TI 2022-062/I
Tinbergen Institute Discussion Paper

Occupational sorting on genes

Revision: February 2023

Thomas Buser[1]
Rafael Ahlskog[2]
Magnus Johannesson[3]
Sven Oskarsson[2]

1 University of Amsterdam and Tinbergen Institute

2 Department of Government, Uppsala University

3 Department of Economics, Stockholm School of Economics

Occupational sorting on genes

Thomas Buser, Rafael Ahlskog, Magnus Johannesson, Sven Oskarsson*

February 2023

Abstract

Are genetic differences between people associated with their career choices? We link data from a large sample of genotyped individuals to Swedish government register data on study major and occupation. Our data contains polygenic indices that summarize genetic variants linked to several components of human capital: cognitive skills, personality traits, mental health, and physical health. We present a detailed mapping of these genetic indices by occupation and study major. We show that differences in genes associated with human capital across careers are highly statistically significant. Rankings of majors and occupations differ strongly across indices, meaning that genes associated with different traits predict entry into different careers. Our results shed new light on the determinants of some of the most impactful decisions people must make in their lives.

*Thomas Buser (corresponding author): School of Economics, Amsterdam School of Economics, University of Amsterdam, and Tinbergen Institute; t.buser@uva.nl. Rafael Ahlskog: Department of Government, Uppsala University. Magnus Johannesson: Department of Economics, Stockholm School of Economics. Sven Oskarsson: Department of Government, Uppsala University.

With rapid progress in social science genetics, it has become increasingly clear that genetic differences play a significant role in shaping education and career outcomes. Recent studies have identified many genetic variants that are associated with educational attainment and income (Rietveld et al., 2013; Okbay et al., 2016; Kweon et al., 2020; Buser et al., 2021). The aim of this paper is to dive deeper and look into the career choices that underlie these associations. Linking genetic data from more than 29,000 genotyped individuals in the Swedish Twin Registry (STR) to government register data, we provide a detailed mapping of how genetic differences associated with determinants of human capital – cognitive skills, personality traits, and (mental) health – vary across different professional careers and education majors.

The STR provides improved polygenic indices (PGIs) for many individual traits based on a recently released repository (Becker et al., 2021). While most individual characteristics are at least partly heritable (Turkheimer, 2000), complex traits tend to be influenced by many individual genetic variants, each with a very small effect (Chabris et al., 2015). PGIs summarize these small correlations between each genetic variant and a given trait in a single number (Harden and Koellinger, 2020). The new repository contains PGIs for a range of traits – including cognition, personality, health and health behaviors, and wellbeing – constructed using a consistent methodology. Using government register data also allows us to control for socioeconomic background based on the socioeconomic data of the parents of the genotyped individuals. This is important because while genes are fixed at conception, they may still be correlated with family background and geographic origin (Plomin and Bergeman, 1991; Hamer and Sirota, 2000; Kong et al., 2018). Much of the resulting selection bias can be eliminated by controlling for socioeconomic background (Selzam et al., 2019; Houmark, Ronda, and Rosholm, 2020; Dawes et al., 2021).

Economic theory predicts that people choose the career path that maximizes their expected utility given their skills, preferences, and beliefs. Empirical studies have uncovered many factors that influence this calculus, including differences across individuals in preferences for job amenities and beliefs about economic returns (Stinebrickner and Stinebrickner, 2014; Wiswall and Zafar, 2015, 2018). An important branch of this literature investigates how people sort into study majors and occupations based on their skills. and personality traits. For instance, cognitive skills and math ability have been shown to predict the choice of college major (Humburg, 2017; Wiswall and Zafar, 2015) and professional career (Coenen, Borghans, and Diris, 2021; Heckman, Stixrud, and Urzua, 2006), such that higher-skilled individuals more often sort into occupations with higher future earnings (Arcidiacono, 2004; Berger, 1988).

A growing number of studies find evidence that non-cognitive traits – such as personality traits, economic preferences, and (mental) health – are important determinants of occupational sorting as well. The most commonly used division of personality factors is the so called Five-Factor model, which reduces variation in personality to five dimensions, often simply called the Big Five: openness to experience, conscientiousness, extraversion, agreeableness and neuroticism (McCrae and Costa, 2008). Many studies in personality psychology additionally argue for the importance of "dark" traits such as narcissism (Emmons, 1987). These personality traits are strong predictors of employment and wages (Heckman, Stixrud, and Urzua, 2006; Cobb-Clark and Tan, 2011; Fletcher, 2013; Deming, 2017), management positions (Moutafi, Furnham, and Crump, 2007; Lounsbury et al., 2016) , as well as occupational sorting (Holland, 1978, 1997; Barrick, Mount, and Gupta, 2003; Kristof-Brown, Zimmerman, and Johnson, 2005; Ham, Junankar, and Wells, 2009; Nieken and Störmer, 2010; Almlund et al., 2011; John and Thomsen, 2014; Viinikainen et al., 2020). Personality is also associated with study choices (Humburg, 2017; Coenen, Borghans, and Diris, 2021). On top of the classic personality traits, economic preferences – such as risk and time preferences – are also related to educational attainment and occupational choice (Bellante and Link, 1981; Dohmen et al., 2011; Fouarge, Kriechel, and Dohmen, 2014; Sutter et al., 2013; Golsteyn, Grönqvist, and Lindahl, 2014; Koudstaal, Sloof, and Van Praag, 2016; Alan and Ertac, 2018; Buser, Niederle, and Oosterbeek, 2021; Angerer et al., 2021). In particular, Brenner (2015) finds that senior managers are significantly less risk averse compared to non-senior executives. Finally, many studies document differences in health and health behaviors across level of education

(Cutler and Lleras-Muney, 2010; Galama, Lleras-Muney, and Van Kippersluis, 2018), study majors (Lipson et al., 2015; Montez et al., 2018), and occupations (Llena-Nozal, Lindeboom, and Portrait, 2004; Volkers, Westert, and Schellevis, 2007; Ravesteijn, van Kippersluis, and van Doorslaer, 2013; Rietveld, van Kippersluis, and Thurik, 2015).

All of these factors of human capital – cognitive skills, personality traits, economic preferences, and health – are heritable to some degree (Jang, Livesley, and Vemon, 1996; Romeis et al., 2000; Cesarini et al., 2009; Zyphur et al., 2009; Vukasović and Bratko, 2015). This makes it plausible that genetic differences between individuals that are associated with human capital are linked to sorting into different study majors and professional careers. In this paper, we present detailed mappings of a wide range average genetic indices across 46 educational specializations and 72 occupations (including executives and managers). We show that many of the human capital PGIs in our data vary significantly across study majors and occupations, even after strict corrections for multiple testing. The ranking of careers across different PGIs is very different, showing that genes linked to different traits and skills predict entry into different careers.[1]

The choices of education and occupation are some of the economically most important decisions people must make in their lives, yet we understand them poorly. Our results advance our knowledge by showing that genetic differences are strongly associated to occupational sorting. We do not conduct our analysis at the level of single genetic variants but use polygenic traits for human capital-related traits. Our findings therefore do not only show that genetic differences predict career choices but also give us indications for the traits through which genes affect sorting.

The associations between genes and career sorting we document are not necessarily causal. PGIs can be correlated with cultural and socioeconomic background and upbringing (Plomin and Bergeman, 1991; Hamer and Sirota, 2000; Abdellaoui et al., 2013; Kong et al., 2018). However, controlling for family SES can eliminate most of this bias (Selzam et al., 2019; Houmark, Ronda, and Rosholm, 2020; Dawes et al., 2021) and thanks to using government registry data that can be linked to people's parents we have detailed and precise information on people's socioeconomic background. Our results show significant variation in the polygenic indices across study majors and occupations conditional on controls for socioeconomic background and geographic sorting. The fact that a particular PGI predicts sorting into particular careers does not necessarily mean than this sorting in associated to the trait proxied by the PGI. Many genes affect several traits – a phenomenon known as pleiotropy – and are therefore included in several PGIs.[2] The fact that the rankings of career options vary across different PGIs indicates, however, that the different indices capture different genetically influenced traits that have different effects on occupational sorting.

Method

Polygenic Indices

Human DNA is composed of a sequence of approximately 3 billion pairs of nucleotide bases. These nucleotide base pairs are one of two types (alleles): adenine paired with thymine (i.e. AT), or cytosine paired with guanine (CG). At the overwhelming majority of these 3 billion locations in the genome (approximately 99.9%), there is practically no variation in the nucleotide base pairs across individuals. The segments of DNA in which individuals do differ are called genetic polymorphisms. What we here sometimes label genetic variants refer to the simplest and most common kind of genetic polymorphism, called a single-nucleotide polymorphism (SNP). SNPs are locations in the DNA sequence in which individuals differ from each other in terms of a single nucleotide base pair. The most common allele at a certain locus (that is, AT or CG) in a population is called the major allele, and the nucleotide base pair that is less common is called the minor allele. At conception, each individual inherits half of her DNA

[1]In a separate paper (Buser et al., 2021), we show that a pre-registered selection of PGIs linked to cognitive skills and personality traits affect unidimensional career outcomes including income, occupational prestige, and educational attainment.

[2]Figure S1 shows the pairwise correlations between the PGIs.

from her mother and half from her father. For a given SNP, one allele is transmitted from each parent. Therefore, for a particular SNP, there are three possibilities: An individual has zero minor alleles, one minor allele, or two minor alleles. This number (0-2) is called the individual's genotype for this particular SNP. Longer strings of base pairs (from a few hundred to a few million in a row) are what forms a gene, i.e. the instructions for the cell to produce a particular protein. When a gene contains SNPs, i.e. base pairs that differ between people, there are, in effect, different versions of the gene, that may result in slightly different versions of the protein. Due to the very large number of SNPs that are potentially relevant for human behavior and economic outcomes, it is difficult to incorporate them jointly in an econometric model. Instead, the established way of exploiting the SNP data is to construct a polygenic index (PGI) that additively summarizes the effects of a very large number of, often several million, SNPs.

