

---

## 14. Sociogenomics: theoretical and empirical challenges of integrating molecular genetics into sociological thinking\*

Melinda C. Mills

---

### 1. INTRODUCTION

As of November 1, 2020, a total of 4,761 discoveries have been made about the genetic basis of many traits, ranging from when one has their first child, predilection to caffeine consumption to type 2 diabetes (Mills & Rahal 2020). With the advent of genome-wide association studies (GWAS) in the early 2000s, we have moved from knowing that a particular outcome might have a 10 or 20 percent genetic basis to isolating the actual genetic loci and their potential biological function. A GWAS is a hypothesis-free search across the entire genome, examining each genetic locus (or region) one by one to see if there is a statistical relationship (association) with a particular outcome. This burst of knowledge has been a revelation, but also a challenge and opportunity for sociological and social science thinking.

In the past decade the sub-discipline of sociogenomics has emerged, or some would argue re-emerged, which is the integration of social science and molecular genetic research. Although previous work has extensively reviewed the challenges and possibilities of infusing genetics into sociological research (Conley 2016; Freese & Shostak 2009; Mills & Tropf 2020), several questions deserve further introspection. The aim of this chapter is to explore the implications of recent advances in molecular genetic research for theory, data and methodological applications in rigorous empirical sociology. I first chronicle a brief history of social status and mobility research in the late nineteenth century by Galton and work by Sorokin in the 1920s on social mobility. This is the birth of the nature versus nurture debate, but at the time was largely used as a means to reinforce existing social status. I then present an abridged history of contemporary genetic research from behavioral genetics and twin studies on a multitude of sociological topics from the early twentieth century to the use of whole genome molecular genetic data, the GWAS, the application of polygenic scores and advanced causal modeling. The chapter then explores how genetics fits with agency and rational choice-based models, examining communalities and challenges. Here, I also explore how genetically informed research can inform rational choice theory as a window to preferences and interrelated and latent genetic elements that may drive behavior. This is followed by an explanation of gene-environment interaction (G×E), outlining the main theories and exemplary studies. I conclude with a reflection of this area of research and

---

\* K. Gërkhani, N.D. de Graaf and W. Raub (eds) (2022), *Handbook of Sociological Science: Contributions to Rigorous Sociology*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing. ISBN: 978 1 78990 942 5; <http://dx.doi.org/10.4337/9781789909432>.

highlight an astounding lack of diversity in data and acknowledgment of macro-level contextual factors.

This chapter draws on the introduction of basic concepts and review of contemporary genetic research from a recent article I have written in the *Annual Review of Sociology* (Mills & Troup 2020), a textbook on the topic (Mills et al. 2020) and articles summarizing all genetic discoveries to date (Mills & Rahal 2019, 2020). Given that this chapter weaves together molecular genetics into sociological thinking, it is inevitable that readers may encounter new terminology. Although I strive to define them when introduced, the above-mentioned sources provide additional explanation of the main terms and concepts. I have likewise drawn from several GWAS studies that I have led, uncovering the genetics of sexual and reproductive behavior (Barban et al. 2016; Mills et al. 2021) and linking the fitness trait of number of children ever born to ancient genome data to study evolutionary aspects of reproductive behavior (Mathieson et al. 2020).

## 2. NATURE VERSUS NURTURE AND SOCIAL STATUS

### 2.1 Galton's Great Men of Science

The legacy of resistance to integrating biological and genetic thinking into sociology emanates from the original framing of 'nature versus nurture'. It insinuates a false dichotomy of biological drivers pitted against social behavior and phenomenon. We have increasingly learned that the complex behavioral outcomes sociologists study such as social mobility, education, employment, fertility and family processes are a combination of both nature *and* nurture (Tabery 2014). Beyond this dichotomy, early nineteenth century work also had an ugly history that linked biology to the study of intelligence and criminality to reinforce the superiority of certain groups (Martschenko et al. 2019). Although nature versus nurture debates emerged throughout history, the first scientific mention is often attributed to Galton's Royal Institution speech of February 27, 1874 (Galton 1874). Galton examined the autobiographical notes of 180 scientific Fellows and established several key traits of what he termed the 'great men of science'. The first was energy of the body and mind, present in 75 percent of men, where he noted that one: 'rowed 105 miles in 21 hours,' another 'walked 50 miles a day without fatigue' or when captivated by a particular subject 'I have written for seven or eight hours without interruption' (p. 228). He also identified other traits ranging from head circumference<sup>1</sup> to steadiness of pursuit. Independence of character was crucial with many 'possessing it in excess' and finally, a strong innate taste for science. Here he also pitted the influence of heredity against the environment, suggesting a strong role of heredity and nature underlying these great men.

Galton was best known as the father of eugenics and focused on family studies, particularly in his work *Hereditary Genius* (Galton 1869) where he collected data from multiple generations of core occupations, namely scientists, judges and military commanders.

---

<sup>1</sup> As the father of eugenics one might have assumed that Galton would attribute large head circumference to superiority, but in fact during this speech he noted that 'the general scientific position of the small-headed and large-headed men seems equally good' (p. 229).

His work also formed the foundation of twin studies, comparing monozygotic (identical) twins that were raised either together or apart with fraternal twins or siblings. Galton's interest in the nature versus nurture divide was, however, nefarious, with the aim to develop eugenic interventions. These perverse interventions, such as encouraging so-called 'more intelligent' and 'less criminal' individuals to marry and produce more progeny, were planned to eliminate undesirable traits and increase the prevalence of desirable ones in the gene pool.

## 2.2 Sorokin's Social Mobility and Studies of Intelligence

This type of thinking has resurfaced various times throughout history and again in sociology in 1927 with Sorokin's book *Social Mobility*, which focused considerable attention on the nature versus nurture debate in social status and mobility (Sorokin 1927). In his chapter on occupational stratification, he distinguishes between two factors: the importance of occupation in distinguishing the survival and well-being of the group and the degree of intelligence required to perform different occupational tasks. Sorokin immediately received disapproval by some who questioned why some occupations requiring higher intelligence should be rated more highly than those demanding low intelligence (Joslyn 1927).

Sorokin examined what he termed objective and subjective aspects, the latter he defined as the qualities and traits of individuals that determine selection into one's position in society. To do so he examined differences in social strata in relation to physique, vitality and intelligence. His intention was not to empirically focus on whether these aspects were inherited or acquired. The chapter draws on various data sources, comparing different social classes on differences in height (then termed as stature), weight, cranial capacity, vitality and general health. He concluded that the upper classes are taller, heavier, greater cranial capacity and have superior health and vitality. Sorokin was, however, disparaged for assuming causality from correlations and ignoring that these differences were due to better nutrition and the physical environment of the upper classes and that larger cranial capacity was merely a reflection of being taller and larger (Joslyn 1927). The 'superior' physical traits of the upper strata were thus likely the outcome rather than the cause. Sorokin's work seemed as a one-sided interpretation of nature and nurture, interpreting the correlation of social status with general health and vitality he stated:

The above correlation means that . . . physical superiority has been the condition which has favored the social promotion of individuals and has facilitated their social climbing, while physical inferiority has facilitated the social sinking of individuals and their location in the lower social strata. (Joslyn 1927, p. 275)

And he asserted: 'Aside from the problem whether the result is due to heredity or to environment, the higher social classes, on the whole, are more intelligent than the lower ones.' (p. 292). After establishing this correlation, he then asked whether it is nature or nurture, concluding that it is the outcome of heredity and environment, selection and adaptation, but largely defended the position that heredity accounts for differences in educational achievement.

In the late 1960s, Eckland once again revived the study of intelligence in sociology, arguing that variations in intelligence between the social classes are attributed to

assortative mating and heredity (Eckland 1967). These debates continued throughout history and are well documented elsewhere (Kevles 1995). In the 1970s, research claiming that the intelligence score gaps between black and white Americans was attributed to genetic differences between blacks and whites (nature) even led to calls to eliminate social policy and public funding to compensate minority groups from poorer socioeconomic environments. This kind of thinking has been extended by others such as the notorious work of Murray in *The Bell Curve* or the recent book *Coming Apart* on so-called ‘solutions’ to help disadvantaged Americans. In *The Bell Curve*, Murray and Herrnstein conclude:

The technically precise description of America’s fertility policy is that it subsidizes births among poor women, who are also disproportionately at the low end of the intelligence distribution. We urge generally that these policies, represented by the extensive network of cash and services for low-income women who have babies, be ended. (Murray and Herrnstein p. 548)

The central conclusion is that policies should not focus on improving poorer children’s living standards due to the fact that they operate to encourage poor, low-IQ women to reproduce. This brief review provides a short description of why the integration of biology and genetics into sociological thinking has been tainted by repugnant work that often used intelligence as a proxy for biological differences between groups in order to justify and reinforce inequalities and segregation (Martschenko et al. 2019; Herd et al. 2021).

