How Genome-Wide Association Studies on Educational Attainment Reify Eugenic Ideologies

Dawn Warfield*

Mentor: Dr. Emily Merchant
Science and Technology Studies, UC Davis

Abstract

Genomic science and data are becoming ubiquitously entwined with how we describe the human condition in its physical, psychological, and social expression. Behavioral heredity research delves into genetic data science to reveal possible conditions, correlates, and determinants for particular behaviors and aptitudes that shape our individual and collective human experience. This article surveys the history, research culture, and methods of behavioral heredity science to investigate the techno-social interaction and motivations for recent genome-wide association studies (GWAS) performed to identify genomic biomarkers associated with educational attainment.

Introduction

The knowledge regime of behavioral heredity began in the late 19th century with Mendelian inference and extrapolated findings from animal breeding. As the technology of molecular biology enabled new understandings about the relationship between the heritability of traits and medical conditions through genes, social scientists began to explore genetics as a means to biologically explain human behavior and socioeconomic potential. In the most recent iteration of behavioral heredity, sociogenomic researchers now have access to both genomic and empirical survey data derived from longitudinal studies with broadening population samples, engendering new possibilities for finding associations.

A Historical Perspective on Sociogenomics

The longer history of sociogenomics begins with Francis Galton, who coined the word "eugenics," literally meaning "well borne" in Greek, in 1883 (Müller-Wille and Rheinberger). Galton's belief that human society could be perfected through selective breeding and social divestment of those considered to be a threat to the gene pool proved popular among scientists at the turn of the twentieth century. Eugenics had a synergistic relationship with the era's scientific racism, epitomized in Harvard anthropologist Louis Agassiz, who insisted that black brains had lower capacities and that educating them could cause brain damage or even death (Beckwith). Eugenics gained momentum to shape US immigration policy. Hereditarian studies purporting that immigration would pollute white progeny with deleterious results like lower birth rates, worse infant health, and impaired intelligence held sway in enacting public policies that restricted immigration from non-northern European countries (Ngai). From 1907 until 1979, over 30 states passed eugenic sterilization laws and about 60,000 procedures were done (Stern). The 1927 Supreme Court case *Buck v. Bell* added legitimacy to eugenics. The forced sterilization of poor, black, and immigrant women was often rationalized by "feeblemindedness," a slippery category that claimed a basis in heredity (Stern). The popularity of eugenics waned after WWII, when it became indelibly associated with Nazism, but it didn't exactly disappear. To recover the legitimacy of the eugenic project, its American supporters turned their attention to medical genetics.

Beginning in the middle of the twentieth century, molecular biology engendered new technologies for empirical study of the gene. These empirical tools, however, did not make their way into behavior genetics, which emerged in the 1960s to quantify the "heritability" of given social traits and outcomes through phenotype analysis in twins (Panofsky). As American society underwent great changes in the late 1960s and early 1970s, marked by second wave feminism and the civil rights movement, supporters of the status quo cited research in behavior genetics to assert that inequality of socio-economic outcomes was rooted in biological differences in ability. In 1969, educational psychologist Arthur Jensen extrapolated from his research to claim that programs like Head Start would never overcome IQ differences between white and black children because the majority of the difference was genetic (Jensen). In 1994, the controversial book "The Bell Curve," written by political scientist Charles Murray and psychologist Richard Hernstein, purported that there is great disparity between the races in their biologically determined abilities. Murray and Heinstein utilized biological determinism to provide rationale for fiscal conservatives to argue for the divestment of social welfare programs that they claim encourage low IQ women to breed and be supported by society (Hernstein and Murray). Behavior heredity also sought evidence of genetic determinism to explain gender inequalities. In 1979 educational psychologists Camilla Benbow and Julian Stanley hypothesized "superior male mathematical ability, which may in turn be related to greater male ability in spatial tasks" (Benbow and Stanley). This claim received considerable media coverage (Williams and King). Supporters saw in Benbow and Stanley's work a "math gene" that men had and women did not, while critics pointed to differences in exposure to math between girls and boys (Beckwith).

