Wrestling with Social and Behavioral Genomics: *Risks, Potential Benefits, and Ethical Responsibility*

BY MICHELLE N. MEYER, PAUL S. APPELBAUM, DANIEL J. BENJAMIN, SHAWNEEQUA L. CALLIER, NATHANIEL COMFORT, DALTON CONLEY, JEREMY FREESE, NANIBAA' A. GARRISON, EVELYNN M. HAMMONDS, K. PAIGE HARDEN, SANDRA SOO-JIN LEE, ALICIA R. MARTIN, DAPHNE OLUWASEUN MARTSCHENKO, BENJAMIN M. NEALE, ROHAN H. C. PALMER, JAMES TABERY, ERIC TURKHEIMER, PATRICK TURLEY, AND ERIK PARENS

Executive Summary

Social and behavioral scientists are increasingly frequently collaborating with geneticists or adapting the methods of genetics research to investigate how genomic differences are associated with differences in a wide variety of behavioral and social phenotypes. The huge and varied range of phenotypes investigated in social and behavioral genomics (SBG) research, broadly construed, includes smoking and eating behavior, schizophrenia, attention deficit-hyperactivity disorder (ADHD), a sense of well-being, introversion, risk-taking preferences, income, intelligence, and educational attainment. Researchers study these phenotypes because they believe that doing so can, among other benefits, contribute to more rigorous social (and health) science, which can in turn contribute to more just social policies.

Because, as we detail in part 1, there is such a long history of attempts to use claims about genetic differences in such phenotypes to advance unjust social policies—and because of the potential of such claims to undercut efforts at creating more just social policies—SBG research can be deeply controversial. In this report, we seek to convey what our working group, composed both of scientists who conduct SBG research and of scholars who think critically about it (see boxes 1 and 2), learned from three years of wrestling with the historical, social, and scientific facts relevant to the ethics of SBG research. More specifically, we seek to articulate where a majority of our working group did—and did not—achieve consensus about the issues with which we wrestled.

To understand the risks and potential benefits of SBG research in more depth, our group first had to wrestle with the scientific question, what can genetics tell us about social outcomes as complex as, for example, educational attainment? Addressing that question required us to put SBG research in the context of genetics research more

Michelle N. Meyer, Paul S. Appelbaum, Daniel J. Benjamin, Shawneequa L. Callier, Nathaniel Comfort, Dalton Conley, Jeremy Freese, Nanibaa' A. Garrison, Evelynn M. Hammonds, K. Paige Harden, Sandra Soo-Jin Lee, Alicia R. Martin, Daphne Oluwaseun Martschenko, Benjamin M. Neale, Rohan H. C. Palmer, James Tabery, Eric Turkheimer, Patrick Turley, and Erik Parens, "Wrestling with Social and Behavioral Genomics: Risks, Potential Benefits, and Ethical Responsibility," in *The Ethical Implications of Social and Behavioral Genomics*, ed. Erik Parens and Michelle N. Meyer, special report, *Hastings Center Report* 53, no. 2 (2023): S2-S49. DOI 10.1002/hast.1477

Article Table of Contents

generally. It is easy to forget today how significant it was in the 1960s and '70s for psychologists and behavioral geneticists to demonstrate that genetic differences were making a difference with respect to complex phenotypes like autism and schizophrenia. For the rest of the twentieth century and into the second decade of the twenty-first, there were concerted efforts to identify *which* genetic variants were making a difference and *how* they were doing that. Many of those efforts at explaining how genetic differences were making a phenotypic difference were frustrating to those who undertook them. Beginning in the late 2000s, however, a new tool that is the focus of much of our discussion—polygenic indexes (PGIs)—made it possible at least to begin making *predictions* about future outcomes.

In parts 2 and 3, we suggest that PGIs might not be as useful as some enthusiasts suggest; but neither are they useless, as some critics suggest. Consider one of the most predictive PGIs for a social science phenotype that currently exists, which is based on the fourth in a series of studies of educational attainment, known as "EA4." The predictive power of the EA4 PGI that is attributable to the *causal effects* of genetic variants is only about 5 percent of the total variance in observed differences among individuals. And so, from one perspective, the EA4 PGI explains relatively little about the observed differences among individuals that are attributable to causal genetic effects. But from another, equally important perspective, the predictive power of PGIs can be comparable to environmental variables like family income that researchers commonly use. The predictive power of the EA4 PGI attributable to *associations* between genetic variants and educational attainment is approximately 15 percent. This predictive power can be useful, for instance, as a relatively strong control variable for social and health scientists.

To discuss the risks and benefits of SBG research, we also had to wrestle with the meaning of a key concept that geneticists use, whether they are studying heart disease, schizophrenia, or educational attainment: the concept of a "population." More specifically, genetics research usually entails identifying people who are genetically similar to some reference population, such as one described in the 1000 Genomes Project. Researchers deploy this concept to increase the likelihood that, when they detect an association between a genetic variant and the phenotype they are investigating, the variant is in fact associated with that phenotype, as opposed to being an artifact of factors like human migration.