### 3. CONTEMPORARY EMPIRICAL SOCIOGENOMIC RESEARCH

Contemporary empirical sociologists from around 2005 and onwards integrate genetic work into their research to examine a plethora of core sociological topics including the role of neighborhood disadvantage, resilience, stress trajectories, fertility, educational attainment, well-being, health and assortative mating. This work regularly overtly distances itself from the type of research described in the previous section. Before the availability of molecular genetics for large groups of unrelated individuals, initial work often examined differences in heritability via the use of twin models (Boardman & Fletcher 2015; Branigan et al. 2013; Conley et al. 2013). Heritability is the proportion of variation in a trait of a particular population attributed to genetic differences (Mills et al. 2020; Visscher et al. 2008). Heritability in twin models is thus determined by comparing the estimates of a trait’s correlation between monozygotic (identical) twins who share almost 100 percent of their genetic material with dizygotic twins (i.e., share around 50 percent of genetic material).

This was then followed by early candidate gene studies, such as the now infamous and debunked Caspi (2002) study which claimed a link between the 5HTTLPR gene and higher sensitivity to stressful life events. Most of these early candidate gene studies and gene  $\times$  environment interaction studies were underpowered false positives, the faults of which are sufficiently detailed elsewhere (Duncan & Keller 2011; Duncan et al. 2014). The majority of these candidate gene studies have been shown as erroneous false positives, such as most early genetic work linked to the study of intelligence (Chabris et al. 2012). Other early research dating back to the 1990s includes the infamous MAOA candidate gene work, coined as the ‘warrior gene’. Brunner studied a large Dutch family of five

generations of men who were extremely violent and found that they all lacked monoamine oxidase A (Brunner et al. 1993). Multiple researchers built on this such as claims that males with MAOA-L were more likely to be gang members (Beaver et al. 2010). It was soon reported however, that around 40 percent of the population carry MAOA-L (Hunter 2010). Although many of these early behavioral candidate genes were false positives, some candidates such as APOE-E for Alzheimer's and FTO for obesity, discussed shortly, did replicate.

Particularly since 2005, genome-wide association studies (GWASs) emerged as the primary technique to identify associations between outcomes and single-nucleotide polymorphisms (SNPs), which measure single base differences in DNA.<sup>2</sup> This chapter does not define and cover genetic concepts in any depth. As noted earlier, readers interested in an introduction to fundamental concepts in genetics, a primer in human evolution, statistical models and estimation of these techniques can refer to other sources or textbooks aimed at social scientists (Conley 2016; Mills et al. 2020). Particularly since 2016, as more data, measures, sample sizes and computational power grew, the types of topics studied with genetic data has exploded to 4,761 GWAS genetic discoveries as of November 1, 2020 (Mills & Rahal 2019, 2020).<sup>3</sup> What became increasingly obvious with the complex behavioral traits that sociologists study is that traits are not monogenic but rather polygenic or, in other words, many SNPs all have tiny associations with traits (Boyle et al. 2017). Polygenic scores are a linear combination of multiple SNP effects across the entire genome, weighted by the GWAS effect size, and represent the genetic variables that contemporary sociologists would integrate into their empirical research.

The majority of GWAS discoveries are in the area of disease and health (Mills & Rahal 2019). Mills & Tropf (2020) recently described myriad genetic discoveries that are likely of interest to empirical sociologists. Many of these traits have successive GWASs, such as the study of educational attainment that has three successive studies, with each increasing the sample size and subsequent number of genetic loci found. Social status measures include educational attainment (Lee et al. 2018), cognitive performance (Davies et al. 2015) and household income (Hill et al. 2019). Other research has examined sexual, fertility and reproductive behavior, including an initial study led by sociologists and demographers studying the age at first birth and number of children ever born (Barban et al. 2016). A recent follow-up to this study found 371 loci related to age at first sex and birth (Mills et al. 2021), which were linked to externalizing behavior. A companion study of number of children ever born in contemporary populations traced evolutionary patterns from ancient genome data, which we return to shortly (Mathieson et al. 2020). There has also been an extensive examination of multiple psychological traits, including the study of well-being (Baselmans et al. 2019). Other relevant research includes study of risk tolerance and behavior (Karlsson Linnér et al. 2019), and many on addiction related traits such as substance use, including alcohol and smoking (Jang et al. 2020).

<sup>2</sup> A SNP is a single nucleotide polymorphism, which is a variation in a single *nucleotide* (i.e., A, C, G, or T) that occurs at a specific position in the genome. A SNP usually exists as two different forms (e.g., A vs. T). These different forms are called *alleles*. A SNP with two alleles has three different genotypes (e.g., AA, AT, and TT).

<sup>3</sup> Readers interested in searching for all genetic traits discovered using GWAS and diversity by population can use the dropdown search menu in the [www.gwasdiversitymonitor.com](http://www.gwasdiversitymonitor.com), the companion dashboard to Mills & Rahal (2020) published in *Nature Genetics*, which is updated daily.

This section focused on describing the analytical approaches and substantive research topics that have been shown to have a genetic basis via GWAS discoveries. The combined polygenic score for the latest study of educational attainment found that it predicted around 11 percent of the total variation (i.e., incremental  $R^2$ ), with the polygenic score for the recent GWAS of age at first sex and first birth predicting 5–6 percent of the variance. This suggests that ignoring a genetic basis of classic outcomes in sociology may be overlooking potentially important predictors. What remains relatively unexplored, however, is how sociological theory and empirical models cope with these new genetic variables.

## 4. GENETICS AND MICRO-MACRO THEORETICAL MODELS IN SOCIOLOGY

### 4.1 Rational Choice Agency-based Theories

Empirical sociology reflects a strain of the discipline that is a fundamentally theory-guided endeavor devoted to describing mechanisms, which lead to testable hypotheses (see the chapter by Raub, De Graaf & Gërkhani on rigorous sociology in this Handbook). Theory can take the form of narrative accounts to more rigorous formal specifications such as in analytical sociology (Coleman 1990; Raub et al. 2011; see also Manzo's chapter on analytical sociology). This is in contrast to the GWAS technique underpinning genetic discovery, which is characterized as a 'hypothesis free' technique of data mining to uncover correlations between a particular outcome with genetic loci. After genetic loci are isolated, however, similar to sociological research, the aim is to isolate the underlying relationships and biological mechanisms. This includes attempting to understand the gene functions or actual causal genes related to the outcome under study to isolate pathways and identify certain tissues or cell types where the genes from the loci that have been found are expressed (Pers et al. 2015). When genetic information is included within sociology and related research, it follows the common precepts of rigorous sociology, often focusing on core causal questions (Conley & Zhang 2018; DiPrete et al. 2018). Sociologically and genetically informed approaches thus acquiesce in the attempt to describe and explain outcomes by examining regularities (see the chapter by Jackson on sociology as a population science) and seeking out potential mechanisms and causal relationships.

Many micro-level sociological theories are based on agency or choice assumptions. Variants of rational choice theory (see, for example, Hechter & Kanazawa 1997 and the showcase-chapter by Breen & Goldthorpe) yield well-known examples of such assumptions, with some such variants echoing choice-based theories in economics. Rational choice theories (see Diekmann's chapter on rational choice sociology) *as such* specify how individuals act, *given* their feasible alternatives on the one hand and, on the other hand, their beliefs and their preferences, including, among others, their time preferences. Typically, then, it is assumed that action follows some variant of utility maximization. Rational choice theory *as such*, however, does not specify actors' beliefs and preferences themselves. Additional assumptions on 'substantive' properties of beliefs and preferences are often needed in addition to assumptions on (some variant of) utility maximization