In the 1990s, the Human Genome Project (HGP) began with the aim to fully sequence the human genome. Supporters expected that the project would produce the key to all things human, but instead it revealed further complexity with huge spectrums of possibilities for genomic contributions and environmental adaptations (Reardon). The epic scale of the project catalyzed an inter-institutional cooperative endeavor, with numerous public and private investors investing millions of dollars into equipment and analysis, but researchers grappled with how to generate a return on investment from the mass of information the HGP produced (Reardon). The HGP made the aspirational and inclusive discovery that humans are all more than 99% the same, but researchers immediately turned to the differences among us, spawning new projects like the Human Genome Diversity Project (HGDP), to probe those differences (Reardon). What the HGP and HGDP did effectively was drive researchers to gather big data on human genetic variation and forge techniques for making knowledge of the genome more functionally predictive. Genomic data, and not hereditarian inference, would prove to be a great advancement in understanding human variation, and subsequently utilized by behavior heredity science to biologically explain how genes contribute to human performance.

GWAS and PGS

Since the completion of the HGP, the promise of genomic data has compelled social scientists to collect genetic samples along with surveys of population outcomes, enabling the current era of data-driven sociogenomics and genome-wide association studies (GWAS). GWAS originated in medical studies, where the technique has succeeded in identifying regions of the genome with genes that have causal effects on health. This research strategy focuses on the less than 1% of the human genome that is known to differ from person to person—single nucleotide polymorphisms—or SNPs for short. GWAS correlate SNPs from across the genome to phenotypes of interest in large sample populations. This strategy has been very successful in expanding our understanding of diabetes and identifying the BRCA gene associated with breast cancer (Billings and Florez; Kraft and Haiman).

Social scientists use GWAS to correlate SNPs with social traits or outcomes. The strength of GWAS SNP correlations is highly dependent on having huge genetic samples with information on measurable traits that are of interest to the researcher. Because GWAS requires samples of 100,000 or more, they are typically done by consortia, which meta-analyze findings from separate studies spanning the social, medical, and commercial realms. The studies of educational attainment that are the focus of this article were all done by the Social Science Genetic Association Consortium (SSGAC), which was founded in 2011 by a group of economists. In addition to promoting research, the SSGAC works to build the field of sociogenomics through workshops for postdocs and young faculty members at the Russell Sage Foundation's Summer Institute for Social-Science Genomics.
Between 2013 and 2018, The SSGAC produced three GWAS to identify SNP correlations for educational attainment. These studies focused on educational attainment, as this was a phenotype that was consistently available across all samples, but they treat educational attainment as a proxy for socioeconomic success, which was the original target of eugenics. As these studies grew in size from just over 100,000 participants in 2013 to over 1.1 million in 2018, the number of SNPs found to be statistically associated with educational attainment expanded from 3 to 1,271. The SSGAC used the findings from these GWAS to construct polygenic scores (PGS), which quantify an individual's genetic association with a measurable phenotype expression, in this case educational attainment. The headlines were flashy, announcing that more than a thousand genetic loci are related to educational

attainment (Rayner et al.), but the fine print was less exciting. The SNPs with the strongest association corresponded to a difference of about three weeks of additional education (Alonso). Summed into a PGS, these findings could explain only about 12% of the variance in a sample of European Americans, and very little of the variance among nonwhites (Okbay et al.). Subsequent research found that parental income was a better predictor of educational attainment than the PGS for educational attainment (Martin). On an individual level, the PGS is not useful for predicting the educational attainment of any given person (Zimmer).