As is obvious from the fact that geneticists themselves have conceived of and labeled the populations they study in very different ways over the last seventy years, those groupings are not written in nature. Indeed, because there is continuous variation within and between groups, there are no clear breaks between populations. Today, the "populations" that result from this practice of including or excluding people from a given analysis are often, very imperfectly, called "genetic ancestries"—whether they are defined at the continental level (for example, with "European ancestries," or "EUR") or more granularly (for example, as "the Finnish," or "FIN").

Risks and Benefits of SBG Research

Having clarified to some extent what genetic information can—and cannot—teach people about complex phenotypes and having clarified what geneticists mean by "genetic ancestry," we intend for the reader of the report to be better prepared to discuss the risks and potential benefits of SBG research. In part 4, we catalogue risks of SBG research (though not only SBG research) at the individual, group, and social levels. These risks include stigmatization, discrimination in a range of domains, the reification of race and ethnicity as biological concepts, scientifically or ethically inappropriate applications of SBG research, genetic fatalism (the generally inaccurate belief that genetic predispositions make environmental interventions futile), and genetic distractionism (the risk that attention and resources devoted to genes and genomic research will displace attention and resources devoted to more effective environmental interventions).

By contrast, in part 5, we recognize several potential benefits of SBG research, beyond the intrinsic value of better understanding the world and humans' place in it. At a broad level, we note that because genetic (and environmental) factors are important for variation in virtually every human phenotype, failing to consider the role of genes in *some* way means that the scientific record on important phenomena is likely to be incomplete at best and inaccurate at worst. In particular, potential benefits include better understanding *environmental* causes and the *limits* of genomic influence, improving social science and clinical trials by using PGIs as control variables, advancing health research, and, through these more direct benefits, indirectly improving policies.

Distinguishing Justifiable and Unjustifiable SBG Research

Our assessment of the risks and potential benefits of SBG research leads us to conclude that SBG research on a wide range of phenotypes can be worth conducting, funding, and publishing. However, we also articulate two levels of concern in part 6 (for an overview of the working group's findings, see figure 4, "Responsible Behavior in the Context of Sociobehavioral Genomics [SBG] Research," on p. S30). First, we consider SBG research involving "sensitive phenotypes" to be *SBG research of heightened concern*. At a minimum, heightened obligations of responsible conduct and communication apply to research of heightened concern.

Although nearly any social or behavioral phenotype has some potential to be sensitive, we have more concern about studying (and creating PGIs for) some phenotypes than others. These include phenotypes that can be viewed in a society (rightly or wrongly) as being very consequential to social status (for example, obesity, substance-use disorders, intelligence-test scores, educational attainment, income, and criminalized behaviors), phenotypes that are or have historically been part of harmful stereotypes about minoritized groups and threaten to reify the biologization of social identities (such as financial prowess, academic diligence, hysteria, hypersexuality, musical beat synchronization, and athleticism), and phenotypes that are central to a minoritized group's identity (such as sexual orientation, sexual behavior, and gender identity). Recognizing that sensitivity is contingent on time and place and that our working group is limited by its United States-based perspective, we recommend that those assessing SBG research phenotypes do their best to attend to current and likely near-term future factors affecting the sensitivity of phenotypes.

Second, we consider SBG research of the greatest concern to be research (1) on sensitive phenotypes that (2) compares groups defined by (a) race, (b) ethnicity, or (c) genetic ancestry, where genetic ancestry could easily be misunderstood as race or ethnicity ("group-comparison research," for short). All members of the working group have serious doubts about the scientific validity of group-comparison research today regarding SBG phenotypes. Such comparisons would be confounded by different allele frequencies and linkage disequilibrium patterns and by the different environments in which groups live. By "environments," we mean not only the differences of living in, say, China and Finland, but also the different sociopolitical forces even within a geographic area that shape behavioral and social phenotypes. And we all agree that—considering the social risks of group-comparison research—scientific validity should be an ethical precondition of conducting, funding, or publishing it.

However, we disagree both about the likelihood that group-comparison research will ever be sufficiently valid to yield meaningful results and, if it were, about whether scientific validity alone would be enough to justify such research ethically. We note that, even as they disagree on this point, working-group members nevertheless can and do marshal the same commitments to improving human welfare and justice to support their positions.

For some members of our working group, scientific validity is "compelling justification" enough. Those who adopt this perspective have different reasons: Some view the pursuit of scientific knowledge as an absolute value. Others who take the position that scientific validity is jus-

Box 1. About This Project

This consensus report is the result of three years of mutual learning, deliberations, and debates among nineteen scholars with very diverse views about social and behavioral genomics research. The co-principal investigators chose to include in this working group both scholars who conduct social and behavioral genomics research and those who think critically about such research. This disciplinary and perspectival diversity made consensus more challenging to achieve—and indeed, on some points, consensus turned out to be out of reach—but the diversity also helped ensure that the considerable consensus that the group did achieve reflects a sophisticated, multifaceted understanding of both the scientific and ethical issues at stake.

The project was funded by the Robert Wood Johnson Foundation, the Russell Sage Foundation, and the JPB Foundation. Additional support comes from Open Philanthropy and The Gil Omenn and Martha Darling Fund for Trusted and Trustworthy Scientific Innovation, a Hastings Center fund. The content is solely the responsibility of the authors and does not necessarily represent the views of any of the funders.

tification enough do so on pragmatic grounds. In other words, although these members are open, in theory, to the idea that some social science research is too dangerous to justify, they believe that, in practice, it would be difficult or impossible fairly and reasonably to draw such lines. They further suspect that attempts to set limits on research are likely to cause more overall harm than good and that justice will sometimes be advanced by group-comparison research—for instance, by better understanding how groups differ genetically so that environmental interventions can be tailored to these differences.