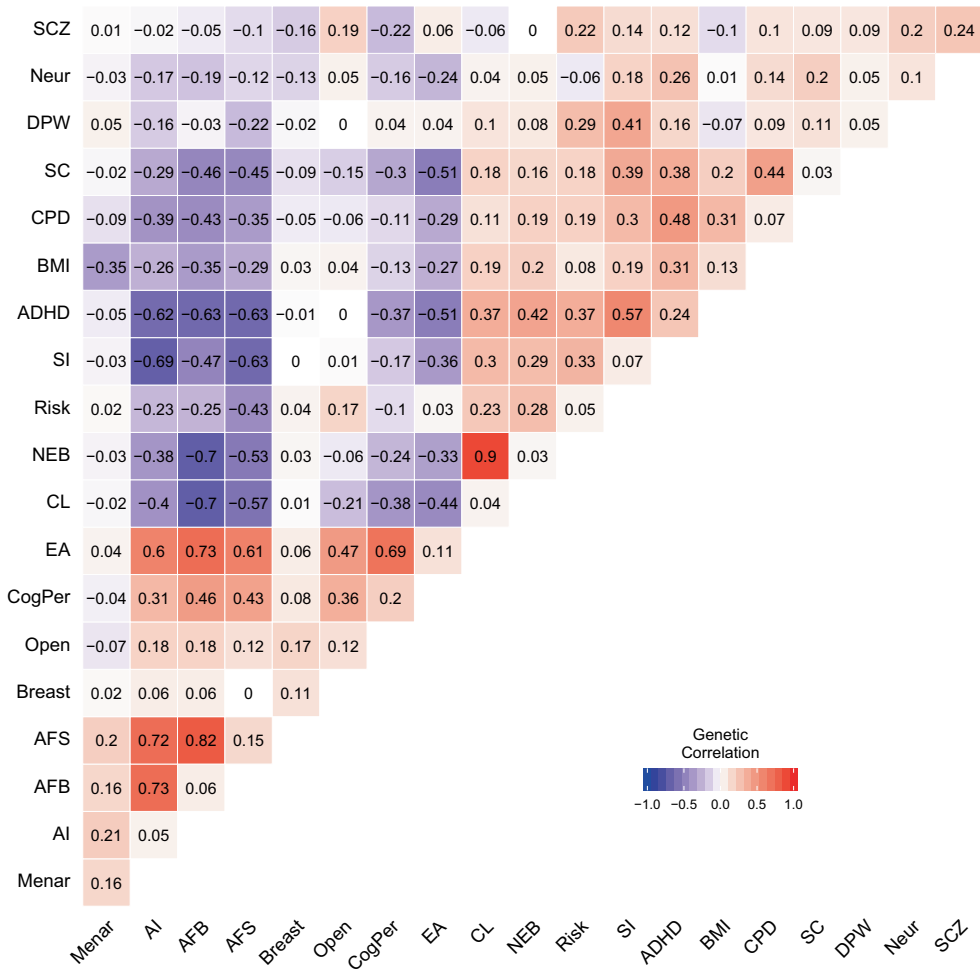
when applying rational choice theory in social science explanations. There are arguments, too, that individual preferences are important kinds of explananda for sociology (Boudon 1977) and a similar point could be made with respect to actors' beliefs. This is one of the areas where synergies can be gained for sociological theory and empirical research by combining assumptions on agency, such as assumptions on rational choice, with insights from genetics. Specifically, this includes progress in understanding behaviors that seem counter-intuitive from an agency perspective, at least at first sight, such as behaviors with detrimental outcomes like obesity or opioid addiction, and more generally behaviors often referred to in the psychological literature as lack of self-control, risk-taking, externalizing behavior or behavioral disinhibition (Mills et al. 2021).

## 4.2 Self-control and Behavioral Disinhibition

To illustrate how recent advances in our genetic understanding of traits can inform our knowledge, consider the two examples of outcomes that have radically changed in many societies across the past decades – obesity and age at reproductive onset (age at first sex and birth). Turning first to obesity, although most understand that daily exercise is beneficial and being obese is detrimental, many industrialized countries are now facing an obesity epidemic (James et al. 2001). Biological and genetic research demonstrates that many humans have weak physiological mechanisms to resist excessive food intake when food is abundant. Research about the renowned FTO or 'fat gene' found that carriers of a certain variant of the FTO gene have reduced activity in areas of the brain that are associated with impulse control (Church et al. 2010). This in turn affected changes in their weight, impulse eating and brain function as they aged. Individuals homozygous for the at-risk A allele were shown to weigh around 3 kg (6.6 lb) more than those with the low risk T allele. Other studies demonstrated that people with the detrimental variant gained more weight as they aged, likely experiencing changes in their brain function related to impulsivity with age.

Differences in behavioral disinhibition or externalizing has often been linked with psychiatric disorders and substance use (Iacono et al. 2008). A recent study linked behavioral disinhibition and risk taking to earlier age at first sex and teenage birth, suggesting that there may be underlying genetic factors linking multiple behaviors (Mills et al. 2021). Figure 14.1 provides an excerpt of these results in the form of a heatmap, with dark purple representing negative genetic correlations and dark red, positive genetic correlations. Age at first sex (AFS) and age at first birth (AFB) are coded in a chronological direction (i.e., higher AFS/AFB positive correlation). Genetic correlations are examined via a method called LD-score regression, which exploits the linkage disequilibrium structure of the genetic data<sup>4</sup> to show the genetic overlap or in other words the extent to which different traits share the same genetic loci (Finucane et al. 2015). Here we see, logically, the strongest positive genetic correlations between the fertility variables, such as childlessness with a lower number of children ever born (0.90). Strikingly, educational attainment has a

<sup>4</sup> As a result of genetic recombination, polymorphisms are inherited together through what is called linkage disequilibrium (LD), which is the non-random occurrence of the combinations of two or more linked genomic loci. For instance, if a T at one SNP locus is generally observed with a G at another SNP locus, these two SNPs are said to be in LD.



*Note:* Calculated by LD score regression, with SNP heritabilities along the diagonal. Menarche = Age at menarche; AI = age at initiation of smoking; AFB = age at first birth; AFS = age at first sex; Breast = ever had breast cancer; Open = openness (personality); CogPer = cognitive performance; EA = educational attainment in years; CL = childlessness; NEB = number of children ever born; SI = ever smoked; ADHD = attention deficit hyperactivity disorder; BMI = Body mass index; CPD = cigarettes per day; SC = smoking cessation; DPW = alcoholic drinks per week; Neur = neuroticism (personality); SCZ = Schizophrenia.

*Source:* Mills et al. (2021).

**Figure 14.1** Genetic correlations and SNP heritabilities between and among reproductive, behavioral, psychiatric, substance use, personality and anthropometric traits

remarkably high genetic correlation of 0.73 with AFB and 0.61 with AFS (i.e., more years of education; later AFS/AFB). This suggests a high genetic overlap in the loci that predict both educational attainment and reproductive behavior (AFS, AFB). We also see a negative correlation (−0.43) of those less genetically prone to risk in adulthood likewise less

prone to early teenage sex and pregnancy. One of the strongest genetic associations is also ADHD ( $-0.63$ ), which has been phenotypically related to risky sexual behavior and substance use, which suggests an underlying link with early sexual debut and teenage pregnancy. Addictive and substance use behavior also had striking correlations, particularly early age at onset of smoking (not shown here, but  $0.73$ ), which provides a window into adolescence, coupled with early AFS/AFB.

These examples provide some evidence that there may be an underlying genetic aspect to many related behaviors, likely linked to externalizing or behavioral disinhibition. These hitherto unmeasured or latent genetic elements represent a constellation of behaviors and challenge our thinking of individuals' ability for transitivity (ability to rank) and restrictions on behavior in behavioral choice set models. When genetics, such as predisposition to a lack of self-control is introduced, it reveals potential mechanisms and reasons as to why there might be a discrepancy between long-term goals and detrimental short-term behavior.

## 5. GENE X ENVIRONMENT INTERPLAY AND MACRO-LEVEL CONTEXT

A unique strength of sociology is the ability to describe behavior or outcomes not only at the level of the individual, but within the environmental context of meso- and macro-level systems, including organizations, families, regions and nations. The backbone of rigorous empirical sociology is a mechanism-based explanation that positions individuals' micro-conditions as being influenced by macro-conditions that in turn result in the aggregated macro-level outcomes that we observe (see the chapter by Raub, De Graaf & Gërkhani). For this reason, when sociologists examine genetic data, one of the first inclinations is to understand how context – environment in genetic research – interacts to impact outcomes. In sociogenomics, the macro-level context is often studied in gene-environment ( $G \times E$ ) studies to examine how it modifies associations between genetic predisposition and behavioral outcomes, whereas gene-environment correlation ( $rGE$ ) is the process by which an individual's genotype influences or is associated with exposure to the environment (i.e., genes and the environment operate in tandem).

