeld et al., 2018; Lin, 2020) or unfair (e.g. Kweon et al. 2020; Conley and Fletcher, 2018; Harden, 2021)

DEFINITIONS

Circumstances: factors that an individual cannot possibly affect.

Luck: random events which the individual affects, even if inadvertently.

Effort: choices that do not involve a random factor.

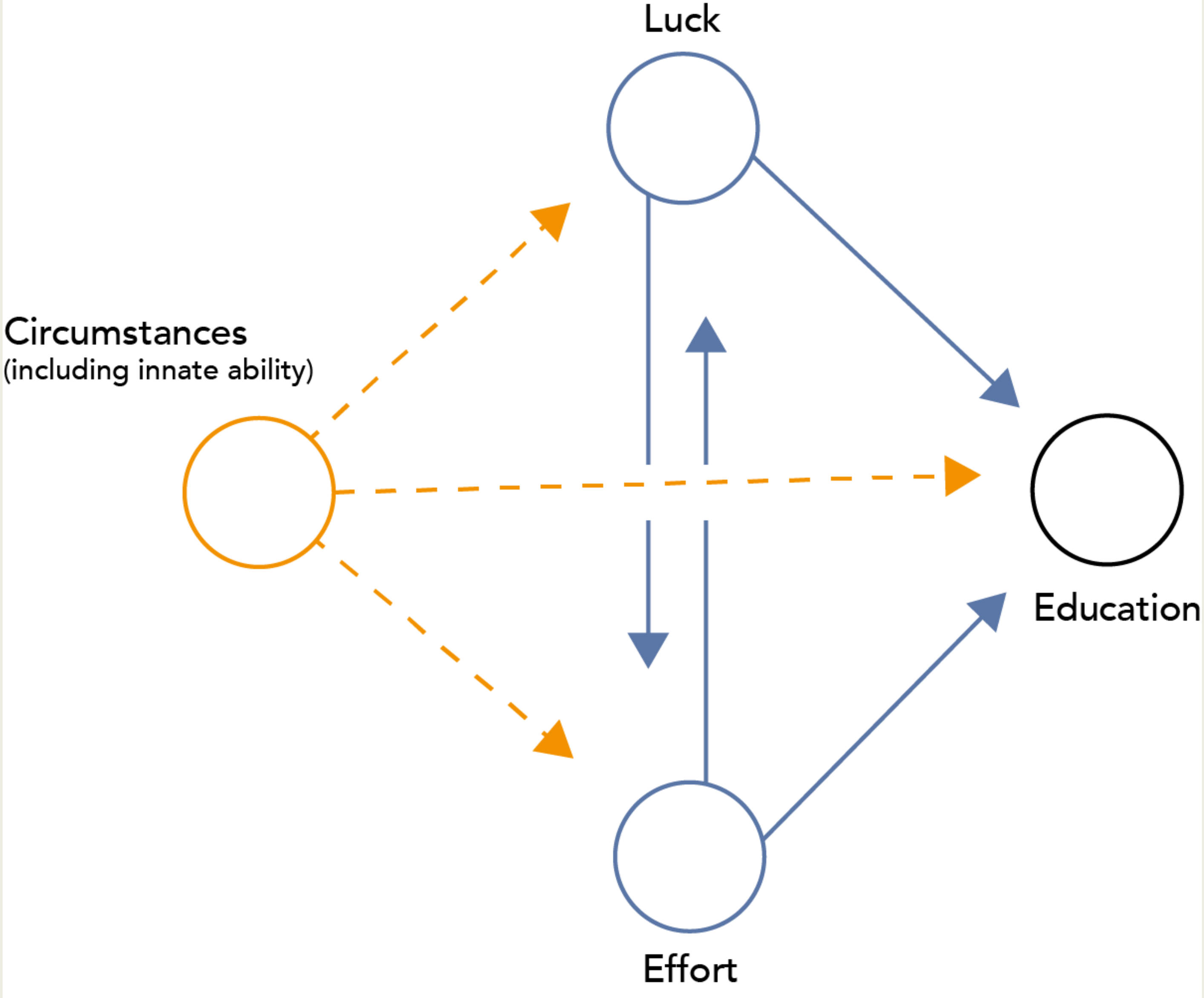


DEFINITION EOP 1

Genes are an unfair source of advantage in education

$$EOp\ is\ satisfied\ iff$$
$$\forall(c, c'), \forall e, F(.|c, e) = F(.|c', e)$$