Like earlier iterations of behavioral heredity science, the inclusion of nonwhites in the process of establishing normative genomic metrics – or in this case, their non-inclusion – continues to be problematic. Informed by the field of population genetics, GWAS are typically done on samples that are homogeneous in terms of genetic ancestry. All of the educational attainment GWAS were limited to people of European descent, and efforts to apply their findings to nonwhites have largely failed. The GWAS catalogue reports that 79% of all GWAS participants are of self-reported European descent, even though they make up only 16% of the global population (Martin et al.). Further complicating the issue, geneticist have found that studies on Hispanic/Latino and African American individuals show that these populations "contribute an outsized number of associations relative to studies of similar sizes in Europeans" (Martin et al.). Inclusive studies would have to be done on large samples of people of color in order to generate comparable results, and large-enough samples currently don't exist ("Genetics for All"). Therefore, whatever these studies identify as optimal genomic metrics for educational attainment are modeled by and only relevant to homogenous white populations.

Social Implications

Scientific findings don’t need to be substantiated to take hold in society. Let us reconsider Benbow and Stanley’s “math gene.” Media coverage of this spurious finding affected parental perceptions and expectations about the mathematical abilities of their children. A subsequent study by Jaqueline Eccles and Janis Jacobs in 1995 on the effect of exposure to the math gene media found that fathers exposed to misinformation felt confirmed that math wasn't as important for their girls as it was their boys and continued and/or lowered their expectations for their daughters' math ability. Mothers absorbed the information personally, believing they weren't as good at math and/or as intelligent as men and projected these beliefs onto their daughters (Eccles and Jacobs). Research in social psychology has shown that women will do less well on a math test if they are reminded of their ‘womanness’ before taking the test, simply by being asked to mark ‘M’ for male or ‘F’ for female on the test, or by being in a room full of men, whereas men will do better on the test if they are reminded of their "maleness" (Fine).

In review of the history of behavioral heredity science, a few patterns emerge. Claims of behavioral heredity are often used by political agents who wish to reinforce negative stereotypes of non-whites, women, and the poor in order to support policy arguments for divestment from social welfare programs that are implemented as safety nets and to mediate equity of opportunity. Genomic studies surface new claims of determinative genetics and the media embeds them into the communal social ethos. Once an idea is popularized, it's very difficult to get rid of. For example, the Graduate Records Exams (GRE) are known to have very little power to assess student intelligence or predict how well you will do in graduate school and the College Board isn't even a formal educational or governmental institution—but there continues to be a growing industry that's literally banking on the fears of students that their GRE score determine their future success.

The inference that genomic associations can determine socio-economic outcomes fuels the development of new products for people seeking to explain their behavior through science, gain a competitive edge, or even select mates that will produce desired traits in their progeny (Conley and Fletcher). At-home DNA test companies share your data with genomic consortiums to advance the field of genomic prediction, which helps to create new curious customers and products. There are new products in development seeking to capitalize on the idea of genetic prediction, like EduCred, which is positioning itself to apply the model of utilizing genetic data for precision medicine to create products and methods for "precision academics" (EduCred). There are institutional, financial, and political stakeholders seeking gains through genomic data, making the temptation of the low-hanging fruit of philosophical fallacy a real peril.

Conclusion

Philosopher John Dewey warned that "the most pervasive fallacy of philosophical thinking goes back to neglect of context" (Dewey). In light of our current political climate and finding ourselves on the precipice of CRISPR-enabled genomic interventions, George Pappas's analysis of John Dewey's ethics illuminates four aptly important philosophical fallacies (Pappas). The first, "The Analytic Fallacy," is where results are interpreted as complete within themselves (Pappas). Polygenic scores should not stand alone to predictively determine an individual's abilities or potential. The second, "Unlimited Universalization," is when context is ignored and conclusions are elevated to imply universal application (Pappas). Polygenic scores do not include people of color and are thus far too limited in their scope and specificity to be meaningfully applied to entire populations. Media that reports findings as if they apply to all with little attention to the fine print can mislead the public. Pappas's third fallacy, "Selective Emphasis," happens when the selectivity and purpose of selection is ignored in the context of an inquiry (Pappas). The limitations of the SSGAC's samples and methodological strategies, as well as their research motivations, are not wholly transparent in their articles. Their studies and training activities are funded by organizations including the Russell Sage Foundation and the Ragnar Söderberg Foundation, which provide capital to support their social policy aims. This information provides important contextualization for their scientific deliverables. The last is "The Intellectualism Fallacy," a combination of the previous three fallacies, where the pre-cognitive and non-cognitive are dismissed (Pappas). Dewey advises us to remember that "the world in which we immediately live, that in which we strive, succeed, and are defeated is preeminently a qualitative world" (Dewey). When researchers attempt to quantify that which is qualitative, like home environment and educational differences, and merely plug in numbers to represent abbreviated personal histories, they diminish the role of environment and exaggerate the role of biological determinism.