The fact that we are interested in or that we include genetics into sociological models does not mean that there is a diminished interest in the study of socio-environmental factors. In fact, most researchers in the field of sociogenomics would argue that the social structural and micro-level predictors of many of the complex behavioral outcomes we study still remain as the most important predictors. The introduction of genetic information is thus a complement not a replacement. Gene-environment interaction ( $G \times E$ ) is one of the most compelling, but challenging areas of research. The environment is characterized by an upstream process that may influence the trait of study and is exogenous, challenges which I return to in the final methodological section. For a detailed history of  $G \times E$  research, definitions of the environment and a detailed description of the various theoretical models see Chapter 6, and Chapter 11 for programming examples to empirically estimate these models (Mills et al. 2020). A comparison of the primary theories and examples of applications is shown in Table 14.1, with additional applications described in the chapter by Hopcroft, Dippong, Liu & Kail on evolution, biology, and society.

Table 14.1 Summary of theoretical and conceptual models underlying G×E studies

Theory	Brief summary	Original article and further reading	Example of an empirical article that uses this theory
Diathesis-stress, also known as Vulnerability, Contextual Triggering	A predisposition (i.e., a diathesis) for the phenotype lies dormant until triggered by environment (e.g., stressor)	Monroe & Simons (1991)	South & Krueger (2008)
Bioecological or social compensation model	Genetic influences are maximized in stable and adaptive environments that permit positive, stable interactions (proximal processes) between individuals and their environment enabling them to reach their genetic potential  Social compensation: environment is free of stress or has positive enriching properties	Bronfenbrenner & Ceci (1994)  Shanahan & Hofer (2005)	Turkheimer et al. (2003)
Differential susceptibility	Plasticity varies by individual with some (known as orchids) being more susceptible to (i.e., genetically influenced) by the effect of both positive and negative environments whereas others (known as dandelions) are more resilient	Belsky & Pluess (2009)	South & Krueger (2013)
Social control or social push	Genetic influences are filtered and dampened in particular environments.  Social control: social norms and structural constraints	Shanahan & Hofer (2005)	Boardman (2009) Dick et al. (2007) Liu & Guo (2015)

Source: Taken from Table 6.1 Mills et al. (2020).

The majority of research in this area applies the diathesis-stress model, which hypothesizes that genetic differences are associated with negative outcomes in risky environments. Here, research often focuses on particular stressors or triggers (e.g., death of spouse) or contextual triggering. When scrutinized, the social compensation model is conceptually the flipside or mirror of the diathesis-stress model. It focuses on protective aspects of the environment that enable individuals to realize their genetic potential, such as a supportive home environment or parents who do not smoke. Relevant early work tested the Scarr-Rowe hypothesis (Scarr-Salapatek 1971) that families with more resources would be more able to help their children reach their genetic potential. Seminal work demonstrated

that the heritability of IQ was the highest in families with high socioeconomic status (Turkheimer et al. 2003). Interestingly, this finding was replicated in the US, but not in some European countries or Australia, likely due to the stronger safety net for disadvantaged households in these contexts compared with the US (Tucker-Drob & Bates 2016).

The differential susceptibility model is a variation of the diathesis-stress model which hypothesizes that some individuals are sensitive to negative and positive environments, often referred to as the orchid-dandelion hypothesis. A low socioeconomic environment, for instance, allows for the expression of genetic vulnerabilities related to poor outcomes (e.g., substance use, addiction). A highly enriched environment allows positive genetic predispositions to be realized. Individuals are assumed to have a certain level of plasticity with crossover interactions of some being more susceptible to positive and negative environments. Finally, the social control model hypothesizes that genetic associations are attenuated or dampened in the presence of socially restrictive environmental contexts. This could include parental monitoring or religious norms that restrict certain behavior and exposures to drinking or smoking. One study, for instance, demonstrated that an educational reform in the UK that forced students to remain in school longer had positive health effects, particularly for those with a higher polygenic score associated with obesity (Barcellos et al. 2018). A theoretical extension of the typical  $G \times E$  framework examined how environmental changes jointly modified the genetic correlation between educational attainment and smoking over the past century (Wedow et al. 2018).

A sizeable amount of twin study research focused on decomposing variance into genetic and environmental components, often concentrating on socioeconomic differences (Purcell 2002; Turkheimer et al. 2003). A commonly used approach to examine  $G \times E$  is to study how genetic effects vary as a function of the environment, proxied by birth cohort or country.

Birth cohort has been used as a proxy for exposure to obesogenic environments. A recent study demonstrated that birth cohort modified the association between FTO and BMI, suggesting that genotype–phenotype (outcome) correlations are likely highly dependent on the time period or birth cohort of individuals (Rosenquist et al. 2015). Using a polygenic score for BMI, another study found that the magnitude of associations of the polygenic score for BMI were larger for more recent cohorts (Walter et al. 2016). This finding also holds for other outcomes. Another genetic study of age at first birth found that the SNP-heritability<sup>5</sup> of birth cohort for women changed from 9 percent [CI = 4–14] for those born in 1940, climbing to around 22 percent [CI=19–25] for those born in 1965 (Mills et al. 2021).

An individuals' genetic makeup is not only exogenous, but also the product of a long-term evolutionary process, covered in the chapter by Hopcroft, Dippong, Liu & Kail. When thinking about overeating and obesity, evolutionary biology supposes that the process of natural selection endowed humans with the ability to store energy as body fat

---

<sup>5</sup> SNP-heritability is the proportion of phenotypic variance jointly accounted for by all variants on standard GWAS chips and calculated using the statistical method called GREML (Genome-based Restricted Maximum Likelihood), which estimates the variance component to quantify the total narrow-sense (additive) contribution of a trait's heritability. See Mills et al. (2020) for detailed description of differences between Twin, SNP- and GWAS-based heritability (Section 1.6.3).

to survive regular famines. Geneticist Neel (Neel 1962) proposed the ‘thrifty genotype’ hypothesis that modern human populations were genetically maladapted to the modern diet and lifestyle. Although caloric energy was advantageous during a famine, it was no longer efficient in modern environments. Others have examined how genetic determinants of reproductive success highlight mechanisms underlying fertility and identify alleles under present-day selection (Mathieson et al. 2020). In a recent GWAS of number of children ever born (NEB) we found that NEB-increasing alleles increased in frequency over the past two generations. Comparing our NEB GWAS on contemporary populations with ancient genome data, we furthermore identified that the *FADS1/2* locus that has been under selection for thousands of years, still remains under selection today.

## 6. DATA AND METHODOLOGICAL CHALLENGES

Until now, this chapter has explored substantive and theoretical questions of the challenges and prospects of integrating molecular genetic data and thinking with sociology. In this section, I first outline the core challenges and potential solutions for G×E analysis followed by describing the striking lack of multiple types of diversity in genetic data with implications for sociological and other research.

### 6.1 Challenges and Potential Solutions for Studying G×E Analysis

Although G×E research is inherently appealing, it is technically very challenging to carry out, as witnessed by the early failures of candidate gene research discussed earlier. As documented previously, there are several main challenges, described in detail in Table 6.2 of Mills et al. (2020). First, the theoretical model underlying G×E interaction is not formalized, leading to black box explanations and some of the central theories lead to similar outcomes. As alluded to in the previous section, many of the theories are not mutually exclusive nor are they formalized in the way familiar to analytical or formal sociological theory. G×E theories often articulate plausible biological pathways, yet generally lack the downstream biological analyses related to G.

A second problem is the inability to find environmental conditions that are actually exogenous and the related problems in measuring environmental exposure (Conley 2017). It is often the case that there is both an inconsistent measure of the environmental exposure or that environmental conditions are time-varying and manifold. A third conundrum is that when the environment is viewed as a proximate moderator, it is actually considerably downstream from the social environmental factors that supposedly structure the actual exposure (Boardman et al. 2013). Environments such as obesogenic environments or maternal smoking are often used, but very downstream from the social environmental factors that structure that exposure. The fourth shortcoming is the lack of statistical power and enormous samples required to detect any real G×E. To estimate the magnitude and strength of association, a sufficiently large sample is required.