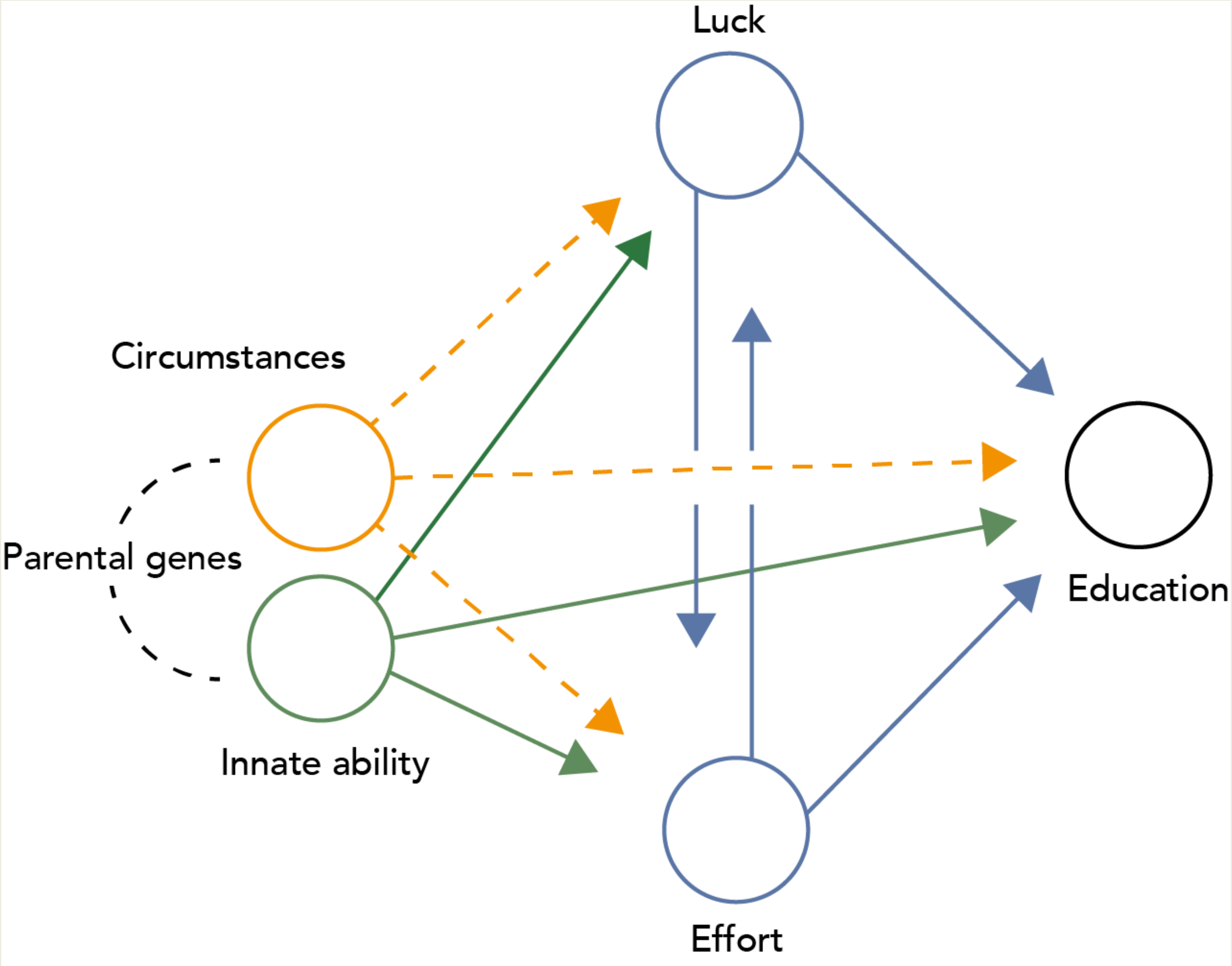
(Lefranc et al., 2009)



DEFINITION EOP2

Genes are a fair source of advantage in education

$$EOp\ is\ satisfied\ iff$$
$$\forall(c,c'),\forall a,\forall e$$
$$F(.|c,a,e) = F(.|c',a,e)$$



TESTABLE IMPLICATIONS DEF.1

Genes are an unfair source of advantage in education

$$EA_i = \alpha + \beta . Circumstances_{\ i} + r_i \longrightarrow \boldsymbol{\eta} = R^2 = 0$$

TESTABLE IMPLICATIONS DEF.2

Genes are a fair source of advantage in education

$$EA_i = \alpha + \beta.IAbility_{\ i} + r_i \longrightarrow R^2$$

$$EA_i = \alpha + \beta.IAbility_{\ i} + \phi.Circumstances_i + r_i \longrightarrow R'^2$$

$$\boldsymbol{\eta} = R'^2 - R^2 = 0$$

THE EDUCATIONAL ATTAINMENT POLYGENIC SCORE

- 4 Genome Wide Association (GWAS) studies that study educational attainment (Rietveld et al., 2013; Okbay et al., 2016; Lee et al., 2018; Okbay et al., 2022)
- $R^2 = 16\%$. Lead variants in the Educational Attainment Polygenic Score (EA PGS) are involved in brain development and neuron-to-neuron communication (Lee et al., 2018)
- Within family and mediation studies show that it captures cognitive and non-cognitive abilities (Belsky et al., 2018; Lee et al., 2018; Papageorge and Thom, 2020; Mõttus et al., 2017; Belsky et al., 2016)
- Predicts upward social mobility (Belsky et al., 2016, 2018) and educational attainment within families (Domingue et al., 2015)
- It is <u>not deterministic</u>! But having a high EA PGS gives you an <u>advantage</u>

LIMITATIONS – GENETIC NURTURE

- Controlling for parental PGS decreases the predictive power of PGS by ~50% (Kong et al. 2018; Selzam et al., 2019; Cheesman et al., 2020)
- The indirect effect of the EA PGS is a standard circumstance
- I obtain direct effect estimates using offspring phenotypic information (Wu et al., 2020) and by controlling for childhood SES (Selzam et al., 2019)
- The direct effect estimates on the Health and Retirement Survey (HRS) are 54% (offspring phenotype) and <u>53.5% (childhood SES)</u> for years of education

SUMMARY STATISTICS

	Baseline sample		1920-1929		1950-1959	
	mean	sd	mean	sd	mean	sd
Years of education	13.2	2.5	12.7	2.7	13.9	2.2
Education (categories)	3.5	1.4	3.3	1.5	3.9	1.2
Educational attainment polygenic score	-0.2	0.1	-0.2	0.1	-0.2	0.1
Childhood financial capital	0.0	1.0	-0.1	1.0	0.1	1.0
Childhood human capital	0.4	0.8	0.1	0.8	0.8	0.8
Childhood social capital	-0.0	1.1	0.1	1.1	-0.0	1.1
Childhood SES index	0.3	0.9	0.0	0.8	0.6	0.8
Gender	0.6	0.5	0.6	0.5	0.6	0.5
N	8,197		1,622		1,232	