Formally, a PGI s_i is a weighted sum of SNPs:

$$s_i = \sum_{j=1}^{J} \hat{\beta}_j x_{ij}$$

where x_{ij} is individual i's genotype at SNP j. The weights $\hat{\beta}_j$ are estimated in a genome-wide association study (GWAS) which tests all measured SNPs for associations with the outcome of interest. Since the number of SNPs J is typically orders of magnitude greater than the number of individuals in the sample, it is impossible to fit all SNPs simultaneously in a multiple regression. Instead, the outcome is regressed on each SNP separately, resulting in J regressions in total.

As a simplified example, imagine there are just two SNPs in the genome, for which a given individual can have either zero, one or two minor alleles. A GWAS for educational attainment shows that each additional minor allele in the first SNP is associated with a five days increase in educational attainment and each additional minor allele in the second SNP is associated with a 10 days increase. The resulting PGI for educational attainment would then consist of adding the number of minor alleles for the first SNP multiplied by 5 and the number of minor alleles for the second SNP multiplied by 10. Results from a GWAS, extending the example with two SNPs, can be used to construct PGIs based on the correlation between each of millions of genotyped SNPs and an outcome of interest.

The polygenic indices we use stem from the work of the Social Science Genetic Association Consortium (SSGAC) . They use several datasets and employ a unified approach to estimate new, more predictive PGIs for a large range of traits and outcomes. Becker et al. (2021) provide a detailed description of the methods and data sources. Importantly, the PGIs for a specific cohort (such as the Swedish Twin Registry) are based on GWA analyses excluding that particular cohort. That is, the GWAS discovery for the PGIs we use was conducted on independent data. The Swedish Twin Registry includes all PGIs for which Becker et al. reported a predictive capacity – as measured by the incremental R^2 (ΔR^2) – of at least one percent. The ΔR^2 for a PGI is obtained by subtracting the R^2 from a baseline model only including sex, age dummies, interactions between sex and the age dummies and the first 20 principal components of the genetic-relatedness matrix as predictors from the R^2 from a model that also includes the PGI as a regressor. The main data sources used by Becker et al. to construct the PGIs are the UK Biobank (UKB) and 23andme, an online direct-to-consumer DNA testing service. For many traits, published meta-analysis results that included other samples were also included. The exact trait measure used for the same PGI can vary across datasets.

The PGIs included in the STR data cover a wide range of traits, all of which are conceivably indicators of an individual's human capital: cognition – cognitive performance, self-rated math ability, age started reading – personality and mental health – ADHD, adventurousness, depressive symptoms, extraversion, morning person, narcissism, neuroticism, openness, religious attendance, and risk tolerance – substance use – alcohol use disorder, cannabis use, cigarettes per day, drinks per week, ever smoked – health conditions – asthma/eczema/rhinitis, asthma, hay fever, migraine, nearsightedness, self-rated health – wellbeing – satisfaction with family, satisfaction with friends, being left out of social activity, subjective wellbeing – and anthropometric and fertility indicators – physical activity, age at first birth, BMI, height. The STR data contain a PGI for educational attainment (EA). The EA PGI likely captures a variety of cognitive and non-cognitive skills which influence educational attainment and can therefore be interpreted as a measure for having won the "genetic lottery" for doing well in one's career.

Register data

The Swedish Twin Registry (STR) – the world's largest twin registry containing all twins born in Sweden from 1886 onwards (Lichtenstein et al., 2006) – is in several ways ideal for answering our research questions. Approximately 43,000 of the twins in the registry are genotyped. STR data can be linked to administrative registry data through Statistics Sweden. Swedish registry data contains indicators of education major and occupation reaching back many decades. Individuals can also be linked to the administrative data of their parents, allowing us to construct an indicator of parental socioeconomic status (SES).

Our classification of study majors in academic and vocational education is based on the Swedish educational nomenclature (SUN), which is a national version of the International Standard Classification of Education (ISCED). The SUN classification is available annually from the year 1990 onwards. We use the most recent available observation for each individual. We exclude individuals who only completed compulsory secondary schooling as they do not have a study major. Our classification of occupations is based on the 1996 version of the Swedish Standard for Occupational Classification (SSYK). This classification is available in the quinquennial census data for the years 1960-1990 and annually from 2001 to 2013. From 2014, a new and very different classification was used. We use the available observation that is closest to the year an individual turned 35 years old. For both classifications, we divide people into categories based on the most fine-grained four-digit level information when feasible, using higher levels or combining categories where this is necessary to avoid an excessively small number of observations in a given category. Supplementary Tables S2 and S3 show the resulting categorizations including the number of observations as well as the gender ratio and proportion of people who graduated from college in each cell. Because this process necessarily involves subjective judgements, we will also present results based on the two-digit level classification in the Supplementary Information.

We control for socioeconomic background using a measure of family SES that is constructed as an additive index of two items: highest parental education (measured as years of full-time schooling) and average parental earnings. We use parental earnings data for the closest available year to the parent being aged 55, i.e. ten years before retirement. To adjust for differences in scales between the two variables, we initially standardize the two subitems to have a mean of 0 and a standard deviation of 1. We use population data to obtain means and standard deviations for each parental birth cohort and then carry out the standardization separately within each cohort in order to take into account changes in average income and education levels over time. Consequently, our measure of family SES takes a value of 0 for an individual whose parents score average on each of the two items relative to other parents born the same year. For individuals where parental education is missing, we use parental income only and vice versa. For a similar approach to measuring family SES using Swedish register data see Lindgren, Oskarsson, and Persson (2019).

Analysis

GWAS results and, consequently, the resulting PGIs, may contain environmental confounds. This can be due to "genetic nurture" (Plomin and Bergeman, 1991; Kong et al., 2018). That is, the environment provided by parents might be correlated with and influenced by their genes (and therefore the genes of their children). Another potential confounder is assortative mating. If individuals with certain genetic tendencies select mates who have particular genetically influenced traits, this can induce spurious genetic correlations (Hartwig, Davies, and Davey Smith, 2018). Furthermore, different subgroups in a population that have different allele frequencies may have different outcomes due to other non-genetic factors such as cultural norms, policies, geographic environments, or economic circumstances. This can induce bias known as population stratification (Hamer and Sirota, 2000; Abdellaoui et al., 2013). At the GWAS stage, researchers typically try to limit bias from population stratification by restricting samples to a relatively homogenous population – usually by limiting the study sample to individuals of European descent – and by controlling for the leading principal components in the genetic-relatedness matrix to capture the possible confounding influence of population stratification (Price et al., 2006).

Nevertheless, PGIs can be correlated with socioeconomic background and differences in average PGI levels across study major or occupations may therefore reflect sorting on background rather than the genetic variants summarized by the PGIs. Controlling for family SES can eliminate most of this bias (Selzam et al., 2019; Houmark, Ronda, and Rosholm, 2020; Dawes et al., 2021). To do this as thoroughly as possible, we control for parental SES, dummies for municipality of residence at age 16, and birth-year dummies interacted with gender, as well as the first 20 components of the genetic-relatedness matrix in the genetic-relatedness matrix. We have access to annual information on municipality of residence from 1968 and onwards. For 1960 and 1965 we can retrieve corresponding information from the quinquennial censuses. We use the information on municipality of residence from the census closest in time to the 16th birthday and use the information from the 1960 census for anyone born in or before 1946. We use the contemporary division into 290 municipalities.

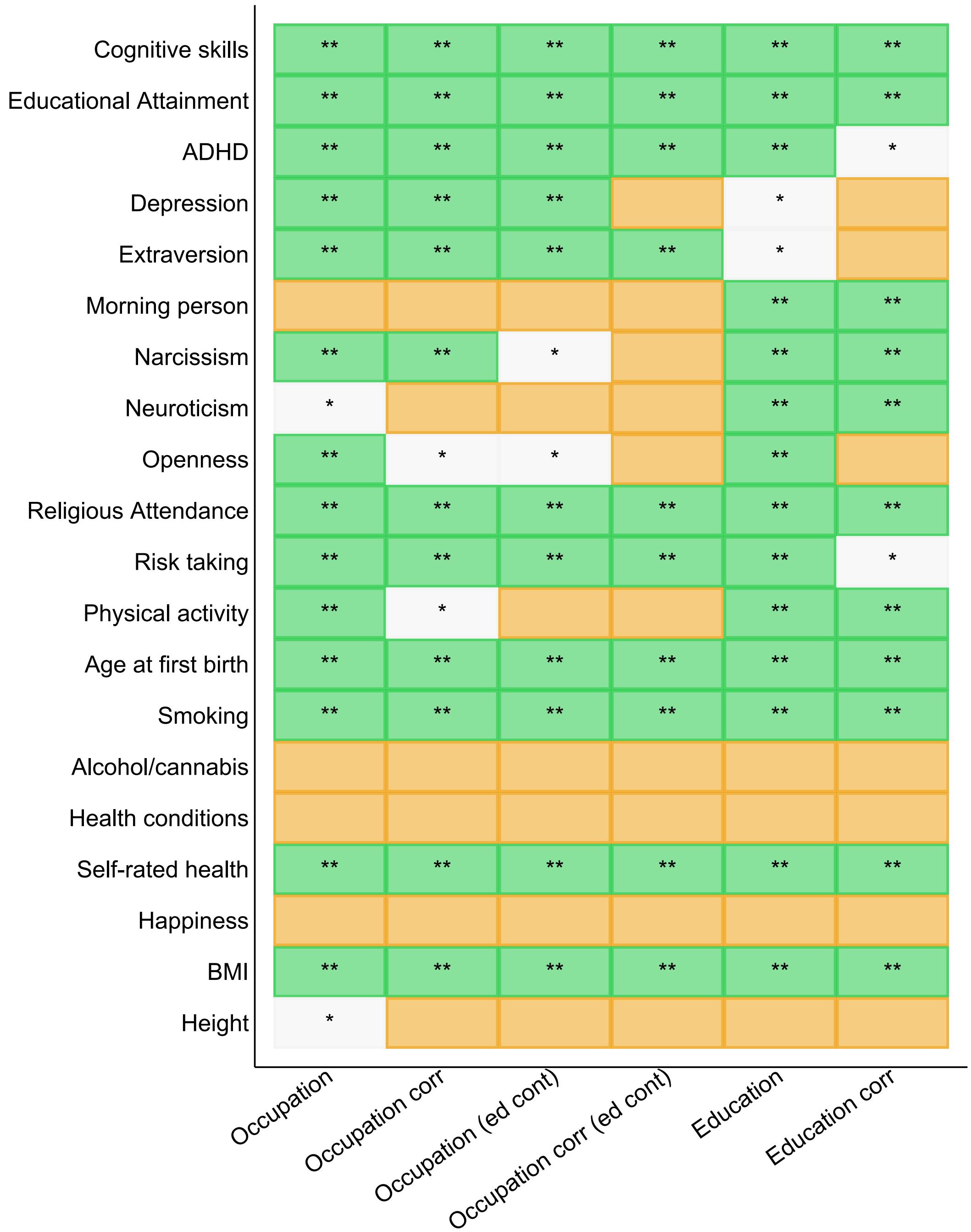
We will follow an exploratory approach whereby we first determine which PGIs vary statistically significantly across careers and then show detailed mappings of these PGIs across occupations and study majors. Many of the traits and behaviors for which we have PGIs are strongly related to each other. To make our analyses more manageable, we start by combining some of them into a single index by calculating the first principal component of groups of related traits. We combine cognitive performance, self-rated math ability, and age started reading into a single cognitive skills index; we combine risk tolerance and adventurousness into a single risk taking index; we combine cigarettes per day and ever smoked into a single smoking index; we combine alcohol use disorder, drinks per week, and cannabis use into a single substance use index; we combine asthma/eczema/rhinitis, asthma, hay fever, migraine, and nearsightedness into a single health conditions index; and we combine satisfaction with family, satisfaction with friends, being left out of social activity, and subjective wellbeing into a single happiness index.