A fifth problem is several methodological mistakes during the regression analysis that regularly occur. The first is when the detected G×E interactions are actually an artefact that is driven by confounders such as sex, age or socioeconomic status rather than the

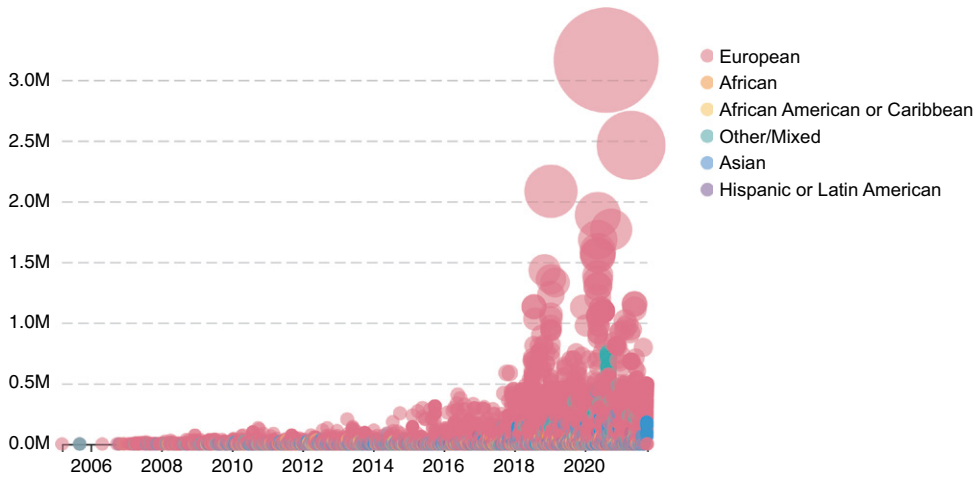
G or E. Researchers often incorrectly enter cofounders as covariates in general linear models. A simple solution is that they actually need to enter the covariate-by-environment and covariate-by-gene interaction in the same model that tests the  $G \times E$  interaction (Keller 2014). Another issue is when the interaction is sensitive to the scaling of the metric, but the researcher only includes the multiplicative scale. Again, a simple solution is to evaluate the interaction on both the multiplicative and additive scales. Another problem is collider bias, which occurs when controlling for heritable covariates in regression models with polygenic scores introduces endogenous selection bias when one or more of these covariates is dependent on unmeasured factors that also affect the outcome (Akimova et al. 2021). This in turn leads to spurious associations and effect sizes, suggesting that more attention needs to be placed on causal inference, and interpretation of models with polygenic scores and non-exogenous environments.

A sixth problem is that by ignoring the macro-context, which is often the case in most GWA studies, they mask  $G \times E$ . To achieve a large sample size, GWAS often combine highly heterogeneous data from multiple countries, across different birth cohorts and individuals of varying ages and conduct a meta-analysis. By doing so, they assume that the influence of genetics amongst individuals is universal across time and place. We examined the genetics of height, BMI, educational attainment, age at first birth and number of children ever born across various countries and several historical periods. What we found was that for highly complex and socio-environmentally influenced traits such as educational attainment, fertility and BMI, between 40 to 75 percent of the genetic effects were ‘hidden’ or ‘watered down’ when data were combined; genetic effects were thus highly country and time specific (Tropf et al. 2017).

## 6.2 Lack of Diversity and Non-representative Samples

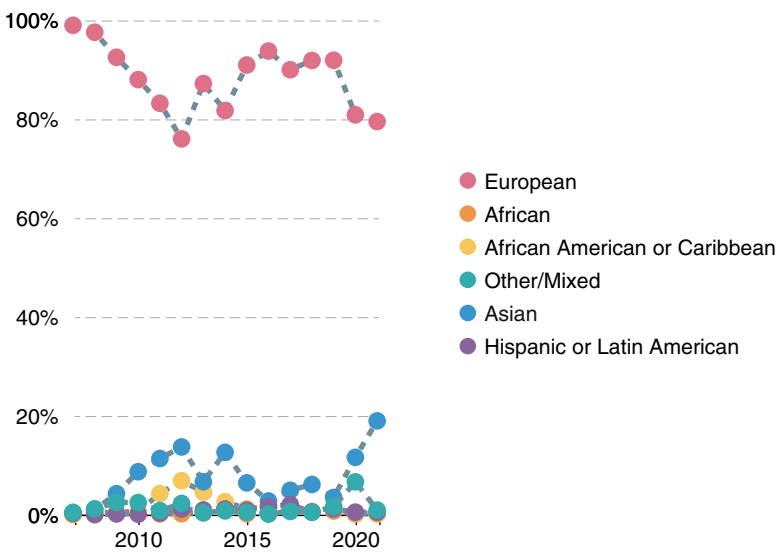
A main interest of sociologists has and will continue to be the study of gene–environment interaction or, in other words, how environments modify genetic influence. Although several detailed and technical aspects were listed above, perhaps the gravest threat is the lack of diversity in the data in terms of the representativeness of the sample, geographical, ancestral and socioeconomic status.

A hallmark of most rigorous empirical sociological research is well-designed representative samples to draw systematic conclusions about group differences. Yet within genetics there is increased recognition that the data lack ancestral diversity (Martin et al. 2017). We recently conducted a systematic, computational analysis of all GWAS discoveries for 13 years from 2005 until late 2018 (Mills & Rahal 2019), later updated in an interactive online dashboard that updates GWAS discoveries on a daily basis, derived from the NHGRI-EBI GWAS Catalog (Mills & Rahal 2020). Although there has been an explosion in the number of GWAS studies over time, Figures 14.2 and 14.3 illustrate that as of November 5, 2021, 84.2 percent of studies still come from individuals of European ancestry, a number which has remained relatively stable over time. In particular, the large studies that draw on direct-to-consumer commercial data such as 23andMe have very large samples. A study from 2019 on well-being with 2.3 million cases (Baselmans et al. 2019) was eclipsed in August 2020 with a large study of sex-differentiation in participation bias (Pirastu et al. 2021) and one on substance use and psychiatric disorders in over 3 million individuals (Jang et al. 2020).



Source: Mills & Rahal (2020) [www.gwasdiversitymonitor.com](http://www.gwasdiversitymonitor.com).

Figure 14.2 GWAS participant ancestry, 2005 to November 5, 2021



Source: Mills & Rahal (2020) [www.gwasdiversitymonitor.com](http://www.gwasdiversitymonitor.com).

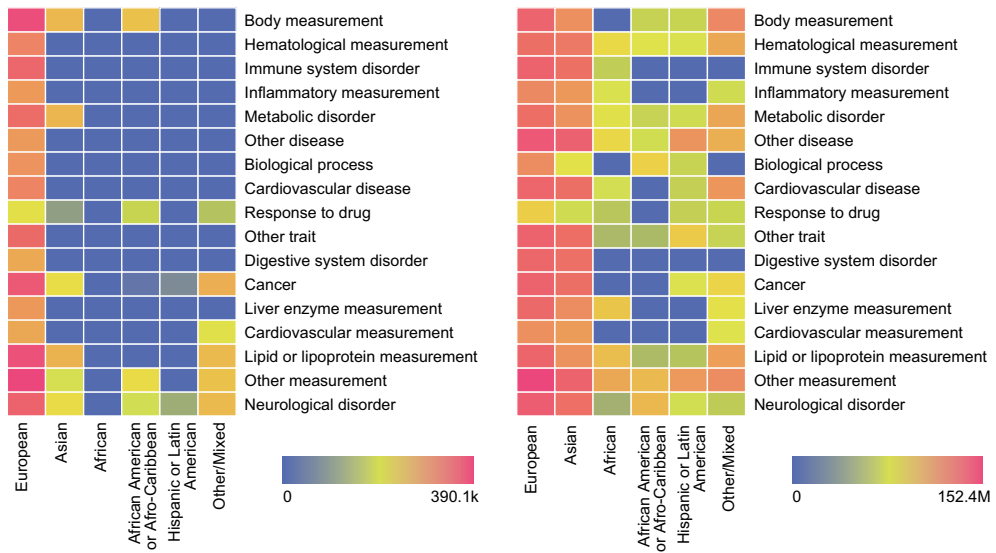
Figure 14.3 Percentage of participants from all genetic GWAS discoveries 2005 to November 5, 2021

6.3 Why is Lack of Diversity a Problem?

Beyond obvious ethical issues such as the exacerbation of structural inequalities, the transferability of GWAS results across different ancestral groups depends on multiple factors such as allele frequency, linkage disequilibrium, genetic architecture, epistasis and gene-environment interaction. Vitrally, single-ancestry GWAS have limited portability to other groups (Duncan et al. 2019; Martin et al. 2017). In other words, since around 90 percent of genetic discoveries and the polygenic scores derived from them are from European ancestry populations, they are only applicable for research on those groups. This has serious implications for the reinforcement of existing inequalities, particularly in the realm of health, where genetic research can also be linked to pharmaceutical interventions (Martin et al. 2019).