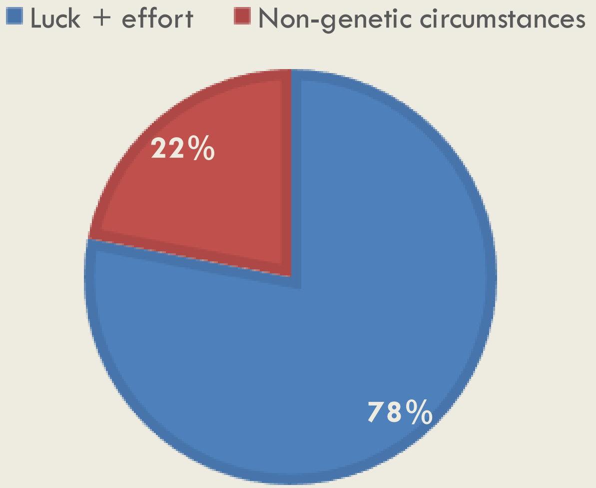
Table 1 Descriptive statistics of the baseline sample.

RESULTS

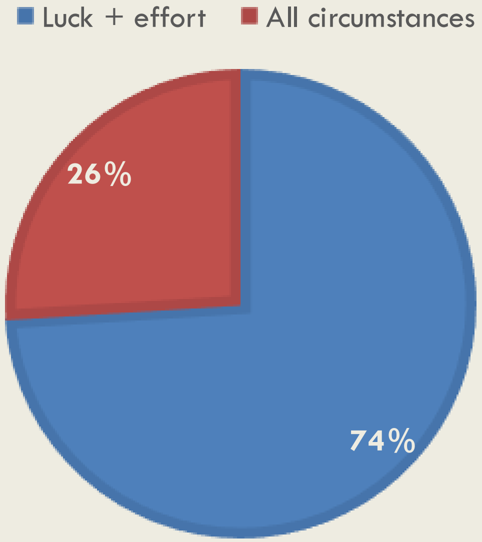
Health and Retirement survey, Cohorts 1920-1959; N= 7087