A second group of working-group members, while acknowledging the importance of scientific knowledge and freedom, emphasizes that these are not absolute values but must be balanced with others, including welfare and justice. These members think that, in almost all cases, the social risks of the research would outweigh the potential benefits and that justice would best be served by abstaining from group-comparison research concerning sensitive SBG phenotypes. However, they leave open the possibility that, in rare cases, the ethical analysis could turn out differently. Assuming that a study could meet the precondition of yielding scientifically valid conclusions, these members would require that the study have a sufficiently favorable risk-benefit profile. Given the highly contextual nature of research risks and potential benefits, a case-by-case assessment would be required, especially in light of the fact that research risk is generally assessed not in isolation but, rather, in comparison to existing risks (of stigma and discrimination, for example), which are not static.

In summary, and as depicted in figure 4, we all agree that group-comparison research requires a compelling justification of the study's scientific validity. While some of us believe that researchers should be free to pursue any scientifically valid research, others of us would additionally require a compelling justification of the study's risk-benefit profile. We all recommend that, absent the relevant compelling justification(s)—a criterion that some of us think will never be met—researchers not conduct, funders not fund, and journals not publish research on sensitive phenotypes that compares groups defined by race, ethnicity, or genetic ancestry, where genetic ancestry could easily be misunderstood as race or ethnicity.

Responsible Conduct and Communication of SBG Research

Although all research, especially with human participants, should be responsibly conducted and communicated, in part 7, we call for special attention to these matters in SBG research of heightened concern (see figure 4). Needless to say, stakeholders involved with any SBG research of greatest concern—including funders, journal editors and reviewers, and the media—should also adopt these practices, as applicable. But because researchers are central agents in all aspects of research, we direct our recommendations to them in the first instance.

With respect to responsible research conduct, we recommend that researchers engage people about whom the study pertains to (including but not limited to those who provide study data); be clear in their own minds about why they are using membership in a group or "population" as an inclusion or exclusion criterion (or otherwise) in their studies and make sampling choices that reflect that intention; justify how they define and measure the phenotypes under study; conduct only adequately powered studies; replicate their findings in hold-out samples; whenever possible, conduct adequately powered within-family analyses; and work with the rest of the research community to ensure that any benefits of SBG research and PGIs extend to all.

With respect to responsible communication of SBG research in scientific papers, we recommend that researchers, in either the main text or, as necessary, a supplement, explicitly describe and justify their methods for defining any groups or "populations," including whether (and if so, how) the researchers controlled for confounding variables, and explicitly distinguish among race, ethnicity, genetic ancestry, and other group or population terms; work toward language for describing human populations that reflects the continuum of genetic diversity and makes these populations less easily conflated with race or ethnicity; report effect sizes in the abstract and avoid graphs that exaggerate them; and embed caveats and context in graphs and tables to make it more difficult for them to be misappropriated; and develop a salient "key-points" box that conveys how the results should—and should not—be interpreted and used.

Finally, scientific results are often communicated in other ways and to other audiences, including via press releases, frequently-asked-questions (FAQs) documents, websites, videos, and social media. In whatever form dissemination takes, research results should not be hyped, and warnings should be included against misinterpretation and misuse by other scientists and nonscientists, including the media, policy-makers, practitioners, and members of the public.

As difficult as the problems of science literacy and clear communication about complex science are, we end part 7—and our report—by acknowledging a further problem that is at least as difficult: different people can and do bring different values to the same set of facts. Therefore, once researchers have fulfilled their duty of responsibly conducting and communicating SBG research, there remain the potential harms that do not result from a misrepresentation, misinterpretation, or misunderstanding of facts but from invidious values.

Much as we cannot offer a simple algorithm for weighing the potential harms and benefits of any given SBG protocol, we cannot offer a simple solution to the hard problem regarding invidious values. But recognizing why and how SBG research raises questions that demand to be wrestled with is, we think, in itself an important step in the right direction. Meanwhile, we hope that our description of the historic, scientific, and ethical terrain and our recommendations for the responsible conduct and communication of SBG research will be useful to others as they wrestle with social and behavioral genomics research.

Introduction to the Working Group and the Aims of This Report

Social and behavioral scientists are increasingly frequently collaborating with geneticists, or adapting the methods of genetics research, to investigate how genomic differences are associated with differences in a wide variety of behavioral and social phenotypes. We take social and behavioral genomics (SBG) research to include not only phenotypes such as sense of well-being, introversion, risk-taking preferences, income, intelligence, and educational attainment, but also what some would consider "clinical phenotypes" that include substantial social or behavioral components or that are defined via behavioral checklists, such as smoking and eating behavior, schizophrenia, and ADHD.[1] Researchers study these phenotypes because they believe that doing so can, among other benefits, contribute to more rigorous social (and health) science, which can in turn contribute to more just social policies.[2]