Results

We will proceed with our analysis in several steps. First, we will select those PGIs which vary significantly across occupations or study majors. Second, we will show detailed mapping of those PGIs across all 46 educational specializations and 72 occupations. We will also show rankings of majors and occupations according to each of the trait PGIs. Prior to these analyses, we correct the PGIs for socioeconomic background and geographic sorting by regressing them on our socioeconomic status index, municipality of origin dummies, the first 20 components of the genetic-relatedness matrix, and birth-year dummies interacted with gender, and then using the standardized residuals for our analyses. Throughout, we use statistical significance at a strict 0.5% as our threshold for designating a result as "statistically significant" and 5% as our threshold for "suggestive evidence" (Benjamin et al., 2018).

To check whether a PGI varies across individuals who choose different careers, we regress each of the 20 SES-corrected PGI on the full set of 72 occupation dummies and 46 educational specialization dummies and then check the joint significance. Figure 1 shows a heatmap of p-values. For each PGI, we show six tests: variation across occupations, variation across occupations corrected for multiple testing using a conservative Bonferroni correction, variation across occupations conditional on education level, variation across occupations conditional on education level corrected for multiple testing, variation across education majors, and variation across education majors corrected for multiple testing. We will retain those PGIs which vary statistically significantly either across occupations or education majors after applying the Bonferroni correction. This leads us to drop four PGIs: alcohol/cannabis

Figure 1: Significance of variation in PGIs across occupations and study majors



Note: **p<0.005; *p<0.05. P-values are from test of joint significance after regressions of each PGI on the full set of 72 occupation dummies and 46 educational specialization dummies. P-values in the second, fourth, and sixth columns are Bonferroni-corrected for multiple testing. PGIs are corrected for socioeconomic background and geographic sorting by regressing them on our socioeconomic status index, municipality of origin dummies, the first 20 components of the genetic-relatedness matrix, and birth-year dummies interacted with gender, and then using the standardized residuals.

consumption, the combined index for health conditions, happiness, and height.

The aim of the following analyses is to document how the remaining 16 PGIs vary across people who chose different study majors and work in different occupations. It is also interesting to rank study majors and occupations by their overall distinctiveness across all PGIs. We use two approaches to do this. First, we calculate the overall dissimilarity of each study major and occupation across all PGIs relative to all other majors or occupations. Using Euclidean distance as our measure of dissimilarity, we first calculate how dissimilar each career option is from each other career option across the 16 PGIs and then take the average across the other career options. Second, we use principal components analysis to summarize the 16 PGIs into two principal components which will allow us to show differences across careers using two-dimensional plots. Table S1 in the appendix shows the first two rotated principal components, leaving cells with factor loadings below 0.25 blank. The first principal component picks up a mix of cognitive skills, forward-looking behavior and health, with high positive loadings on self-rated health, age at first birth, cognitive skills, educational attainment, and religious attendance, and high negative loadings on BMI, smoking, depression, and ADHD. The second component picks up on personality, with high positive loadings on risk seeking, extraversion, and openness, and a high negative loading on neuroticism. The first two principal components each represent important aspects of human capital: health and cognitive skills on the one hand and non-cognitive skills on the other hand.

Figure 2 shows a heat map of the rank of each education major in terms of the average value of each standardized PGI.[3] We divide study majors into 46 categories based on the Swedish educational nomenclature (SUN). The ranking of majors is very distinct across trait PGIs. Some study majors stick out by showing up at the extremes of several PGI rankings. For example, people who studied psychology on average score high on the cognitive skills, educational attainment, openness and risk seeking PGIs, but also on the depression, neuroticism, and narcissism PGIs. People who studied science or math score high on the cognitive skills, educational attainment, and age at first birth PGIs, and score low on the extraversion, neuroticism, risk taking, smoking and BMI PGIs. People who studied art rank high on the openness, risk taking, ADHD, depression, and smoking PGIs, and low on the self-rated health and morning person PGIs. People who studied medicine are quite remarkable in that they rank at the extremes for each trait except extraversion (they are also remarkable in that they rank very low on neuroticism but very high on narcissism).[4]

The last three columns of the heatmap in Figure 2 shows the rank of each study major according to the two principal components and overall dissimilarity. Chemical and bio engineering, medicine, dentistry, secondary-school teaching, and science and math rank highest on the first principal component (cognitive skills and health). Basic nursing, child care, and transport services rank lowest. The ranking of education specializations across the second principal component is quite different. The highest ranking majors are medicine, management, pedagogy, art, and policing and other security. The lowest-ranking options are medical analysts and technicians, media production, and general nursing. According to measure of overall dissimilarity, the most genetically distinct education specializations are chemical and bio engineering, medicine, housekeeping and cleaning, dentistry, and basic nursing.

In Figure 3, we use the same method to map differences in average PGIs across occupations.[5] We divide occupations into 72 categories based on the Swedish Standard for Occupational Classification (SSYK). The rankings of occupations differ strongly across traits. For example, people in occupations with a high math content and in teaching-oriented occupations – scientists, doctors, pedagogy professionals, engineers and architects, lab technicians, professors – on average have a high cognitive skills PGI. The top of the extraversion ranking is dominated by professions that require frequent and intense personal contact including executives, restaurant personnel, security

[3]Figure S3 in the appendix shows average values instead of ranks.

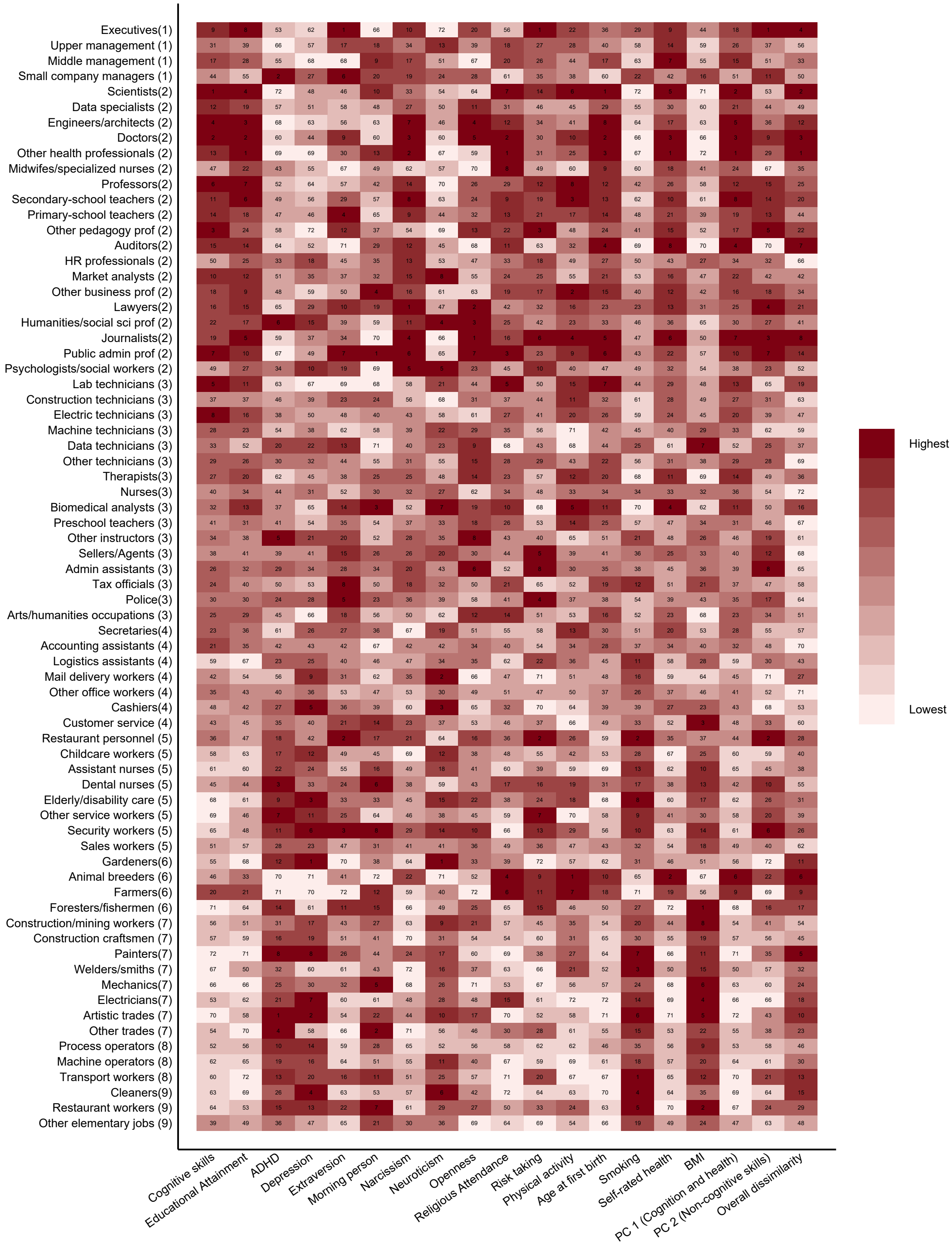
[4]Supplementary Figure S2 presents analogous analyses using a classification based on the second digit of the Swedish Educational Terminology. This classification is much less fine-grained – and consequently less informative – than the one used in Figure 2 that is based on four digit level information as much as possible. The advantage of the cruder classification used in Supplementary Figure S2 is that it does not involve researcher decisions on which categories to merge in case of small cell counts.