<u>Outcome: </u>Variance in years of education

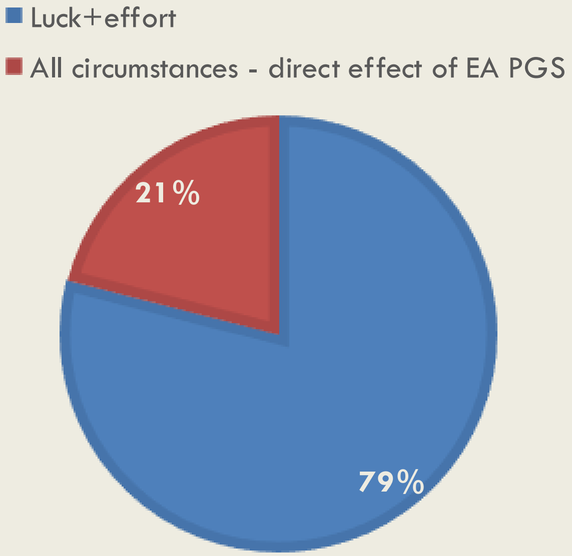
NAÏVE ESTIMATION OF INEQUALITY OF OPPORTUNITY



EA PGS IS AN UNFAIR SOURCE OF ADVANTAGE



DIRECT EFFECT OF PGS IS A FAIR SOURCE OF ADVANTAGE



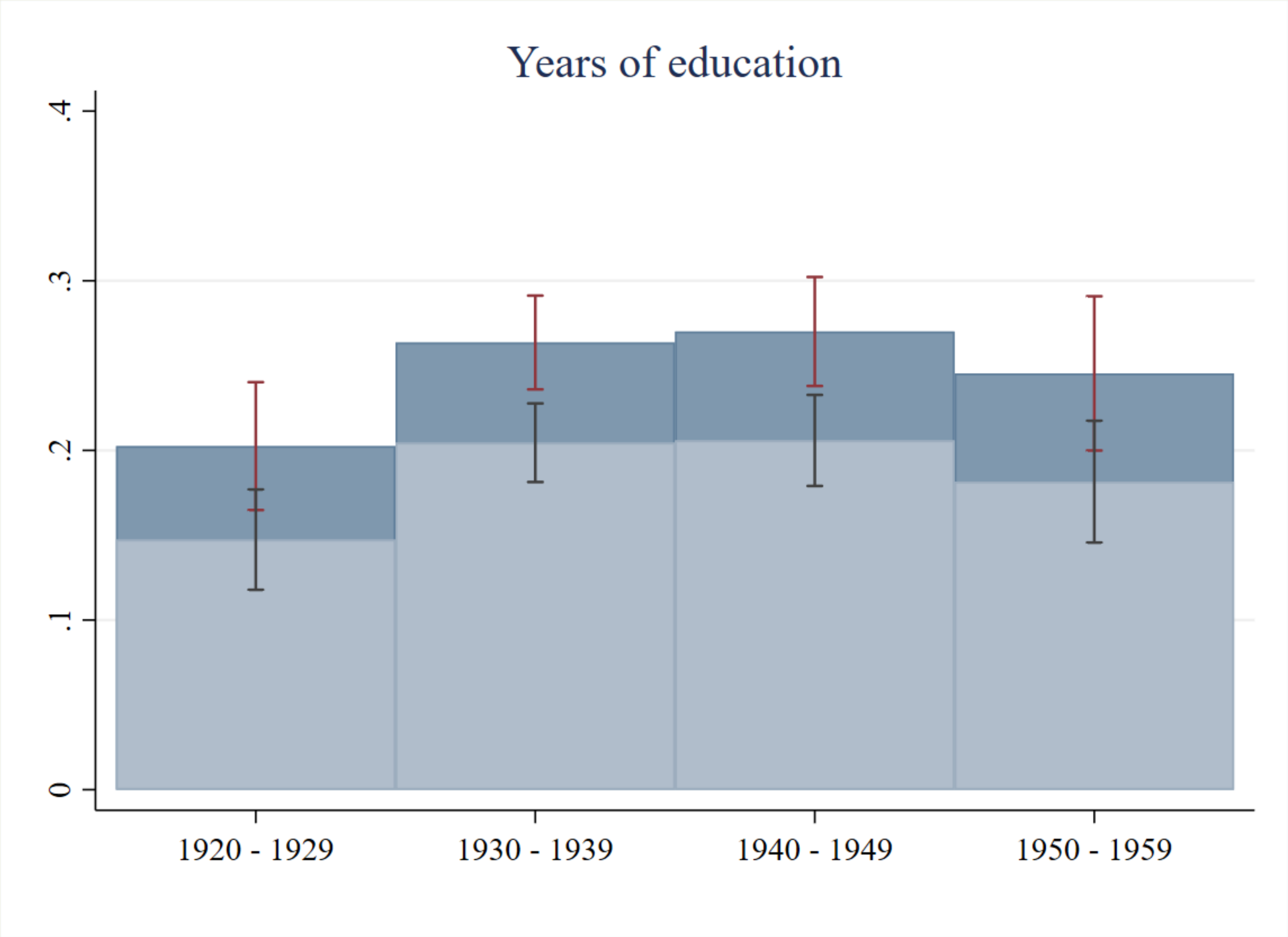
RESULTS

η estimates for **EOp Def. 1** and **EOp Def. 2**

Circumstances: childhood SES, Gender, Genetic nurture and (direct effect of EA PGS)

EOp Def. 1 20.2% to 24.5% (4.3pp increase)

EOp Def. 2 14.7% to 18.% (3.4pp increase)



RESULTS

High school completion

EOp Def. 1 11.9% to 8.7% (3.2 pp increase)

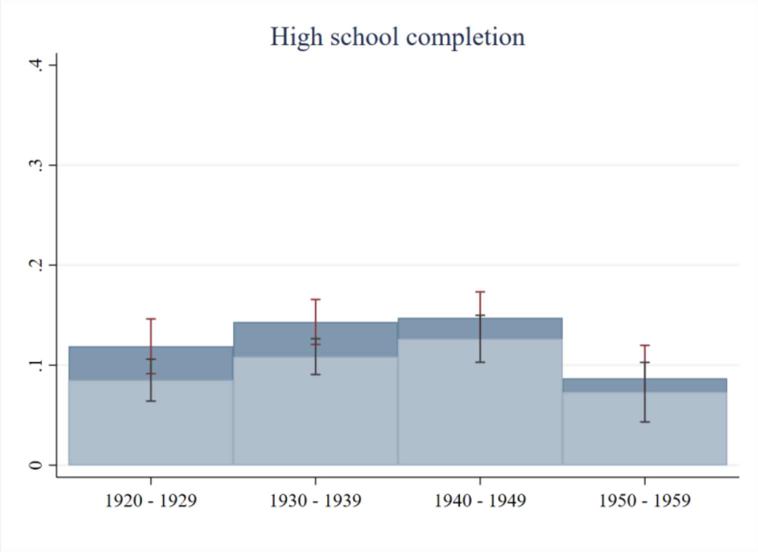
EOp Def. 2 8.5% to 7.3% (1.2 pp increase)

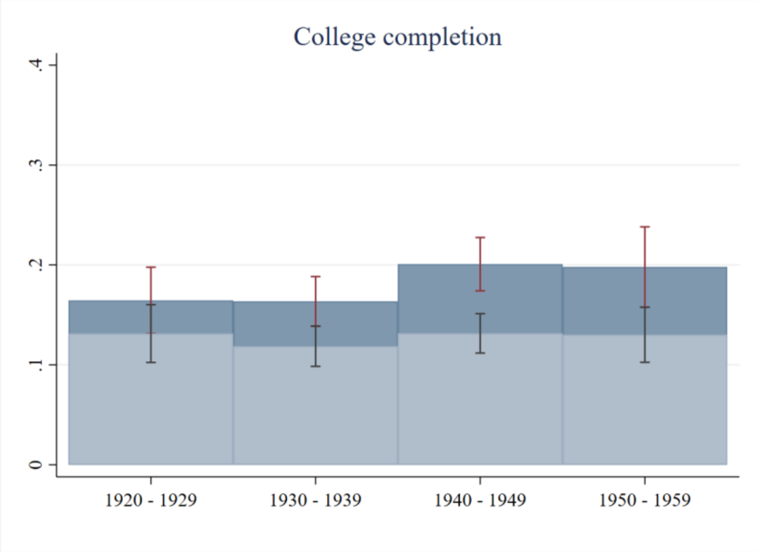
College completion

EOp Def. 1 16.4% to 19.8% (3.4 pp increase)

EOp Def. 2 13.1% to 13.0% (0.1 pp decrease)

Direct effect of the EA PGS 3.3% to 6.8%





CONCLUSION

- Genetic advantage is an important determinant of educational outcomes
- Estimates of inequality of opportunity change considerably depending on whether genetic advantage is considered a fair or unfair source of advantage
- Results suggest that genetic advantage is becoming increasingly important in higher levels of education
- Relevant for policy makers interested in fairness in education and in the economic returns that it warrants
- This work highlights the need for a more open and informed discussion of the true meaning of merit in education

THANK YOU

E-mail: diaspereira@ese.eur.nl

Twitter: @RitaDiasPereira

REFERENCES

Belsky, D. W., Moffitt, T. E., Corcoran, D. L., Domingue, B., Harrington, H., Hogan, S., ... & Caspi, A. (2016). The geneticsof success: How single-nucleotide polymorphisms associated with educational attainment relate to life-course development. Psychological science, 27(7), 957-972.