Figure 14.4 shows the type of outcome studied by different ancestry groups in 2008 and 2021 to identify the most imbalanced areas. Blue means that no GWAS have been conducted in these groups, with red representing the most participants. This imbalance has changed over time. At the initial genetic discovery phase in 2019, for instance, the most imbalanced area is cancer research where European ancestry groups made up 96.3 percent of all participants studied compared with 0.5 percent Hispanic or Latin American, 0.1 percent African and 0 percent African American or Afro-Caribbean. In 2021, this imbalance in cancer research amongst African and African American or Afro-Caribbean groups remains.

The categorization of ancestry groups often relies on genomic analyses which infer ancestry via principle components analysis and author reporting, with a detailed

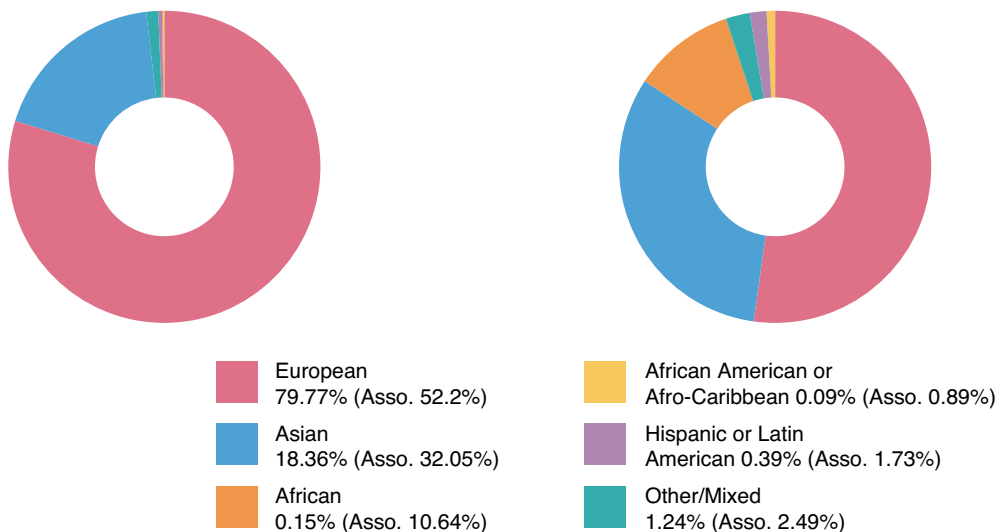


Source: Mills & Rahal (2020) [www.gwasdiversitymonitor.com](http://www.gwasdiversitymonitor.com).

Figure 14.4 Heatmap of type of GWAS outcome studied by Ancestry group, 2008 (left) and 2021 (right)

discussion provided in the Supplementary Material of Mills & Rahal (2020), derived from Morales et al. (2018). It is important to understand that ancestry as it is used in genetics is not the same as race or ethnicity as used in sociology. Race or ethnicity is not a biological category, but rather a term that is socially, culturally and politically constructed (Herd et al. 2021). Ancestry is the measure of genetic variation traced to geographical location and migration and does not map directly into the social science categorization of race and ethnicity. As I and others have described in more detail elsewhere and reiterate here (Herd et al. 2021; Mills & Rahal 2020; Mills et al. 2020), populations are the product of repeated genetic mixing over tens of thousands of years. The concentration of certain genetic alleles or differences between ancestral groups is related to where they descended from. Most evidence places the origins of *Homo sapiens* in sub-Saharan Africa, where the patterns of DNA sequence variation are the greatest. As humans migrated, they took progressively smaller amounts of genetic variation in the gene pool with them. Each new and younger population thus has less time to accumulate new mutations. As Figure 14.5 illustrates, although in 2021 the ‘young’ European ancestry populations represented 79.8 percent of participants in GWAS genetic discoveries, they were only responsible for 52.2 percent of the genetic loci that were uncovered. Although African American or Afro-Caribbean ancestry groups were 0.24 percent of the participants, they provided 11.5 percent of the discovered associations.

Sequencing of Khoi-San bushmen showed that even two people from adjacent villages were as different from one another as any two European or non-African ancestry individuals (Schuster et al. 2010). A second issue is that due to patterns of migration, evolution and population shifts, characterizing populations into the ancestry groups shown



Source: Mills & Rahal (2020) [www.gwasdiversitymonitor.com](http://www.gwasdiversitymonitor.com).

**Figure 14.5** Participants by ancestry (left) and count of all genetic associations discovered (right) for all genetic discoveries, 2021

here using the classic tools and techniques of genetics is becoming increasingly challenging and complex. The classic standard was to rely on reference populations or markers of ancestry that allow populations to be easily distinguished, with methods developed to adjust for population stratification and clustering of samples. Yet current societies are more marked by mixing and intermarriage, which has demanded new tools. As migration and admixture increase, admixed techniques have already and will undoubtedly continue to emerge in the future (Hellenthal et al. 2014). This classification and development of tools to reflect these admixed populations at a more precise and granular level presents a growing area of new research.

#### 6.4 Non-representative Samples

Beyond issues of sample selection related to ancestry, the majority of data in which genetic discoveries are derived often comes from non-representative samples. One of the most used openly available datasets is the UK Biobank, with a sample size of 500,000. The data, however has a 5 percent response rate and is over-represented by those from a higher socioeconomic status, healthier, non-smokers, and has significant geographic clustering (Fry et al. 2017). A selective analytical sample narrows the variation in the environment and thereby reduces statistical power (Domingue et al. 2020). For empirical sociologists, the implications of such selection are immediately obvious for making larger generalizations. Surprisingly, our work in 2019, appeared to our knowledge to be the first to trace the overall geographical distribution of participants used in genetic discoveries. There we showed that 72 percent of genetic discoveries came from participants from just three countries, namely the US (41 percent), UK (40.5 percent) and Iceland (11.5 percent), the latter of which has a population of around 334,000 people. Given what we know about the impact of country on genetic outcomes (Tropf et al. 2017) discussed previously, this concentration of research into such a selective group is striking. It is for this reason that we developed a ‘GWAS Diversity Monitor’ that tracks these trends in real time (Mills & Rahal 2020).

### 7. CONCLUSION

There have been considerable gains in sociogenomic research and genetic discoveries in many of the topics that are at the core of sociology such as education, fertility and well-being. As the polygenic scores derived from these genetic GWAS discoveries continue to increase in their explanation of variance, they become impossible to ignore. Genetics may provide a complement to rational choice agency-based theories to explain the black box of preferences and underlying penchants driving behavior such as self-control, risk aversion or behavioral disinhibition. As many in the field of sociogenomics anticipated, genetics will only ever explain a very small proportion of the complex and interrelated outcomes that sociologists study. Theory and measurement of core sociological variables such as socioeconomic status, childhood and family traits, educational level, birth cohort and national context remain the strongest predictors of the behaviors and regularities empirical sociologists observe. How the environment modifies the relationship between genetic predispositions and the outcomes in G×E research remains a promising, yet challenging future area of research.

Empirical sociology likewise has considerable input and value to offer genetic research. Theoretically informed and well specified formal and analytical models could greatly benefit this field, particularly in the realm of loosely formulated G×E models. Sociologists are also the masters of measurement of socioeconomic variables such as education, social status, intergenerational, family and employment-related variables, many factors that drive and stratify health and other outcomes. Attention and commitment to the value of representative samples or weighting to deal with shortcomings is another strength of this area. Expertise in causality, discussed elsewhere in this book, also has much to offer (see Breen's chapter on causal inference). Most sociogenomic papers give considerable attention to causality and estimate causal models such as Genetic Instrumental Variables (DiPrete et al. 2018), Genomic SEM (Structural Equation Modeling, Grotzinger et al. 2019) and bi-Directional Mendelian Randomization (Davey Smith 2005). Sociologists likewise approach causality in unique manners as well (see Breen's chapter). Our recent work, for instance, scrutinizes the largely unrecognized impact of gene-environment dependencies on the identification of the effects of genes and their variation across environments (Akimova et al. 2021).

Perhaps the most important contribution of sociology is the explicit recognition of the importance of context in the ability to generalize findings. Although sociology is also liable for a focus on Western countries and topics, the discipline would likely not accept that 90 percent of the discoveries come from one population (European ancestry) with little attention to context. The fact that results highly vary by context is common-sense to sociologists. Yet contextual variation has been frequently misinterpreted as a 'fault', such as the inability to replicate research and lack of irreproducibility. For instance, a study that attempted to evaluate the replicability of social science experiments published in *Nature* and *Science* from 2010 to 2015 claimed that many did not stand up to replication (Camerer et al. 2018). Their conclusion was a concern about the 'reliability' of the social science literature. Although there were some potentially dubious studies included, there was little to no attention to whether studies and experiments across different contextual environments *should* in fact replicate. With ever expanding theories, data and methods to grapple with the complexity of behavior and the environment, the future of sociogenomics remains an exciting new frontier.

