Online Appendix

"It Matter What and Where We Measure"

Rafael Ahlskog

A brief primer on polygenic indices

For readers without previous experience with polygenic indices (PGIs), this section contains a slightly more elaborate description of how PGIs are constructed and how their results should be interpreted, as well as some more information on where and how they can be accessed.

Complex behavioral traits, like the ones under consideration in this study, are now known to be highly polygenic – that is, they are not influenced by a single or a handful of genetic variants, but by a very large number (thousands or millions) of variants, each with very small effect sizes. The approach in the earlier generation of studies looking into the effect of single genetic markers (the "candidate gene"-era) has therefore effectively been abandoned.

A more robust approach is to look at the combined effect of a much larger set of markers. This is done by relying, in a first step, on a genome-wide association study (or GWAS), where correlations between measured (typically around 1 million or more) genetic variants (single nucleotide polymorphisms, or SNPs) and some outcome are calculated separately, and tested against a very stringent significance threshold (typically $p < 5 \times 10^{-8}$). The output of a GWAS is thus a set of coefficients for all SNPs. In a subsequent step, we can use this output out of sample by multiplying the number of alleles an individual has for a particular genetic variant by the GWAS coefficient for that variant, and in this fashion additively summing up all of the previously discovered variants. This produces a single index – a polygenic index – which can be interpreted as some fraction of that individual's genetic propensity for the trait under study, where the size of that fraction is a function of the precision of the weights for each SNP and thus the sample size used in the original GWAS.

The PGI will capture a number of other things, apart from genetic signal, however – the most important arguably being so-called *population stratification*. That is, groups of individuals who differ on average in their genetic makeup may also face different environmental circumstances. The coefficient uncovered in the GWAS therefore captures both possible genetic effects *and* the effects of environmental factors that are correlated with genetic variation. The typical way of addressing this is to include a number of principal components of the genetic relatedness matrix, thus capturing at least some of this stratification (Price et al. 2006). This has been shown to be insufficient, however, with remaining stratification bias being non-negligible especially for socially relevant traits like education (e.g. Selzam et al. 2019).

A more reliable design-based way of controlling for this is to rely on the random process of recombination: genetic differences between siblings are random. When using a sample of siblings, and adding pair fixed effects, the problem of stratification therefore disappears. When using data from dizygotic twins, who are essentially siblings born at the same time, one also removes (non-confounding) noise resulting from cohort effects, resulting in increased precision.

When interpreting the results from family-based PGI designs, it is important to remember that although downstream differences have a causal interpretation, the causal effect of a PGI for e.g. education is not necessarily exclusively mediated by actual trait education – many of the same genetic variants that predict education may also influence (or be genetically confounded with variants for) other traits in parallell to education. That is, even though a genetic effect can be established, this effect can be transmitted through a wider range of intermediary traits than education.

Polygenic indices can theoretically be constructed in any study that has genotype information (i.e. where participants have contributed DNA samples), as long as one has access to a set of GWAS estimates. One of the strengths of this approach is, therefore, that as GWA studies become larger and more precise, better polygenic indices can be constructed in any given genotyped sample without having to gather any extra information on the participants.

Consistently constructed olygenic indices are available in a number of large cohorts, in large part thanks to the Polygenic Index Repository Project (Becker et al. 2021). This repository contains ready-made PGIs for a wide range of traits (health-related, anthropometric, psychological and behavioral) in a total of 13 genotyped samples ranging in sample size from the hundreds (e.g. Texas Twins, Dunedin study) to the hundreds of thousands (UK BioBank). Apart from the Swedish Twin Registry, several other genotyped cohorts also contain variables relevant for political scientists, like Add health, the Wisconsin Longitudinal Study, the English Longitudinal Study of Aging, and the Minnesota Twin and Family Study. Access to these datasets can be obtained after getting approval from an ethical review board. Since the cost of genotyping has decreased dramatically in the last decade, many other genotyped samples can also be expected to become available in the near future.

A longer note on methods

This section contains a more elaborate description of the two methods used in the paper.