Belsky, D. W., Domingue, B. W., Wedow, R., Arseneault, L., Boardman, J. D., Caspi, A., Conley,D., Fletcher, J. M., Freese, J., Herd, P., et al. (2018). Genetic analysis of social-class mobility in five longitudinal studies. Proceedings of the National Academy of Sciences, 115(31):E7275–E7284.

Cheesman, R., Hunjan, A., Coleman, J. R., Ahmadzadeh, Y., Plomin, R., McAdams, T. A., Eley,T. C., and Breen, G. (2020). Comparison of adopted and non-adopted individuals reveals gene–environment interplay for education in the UK biobank. Psychological Science, 31(5):582–591.

Conley, D. and Fletcher, J. (2018).The Genome Factor: What the social genomics revolution reveals about ourselves, our history, and the future. Princeton University Press.

Davies, G., Marioni, R. E., Liewald, D. C., Hill, W. D., Hagenaars, S. P., Harris, S. E., ... & Deary, I. J. (2016). Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N= 112 151). *Molecular psychiatry*, *21*(6), 758-767.

Domingue, B. W., Belsky, D. W., Conley, D., Harris, K. M., and Boardman, J. D. (2015). Poly-genic influence on educational attainment: New evidence from the national longitudinal study of adolescent to adult health. AERA open, 1(3):2332858415599972.

Dworkin, R. (1981). What is equality? Part 2: Equality of resources. Philosophy & Public Affairs, pages 283–345.

REFERENCES

Harden, K. P., Domingue, B. W., Belsky, D. W., Boardman, J. D., Crosnoe, R., Malanchini, M.,Nivard, M., Tucker-Drob, E. M., and Harris, K. M. (2020). Genetic associations with mathematics tracking and persistence in secondary school. NPJ Science of Learning, 5(1):1–8.

Kathryn Paige Harden. (2021). *The genetic lottery why DNA matters for social equality*. Princeton Oxford Princeton University Press.

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., Young, A. I., Thorgeirsson, T. E.,Benonisdottir, S., Oddsson, A., Halldorsson, B. V., Masson, G., et al. (2018). The nature of nurture: Effects of parental genotypes.Science, 359(6374):424–428.

Kweon, H., Martschenko, D. O., Harden, K. P., DiPrete, T. A., Koellinger, P. D., et al. (2020). Genetic fortune: Winning or losing education, income, and health. Technical report, Tinbergen Institute.

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T. A., Bowers, P., Sidorenko, J., Linnér, R. K., 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):1112–1121.

Lee, S. Y. and Seshadri, A. (2018). Economic policy and equality of opportunity. The Economic Journal, 128(612):F114–F151.

Lefranc, A., Pistolesi, N., and Trannoy, A. (2009). Equality of opportunity and luck: Definitions and testable conditions, with an application to income in France.Journal of Public Economics,93(11-12):1189–1207.

REFERENCES

Lin, M. J. (2020). The social and genetic inheritance of educational attainment: Genes, parental education, and educational expansion. *Social science research*, *86*, 102387.

Mõttus, R., Realo, A., Vainik, U., Allik, J., and Esko, T. (2017). Educational attainment and personality are genetically intertwined. Psychological Science, 28(11):1631–1639.

Muslimova, D., Meddens, S. F. W., Rietveld, C. A., Von Hinke, S., Dias Pereira, R., and Van Kippersluis, H. (2020a). The relationship between the predictive power of polygenic scores and the genetic ranking of individuals across PGS construction methods. Mimeo

Nozick, R. (1974). Anarchy, state, and utopia, volume 5038. New York: Basic Books.

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., Turley, P., Chen,G.-B., Emilsson, V., Meddens, S. F. W., et al. (2016). Genome-wide association study identifies 74loci associated with educational attainment. Nature, 533(7604):539–542.

Papageorge, N. W. and Thom, K. (2020). Genes, education, and labor market outcomes: evidence from the health and retirement study. Journal of the European Economic Association, 18(3):1351–1399.

Rawls, A. (1971).Theories of Social Justice. Harvard University Press.

REFERENCES

Rimfeld, K., Krapohl, E., Trzaskowski, M., Coleman, J. R., Selzam, S., Dale, P. S., Esko, T., Metspalu,A., and Plomin, R. (2018). Genetic influence on social outcomes during and after the soviet era in Estonia. Nature Human Behaviour, 2(4):269–275.

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., ... & Preisig, M. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. science, 340(6139), 1467-1471.

Howe, L. J., Nivard, M. G., Morris, T. T., Hansen, A. F., Rasheed, H., Cho, Y., ... & Within Family Consortium. (2021). Within-sibship GWAS improve estimates of direct genetic effects. *BioRxiv*.

Becker, J., Burik, C. A., Goldman, G., Wang, N., Jayashankar, H., Bennett, M., ... & Okbay, A. (2021). Resource profile and user guide of the Polygenic Index Repository. *Nature human behaviour*, 1-15.

van Kippersluis, H., Biroli, P., Galama, T. J., von Hinke, S., Meddens, S. F. W., Muslimova, D., Pereira, R., & Rietveld, C. A. (2021). Using Obviously-Related Instrumental Variables to Increase the Predictive Power of Polygenic Scores. *bioRxiv*.

Roemer, J. E. (1993). A pragmatic theory of responsibility for the egalitarian planner. Philosophy &Public Affairs, pages 146–166.