[5]Figure S5 in the appendix shows average values instead of ranks.

Figure 2: Differences in trait PGIs across study majors

Note: The figure presents a heat map of average standardized PGIs across education majors. PGIs are first corrected for socioeconomic background by regressing them on socioeconomic status, municipality dummies, the first 20 components of the genetic-relatedness matrix, and birth-year dummies interacted with gender. P-values indicate the statistical significance of the difference in average residualized PGIs across study majors (from OLS regressions of each PGI on study major dummies; standard errors clustered at the family level). The sample consists of genotyped individuals born from 1935 to 1994 (N=29,433).

Figure 3: Differences in trait PGIs across occupations



Note: The figure presents a heat map of average standardized PGIs across occupations. PGIs are first corrected for socioeconomic background by regressing them on socioeconomic status, municipality dummies, the first 20 components of the genetic-relatedness matrix, and birth-year dummies interacted with gender. P-values indicate the statistical significance of the difference in average residualized PGIs across study majors (from OLS regressions of each PGI on study major dummies; standard errors clustered at the family level). We added the first digit of the SSYK code to the labels. The first-digit level closely corresponds to the first digit of the International Standard Classification of Occupations (ISCO). Occupations are divided into nine broad categories: (1) Managers; (2) Professionals; (3) Technicians and associate professionals; (4) Clerical support workers; (5) Service and sales workers; (6) Skilled agricultural, forestry and fishery workers; (7) Craft and related trades workers; (8) Plant and machine operators, and assemblers; (9) Elementary occupations. The sample consists of genotyped individuals born from 1935 to 1988 (N=24,838).

workers, primary school teachers, and police (farmers, auditors, and gardeners rank lowest). The top of the openness ranking is dominated by humanities-oriented professions.[6]

These are just some examples. To get an overall view of occupational sorting on genes related to human capital, we again also map the two principal components and overall dissimilarity across occupations. On the first principal component (cognitive skills and health), health professionals, scientists, doctors, auditors, and engineers rank highest. People in artistic trades, painters, and transport workers rank lowest. Executives, restaurant personnel, journalists, lawyers, and pedagogy professionals rank highest on the second principal component (non-cognitive skills). Gardeners, mail delivery workers, and auditors rank lowest. The genetically most distinctive occupations according to our measure of overall dissimilarity are health professionals, scientists, doctors, executives, and painters.

In Figure 4, we visualize the genetic differences across study majors and occupations in a different way. The graphs show scatter plots of careers along the two principal components. The graphs therefore show how each study major and occupation is situated along the cognitive skills and health axis on the one hand (PC1) and the non-cognitive skills axis on the other hand (PC 2).

In terms of educational specializations, notable outliers include the fields of medicine and management (high on both components); police/security and hotels/restaurants (low on the first and high on the second component); science/math (high on the first and low on the second component); and media production and childcare (low on both components). That is, people who studied policing on average have many genetic variants that predict non-cognitive traits including extraversion and risk tolerance, but comparatively fewer that predict cognitive skills and health. The opposite is true for people who studied science or math. We see similar patterns for occupations: executives are high on both components; auditors are high on the first and low on the second; restaurant personnel are low on the first and high on the second; and gardeners are low on both. The shading on the markers in the scatter plots represents the dissimilarity of each career across all 16 PGIs. In general, careers towards the edges of the graphs are more dissimilar than the ones in the center, indicating that the two principal components succeed in capturing the genetic variation measured by the PGIs.

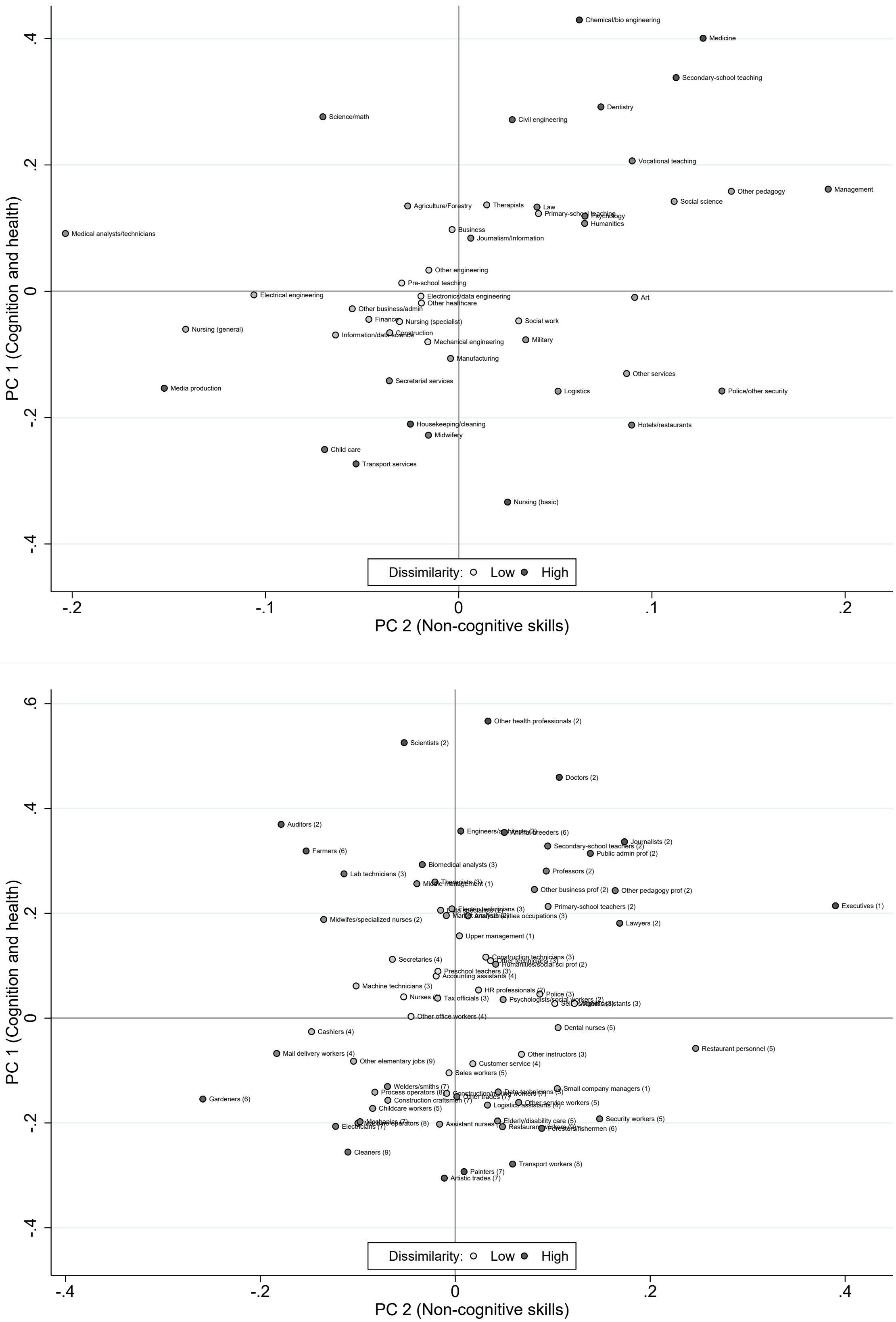
Discussion

We explore how genetic differences across individuals relate to their career choices. We use data from genotyped individuals in the Swedish Twin Registry (STR) that we link to government data on education major and occupation. The STR data contains a range of polygenic indices (PGIs) that summarize the presence of genetic variants associated with a given trait or behavior. We find 16 PGIs related to several aspects of human capital – cognition, personality, and (mental) health – that vary significantly across people who chose different education majors and occupations. We then present a detailed mapping and ranking of these genetic indices across occupations and study majors.

What to study and which professional career to pursue are two of the socially and economically most impactful decisions people must make in their lives. Our study reveals two main insights that can help us better understand who chooses which career and why. First, we show that – controlling for socioeconomic background and geographic origin – there are highly statistically significant genetic differences across people who choose different careers. Second, the ranking of careers differs strongly across genetic indices which summarize genetic variants associated with different traits. That is, different genes (that are associated with different traits) predict entry into different careers. Third, some occupations are nevertheless remarkable in being highly distinctive across several indices.

[6]Supplementary Figure S4 shows an analogous heatmap using the two-digit level of the SSYK to classify occupations. The main conclusion using the two-digit classification is the same: people who sort into different occupations differ genetically and different occupations show up at the extremes of different PGI rankings.

Figure 4: First two principal PGI components across study majors and occupations



References

Abdellaoui, Abdel, Jouke-Jan Hottenga, Peter De Knijff, Michel G Nivard, Xiangjun Xiao, Paul Scheet, Andrew Brooks, Erik A Ehli, Yueshan Hu, Gareth E Davies et al. 2013. "Population structure, migration, and diversifying selection in the Netherlands." *European Journal of Human Genetics* 21 (11):1277–1285.

Alan, Sule and Seda Ertac. 2018. "Fostering patience in the classroom: Results from randomized educational intervention." *Journal of Political Economy* 126 (5):1865–1911.

Almlund, Mathilde, Angela Lee Duckworth, James Heckman, and Tim Kautz. 2011. "Personality psychology and economics." In *Handbook of the Economics of Education*, vol. 4. Elsevier, 1–181.

Angerer, Silvia, Jana Bolvashenkova, Daniela Glätzle-Rützle, Philipp Lergetporer, and Matthias Sutter. 2021. "Children's Patience and School-Track Choices Several Years Later: Linking Experimental and Field Data." CESifo Working Paper.