Given the long history of attempts to use claims about genetic differences in such phenotypes to deprive people of opportunity, liberty, and even life, some readers' initial impulse might be to say that SBG research should simply be halted. That response, however, is neither practical nor wise. First, practically speaking, such research is already under way across the globe, using the same DNA samples and the same methods—including the one we will focus on in this report, the polygenic index (PGI),[3] also known as a polygenic risk score (PRS) or a polygenic score (PGS)[4]—as those used to conduct medical research. Social and behavioral phenotypes such as educational attainment and income are among the social "determinants" of health; understanding these phenotypes, including whether and how genes are associated with or partly influence them, may therefore be important to understanding health. Moreover, educational attainment, intelligence, and personality reflect (among other things) brain function and have genetic and phenotypic correlations with conditions such as schizophrenia, neurocognitive disorders, substance use, and eating disorders. It is therefore difficult to draw a line between medical and SBG phenotypes. Even if some biobanks did draw such a line, others would almost certainly draw it differently, or not at all, leaving avenues open for SBG research. Even if all biobanks prohibited the use

Box 2. The Working Group

Co-principal investigators

- Erik Parens
  The Hastings Center
- Michelle N. Meyer
  Geisinger

Steering committee

- Paul S. Appelbaum
  Columbia University
- Sandra Soo-Jin Lee
  Columbia University
- Daphne Oluwaseun Martschenko
  Stanford University

Additional members of the working group

- Daniel J. Benjamin
  University of California, Los Angeles
- Shawneequa L. Callier
  George Washington University
- Nathaniel Comfort
  Johns Hopkins University
- Dalton Conley
  Princeton University
- Jeremy Freese
  Stanford University
- Nanibaa' A. Garrison
  University of California, Los Angeles
- Evelynn M. Hammonds
  Harvard University
- K. Paige Harden
  University of Texas at Austin
- Alicia R. Martin
  Harvard University
- Benjamin M. Neale
  Harvard University
- Rohan H. C. Palmer
  Emory University
- James Tabery
  University of Utah
- Patrick Turley
  University of Southern California
- Eric Turkheimer
  University of Virginia

of their data for studying SBG phenotypes, many of the major datasets are already in wide circulation, which means that SBG research would be left to those who are least likely to responsibly (or even competently) conduct or communicate the results. Nor would an absolutist position against any and all SBG research be wise. In addition to using SBG PGIs to better understand and predict various health outcomes, as previously noted, social scientists use PGIs to control for genetic differences in research, and they hope that such improved research methods can eventually contribute to creating more effective social programs, at less cost and in less time. Because saying no to SBG research, even if it were practical, would mean forgoing some potential benefits in the context of medical and social science research, we should weigh the potential benefits against the risks of particular SBG studies.

To wrestle with both the risks and the potential benefits of SBG research, the co-principal investigators (co-PIs) (Erik Parens and Michelle Meyer) formed a working group made up of scientists who conduct SBG research and scholars who think critically about such research (see box 2). The co-PIs had reason to hope that, despite the very different disciplinary backgrounds and lived experiences of the working-group members, it would be possible to engage in a productive dialogue. That hope was anchored in the recognition that all of the working-group members, despite their different views about the potential risks and benefits of the science, shared at least one fundamental political view: that our society is afflicted by intolerable discrimination and inequality.

Some people in the working group who are critical of this science tend to base their critique on the grounds that SBG research could exacerbate those conditions. Some people in the working group who tend to be enthusiastic about the science are supportive, at least in part, on the grounds that the science can be used in efforts to better understand and perhaps mitigate those conditions. None of us agrees with those who appeal to genetics as a justification for the status quo or as an excuse for not intervening to mitigate the social injustices that plague our society. We might have different intuitions about how much genotype matters for things like educational attainment, but we all agree that many social structures are, to a huge extent, a function of the desire of some to advance their own interests over the interests of others, and that, to the extent that genetic luck does exert an effect, its effects are refracted through, and contingent upon, those structures.

The discussions of our working group gravitated toward one area of SBG research that has been in the public eye: the use of PGIs to study educational attainment, where "educational attainment" refers to the number of years people go to school. Beyond the gravitational pull that educational-attainment research had on our attention, there were reasons for us to use it as an example throughout our discussions. First, among SBG studies, educational attainment is among the most extensively investigated phenotypes. Second, it is among the most fundamental social science variables, especially for researchers interested in social inequality. Finally, because some of the researchers who have led the study of educational attainment are members of the working group, those working-group members who were not SBG researchers could be helped to acquire a solid-enough understanding of the research to be able to offer an informed analysis of its potential risks and benefits.

In this report, we seek to convey what our working group learned from our three years of wrestling[5] with the historical, social, and scientific facts relevant to the ethics of SBG research. More specifically, we seek to articulate where a majority of our working group did—and did not—achieve consensus about the issues that arise in the context of SBG research.