## REFERENCES

- Akimova, E.T., R. Breen, D.M. Brazel, and M.C. Mills (2021), 'Gene-environment dependencies lead to collider bias in models with polygenic scores', *Scientific Reports*, **11**, 9457.
- Barban, N., R. Jansen, R. de Vlaming, A. Vaez, J.J. Mandemakers, . . ., Hyppönen, E. (2016), 'Genome-wide analysis identifies 12 loci influencing human reproductive behavior', *Nature Genetics*, **48**, 1–7.
- Barcellos, S.H., L.S. Carvalho, and P. Turley (2018), 'Education can reduce health differences related to genetic risk of obesity', *Proceedings of the National Academy of Sciences*, **115**, E9765–E9772.
- Baselmans, B.M.L., R. Jansen, H.F. Ip, J. van Dongen, and A. Abdellaoui, . . ., M. Bartels (2019), 'Multivariate genome-wide analyses of the well-being spectrum', *Nature Genetics*, **51**, 445–451.
- Beaver, K.M., M. DeLisi, M.G. Vaughn, and J.C. Barnes (2010), 'Monoamine oxidase A genotype is associated with gang membership and weapon use', *Comprehensive Psychiatry*, **51**, 130–134.
- Belsky, J. and M. Pluess (2009), 'Beyond diathesis stress: Differential susceptibility to environmental influences', *Psychological Bulletin*, **135**, 885–908.
- Boardman, J.D. (2009), 'State-level moderation of genetic tendencies to smoke', *American Journal of Public Health*, **99**, 480–486.

- Boardman, J.D., J. Daw, and J. Freese (2013), 'Defining the environment in gene-environment research: Lessons from social epidemiology', *American Journal of Public Health*, **103**, 64–72.
- Boardman, J.D. and J.M. Fletcher (2015), 'To cause or not to cause? That is the question, but identical twins might not have all of the answers', *Social Science & Medicine*, **127**, 198–200.
- Boudon, R. (1977), *Effets Pervers et Ordre Social*, Paris: PUF.
- Boyle, E.A., Y.I. Li, and J.K. Pritchard (2017), 'An expanded view of complex traits: From polygenic to omnigenic', *Cell*, **169**, 1177–1186.
- Branigan, A.R., K.J. McCallum, and J. Freese (2013), 'Variation in the heritability of educational attainment: An international meta-analysis', *Social Forces*, **92**, 109–140.
- Bronfenbrenner, U. and S.J. Ceci (1994), 'Nature-nurture reconceptualized in developmental perspective: A bioecological model', *Psychological Review*, **101**, 568–586.
- Brunner, H.G., M.R. Nelen, P. van Zandvoort, N.G. Abeling, A.H. van Gennip, . . . , B.A. van Oost (1993), 'X-linked borderline mental retardation with prominent behavioral disturbance: Phenotype, genetic localization, and evidence for disturbed monoamine metabolism', *American Journal of Human Genetics*, **52**, 1032–1039.
- Camerer, C.F., A. Dreber, F. Holzmeister, T.-H. Ho, J. Huber, . . . , H. Wu (2018), 'Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015', *Nature Human Behaviour*, **2**, 637–644.
- Caspi, A. (2002), 'Role of genotype in the cycle of violence in maltreated children', *Science*, **297**, 851–854.
- Chabris, C.F., B.M. Hebert, D.J. Benjamin, J. Beauchamp, D. Cesarini, . . . , D. Laibson (2012), 'Most reported genetic associations with general intelligence are probably false positives', *Psychological Science*, **23**, 1314–1323.
- Church, C., L. Moir, F. McMurray, C. Girard, G.T. Banks, . . . , R.D. Cox (2010), 'Overexpression of Fto leads to increased food intake and results in obesity', *Nature Genetics*, **42**, 1086–1092.
- Coleman, J.S. (1990), *Foundations of Social Theory*, Cambridge, MA: Harvard University Press.
- Conley, D. (2016), 'Socio-genomic research using genome-wide molecular data', *Annual Review of Sociology*, **42**, 275–299.
- Conley, D. (2017), 'The challenges of GxE: Commentary on "Genetic endowments, parental resources and adult health: Evidence from the Young Finns Study"', *Social Science & Medicine*, **188**, 201–203.
- Conley, D., E. Rauscher, C. Dawes, P.K.E. Magnusson, and M.L. Siegal (2013), 'Heritability and the equal environments assumption: Evidence from multiple samples of misclassified twins', *Behavior Genetics*, **43**, 415–426.
- Conley, D. and S. Zhang (2018), 'The promise of genes for understanding cause and effect', *Proceedings of the National Academy of Sciences*, **115**, 5626–5628.
- Davey Smith, G. (2005), 'What can mendelian randomisation tell us about modifiable behavioural and environmental exposures?', *BMJ*, **330**, 1076–1079.
- Davies, G., N. Armstrong, J.C. Bis, J. Bressler, V. Chouraki, . . . , I.J. Deary (2015), 'Genetic contributions to variation in general cognitive function: A meta-analysis of genome-wide association studies in the CHARGE consortium ( $N=53\,949$ )', *Molecular Psychiatry*, **20**, 183–192.
- Dick, D.M., J.C. Wang, J. Plunkett, F. Aliev, A. Hinrichs, S. Bertelsen, J. P. Budde, E.L. Goldstein, D. Kaplan, H.J. Edenberg, J. Nurnberger Jr, V. Hesselbrock, M. Schuckit, S. Kuperman, J. Tischfield, B. Porjesz, H. Begleiter, L. J. Bierut, and A. Goate (2007), 'Family-based association analyses of alcohol dependence phenotypes across DRD2 and neighboring gene ANKK1', *Alcoholism: Clinical and Experimental Research*, **31**, 1645–1653.
- DiPrete, T.A., C.A.P. Burik, and P.D. Koellinger (2018), 'Genetic instrumental variable regression: Explaining socioeconomic and health outcomes in nonexperimental data', *Proceedings of the National Academy of Sciences*, **115**, E4970–E4979.
- Domingue, B., S. Trejo, E. Armstrong-Carter, and E.M. Tucker-Drob (2020), 'Interactions between polygenic scores and environments: Methodological and conceptual challenges', *SocArXiv Prepr*.
- Duncan, L., H. Shen, B. Gelaye, J. Meijsen, K. Ressler, . . . , B. Domingue (2019), 'Analysis of polygenic risk score usage and performance in diverse human populations', *Nature Communications*, **10**, 3328.
- Duncan, L.E. and M.C. Keller (2011), 'A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry', *American Journal of Psychiatry*, **168**, 1041–1049.
- Duncan, L.E., A.R. Pollastri, and J.W. Smoller (2014), 'Mind the gap: Why many geneticists and psychological scientists have discrepant views about gene-environment interaction (G×E) research', *American Psychologist*, **69**, 249–268.
- Eckland, B.C. (1967), 'Genetics and sociology: A reconsideration', *American Sociological Review*, **32**, 173–194.
- Finucane, H.K., B. Bulik-Sullivan, A. Gusev, G. Trynka, Y. Reshef, P.R. Loh, . . . , A.L. Price (2015), 'Partitioning heritability by functional category using GWAS summary statistics', *bioRxiv*.
- Freese, J. and S. Shostak (2009), 'Genetics and social inquiry', *Annual Review of Sociology*, **35**, 107–128.