Arcidiacono, P. 2004. "Ability sorting and the returns to college major." *Journal of Econometrics* 121 (1-2).

Barrick, MR, MK Mount, and R Gupta. 2003. "Meta-analysis of the relationship between the five-factor model of personality and Holland's occupational types." *Personnel Psychology* 56.

Becker, Joel, Casper AP Burik, Grant Goldman, Nancy Wang, Hariharan Jayashankar, Michael Bennett, Daniel W Belsky, Richard Karlsson Linner, Rafael Ahlskog, Aaron Kleinman et al. 2021. "Resource Profile and User Guide of the Polygenic Index Repository." *Nature Human Behaviour* :1–15.

Bellante, Don and Albert N Link. 1981. "Are public sector workers more risk averse than private sector workers?" *ILR Review* 34 (3):408–412.

Benjamin, Daniel J, James O Berger, Magnus Johannesson, Brian A Nosek, E-J Wagenmakers, Richard Berk, Kenneth A Bollen, Björn Brembs, Lawrence Brown, Colin Camerer et al. 2018. "Redefine statistical significance." *Nature Human Behaviour* 2 (1):6–10.

Berger, MC. 1988. "Predicted Future Earnings and Choice of College Major." *ILR Review* 41 (3).

Brenner, Steffen. 2015. "The risk preferences of US executives." *Management Science* 61 (6):1344–1361.

Buser, Thomas, Rafael Ahlskog, Magnus Johannesson, Philipp Koellinger, and Sven Oskarsson. 2021. "Using genes to explore the effects of cognitive and non-cognitive skills on education and labor market outcomes." Working paper.

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek. 2021. "Can competitiveness predict education and labor market outcomes? Evidence from incentivized choice and survey measures." *NBER Working Paper* (28916).

Cesarini, David, Christopher T Dawes, Magnus Johannesson, Paul Lichtenstein, and Björn Wallace. 2009. "Genetic variation in preferences for giving and risk taking." *The Quarterly Journal of Economics* 124 (2):809–842.

Chabris, Christopher F, James J Lee, David Cesarini, Daniel J Benjamin, and David I Laibson. 2015. "The fourth law of behavior genetics." *Current Directions in Psychological Science* 24 (4):304–312.

Cobb-Clark, DA and M Tan. 2011. "Noncognitive skills, occupational attainment, and relative wages." *Labour Economics* 18 (1).

Coenen, J, L Borghans, and R Diris. 2021. "Personality traits, preferences and educational choices: A focus on STEM." *Journal of Economic Psychology* 84.

Cutler, David M. and Adriana Lleras-Muney. 2010. "Understanding differences in health behaviors by education." *Journal of Health Economics* 29 (1):1–28. URL `http://dx.doi.org/10.1016/j.jhealeco.2009.10.003`.

Dawes, Christopher T, Aysu Okbay, Sven Oskarsson, and Aldo Rustichini. 2021. "A polygenic score for educational attainment partially predicts voter turnout." *Proceedings of the National Academy of Sciences* 118 (50):e2022715118.

Deming, David J. 2017. "The growing importance of social skills in the labor market." *The Quarterly Journal of Economics* 132 (4):1593–1640.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G Wagner. 2011. "Individual risk attitudes: Measurement, determinants, and behavioral consequences." *Journal of the European Economic Association* 9 (3):522–550.

Emmons, Robert A. 1987. "Narcissism: Theory and measurement." *Journal of Personality and Social Psychology* 52 (1):11.

Fletcher, JM. 2013. "The effects of personality traits on adult labor market outcomes: Evidence from siblings." *Journal of Economic Behavior & Organization* 89.

Fouarge, D, B Kriechel, and T Dohmen. 2014. "Occupation sorting of school graduates: The role of economic preferences." *Journal of Economic Behavior & Organization* 106.

Galama, Titus J, Adriana Lleras-Muney, and Hans Van Kippersluis. 2018. "The Effect of Education on Health and Mortality: A Review of Experimental and Quasi-Experimental Evidence." .

Golsteyn, Bart HH, Hans Grönqvist, and Lena Lindahl. 2014. "Adolescent time preferences predict lifetime outcomes." *The Economic Journal* 124 (580):F739–F761.

Ham, R, PN Junankar, and R Wells. 2009. "Occupational Choice: Personality Matters." IZA Discussion Paper No. 4105.

Hamer, D and L Sirota. 2000. "Beware the Chopsticks Gene." *Molecular Psychiatry* 5 (1):11–13.

Harden, K Paige and Philipp D Koellinger. 2020. "Using genetics for social science." *Nature Human Behaviour* 4 (6):567–576.

Hartwig, Fernando Pires, Neil Martin Davies, and George Davey Smith. 2018. "Bias in Mendelian randomization due to assortative mating." *Genetic Epidemiology* 42 (7):608–620.

Heckman, James J, Jora Stixrud, and Sergio Urzua. 2006. "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior." *Journal of Labor Economics* 24 (3):411–482.

Holland, JL. 1978. *Manual for the vocational preferences inventory.* Palo Alto CA: Consulting Psychologist Press.

———. 1997. *Making occupational choices: A theory of vocational personalities and work environments.* Englewood Cliffs, NJ: Prentice Hall.

Houmark, Mikkel, Victor Ronda, and Michael Rosholm. 2020. "The nurture of nature and the nature of nurture: How genes and investments interact in the formation of skills." IZA Discussion Paper.

Humburg, M. 2017. "Personality and field of study choice in university." *Education Economics* 25 (4).

Jang, Kerry L., W. John Livesley, and Philip A. Vemon. 1996. "Heritability of the Big Five Personality Dimensions and Their Facets: A Twin Study." *Journal of Personality* 64 (3):577–592. URL `http://dx.doi.org/10.1111/J.1467-6494.1996.TB00522.X`.

John, K and SL Thomsen. 2014. "Heterogeneous returns to personality: The role of occupational choice." *Empirical Economics* 47.

Kong, Augustine, Gudmar Thorleifsson, Michael L Frigge, Bjarni J Vilhjalmsson, Alexander I Young, Thorgeir E Thorgeirsson, Stefania Benonisdottir, Asmundur Oddsson, Bjarni V Halldorsson, Gisli Masson et al. 2018. "The nature of nurture: Effects of parental genotypes." *Science* 359 (6374):424–428.

Koudstaal, Martin, Randolph Sloof, and Mirjam Van Praag. 2016. "Risk, uncertainty, and entrepreneurship: Evidence from a lab-in-the-field experiment." *Management Science* 62 (10):2897–2915.

Kristof-Brown, AL, RD. Zimmerman, and ED Johnson. 2005. "Consequences of individuals' fit at work: A meta-analysis of person-job, person-organization, person-group, and person-supervisor fit." *Personnel Psychology* 58.

Kweon, Hyeokmoon, Casper Burik, Richard Karlsson Linnér, Ronald De Vlaming, Aysu Okbay, Daphne Martschenko, Kathryn Harden Harden, Thomas A DiPrete, and Philipp Koellinger. 2020. "Genetic Fortune: Winning or Losing Education, Income, and Health." Tinbergen Institute Discussion Paper.

Lichtenstein, Paul, Patrick F Sullivan, Sven Cnattingius, Margaret Gatz, Sofie Johansson, Eva Carlström, Camilla Björk, Magnus Svartengren, Alicja Wolk, Lars Klareskog et al. 2006. "The Swedish Twin Registry in the third millennium: an update." *Twin Research and Human Genetics* 9 (6):875–882.

Lindgren, Karl-Oskar, Sven Oskarsson, and Mikael Persson. 2019. "Enhancing electoral equality: can education compensate for family background differences in voting participation?" *American Political Science Review* 113 (1):108–122.

Lipson, Sarah Ketchen, Sasha Zhou, Blake Wagner, III, Katie Beck, and Daniel Eisenberg. 2015. "Major Differences: Variations in Undergraduate and Graduate Student Mental Health and Treatment Utilization Across Academic Disciplines." *Journal of College Student Psychotherapy* 30 (1):23–41. URL `http://dx.doi.org/10.1080/87568225.2016.1105657`.

Llena-Nozal, Ana, Maarten Lindeboom, and France Portrait. 2004. "The effect of work on mental health: does occupation matter?" *Health Economics* 13 (10):1045–1062. URL `http://dx.doi.org/10.1002/HEC.929`.

Lounsbury, John W, Eric D Sundstrom, Lucy W Gibson, James M Loveland, and Adam W Drost. 2016. "Core personality traits of managers." *Journal of Managerial Psychology* 31 (2):434–450.

McCrae, RR and PT Costa. 2008. "The Five-Factor Theory of Personality." In *Handbook of Personality: Theory and Research*. The Guilford Press.

Montez, Jennifer Karas, Wencheng Zhang, Anna Zajacova, and Tod G. Hamilton. 2018. "Does college major matter for women's and men's health in midlife? Examining the horizontal dimensions of educational attainment." *Social Science & Medicine* 198:130–138. URL `http://dx.doi.org/10.1016/j.socscimed.2018.01.005`.

Moutafi, Joanna, Adrian Furnham, and John Crump. 2007. "Is managerial level related to personality?" *British Journal of Management* 18 (3):272–280.

Nieken, P and P Störmer. 2010. "Personality as Predictor of Occupational Choice: Empirical Evidence from Germany." Working paper.

Okbay, Aysu, Jonathan P Beauchamp, Mark Alan Fontana, James J Lee, Tune H Pers, Cornelius A Rietveld, Patrick Turley, Guo-Bo Chen, Valur Emilsson, S Fleur W Meddens et al. 2016. "Genome-wide association study identifies 74 loci associated with educational attainment." *Nature* 533 (7604):539–542.

Plomin, Robert and Cindy S Bergeman. 1991. "The nature of nurture: Genetic influence on "environmental" measures." *Behavioral and Brain Sciences* 14 (3):373–386.

Price, Alkes L, Nick J Patterson, Robert M Plenge, Michael E Weinblatt, Nancy A Shadick, and David Reich. 2006. "Principal components analysis corrects for stratification in genome-wide association studies." *Nature Genetics* 38 (8):904–909.