PART 1: HISTORICAL BACKGROUND

Francis Galton, nineteenth-century polymath and half-cousin of Charles Darwin, sought to weigh and separate the effects of nature and nurture on human social behavior. Galton, a pioneer of population studies, and his students made fundamental contributions to important concepts in statistics still in wide use today, including regression, correlation, standard deviation, and tests of statistical significance.[6] After the publication in 1859 of Darwin's *Origin of Species*, Galton became fascinated with questions of heredity. He was most drawn to what today are called "complex"—that is, caused by multiple genetic and nongenetic factors—and continuously varying traits: the classic example is physical stature, but he considered personality traits such as intelligence, talent, and character particularly interesting and socially relevant. He was among the first to conceive of teasing apart the effects of nature from those of nurture—separating heredity and environment experimentally (or at least statistically). His concept of the "stirp" or "root" of heredity is the origin of the contemporary notion of the germline. Hereditary traits, he argued, are passed from one generation of gametes to the next. Only the "somatic," or body, tissues are affected by the world outside; characteristics acquired during one's life do not enter the stirp and hence are not passed down.[7]

Galton is also known as the "father of eugenics." Like many intellectuals in imperial countries, he feared the English were "degenerating" in terms of intelligence, courage, leadership, and strength. But whereas his contemporary, Herbert Spencer, favored a laissez-faire approach—eliminating social services, safety nets, welfare, and taxes so that the "least fit" would be less likely to survive and reproduce (an approach called "social Darwinism")—Galton devised a scheme of incentives to encourage society's "fittest" to reproduce more.

In the mid-nineteenth century, the belief became widespread that the healthiest, strongest, brightest, and most successful were reproducing at far lower rates than the feeble, foolish, and immoral. Many social thinkers of the time worried that, without intervention, society would degenerate into mindless, violent chaos. Galton's solution to this was a voluntary program in which society's wealthiest and most brilliant, distinguished, and moral individuals would receive incentives to have more children, while those perceived as "unfit" would be discouraged from reproducing. In 1883, he called his scheme "eugenics," from the Greek for "well-born."[8] "What an extraordinary effect might be produced on our race, if its object was to unite in marriage those who possessed the best and most suitable natures, mental, moral, and physical!" he exclaimed in 1865. "What a galaxy of genius might we not create!"[9]

It is common to try to separate Galton's genuinely pioneering scientific" work on statistics and heredity from his benighted "political" studies on eugenics. But all of Galton's most important work on populations was intimately tied to his ideas about eugenics, and vice versa.[10] For example, true to the colonial British empire of which he was an avid exponent, Galton believed that there were natural and fixed hierarchies between the people of different nations—such as between the British and the people they colonized in South Asia and Africa. As a young man, he measured and counted his way across southern Africa, cataloguing the differences he encountered in height, weight, facial and body structure, and so forth. He took it as self-evident that the British "race" was more intelligent, industrious, gentle, and honorable than any other. He took it for granted that an educated British gentleman was the pinnacle of humanity, as did nearly every other educated British gentleman of the time.[11]

In 1900, the rediscovery of Gregor Mendel's breeding experiments with garden peas was a huge boon for eugenics. It led quickly to the development of a new science of heredity, called "genetics" in 1905 by William Bateson.[12] Whereas Galton had studied complex, continuously varying traits in panmictic (freely interbreeding) human populations, Mendel studied binary, yes-or-no traits (wrinkled seeds or round, yellow seeds or green, and so forth) under conditions of controlled breeding. The discovery of Mendelian analysis, wrote Bateson in 1912, "opens up a new world of physiology . . . organisms may be regarded as composed to a great extent of separate factors, by virtue of which they possess their various characters or attributes. These factors are detachable and may be recombined in various ways. It thus becomes possible to institute a factorial analysis of an individual."[13] Hence, the idea of parsing a human being into its constituent traits, identifying the biological substrates of those traits, and constructing a quantitative profile of the individual has roots more than a century deep.

Mendelian Eugenics in the Progressive Era

The one gene–one trait idea held enormous appeal for those who wanted to apply hereditary science to human amelioration. It helped to create a zealous new eugenics that found fuel in American nativism and Progressive Era reform ideals.[14] In 1910, the zoologist Charles Davenport launched the first scientific research institute for eugenics, the Eugenics Record Office at Cold Spring Harbor, on New York's Long Island.[15] Through the 1910s and early 1920s, many if not most professional geneticists supported eugenics as a practical application of their science, as did many political progressives, including Teddy Roosevelt, W. E. B. DuBois, and Margaret Sanger.[16] Eugenics became a fad, with exhibits in museums and at exhibitions and "better baby" and "fitter family" contests at state fairs.

Social risks abound when using genetics to understand human behavior. Critical history provides a powerful set of tools to orient us and to avoid the social dangers as science navigates these shoal waters.

Like Galton's Victorian eugenicists, the Progressive Era eugenicists were interested in a broad range of human traits, and their greatest interests, like Galton's, lay in social (or rather, antisocial) behavior and, above all, intelligence. The IQ test had just been introduced in the United States from France, and eugenicists latched onto it. The psychologist Henry Herbert Goddard concluded that "feeblemindedness," a broad term that covered all degrees of perceived subnormal intelligence, was a "unit character"—Bateson's term for a trait with simple Mendelian inheritance. A fixation with intelligence, spurred in part by the invention and spread of IQ tests, made it easy for eugenicists to convince themselves that almost all social problems—which for them included criminality, poverty, promiscuity, homosexuality, drunkenness, and so forth—stemmed from feebleminded-ness. A concerted effort to eliminate the "feeblemindedness gene," it seemed, would reduce suffering and improve the quality of life across American society.[17] No federal eugenic sterilization law was ever passed in the United States, but the 1927 Supreme Court decision in *Buck v. Bell* established the constitutionality of state eugenics laws.[18] By the end of the 1930s, well over half of American states had eugenics laws on the books. A signal feature of most of them was the provision for the sexual sterilization, coercively, if necessary, of anyone deemed "unfit." Tens of thousands were sterilized under American eugenic laws.[19] In the 1920s and 1930s, German and American eugenicists admired one another's "progress" in improving the "racial hygiene" of their respective populaces. Adolf Hitler claimed to have profited from the American eugenics literature, and in 1932, he implemented a national law that set up eugenics courts to determine whether a given person was to be sterilized, while the *Lebensborn* program sought to improve the Aryan germ plasm. Nazi activity was the apogee of national eugenic activism, although eugenic fervor, in its heyday, was surely as high in the United States as in Germany.