Context is extremely important when gauging whether the risks are worth the gains when undertaking studies and disseminating claims. Given that PGS on educational attainment are only derived from populations with European origins, and that the scholars who perform GWAS are not following up by investigating how particular genomic loci are associated with learning outcomes, it becomes quite credible to question the purposes and benefits of GWAS on educational attainment. The PGS does not provide medical or social information towards ameliorating educational deficits. From a practical standpoint, it would follow that institutions interested in improving learning outcomes could support social welfare and education programs with the resources and capital spent on crafting normative polygenic scores that have little individual predictive value. Behavior heredity science has fallen prey to all of the above pitfalls throughout its history, and sociogenomics must be conscientious and clearly acknowledge and communicate the limitations and motivations of genomic data analysis performed to derive predictors of socio-economic outcomes.

The future of sociogenomics is rife with potential medical, social, and political implications (Bliss, Social by Nature). Genetic technology is currently ahead of ethical policy development. Sociogenomics may contribute to racialization and genetic stereotyping. Polygenic scores could become a new metric for societal marginalization and increase stress on currently marginalized sectors if they are used to provide rationale for allocating a greater share of resources to people who are already advantaged. Sociogenomics has the potential to impact the personal identity and psychological well-being of people diagnosed with genetically determined traits. Being labeled with scores may inculcate ideas about genetic superiority and inferiority and cause people to think they are either better than others, or that they can't learn or achieve beyond the genomic potentials ascribed to them (Bliss, Race Decoded). There is clearly a need for further research on genome-wide association studies, including: 1) the history of the field's cultural and methodological development through eugenics and behavior genetics and assumptions that may be embedded within methods for constructing polygenic scores that predict educational attainment and economic success; 2) the effects of transmission and re-situation of behavioral genomic correlations from research settings into knowledge and products for the public, and; 3) the personal and sociopolitical benefits and implications of polygenic scores that claim to predict an individual or group's educational potential.

References

Alonso, Jose R. "Genetics of Educational Attainment." *Mapping Ignorance*, https://mappingignorance.org/2018/09/26/genetics-of-educational-attainment/. Accessed 24 Nov. 2019.

Beckwith, Jon. "Gender and Math Performance: Does Biology Have Implications for Educational Policy?" *The Journal of Education*, vol. 165, no. 2, 1983, pp. 158–74, doi:10.2307/42772831.

Benbow, C., and J. Stanley. "Sex Differences in Mathematical Ability: Fact or Artifact?" *Science*, vol. 210, no. 4475, Dec. 1980, pp. 1262–64, doi:10.1126/science.7434028.

Billings, Liana K., and Jose C. Florez. "The Genetics of Type 2 Diabetes: What Have We Learned from GWAS?" *Annals of the New York Academy of Sciences*, vol. 1212, no. 1, Nov. 2010, pp. 59–77, doi:10.1111/j.1749-6632.2010.05838.x.

Bliss, Catherine. *Race Decoded : The Genomic Fight for Social Justice - 01UCD*. Stanford University Press, 2012, https://search.library.ucdavis.edu/primo-explore/fulldisplay?docid=01UCD_ALMA21192436640003126&context=L&vid=01UCD_V1&search_scope=everything_scope&isFrbr=true&tab=default_tab&lang=en_US.
---. *Social by Nature: The Promise and Peril of Sociogenomics*. 2018.