Ravesteijn, Bastian, Hans van Kippersluis, and Eddy van Doorslaer. 2013. *The Contribution of Occupation to Health Inequality.* Emerald Group Publishing Limited, 311–332. URL `http://dx.doi.org/10.1108/S1049-2585(2013)0000021014`.

Rietveld, Cornelius A, Sarah E Medland, Jaime Derringer, Jian Yang, Tõnu Esko, Nicolas W Martin, Harm-Jan Westra, Konstantin Shakhbazov, Abdel Abdellaoui, Arpana Agrawal et al. 2013. "GWAS of 126,559 individuals identifies genetic variants associated with educational attainment." *Science* 340 (6139):1467–1471.

Rietveld, Cornelius A, Hans van Kippersluis, and A Roy Thurik. 2015. "Self-employment and health: Barriers or benefits?" *Health economics* 24 (10):1302–1313.

Romeis, James C, Jeffrey F Scherrer, Hong Xian, Seth A Eisen, Kathleen Bucholz, Andrew C Heath, Jack Goldberg, Michael J Lyons, William G Henderson, and William R True. 2000. "Heritability of self-reported health." *Health services research* 35 (5 Pt 1):995.

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):351–363.

Stinebrickner, Ralph and Todd R Stinebrickner. 2014. "A major in science? Initial beliefs and final outcomes for college major and dropout." *Review of Economic Studies* 81 (1):426–472.

Sutter, Matthias, Martin G Kocher, Daniela Glätzle-Rützler, and Stefan T Trautmann. 2013. "Impatience and uncertainty: Experimental decisions predict adolescents' field behavior." *American Economic Review* 103 (1):510–31.

Turkheimer, Eric. 2000. "Three laws of behavior genetics and what they mean." *Current Directions in Psychological Science* 9 (5):160–164.

Viinikainen, J, P Böckerman, M Elovainio, C Hakulinen, M Hintsanen, M Kähönen, J Pehkonen, L Pulkki-Råback, O Raitakari, and L. Keltikangas-Järvinen. 2020. "Personality, occupational sorting and routine work." *Employee Relations* 42 (6).

Volkers, Anita C, Gert P Westert, and Francois G Schellevis. 2007. "Health disparities by occupation, modified by education: a cross-sectional population study." *BMC Public Health* 7 (1). URL `http://dx.doi.org/10.1186/1471-2458-7-196`.

Vukasović, Tena and Denis Bratko. 2015. "Heritability of personality: A meta-analysis of behavior genetic studies." *Psychological Bulletin* 141 (4):769–785. URL `http://dx.doi.org/10.1037/bul0000017`.

Wiswall, M and B Zafar. 2015. "Determinants of College Major Choice: Identification using an Information Experiment." *The Review of Economic Studies* 82 (2).

Wiswall, Matthew and Basit Zafar. 2018. "Preference for the workplace, investment in human capital, and gender." *The Quarterly Journal of Economics* 133 (1):457–507.

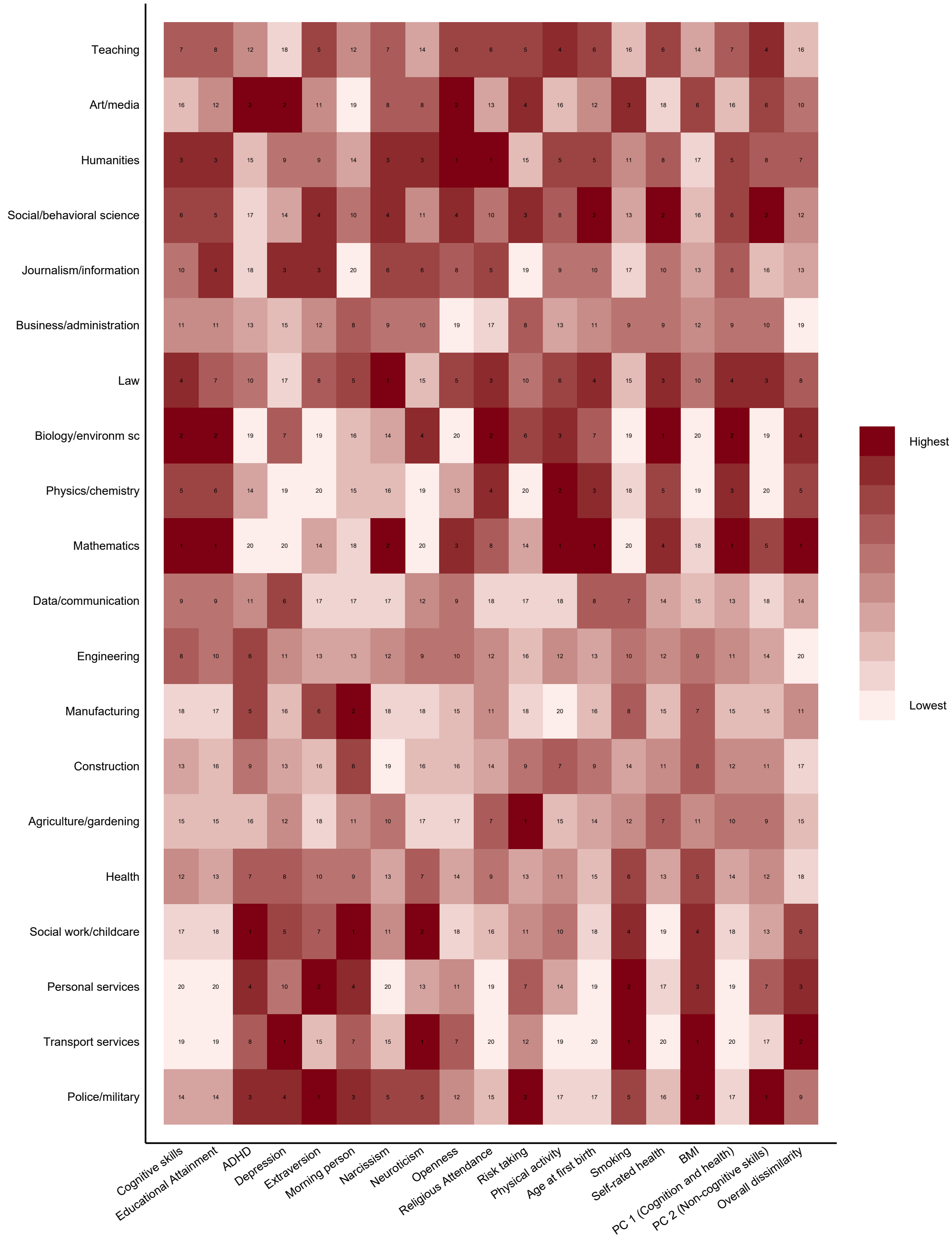
Zyphur, Michael J., Jayanth Narayanan, Richard D. Arvey, and Gordon J. Alexander. 2009. "The genetics of economic risk preferences." *Journal of Behavioral Decision Making* 22 (4):367–377. URL `http://dx.doi.org/10.1002/BDM.643`.