Hereditarianism after World War II in the United States

It is widely believed that American eugenics ended after the Second World War, due to repulsion at Nazi eugenics and the development of "legitimate" medical genetics. However, this is far from true. In North Carolina and Virginia, for example, sterilization rates *increased* in the 1950s.[20] Eugenics in the 1950s was fraught with moral weight, but there was scarcely a pioneer of the young science of medical genetics who did not believe that eugenics would be a good thing, in principle, if a moral way could be found to go about it.[21]

Further, the control of reproduction is not the only way to invoke genetics to justify prejudice or subjugate others. In 1969, the psychologist Arthur Jensen argued that compensatory education programs, begun in the mid-sixties to redress racial inequalities, had "failed." Black people, he hypothesized, were genetically deficient in powers of abstract reasoning.[22] This innately lower ability, Jensen argued, led to lower achievement, lower income, and lower overall social status. Hence, Jensen argued that efforts to improve the lot of (disproportionately Black) disadvantaged children by treating them as if they were "like" (disproportionately White) more advantaged children were doomed to fail. In the 1970s, the psychologist Richard Herrnstein made a similar argument about IQ and social class.[23] In 1994, he and the political scientist Charles Murray published *The Bell Curve: Intelligence and Class Structure in American Life*.[24] "Bell curve" was an unmistakable reference to Galton, whose work on complex traits presumed a "normal," bell-shaped distribution. Its infamous chapter 13, while acknowledging that racial IQ gaps were not solely caused by genetics and that existing evidence did not resolve debates over the respective roles of genes and environments, engaged in lengthy arguments suggesting a partial genetic cause and did not adequately acknowledge the profoundly different environments society has created for Black and White individuals. Such academic speculation about a partial genetic cause of racial IQ gaps is used today by White supremacists.[25]

Hence, social risks abound when using genetics to understand human behavior—particularly when trying to use that knowledge to solve social problems. Critical history provides a powerful set of tools to orient us and to avoid the social dangers as science navigates these shoal waters. Although the authors of this report differ in how skeptical or enthusiastic we are about the potential social benefits of SBG research, we are unanimous in the view that, as the research proceeds, it must be accompanied by continual

reflection and critical discussion to maximize the benefits and minimize the risks.

We wish to be clear that studying the genetic aspects of human behavior is not an inherently status quo-justifying activity.[26] Nor does the desire to understand the genetic contribution to behavior entail a belief that genes are what matter most. Before we offer our analysis of the potential risks and benefits associated with SBG research (in parts 4 and 5), in the next two parts we will describe, in broad strokes, how geneticists and social scientists today go about trying to understand the nature of the relationship between genes and behavior—and even social outcomes like educational attainment. This will include a very brief introduction to twin studies, which do not include analyses at the level of DNA, through to genome-wide association studies and the creation of PGIs, which do include analyses at the level of DNA.

PART 2: FROM TWIN STUDIES TO POLYGENIC INDEXES (PGIs): *Seeking to Understand the Relationship between Genetic Differences and Phenotypic Differences*

In part 3 of this report, we will describe how contemporary SBG researchers use PGIs to investigate the relationships among genetic variants, environments, and phenotypic differences in educational attainment. First, though, in this second part of the report, we briefly describe the history of scientific efforts over the last half century that have led to the creation of PGIs. By the time we begin to discuss the potential risks and benefits of using PGIs to study social outcomes in part 4, we intend to have given the reader a sense of what it is—and is not—reasonable to expect PGIs to reveal about the relationship between genetic differences and observed differences with respect to any trait, from coronary artery disease to educational attainment.

To help the reader track this historical description leading up to the creation of PGIs, we should be explicit about a pair of temptations we aim to resist. These temptations are common but pull in opposite directions. The first, to which enthusiasts of SBG research can succumb, is to mistakenly slip from evidence that genes have *at least some* causal effects to the conclusion that there is knowledge about *which* genes have those effects or *how* those genes have an effect. In fact, evidence that genes are making a difference tells people nothing about the *mechanism* by which genes have that effect.

The second temptation, to which critics of SBG research can succumb, is to suggest that the correlations detected by the methods we discuss are of little or even no legitimate scientific use or, worse, that they are a kind of pseudoscience. In fact, as we will show, although identifying which genetic variants are making a difference with respect to outcomes does not by itself identify the causal mechanism by which they matter, it can contribute to the search for causal mechanisms responsible for outcomes, and those correlations can also, even with the causal mechanisms unidentified, help to make *predictions* of outcomes.