Conley, Dalton, and Jason Fletcher. *The Genome Factor : What the Social Genomics Revolution Reveals about Ourselves, Our History, and the Future - 01UCD*. Princeton University Press, 2017, https://search.library.ucdavis.edu/primo-explore/fulldisplay?docid=01UCD_ALMA21283994450003126&context=L&vid=01UCD_V1&search_scope=everything_scope&tab=default_tab&lang=en_US.

Dewey, John. *The Later Works, 1925-1953*. Edited by Jo Ann Boydston, Carbondale : Southern Illinois University Press, 1981.

EduCred. *What Precisely IS Precision Academics? — EduCred Services*. http://educredservices.com/blog/2017/4/27/what-precisely-is-precision-academics. Accessed 10 Nov. 2019.

Fine, Cordelia. *Delusions of Gender : How Our Minds, Society, and Neurosexism Create Difference*. 1st ed.., New York : W. W. Norton, 2010.

"Genetics for All." *Nature Genetics*, vol. 51, no. 4, 2019, p. 579, doi:10.1038/s41588-019-0394-y.

Jensen, Arthur. "How Much Can We Boost IQ and Scholastic Acheivement?" *Harvard Educational Review*, edited by Arthur Jensen, vol. 39, no. 1, 1969, pp. 1–123, doi:10.17763/haer.39.1.l3u15956627424k7.

Kraft, Peter, and Christopher A. Haiman. "GWAS Identifies a Common Breast Cancer Risk Allele among BRCA1 Carriers." *Nature Genetics*, vol. 42, no. 10, Nature Publishing Group, 2010, pp. 819–20, doi:10.1038/ng1010-819.

Martin, Alicia R., et al. "Clinical Use of Current Polygenic Risk Scores May Exacerbate Health Disparities." *Nature Genetics*, vol. 51, no. 4, Nature Publishing Group, Apr. 2019, pp. 584–91, doi:10.1038/s41588-019-0379-x.

Martin, Nicholas. "Getting to the Genetic and Environmental Roots of Educational Inequality." *Npj Science of Learning*, vol. 3, no. 1, Nature Publishing Group, Dec. 2018, p. 4, doi:10.1038/s41539-018-0021-1.

Müller-Wille, Staffan, and Hans-Jörg. Rheinberger. *A Cultural History of Heredity*.

Ngai, Mae M. "The Architecture of Race in American Immigration Law: A Reexamination of the Immigration Act of 1924." *The*

Journal of American History, vol. 86, no. 1, 1999, pp. 67–92, doi:10.2307/2567407.

Okbay, Aysu, et al. “Genome-Wide Association Study Identifies 74 Loci Associated with Educational Attainment.” *Nature*, vol. 533, no. 7604, Nature Publishing Group, May 2016, pp. 539–42, doi:10.1038/nature17671.

Pappas, Gregory Fernando. *John Dewey’s Ethics : Democracy as Experience*. Bloomington : Indiana University Press, 2008.
Rayner, Christopher, et al. “Genetic Influences on Treatment-Seeking for Common Mental Health Problems in the UK Biobank.” *Behaviour Research and Therapy*, Pergamon, June 2019, p. 103413, doi:10.1016/J.BRAT.2019.103413.

Reardon, Jenny. *The Postgenomic Condition*. University of Chicago Press, 2017, doi:10.7208/chicago/9780226345192.001.0001.

Stern, Alexandra Minna. “That Time The United States Sterilized 60,000 Of Its Citizens | HuffPost.” *Zocolo Public Square*, 2016.

Williams, Dennis A., and Patricia King. “Do Males Have a Math Gene?” *Newsweek*, 1980, p. 73, http://beck2.med.harvard.edu/week6/Newsweek Time.pdf.

Zimmer, Carl. “Years of Education Influenced by Genetic Makeup, Enormous Study Finds.” *The New York TImes*, 23 July 2018, https://www.nytimes.com/2018/07/23/science/genes-education.html.