Roemer, J. E. (1998). *Theories of distributive justice*. Harvard University Press.

REFERENCES

Roemer, J. E. (2004). Equal opportunity and intergenerational mobility: going beyond intergenerational income transition matrices. Generational income mobility in North America and Europe,pages48–57.

Selzam, S., Ritchie, S. J., Pingault, J.-B., Reynolds, C. A., O'Reilly, P. F., and Plomin, R. (2019). Comparing within-and between-family polygenic score prediction. The American Journal of Human Genetics, 105(2):351–363.

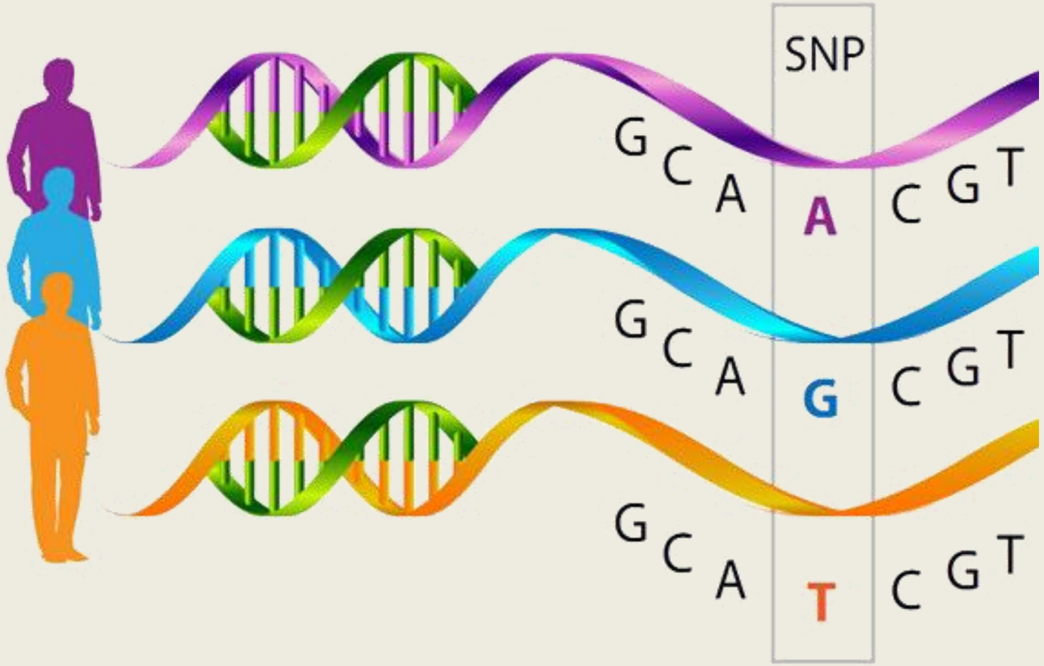
Trannoy, A. (2019). Talent, equality of opportunity and optimal non-linear income tax. The Journal of Economic Inequality, 17(1):5–28.

Tropf, F. C., Lee, S. H., Verweij, R. M., Stulp, G., Van Der Most, P. J., De Vlaming, R., ... & Mills, M. C. (2017). Hidden heritability due to heterogeneity across seven populations. *Nature human behaviour*, *1*(10), *757-765*.

Weale, M. E., Riveros-Mckay, F., Selzam, S., Seth, P., Moore, R., Tarran, W. A., ... & Donnelly, P. (2021). Validation of an Integrated Risk Tool, Including Polygenic Risk Score, for Atherosclerotic Cardiovascular Disease in Multiple Ethnicities and Ancestries. *The American Journal of Cardiology*, *148*, 157-164.

POLYGENIC SCORES (OR INDICES)

- The human genome has ~3.2billion pairs of nucleotides (AT or CG)
- 99.6% of nucleotides are identical
- Single Nucleotide Polymorphism (SNP)
- Polygenic scores aggregate several SNPs that associate strongly with a given trait, for example, educational attainment



THE EDUCATIONAL ATTAINMENT POLYGENIC SCORE

Limitations

- PGS capture parental characteristics that shape the rearing environment. ~50% of predictive power vanishes when "parental genetic nurture" is accounted for (Kong et al. 2018; Selzam et al., 2019; Cheesman et al., 2020)
- The predictive power of PGS is smaller than SNP-based heritability estimates; 12 vs 22-28% (Lee et al., 2018; Davies et al., 2016; Okbay et al., 2016; Tropf et al., 2017)
- Different methods of construction yield different ranking of individuals (Muslimova et al., 2020)
- Focus on European-ancestry populations; polygenic scores have limited portability across ancestry-diverse populations

POLYGENIC SCORES

Future directions

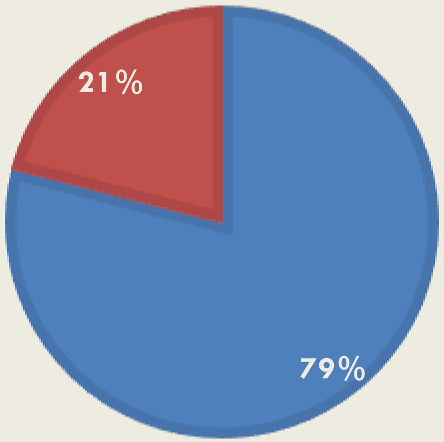
- "Trio" data sets, or sibling data sets allow an estimate of polygenic scores free of "parental genetic nurture" (e.g., Howe et al.,2021)
- Methods that improve predictive power of polygenic scores (e.g., Becker et al. 2021; van Kippersluis, 2021)
- Recent shift towards inclusivity of ancestry-diverse populations (Weale et al. 2021)