In a word, predictions might not be as useful as enthusiasts sometimes seem to be suggesting, but neither are they as useless as critics sometimes seem to be suggesting.

Finding That Genetic Differences Are Making a Difference with Respect to Phenotypic Differences

Given that many of our reflections in this report will concern the potential risks that attend SBG research, it is important to remember a fundamental benefit that geneticists who study complex phenotypes have offered. (By definition, a complex phenotype is one that is influenced by many genetic and many nongenetic factors.) That contribution—the discovery that not only environments but genetic differences, too, matter for complex phenotypes—is so fundamental that today it is thought of as common sense rather than as a fact established by a line of scientific research.

In 1966, the psychologist Leonard Heston published a now-famous article on the role of the environment in mental health.[27] When Heston published his paper, psychoanalysis was in its heyday, and many psychiatrists and psychologists believed that conditions such as schizophrenia and autism were caused by adverse family environments—in particular, in the case of autism, by withholding and cold "refrigerator" mothers (an idea that, in hindsight, was obviously and overtly sexist). Heston was one of many who poked holes in that theory.

Predictions of social and behavioral outcomes might not be as useful as enthusiasts sometimes seem to be suggesting, but neither are they as useless as critics sometimes seem to be suggesting.

In his article, Heston compared two groups of adults who had been separated permanently from their mothers at birth. The first group of adults was born to mothers who had a diagnosis of schizophrenia. The second (control) group was born to mothers who did not have a schizophrenia diagnosis. Whereas some of the children born to mothers with schizophrenia had themselves developed schizophrenia, none of the children born to mothers without schizophrenia had developed it. Heston's results indicated that what was inherited from the biological mothers (and, presumably, fathers) seemed to be part of the explanation for why some people developed schizophrenia and others did not. More precisely, there seemed to be a correlation, or an "association," between genetic differences and observed differences. (In this report, we use the terms "correlation" and "association" interchangeably.)

Twin studies developed the fundamental idea that there is an association between genetic and phenotypic differences across a huge range of phenotypes. Although twin studies were first conducted in the 1920s (with precedents even earlier), they blossomed in the 1980s and 1990s. Twin studies proceeded from the simple fact (unknown to Galton) that, whereas identical twins are essentially 100 percent genetically similar, fraternal twins are on average only 50 percent genetically similar.

The logic of twin studies is simple. A key assumption is that the similarity in rearing environment experienced by identical twins is on average the same as the similarity in rearing environment experienced by fraternal twins. Under this and several other assumptions, if identical twins are more phenotypically similar than fraternal twins, there is reason to believe that genetic differences are causally related to the phenotypic differences—even though, by themselves, traditional twin studies cannot reveal anything about the details of how those causes operate. And that is what twin studies have shown over and over: genetically identical twins are phenotypically more similar (or "concordant") than are fraternal twins with respect to virtually any trait, from cardiovascular disease and diabetes to obesity and religiosity.[28]

Over the years, there have been myriad efforts to challenge the assumptions at work in twin studies and thereby their conclusions regarding the effects of genetic differences on phenotypic differences.[29] But researchers as critical of behavioral genetics as Eric Turkheimer (a behavioral geneticist and a coauthor of this report), have concluded that those assumptions hold well enough to support the most basic inference that behavioral geneticists have made for more than half a century: that genetic differences are making a difference—to varying extents and *contingent upon environments*—with respect to virtually all observed differences in behavior.[30] In other words, as we discuss further in part 3, how strongly (or weakly) genes are associated with a given complex phenotype depends on the environment.

The informal phrase "genetic differences are making a difference" has been distilled into a scientific concept, and ultimately a number, called a "heritability coefficient." The concept was invented in a 1918 paper by R. A. Fisher, which essentially founded the genetic study of complex phenotypes.[31] Fisher showed that Mendelian genetics could be used to calculate the magnitude of the phenotypic correlations that should be expected between relatives, depending on the extent to which the phenotypic differences are caused by genotypic differences. Twenty years later, the plant and animal geneticist Jay Lush named one of the quantities derived by Fisher a "heritability coefficient" and showed that it is central for understanding how quickly and successfully farmers can select for specific plant and animal phenotypes.[32] The heritability coefficients estimated and used by farmers capture causal relationships between genetic variants and phenotypes. However, because estimating heritability coefficients from twin studies and other family studies in humans relies on strong additional assumptions, there has been debate about how well heritability coefficients estimated in humans capture causal associations. Moreover, because there is no ethical way to selectively breed or raise human beings under controlled conditions, there has been debate, ever since Lush's work, about the usefulness of heritability coefficients for understanding phenotypic differences among human beings.[33] Regardless of debates about the assumptions and ultimate usefulness of estimating heritability coefficients in humans, the finding from twin (and adoption and family) studies—that genetic differences are correlated with (and might or might not be causing) at least some portion of the phenotypic differences—is so basic that Turkheimer enshrined it as the "first law of behavioral genetics": "All behavior is heritable."[34] Indeed, he might have called it the "first law of genetics," insofar as virtually all phenotypes are heritable, whether they are matters of anthropometry (such

as height), physical health (such as coronary artery disease), mental health (such as schizophrenia), behavior (such as impulsivity), or social experience (such as educational attainment).