The discordant twin model relies on a simple fact of nature: twins are born of the same parents, at the same time. Furthermore, identical (henceforth monozygotic, or MZ) twins share 100% of their DNA, while fraternal (dizygotic, or DZ) twins share on average about 50% of their segregating genes. When we compare MZ twins to each other, as opposed to comparing unrelated individuals, we are therefore automatically factoring out all confounding factors that the twins share: upbringing, social circles and networks etc., as well as all genetic factors (Vitaro et al. 2009). If the MZ twin with higher education is also systematically more economically right-wing, this correlation *at least* cannot be attributed to any of the aforementioned factors.

A discordant twin model can be estimated in two equivalent ways. One is to include all twin pairs and use structural equation modeling to partial out the direct effect, and how much of the naive correlation is due to genetic vs. environmental confounders. The other approach is to use MZ twin pairs and run conventional regression models with family fixed effects. The original study by Rasmussen et al. (2021) uses the former, implemented as a series of Cholesky decompositions. The focus in this study is to test whether the main effects replicate, rather than disentangling the sources of confounding. For reasons of parsimony, the analysis will therefore rely on the MZ sample and use OLS regression with pair fixed effects. Since all measures are standardized, the methods are equivalent (Turkheimer and Harden 2014).

The complementary method I propose relies on actual molecular genetic data in the form of a PGI. As outlined above, this measure (like any correlational measure) can be confounded. However, thanks to Mendel's First Law, genetic differences between siblings (and therefore DZ twins) are random. Downstream differences between the twins that are related to their genetic differences are therefore causal. The models, thus, are pair-fixed effects models using complete DZ twin pairs, with the PGI as the independent variable, and sex as a control to decrease noise (since DZ twins can be of different genders).[1]

In this study I use a PGI of educational attainment[2] in DZ twins from the PGI Repository Project, to investigate the same question: are differences in education-linked genetics also related to differences in ideology along the two proposed dimensions? This methodological approach contrasts to the discordant MZ twin design in two ways. On the one hand, it has better causal validity since the PGI is randomized between DZ twins by

[1] Note that the common practice of including genomic principal components (see e.g. Price et al. 2006) is not required since population stratification is held constant within DZ twin pairs.

[2] A single-trait PGI is used. In the effective sample, the PGI predicts actual trait education, cross-sectionally, with an incremental R^2 of 8.6%.

nature. On the other hand, as elaborated on above, genetics associated with education may also be associated with other predictors of ideology. Thus, positive results do not necessarily imply that the effect is only mediated by trait education, but can also include the combined effect of other traits that share the same genetic architecture. Another way of putting this, in more familiar technical terms, is to say that while a within-family PGI is causal, it is not a *valid instrument* for education, since it does not satisfy the exclusion restriction.[3]

"Original" scale definitions

The original scale of social/economic conservatism, based on survey questions in the Danish Election Survey, contained seven/five items. Some of them can be translated to items from the SALTY survey fairly easily, while others have to be approximated with similar or related items, and a few left out. The original items with their chosen translations are outlined in table A.1 below. In accordance with the original study, the items are aggregated by simply adding them together (taking their intended direction into account), and standardized. The Cronbach's α for the economic/social scales are .67/.57, respectively, indicating a less than ideal fit.

New scale definitions

The new/alternative scales include a larger selection of items, and a different method of aggregation. They contain all items that have reasonable face validity as measures of the latent constructs, but all items are not weighted equally.

The economic ideology scale includes twelve items in the SALTY survey that are related to economic policy preferences, but not all items indirectly related to public

[3]Since we've moved a step back in the causal chain when using a PGI, however, the distinction between confounding and mediation becomes less clear cut.

Table A.1: Replication definitions of ideological measures

Social ideology

Original item	**Translated item**
1. Violent crimes should be punished much harder than they are today, 3. Crime is best prevented through rehabilitation	Introduce much harder prison sentences for criminals
2. In Denmark we should protect our national traditions	Instate a language test for Swedish citizenship, Increase economic assistance to immigrants to preserve their native culture
4. Homosexuals should enjoy the same rights as everyone else	*No equivalent*
5. Taxes on gas should be increased	Decrease carbon emissions
6. Protecting the environment must not hurt private business	Invest more in preventing environmental degradation
7. Religious extremists should be allowed to have public meetings	*No equivalent*

Economic ideology

1. An individual should be more responsible for himself / The public sector should be responsible for taking care of all, 2. If people who are unemployed don't take the job they are offered they should lose their unemployment benefits / People who are unemployed should be allowed to say no to a job they don't want	Decrease the public sector, Decrease social welfare
3. Competition is healthy. It stimulates people to work hard and develop new ideas / Competition is unhealthy, it brings out the worst in people	Give companies more freedom
4. People with high incomes pay too little in taxes	Taxes should be cut
5. Differences in income are too high in this country. Therefore, people with smaller incomes should have larger increases than others	Decrease economic inequality in society

economics (e.g. defense spending is a cost, as is spending on immigrant culture, but neither is included in the economic ideology dimension, to avoid conflating the economic and social dimensions).