Supplementary Information

Figure S1: Pairwise correlations between PGIs

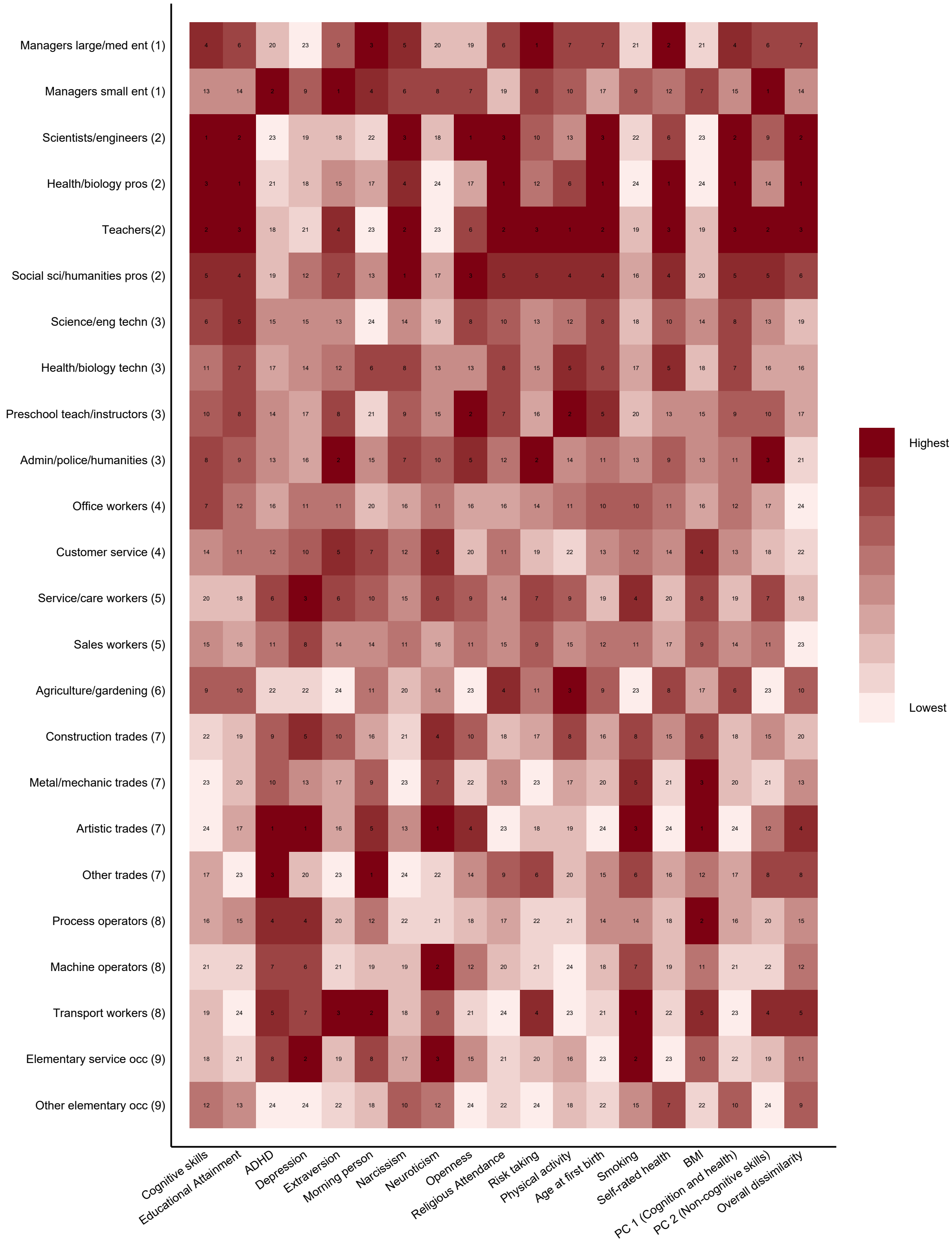
	Cognitive skills	Educational Attainment	ADHD	Depression	Extraversion	Morning person	Narcissism	Neuroticism	Openness	Religious Attendance	Risk taking	Physical activity	Age at first birth	Smoking	Self-rated health	BMI
Cognitive skills		0.42	-0.14	-0.14	-0.00	-0.06	0.10	-0.14	0.08	0.20	0.09	0.05	0.25	-0.13	0.20	-0.07
Educational Attainment	0.42		-0.23	-0.18	0.00	-0.05	0.17	-0.08	0.11	0.33	0.10	0.17	0.44	-0.27	0.34	-0.25
ADHD	-0.14	-0.23		0.19	0.11	0.01	0.01	0.09	0.05	-0.14	0.11	-0.07	-0.26	0.27	-0.22	0.19
Depression	-0.14	-0.18	0.19		-0.06	-0.07	0.03	0.51	0.06	-0.08	-0.07	-0.13	-0.21	0.23	-0.39	0.16
Extraversion	-0.00	0.00	0.11	-0.06		0.04	0.08	-0.13	0.24	-0.01	0.27	0.08	-0.04	0.05	0.05	0.05
Morning person	-0.06	-0.05	0.01	-0.07	0.04		-0.05	-0.07	-0.06	0.00	0.05	0.08	-0.07	-0.00	0.14	-0.01
Narcissism	0.10	0.17	0.01	0.03	0.08	-0.05		0.05	0.09	0.06	0.16	0.06	0.08	0.04	0.10	-0.09
Neuroticism	-0.14	-0.08	0.09	0.51	-0.13	-0.07	0.05		-0.03	-0.04	-0.18	-0.06	-0.11	0.11	-0.24	0.03
Openness	0.08	0.11	0.05	0.06	0.24	-0.06	0.09	-0.03		0.03	0.19	0.04	0.03	0.03	0.00	0.01
Religious Attendance	0.20	0.33	-0.14	-0.08	-0.01	0.00	0.06	-0.04	0.03		-0.07	0.08	0.26	-0.19	0.19	-0.13
Risk taking	0.09	0.10	0.11	-0.07	0.27	0.05	0.16	-0.18	0.19	-0.07		0.17	-0.02	0.10	0.16	0.02
Physical activity	0.05	0.17	-0.07	-0.13	0.08	0.08	0.06	-0.06	0.04	0.08	0.17		0.14	-0.13	0.29	-0.21
Age at first birth	0.25	0.44	-0.26	-0.21	-0.04	-0.07	0.08	-0.11	0.03	0.26	-0.02	0.14		-0.28	0.33	-0.27
Smoking	-0.13	-0.27	0.27	0.23	0.05	-0.00	0.04	0.11	0.03	-0.19	0.10	-0.13	-0.28		-0.32	0.23
Self-rated health	0.20	0.34	-0.22	-0.39	0.05	0.14	0.10	-0.24	0.00	0.19	0.16	0.29	0.33	-0.32		-0.43
BMI	-0.07	-0.25	0.19	0.16	0.05	-0.01	-0.09	0.03	0.01	-0.13	0.02	-0.21	-0.27	0.23	-0.43	

Figure S2: Differences in trait PGIs across study majors (2-digit level classification)



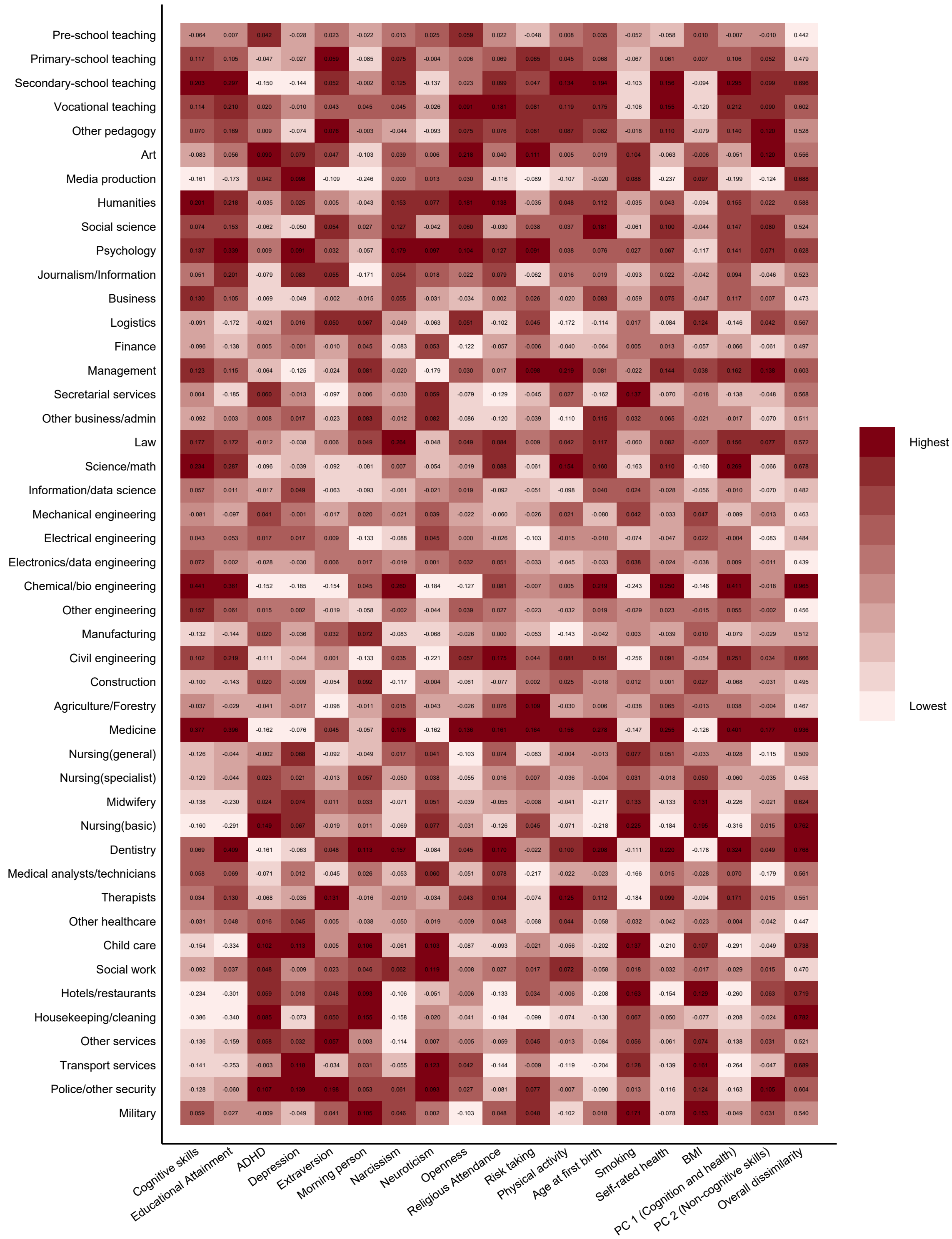
Note: The figure presents a heat map of average standardized PGIs across education majors. PGIs are first corrected for socioeconomic background by regressing them on socioeconomic status, municipality dummies, the first 20 components of the genetic-relatedness matrix, and birth-year dummies interacted with gender. P-values indicate the statistical significance of the difference in average residualized PGIs across study majors (from OLS regressions of each PGI on study major dummies; standard errors clustered at the family level). The sample consists of genotyped individuals born from 1935 to 1994.

Figure S4: Differences in trait PGIs across occupations (2-digit level classification)



Note: The figure presents a heat map of average standardized PGIs across occupations. PGIs are first corrected for socioeconomic background by regressing them on socioeconomic status, municipality dummies, the first 20 components of the genetic-relatedness matrix, and birth-year dummies interacted with gender. P-values indicate the statistical significance of the difference in average residualized PGIs across study majors (from OLS regressions of each PGI on study major dummies; standard errors clustered at the family level). We added the first digit of the SSYK code to the labels. The first-digit level closely corresponds to the first digit of the International Standard Classification of Occupations (ISCO). Occupations are divided into nine broad categories: (1) Managers; (2) Professionals; (3) Technicians and associate professionals; (4) Clerical support workers; (5) Service and sales workers; (6) Skilled agricultural, forestry and fishery workers; (7) Craft and related trades workers; (8) Plant and machine operators, and assemblers; (9) Elementary occupations. The sample consists of genotyped individuals born from 1935 to 1988.