- Fry, A., T.J. Littlejohns, C. Sudlow, N. Doherty, L. Adamska, . . . , N.E. Allen (2017), 'Comparison of sociodemographic and health-related characteristics of UK biobank participants with those of the general population', *American Journal of Epidemiology*, **186**, 1026–1034.
- Galton, F. (1869), *Hereditary Genius: An Inquiry into Its Laws and Consequences*, New York, NY: Macmillan.
- Galton, F. (1874). 'On men of science, their nature and their nurture', *Proceedings of the Royal Institution*, **7**, 227–236.
- Grotzinger, A.D., M. Rhemtulla, R. de Vlaming, S.J. Ritchie, T.T. Mallard, . . . , E.M. Tucker-Drob (2019), 'Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits', *Nature Human Behaviour*, **3**, 513–525.
- Hechter, M. and S. Kanazawa (1997), 'Sociological rational choice theory', *Annual Review of Sociology*, **23**, 191–214.
- Hellenthal, G., G.B.J. Busby, G. Band, J.F. Wilson, C. Capelli, D. Falush, and S. Myers (2014), 'A genetic atlas of human admixture history', *Science*, **343**, 747–751.
- Herd, P., M.C. Mills, and J.B. Dowd (2021), 'Reconstructing sociogenomics research: Dismantling biological race and genetic essentialism narratives', *Journal of Health and Social Behavior*, 002214652110186.
- Hill, W.D., N.M. Davies, S.J. Ritchie, N.G. Skene, J. Bryois, . . . , I.J. Deary (2019), 'Genome-wide analysis identifies molecular systems and 149 genetic loci associated with income', *Nature Communication*, **10**, 5741.
- Hunter, P. (2010), 'The psycho gene', *EMBO Reports*, **11**, 667–669.
- Iacono, W.G., S.M. Malone, and M. McGue (2008), 'Behavioral disinhibition and the development of early-onset addiction: Common and specific influences', *Annual Review Clinical Psychology*, **4**, 325–348.
- James, P.T., R. Leach, E. Kalamara, and M. Shayeghi (2001), 'The worldwide obesity epidemic', *Obesity Research*, **9**, 228S–233S.
- Jang, S.-K., G. Saunders, M. Liu, Y. Jiang, D.J. Liu, and S. Vrieze (2020), 'Genetic correlation, pleiotropy, and causal associations between substance use and psychiatric disorder', *Psychological Medicine*, 1–11.
- Joslyn, C.S. (1927), 'Sorokin on social mobility', *The Quarterly Journal of Economics*, **42**, 130–139.
- Karlsson Linnér, R., P. Biroli, E. Kong, S.F.W. Meddens, R. Wedow, . . . , J.P. Beauchamp (2019), 'Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences', *Nature Genetics*, **51**, 245–257.
- Keller, M.C. (2014), 'Gene  $\times$  environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution', *Biological Psychiatry*, **75**, 18–24.
- Kevles, D.J. (1995), *In the Name of Eugenics: Genetics and the Uses of Human Heredity*, Cambridge, MA: Harvard University Press.
- Lee, J.J., R. Wedow, A. Okbay, E. Kong, O. Maghziyan, . . . , D. Cesarini (2018), 'Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals', *Nature Genetics*, **50**, 1112–1121.
- Liu, H. and G. Guo (2015), 'Lifetime socioeconomic status, historical context, and genetic inheritance in shaping body mass in middle and late adulthood', *American Sociological Review*, **80**, 705–737.
- Martin, A.R., C.R. Gignoux, R.K. Walters, G.L. Wojcik, B.M. Neale, . . . , E.E. Kenny (2017), 'Human demographic history impacts genetic risk prediction across diverse populations', *The American Journal of Human Genetics*, **100**, 635–649.
- Martin, A.R., M. Kanai, Y. Kamatani, Y. Okada, B.M. Neale, and M.J. Daly (2019), 'Clinical use of current polygenic risk scores may exacerbate health disparities', *Nature Genetics*, **51**, 584–591.
- Martschenko, D., S. Trejo, and B.W. Domingue (2019), 'Genetics and education: Recent developments in the context of an ugly history and an uncertain future', *AERA Open*, **5**, 233285841881051.
- Mathieson, I., J.R.B. Perry, and M.C. Mills (2020), 'Genome-wide analysis for reproductive success highlights contemporary natural selection', bioRxiv preprint: <https://www.biorxiv.org/content/10.1101/2020.05.19.104455v1>
- Mills, M.C., N. Barban, and F.C. Tropf (2020), *An Introduction to Statistical Genetic Data Analysis*, Cambridge, MA: MIT Press.
- Mills, M.C. and C. Rahal (2019), 'A scientometric review of genome-wide association studies', *Communications Biology*, **2**, 1–11.
- Mills, M.C. and C. Rahal (2020), 'The GWAS Diversity Monitor tracks diversity by disease in real time', *Nature Genetics*, **52**, 242–243. doi: 10.1038/s41588-020-0580-y.
- Mills, M.C. and F.C. Tropf (2020), 'Sociology, genetics and the coming of age of sociogenomics', *Annual Review of Sociology*, **46**, 553–581.
- Mills, M.C., F.C. Tropf, D.M. Brazel, N. van Zuydam, A. Vaez, . . . , T.H. Pers (2021), 'Identification of 371 genetic variants for age at first sex and birth linked to externalising behaviour', *Nature Human Behaviour*, **5**, 1717–1730.
- Monroe, S.M. and A.D. Simons (1991), 'Diathesis-stress theories in the context of life stress research: Implications for the depressive disorders', *Psychological Bulletin*, **110**, 406–425.

- Morales, J., D. Welter, E.H. Bowler, M. Cerezo, L.W. Harris, . . . , J.A. MacArthur (2018), 'A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS Catalog', *Genome Biology*, **19**, 21.
- Neel, J.V. (1962), 'Diabetes mellitus: A "thrifty" genotype rendered detrimental by "progress"?', *American Journal of Human Genetics*, **14**, 353–362.
- Pers, T.H., J.M. Karjalainen, Y. Chan, H.-J. Westra, A.R. Wood, . . . , L. Franke (2015), 'Biological interpretation of genome-wide association studies using predicted gene functions', *Nature Communications*, **6**, 5890.
- Pirastu, N., M. Cordioli, P. Nandakumar, G. Mignogna, A. Abdellaoui et al. (2021), 'Genetic analyses identify widespread sex-differential participation bias', *Nature Genetics*, **53**, 663–671.
- Purcell, S. (2002), 'Variance components models for gene–environment interaction in twin analysis', *Twin Research and Human Genetics*, **5**, 554–571.
- Raub, W., V. Buskens, and M.A.L.M. van Assen (2011), 'Micro-macro links and microfoundations in sociology', *Journal of Mathematical Sociology*, **35**, 1–25.
- Rosenquist, J.N., S.F. Lehrer, A.J. O'Malley, A.M. Zaslavsky, J.W. Smoller, and N.A. Christakis (2015), 'Cohort of birth modifies the association between FTO genotype and BMI', *Proceedings of the National Academy of Sciences*, **112**, 354–359.
- Scarr-Salapatek, S. (1971), 'Race, social class, and IQ', *Science*, **174**, 1285–1295.
- Schuster, S.C., W. Miller, A. Ratan, L.P. Tomsho, B. Giardine, . . . , V.M. Hayes (2010), 'Complete Khoisan and Bantu genomes from southern Africa', *Nature*, **463**, 943–947.
- Shanahan, M.J. and S.M. Hofer (2005), 'Social context in gene–environment interactions: Retrospect and prospect', *Journal of Gerontology B: Psychological Social Science*, **60**, 65–76.
- Sorokin, P. (1927), *Social Mobility*, New York, NY: Harper and Brothers.
- South, S.C. and R.F. Krueger (2008), 'Marital quality moderates genetic and environmental influences on the internalizing spectrum', *Journal of Abnormal Psychology*, **117**, 826–837.
- South, S.C. and R.F. Krueger (2013), 'Marital satisfaction and physical health: Evidence for an orchid effect', *Psychological Science*, **24**, 373–378.
- Tabery, J. (2014), *Beyond Versus: The Struggle to Understand the Interaction of Nature and Nurture*. Cambridge, MA: MIT Press.
- Tropf, F.C., R.M. Verweij, P.J. van der Most, G. Stulp, A. Bakshi, . . . , M.C. Mills (2017), 'Hidden heritability due to heterogeneity across seven populations', *Nature Human Behaviour*, **1**, 757–765.
- Tucker-Drob, E.M. and T.C. Bates (2016), 'Large cross-national differences in gene  $\times$  socioeconomic status interaction on intelligence', *Psychological Science*, **27**, 138–149.
- Turkheimer, E., A. Haley, M. Waldron, B. D'Onofrio, and I.I. Gottesman (2003), 'Socioeconomic status modifies heritability of IQ in young children', *Psychological Science*, **14**, 623–628.
- Visscher, P.M., W.G. Hill, and N.R. Wray (2008), 'Heritability in the genomics era—concepts and misconceptions', *Nature Reviews Genetics*, **9**, 255–266.
- Walter, S., I. Mejia-Guevara, K. Estrada, S.Y. Liu, and M.M. Glymour (2016), 'Association of a genetic risk score with body mass index across different birth cohorts', *Journal of the American Medical Association*.
- Wedow, R., M. Zacher, B.M. Huibregtse, K. Mullan Harris, B.W. Domingue, and J.D. Boardman (2018), 'Education, smoking, and cohort change: Forwarding a multidimensional theory of the environmental moderation of genetic effects', *American Sociological Review*, **83**, 802–832.