The social ideology measure uses nine items related to crime and punishment, traditional/authoritarian moral values, multiculturalism and migration. Items related to multiculturalism and migration are often excluded from measures of social ideology for reasons of parsimony (e.g. Feldman and Johnston 2014), while others use e.g. a GAL-TAN scale (which explicitly includes "nationalism") as synonymous with social ideology (e.g. Dehdari et al. 2022). Either way, there is one such item in the original study ("In Denmark we should protect our national traditions"), and I include all available items in this cluster in the new measure. The possibility remains, of course, that this cluster of attitudes should theoretically be treated as a separate dimension (see more under Results and Discussion). Instead, I do exclude environmental policy items, of which there is one in the original measure. While there may be a correlation between conservatism and non-environmentalism in many contexts, much like there is between socially and economically conservative positions, environmental concerns could theoretically fit into both a progressive and a conservative world-view (this is to say that they lack face validity) (Pilbeam 2003).

To increase the precision of the scales as measures of the latent constructs, I run out-of-sample[4] structural equation models with the respective items and use the estimated weight of each item when aggregating the final scales.[5] Using this method allows the items to contribute different amounts to the scale, instead of assuming that they have equal importance.

The included items related to economic ideology are the following:

- Decrease the public sector

[4]I.e. in MZ twins for the DZ sample and vice versa. This helps avoid overfitting.

[5]The RMSEA of the two constructs were both .077 (i.e. below the common threshold of "good" fit, $< .08$), while the α values for the items were .73/.67.

- Decrease social welfare
- Taxes should be cut
- Sell publically owned companies to private buyers
- Run more healthcare in the private sector
- Promote more free trade in the world
- More private schools
- Give companies more freedom
- Decrease the influence of the financial market over politics
- Keep property taxes
- Keep maximum fees in public child care
- Decrease economic inequality in society

The included items related to social ideology are the following:

- Give school grades at a younger age
- Ban pornography
- Limit access to abortion
- Introduce much harder prison sentences for criminals
- Instate a language test for Swedish citizenship
- Admit fewer refugees
- Allow more skilled migration

- Increase economic assistance to preserve immigrants' native culture
- Decrease defense spending

Power and sensitivity analysis

Power for this study is calculated using two approaches: first, calculating the actual power based on a given expected effect size, and second, calculating sensitivity (lowest detectable effect size) based on a given power. For the first approach, an expected effect size of .1 was chosen based on two considerations. First, .1 is often considered to be the lower bound of a "small" effect size, meaning that effect sizes smaller than this can generally be considered of marginal practical importance. Second, it is very close to the .092 effect size found in the original study.

Since the data are clustered in pairs, observations are not independent. The following calculations assume that the intra-cluster correlation is approximately .5, and that the *effective* number of observations can therefore be obtained by deflating the total N by 1.5.[6] Furthermore, calculations are reported based on a two-tailed .05 significance threshold, as well as separately Bonferroni-adjusted for eight tests (ie. $.05/8 = .00625$). Note that Bonferroni correction is overly conservative when tests are correlated (which these tests are going to be by design), and the true power/sensitivity is therefore going to lie somewhere between the single-test and Bonferroni-adjusted estimates.

The total number of individuals in the discordant MZ twin models varies between 1682–1908. This gives an estimated power of 92–95%, or 73–80% when Bonferroni-adjusted. The total number of individuals in the PGI models is 2674–3182. This gives an estimated power of 99%, or 93–97% when Bonferroni-adjusted.