IMPORTANCE OF GENETIC ADVANTAGE

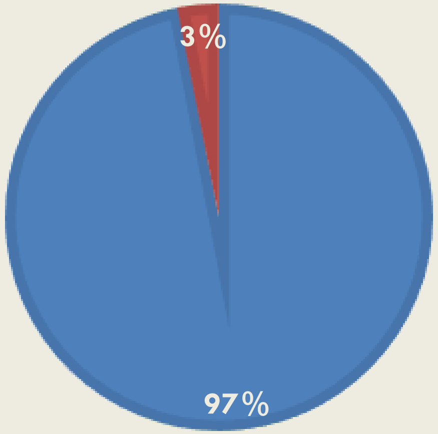
Cohorts 1926-1953; N= 7087

Outcome: Variance in years of education

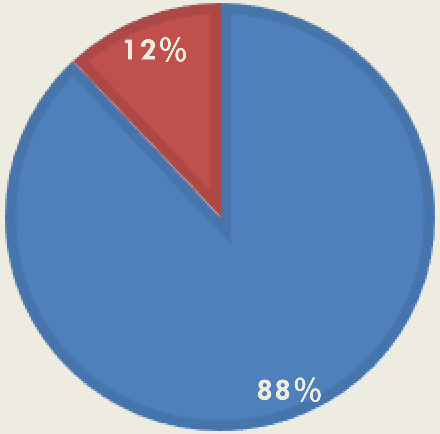
EXPLAINED VARIANCE BY CHILDHOOD SES



EXPLAINED VARIANCE BY GENDER, STATE OF BIRTH, AND MONTH OF BIRTH



EXPLAINED VARIANCE BY THE EDUCATIONAL ATTAINMENT POLYGENIC SCORE



SUMMARY STATS

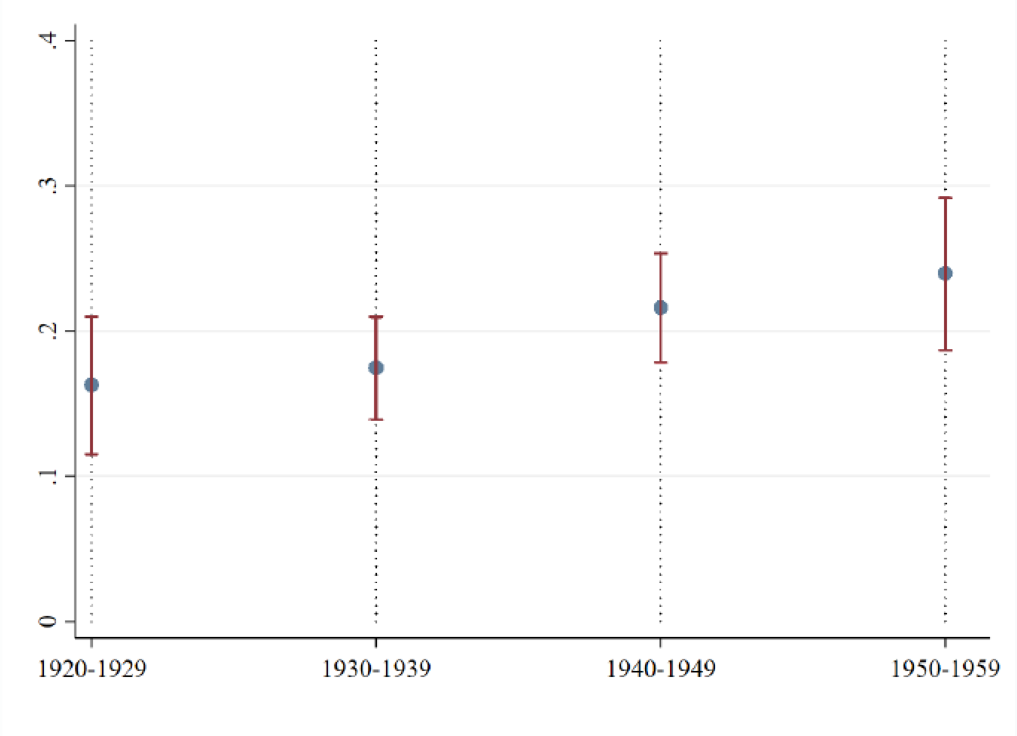


Fig 5 Correlation between the Educational Attainment PGS and childhood SES by cohort.

RESULTS

	Years of education			High school			College		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	-0.29**		-0.27**	0.02*		0.02**	-0.10**		-0.09**
	(0.05)		(0.05)	(0.01)		(0.01)	(0.01)		(0.01)
Childhood financial capital	0.20**		0.17**	0.03**		0.03**	0.03**		0.02**
	(0.03)		(0.03)	(0.00)		(0.00)	(0.00)		(0.00)
Childhood social capital	0.06*		0.04*	0.01**		0.01**	0.01*		0.01+
	(0.02)		(0.02)	(0.00)		(0.00)	(0.00)		(0.00)
Childhood human capital	1.24**		1.11**	0.13**		0.12**	0.16**		0.14**
	(0.03)		(0.03)	(0.01)		(0.01)	(0.01)		(0.01)
EA PGS (standardized)		0.83**	0.61**		0.08**	0.06**		0.13**	0.10**
		(0.03)	(0.02)		(0.00)	(0.00)		(0.00)	(0.00)
R-squared	0.206	0.105	0.262	0.111	0.050	0.137	0.132	0.087	0.184
N	8197	8197	8197	8197	8197	8197	8197	8197	8197

Table 2 Results of the OLS regression explaining educational attainment. Coefficients are displayed with robust standard errors. Columns 1-3 explain years of education. Columns 4-6 explain high school completion, where 1 is having at least a high school diploma and 0 otherwise. Columns 7-9 explain having a bachelor degree, where 1 is having at least a bachelor degree and 0 otherwise.