Figure S3: Differences in trait PGIs across study majors

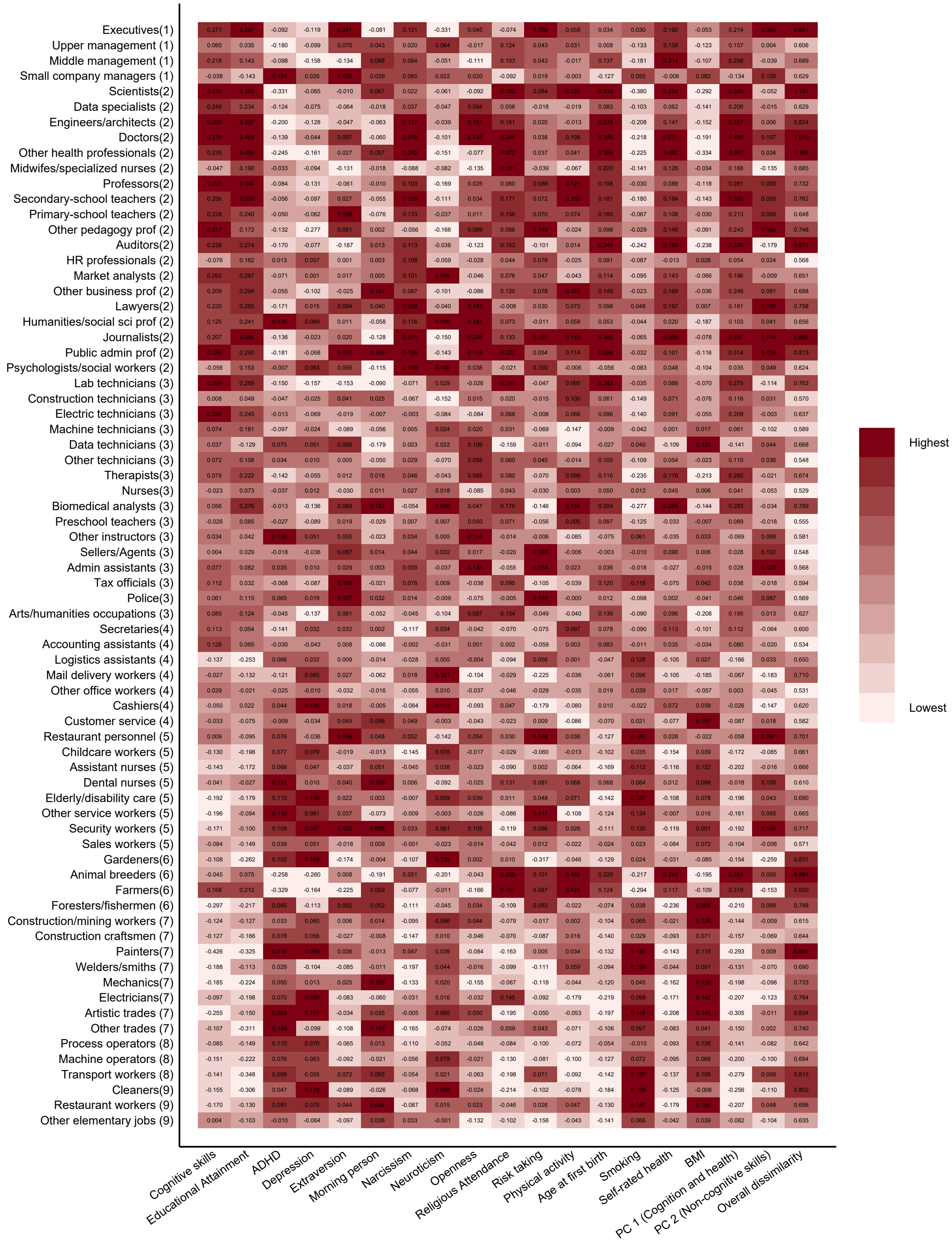


Note: The figure presents a heat map of average standardized PGIs across education majors. PGIs are first corrected for socioeconomic background by regressing them on socioeconomic status, municipality dummies, the first 20 components of the genetic-relatedness matrix, and birth-year dummies interacted with gender. P-values indicate the statistical significance of the difference in average residualized PGIs across study majors (from OLS regressions of each PGI on study major dummies; standard errors clustered at the family level). The sample consists of genotyped individuals born from 1935 to 1994 (N=29,433).

Table S1: Rotated factor loadings

PGI	Component 1	Component 2
Cognitive performance	0.2585	
Educational attainment	0.3815	
ADHD	-0.2874	
Depression	-0.2838	
Extraversion		0.4914
Monring person		
Narcissism		
Neuroticism		-0.3031
Openness		0.3445
Religious Attendance	0.2611	
Risk taking		0.5520
Activity		
Age at first birth	0.3727	
Smoking	-0.3253	
Self-rated health	0.3874	
BMI	-0.3021	

Figure S5: Differences in trait PGIs across occupations



Note: The figure presents a heat map of average standardized PGIs across occupations. PGIs are first corrected for socioeconomic background by regressing them on socioeconomic status, municipality dummies, the first 20 components of the genetic-relatedness matrix, and birth-year dummies interacted with gender. P-values indicate the statistical significance of the difference in average residualized PGIs across study majors (from OLS regressions of each PGI on study major dummies; standard errors clustered at the family level). We added the first digit of the SSYK code to the labels. The first-digit level closely corresponds to the first digit of the International Standard Classification of Occupations (ISCO). Occupations are divided into nine broad categories: (1) Managers; (2) Professionals; (3) Technicians and associate professionals; (4) Clerical support workers; (5) Service and sales workers; (6) Skilled agricultural, forestry and fishery workers; (7) Craft and related trades workers; (8) Plant and machine operators, and assemblers; (9) Elementary occupations. The sample consists of genotyped individuals born from 1935 to 1988 (N=24,838).

Table S2: Our classification of study majors

	N	Percent female	Precent higher education
Pre-school teaching	857	0.88	1.00
Primary-school teaching	601	0.81	1.00
Secondary-school teaching	548	0.54	1.00
Vocational teaching	424	0.55	1.00
Other pedagogy	374	0.76	0.95
Art	585	0.60	0.42
Media production	296	0.53	0.40
Humanities	459	0.62	0.99
Social science	743	0.52	1.00
Psychology	224	0.69	1.00
Journalism/Information	249	0.68	0.99
Business	1873	0.52	0.60
Logistics	247	0.57	0.36
Finance	837	0.69	0.02
Management	224	0.59	0.98
Secretarial services	508	0.89	0.25
Other business/admin	358	0.62	0.45
Law	333	0.45	1.00
Science/math	419	0.49	1.00
Information/data science	492	0.34	0.83
Mechanical engineering	1631	0.08	0.36
Electrical engineering	724	0.07	0.40
Electronics/data engineering	776	0.09	0.49
Chemical/bio engineering	216	0.41	0.87
Other engineering	1647	0.10	0.37
Manufacturing	439	0.41	0.21
Civil engineering	279	0.50	0.89
Construction	1043	0.08	0.33
Agriculture/Forestry	687	0.33	0.29
Medicine	402	0.50	1.00
Nursing(general)	538	0.84	1.00
Nursing(specialist)	1234	0.86	0.70
Midwifery	755	0.96	0.01
Nursing(basic)	622	0.87	0.25
Dentistry	117	0.62	0.97
Medical analysts/technicians	249	0.76	0.88
Therapists	393	0.83	0.96
Other healthcare	346	0.89	0.55
Child care	709	0.79	0.13
Social work	602	0.83	0.83
Hotels/restaurants	479	0.67	0.05
Housekeeping/cleaning	256	0.93	0.01
Other services	444	0.79	0.40
Transport services	383	0.28	0.23
Police/other security	317	0.16	0.83
Military	170	0.06	0.94

Table S3: Our classification of occupations

	N	Percent female	Precent higher education
Executives (1)	96	0.15	0.51
Upper management (1)	174	0.39	0.75
Middle management (1)	266	0.24	0.66
Small company managers (1)	308	0.40	0.32
Scientists (2)	75	0.57	0.95
Data specialists (2)	544	0.21	0.83
Engineers/architects (2)	478	0.27	0.94
Doctors (2)	220	0.45	0.99
Other health professionals (2)	124	0.60	0.99
Midwifes/specialized nurses (2)	252	0.86	1.00
Professors (2)	194	0.39	0.99
Secondary-school teachers (2)	554	0.56	0.94
Primary-school teachers (2)	576	0.75	0.97
Other pedagogy prof (2)	101	0.62	0.86
Auditors (2)	173	0.48	0.86
HR professionals (2)	149	0.49	0.73
Market analysts (2)	132	0.45	0.80
Other business prof (2)	213	0.36	0.86
Lawyers (2)	112	0.41	0.96
Humanities/social sci prof (2)	220	0.53	0.78
Journalists (2)	173	0.55	0.72
Public admin prof (2)	306	0.53	0.82
Psychologists/social workers (2)	276	0.70	0.92
Lab technicians (3)	164	0.60	0.67
Construction technicians (3)	339	0.07	0.50
Electric technicians (3)	309	0.09	0.47
Machine technicians (3)	406	0.06	0.42
Data technicians (3)	181	0.44	0.45
Other technicians (3)	560	0.20	0.50
Therapists (3)	360	0.82	0.91
Nurses (3)	742	0.91	0.99
Biomedical analysts (3)	75	0.95	0.96
Preschool teachers (3)	722	0.90	0.97
Other instructors (3)	245	0.64	0.73
Sellers/Agents (3)	1207	0.34	0.44
Admin assistants (3)	230	0.69	0.70
Tax officials (3)	158	0.72	0.47
Police (3)	181	0.17	0.97
Arts/humanities occupations (3)	176	0.39	0.52
Secretaries (4)	291	0.94	0.32
Accounting assistants (4)	420	0.86	0.30
Logistics assistants (4)	376	0.20	0.21
Mail delivery workers (4)	172	0.38	0.30
Other office workers (4)	1201	0.88	0.23
Cashiers (4)	215	0.79	0.26
Customer service (4)	310	0.83	0.35
Restaurant personnel (5)	288	0.72	0.28
Childcare workers (5)	826	0.93	0.27
Assistant nurses (5)	1234	0.93	0.19
Dental nurses (5)	113	0.99	0.22
Elderly/disability care (5)	1034	0.83	0.34
Other service workers (5)	206	0.81	0.18
Security workers (5)	229	0.25	0.37
Sales workers (5)	921	0.70	0.29
Gardeners (6)	111	0.26	0.18
Animal breeders (6)	69	0.45	0.20
Farmers (6)	316	0.33	0.14
Foresters/fishermen (6)	109	0.04	0.06
Construction/mining workers (7)	594	0.01	0.08
Construction craftsmen (7)	580	0.06	0.10
Painters (7)	171	0.08	0.09
Welders/smiths (7)	286	0.04	0.06
Mechanics (7)	323	0.05	0.06
Electricians (7)	188	0.06	0.15
Artistic trades (7)	149	0.34	0.15
Other trades (7)	114	0.35	0.07
Process operators (8)	296	0.11	0.11
Machine operators (8)	1173	0.30	0.12
Transport workers (8)	793	0.14	0.08
Cleaners (9)	458	0.91	0.09
Restaurant workers (9)	303	0.89	0.24
Other elementary jobs (9)	332	0.34	0.21