Given that the original effect size could be subject to the winner's curse and therefore be smaller in reality, using these estimates to inform a power analysis may produce

[6] In the main results, this is solved by using clustered standard errors by twin pair.

inflated estimates. Thus, it is also advisable to calculate the sensitivity – i.e. the smallest true effect size detectable with a fixed power of 95%. On the same assumptions as before, this gives a smallest detectable effect size of 0.101–0.108 in the MZ models, or 0.123–0.131 when Bonferroni-adjusted, and 0.078–0.085 in the PGI models, or 0.076–0.083 when Bonferroni-adjusted.

Reporting Standards for Experiments

Since this was not an experiment, most of the APSA Recommended Reporting Standards for Experiments do not apply. However, some additional information not reported in the main manuscript is of some relevance.

A. Hypotheses

The research question in this study is outlined in the introduction of the manuscript. The purpose was to replicate the main results from Rasmussen et al. (2021).

B. Subjects and context

The subject pool was all complete genotyped twin pairs in the Swedish Twin Registry who had also answered the SALTY survey. No other criteria were used. The survey was conducted by the Swedish Twin Registry on paper and mailed in in late 2009 and early 2010 and included all genotyped twins born between 1943 and 1958. The response rate was 46% (Magnusson et al. 2012).

C. Allocation method

Not applicable.

D. Treatments

Not applicable.

E. Results

For outcome measures, see above. The only covariate used was participant sex, in the PGI models. CONSORT flow chart not applicable. Details on means/standard deviations/attrition/missing data by group not applicable.

F. Other information

For information about IRB, funding and replication materials, see main manuscript.

Descriptives

Tables A.2 and A.3 contain descriptive statistics for the MZ and DZ subsamples, respectively.

Table A.2: Descriptive statistics, MZ sample

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
Birth year of twin pair	2,716	1,950	4.637	1,943	1,958
Sex of twin	2,716	0.558	0.497	0	1
Education years	2,715	12.02	2.623	7	19
Orig. EI	2,636	7.923	3.542	-1	19
Orig. SI	2,677	-3.319	2.909	-13	6
New EI	2,495	2.800	0.701	1.026	4.796
New SI	2,616	3.392	0.729	1	4.961
earlier_grades	2,701	3.336	1.245	1	5
ban_pornography	2,709	3.907	1.232	1	5
limit_abortion	2,703	1.879	1.163	1	5
harder_punishment	2,716	3.849	1.106	1	5
language_test_citizenship	2,711	3.697	1.172	1	5
fewer_refugees	2,705	3.248	1.214	1	5
more_skilled_immigration	2,692	3.371	1.018	1	5
more_support_immigrant_culture	2,709	3.823	0.995	1	5
decrease_defense_spending	2,694	2.706	1.047	1	5
decrease_public_sector	2,685	2.510	1.185	1	5
decrease_welfare	2,701	2.682	1.082	1	5
lower_taxes	2,696	3.397	1.159	1	5
sell_public_enterprise	2,697	2.418	1.138	1	5
more_private_healthcare	2,691	2.757	1.147	1	5
more_free_trade	2,667	3.594	0.901	1	5
more_freeschools	2,689	2.565	1.051	1	5
more_freedom_companies	2,697	3.285	0.963	1	5
decrease_finmarket_impact	2,662	2.453	0.927	1	5
keep_property_taxes	2,690	3.422	1.252	1	5
keep_maxtaxa	2,681	2.404	1.028	1	5
decrease_economic_inequality	2,706	2.051	0.996	1	5
less_carbondioxide	2,694	4.356	0.747	1	5
decrease_pollution	2,709	4.329	0.750	1	5