RESULTS

	Years of education		High school completion		College completion	
	Direct effect	% of PGS	Direct effect	% of PGS	Direct effect	% of PGS
Pooled	0.057	0.536	0.026	0.506	0.053	0.603
1920-1929	0.055	0.620	0.034	0.635	0.033	0.590
1930-1939	0.059	0.543	0.035	0.558	0.045	0.598
1940-1949	0.064	0.489	0.021	0.387	0.069	0.581
1950-1959	0.064	0.505	0.014	0.420	0.068	0.571

Table A2 Estimates of the direct effect of the EA PGS by cohorts. Columns 1-2 depict the estimates for years of education, columns 3-4 for high school completion and 5-6 for college completion. Column 1, 3 and 5 depict the R-squared of a regression explaining the outcome variable using the direct effect of the EA PGS. Columns 2, 4 and 6 depict the ratio between the R-squared of the direct effect and the R-squared of the EA PGS.

ROBUSTNESS

	EOp Definition 1			*EOp Definition 2*		
	Pooled	1920-1929	1950-1959	Pooled	1920-1929	1950-1959
	(1)	(2)	(3)	(4)	(5)	(6)
Original specification	0.262	0.203	0.245	0.206	0.147	0.182
Adjusted R-squared	0.262	0.200	0.242	0.205	0.145	0.179
Full Model	0.282	0.274	0.288	0.225	0.219	0.224
Cognition PGS	0.224	0.153	0.207	0.203	0.140	0.182
Principal components	0.268	0.209	0.246	0.211	0.153	0.182
HRS survey weights	0.267	0.204	0.245	0.208	0.151	0.180
Mortality survey weights	0.270	0.199	0.278	0.212	0.145	0.203
Birth year survey weights	0.265	0.197	0.270	0.209	0.144	0.197

ROBUSTNESS

	Baseline		1920-1929		1930-1939		1940-1949		1950-1959	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Female	-0.3**	-0.3**	-0.2+	-0.2	-0.3**	-0.3**	-0.3**	-0.3**	-0.3*	-0.3*
	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
EA PGS	0.6**	0.7**	0.7**	0.7**	0.6**	0.7**	0.6**	0.7**	0.6**	0.6**
	(0.0)	(0.0)	(0.1)	(0.1)	(0.0)	(0.1)	(0.0)	(0.1)	(0.1)	(0.1)
CFC	0.2**	0.2**	0.1*	0.1*	0.2**	0.2**	0.2**	0.2**	0.1*	0.1*
	(0.0)	(0.0)	(0.1)	(0.1)	(0.0)	(0.0)	(0.0)	(0.0)	(0.1)	(0.1)
CSC	0.0*	0.0*	0.0	0.0	0.1*	0.1+	0.0	0.0	0.0	0.0
	(0.0)	(0.0)	(0.1)	(0.1)	(0.0)	(0.0)	(0.0)	(0.0)	(0.1)	(0.1)
CHC	1.1**	1.1**	1.1**	1.1**	1.1**	1.1**	1.1**	1.1**	0.9**	0.9**
	(0.0)	(0.0)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
CFC × EA PGS		-0.0		-0.0		0.0		-0.0		0.0
		(0.0)		(0.1)		(0.0)		(0.0)		(0.1)
CSC × EA PGS		-0.0		-0.0		-0.0		-0.0		0.0
		(0.0)		(0.1)		(0.0)		(0.0)		(0.0)
CHC × EA PGS		-0.0		0.0		-0.1		-0.0		0.1
		(0.0)		(0.1)		(0.1)		(0.1)		(0.1)
1.male × EA PGS		0.0		0.0		0.0		0.0		0.0
		(.)		(.)		(.)		(.)		(.)
2.female × EA PGS		-0.1*		-0.2		-0.1		-0.1		-0.2
		(0.0)		(0.1)		(0.1)		(0.1)		(0.1)
R-squared	0.262	0.263	0.203	0.204	0.264	0.264	0.270	0.271	0.245	0.248
N	8,197	8,197	1,622	1,622	2,880	2,880	2,463	2,463	1,232	1,232

Table A9 Results of an OLS regression explaining years of education. Column 1 and 2 show the results for the whole sample and the subsequent columns show the results for each cohort. Odd columns include childhood SES, gender and the EA PGS as regressors. CFC stands for childhood financial capital, CSC for childhood social capital and CHC for childhood human capital. Pair columns include all interactions as regressors. Robust standard errors in parenthesis.

ROBUSTNESS

	Years of education			
	(non-standardized)		(standardized)	
	(1)	(2)	(3)	(4)
Gender	-0.3**	-0.1**	-0.3**	-0.1**
	(0.1)	(0.0)	(0.1)	(0.0)
EA PGS	0.6**	0.2**		
	(0.0)	(0.0)		
Childhood financial capital	0.2**	0.1**		
	(0.0)	(0.0)		
Childhood social capital	0.0*	0.0*		
	(0.0)	(0.0)		
Childhood human capital	1.1**	0.4**		
	(0.0)	(0.0)		
EA PGS (standardized)			0.6**	0.2**
			(0.0)	(0.0)
Childhood financial capital (standardized)			0.2**	0.1**
			(0.0)	(0.0)
Childhood social capital (standardized)			0.1*	0.0*
			(0.0)	(0.0)
Childhood human capital (standardized)			0.9**	0.4**
			(0.0)	(0.0)
R-squared	0.262	0.262	0.262	0.262
N	8,197	8,197	8,197	8,197

Table A10 Results of an OLS regression explaining years of education. In columns 1 and 2 the standardized outcome and in columns 3 and 4 the non-standardized.