Table A.3: Descriptive statistics, DZ sample

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
Birth year of twin pair	6,682	1,950	4.548	1,943	1,958
Sex of twin	6,682	0.531	0.499	0	1
Education years	6,679	11.79	2.620	7	19
Orig. EI	6,432	7.715	3.622	-1	19
Orig. SI	6,590	-3.558	2.986	-13	7
New EI	6,061	2.774	0.704	1	5
New SI	6,414	3.342	0.757	1.110	5
PGI EA	6,682	0.00988	0.997	-3.752	3.715
earlier_grades	6,635	3.274	1.274	1	5
ban_pornography	6,661	3.907	1.230	1	5
limit_abortion	6,650	1.877	1.150	1	5
harder_punishment	6,671	3.800	1.124	1	5
language_test_citizenship	6,670	3.616	1.217	1	5
fewer_refugees	6,651	3.204	1.227	1	5
more_skilled_immigration	6,600	3.353	1.039	1	5
more_support_immigrant_culture	6,678	3.785	0.997	1	5
decrease_defense_spending	6,610	2.653	1.060	1	5
decrease_public_sector	6,588	2.491	1.206	1	5
decrease_welfare	6,604	2.656	1.106	1	5
lower_taxes	6,617	3.297	1.176	1	5
sell_public_enterprise	6,625	2.363	1.150	1	5
more_private_healthcare	6,591	2.734	1.160	1	5
more_free_trade	6,561	3.594	0.919	1	5
more_freeschools	6,618	2.545	1.076	1	5
more_freedom_companies	6,625	3.257	0.981	1	5
decrease_finmarket_impact	6,542	2.415	0.947	1	5
keep_property_taxes	6,589	3.354	1.239	1	5
keep_maxtaxa	6,554	2.399	1.027	1	5
decrease_economic_inequality	6,643	2.015	0.988	1	5
less_carbondioxide	6,622	4.386	0.737	1	5
decrease_pollution	6,672	4.365	0.748	1	5

Table A.4: Discordant twin models, matched samples

VARIABLES	(1) Orig. EI	(2) New EI	(3) Orig. SI	(4) New SI
Education years	-0.0523	-0.0257	-0.0516	-0.129***
	(0.0561)	(0.0577)	(0.0541)	(0.0463)
Constant	0.0757***	0.0651***	0.0428***	0.0138**
	(0.00672)	(0.00691)	(0.00647)	(0.00555)
Observations	1,560	1,560	1,560	1,560
R-squared	0.708	0.731	0.745	0.794
Twin pair FE	YES	YES	YES	YES

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Robustness: matched samples

Tables A.4 and A.5 contain all main models, but with matched samples. This makes sure differences between models are not driven by non-overlapping samples. Results do not change appreciably when this is accounted for.

Table A.5: PGI models, matched samples

VARIABLES	(1) Orig. EI	(2) New EI	(3) Orig. SI	(4) New SI
PGI EA	0.0243	0.0173	-0.133***	-0.153***
	(0.0345)	(0.0346)	(0.0378)	(0.0366)
Sex of twin	-0.172***	-0.108**	-0.119**	0.0397
	(0.0489)	(0.0488)	(0.0512)	(0.0491)
Constant	0.0542**	0.00398	0.0246	-0.0834***
	(0.0257)	(0.0256)	(0.0269)	(0.0258)
Observations	2,526	2,526	2,526	2,526
R-squared	0.636	0.644	0.624	0.638
Twin pair FE	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

References

Becker, Joel et al. (2021). "Resource Profile and User Guide of the Polygenic Index Repository". *Nature Human Behaviour* 5.12, pp. 1744–1758.

Dehdari, Sirus H. et al. (2022). "A Pink Slip for the Blue Reform". *Journal of Political Institutions and Political Economy* 3.1, pp. 1–21.

Feldman, Stanley and Christopher Johnston (2014). "Understanding the Determinants of Political Ideology". *Political Psychology* 35.3, pp. 337–358.

Magnusson, PKE et al. (2012). "TThe Swedish Twin Registry: Establishment of a Biobank and Other Recent Developments". *Twin Research and Human Genetics* 16.1.

Pilbeam, Bruce (2003). "Natural Allies? Mapping the Relationship between Conservatism and Environmentalism". *Political Studies* 51.3, pp. 490–508.

Price, Alkes L. et al. (2006). "Principal Components Analysis Corrects for Stratification in Genome-Wide Association Studies". *Nature Genetics* 38.8, pp. 904–909.

Rasmussen, Stig Hebbelstrup Rye et al. (2021). "Educational Attainment Has a Causal Effect on Economic, But Not Social Ideology". *Political Studies*.

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, pp. 351–363.

Turkheimer, Eric and K. Paige Harden (2014). "Behavior Genetic Research Methods". *Handbook of Research Methods in Social and Personality Psychology*. Ed. by Harry T. Reis and Charles M. Judd. Second. Cambridge University Press, pp. 159–187.

Vitaro, Frank et al. (2009). "The Discordant MZ-twin Method". *International Journal of Behavioral Development* 33.4, pp. 376–382.