Seeking to Identify *Which* Genes Are Making a Difference

The twin, adoption, and family studies of the mid- and late-twentieth century indicated that genetic differences somehow were correlated with differences in phenotypes. Again, however, estimates of heritability coefficients by themselves neither rule in nor rule out a causal relationship between genetic and phenotypic differences. Whether a given estimate of a heritability coefficient implies causation depends on the validity of the statistical assumptions made in a particular context; other evidence, from experimental work in model organisms, for instance, might also be brought to bear. To begin to try to specify which—and ultimately how—differences in genomes were responsible for the phenotypic variation would require the arrival of molecular techniques.

Several developments in the 1990s gave human geneticists reason to be optimistic about the search for the specific genes responsible for various human traits. First, an international team of scientists collaborated on the Human Genome Project, which set out to provide a full sequence of the human genome. Having that sequence could provide a firm foundation for geneticists to determine how many genes humans carried, where they were located, and which locations were the sites of naturally occurring genetic variation. Second, medical geneticists managed to successfully identify a number of single genes or gene deletions that caused rare traits, such as cystic fibrosis, neurofibromatosis, and Huntington's disease.[35]

That gave scientists reason to think that more complex traits, such as depression and diabetes and maybe even general intelligence and criminal behavior, could have equally identifiable genetic causes. And, indeed, the 1990s were marked by a series of genetic studies hailed in the press for finding the "gay gene,"[36] the "intelligence gene,"[37] and the "warrior gene."[38] Some of these putative findings were based on "candidate-gene" studies, which often started by selecting a gene with a known neurochemical activity hypothesized to be relevant to the trait under scrutiny. For example, a gene involved in serotonin regulation, it was hypothesized, could account for variation in who develops depression, and so geneticists set out to see if people with depression were more likely to have a variant of that gene in comparison to people without depression. When they reported a positive result, the "depression gene" was added to the list of exciting findings.[39]

By the start of the 2000s, however, it became clear that (with some exceptions, including rare forms of common diseases like breast cancer and Parkinson's) many of those original positive findings about strong associations between single candidate genes and common phenotypes were illusions. The initial findings routinely failed to replicate when different teams of scientists went looking for the same association.[40] What was initially cast as a series of triumphant discoveries came to be seen as systemic publication bias that favored positive results generated by underpowered studies (that is, from samples with too few individuals). As a result, another strategy was needed to better understand the relationship between genetic variants and complex traits, one that was not so prone to the methodological errors of the candidate-gene approach.

From GWASs to PGIs

In the 2000s, after the disappointing results of those early candidate-gene studies, a new molecular approach emerged: the genome-wide association study (GWAS). The advent of GWASs relied upon the success of the Human Genome Project in sequencing human genomes: that is, in specifying the sequence of each of the three billion nucleotide base pairs—those As, Cs, Ts, and Gs that together constitute the chromosomes, which together constitute the human genome. It also relied on the success of research that built on the Human Genome Project, most notably the HapMap Project, which identified the millions of places in the genome where there are "common" variants called "single nucleotide polymorphisms" ("SNPs"), which are defined as variants appearing in at least 1 percent of the world's population.[41] Instead of a C, for example, there is a T, or instead of an A there is a G.

Different from the early candidate-gene studies, which proceeded from a hypothesis about an association between a single gene and a phenotype or outcome, GWASs did not proceed from hypotheses about any given SNP being correlated with an observed difference. Indeed, being "hypothesis free" was one of the virtues hailed by scientists who used the methodology. It was soon discovered that many of the early GWASs,[42] involving relatively small samples of individuals, turned out to have been too underpowered to produce replicable findings.[43] But geneticists then began using much larger sample sizes and were able to detect myriad replicable associations between SNPs and observed differences. GWASs with several thousand participants were augmented with tens of thousands and then hundreds of thousands, and those larger samples provided finer resolution to identify smaller effects, which turned out to be numerous. This success gave to rise to what Christopher Chabris and colleagues called, in 2015, the "Fourth Law of Behavior Genetics": "A typical human behavioral trait is

Some members of our working group are skeptical that GWASs will ever yield much knowledge about genetic causal mechanisms for behavioral or social phenotypes. Others of us are more optimistic.

associated with very many genetic variants, each of which accounts for a very small percentage of the behavioral variability."[44]

Many researchers believe that GWAS findings can provide important clues about causal pathways, which subsequent work can further investigate. Moreover, GWASs have been of enormous value to pharmaceutical companies in identifying promising targets for drug development[45] and repurposing[46] and are routinely used to generate hypotheses about causal pathways. There has been some limited success in learning about causal pathways from GWASs of disease phenotypes. Even fewer insights about causal mechanisms have emerged from GWASs of psychiatric phenotypes and fewer still from GWASs of nonpsychiatric behavioral phenotypes.[47] Some members of our working group are skeptical that GWASs (even with much improved data and methods) will ever yield much knowledge about genetic causal mechanisms for behavioral or social phenotypes. In their view, behavioral and social phenotypes seem likely to be too far removed from the biology that is directly influenced by genetic variation. Others of us are more optimistic, believing that the length of the causal chain varies across phenotypes and expecting that, as science advances, at least some phenotypes currently understood as "social" or "behavioral" (such as schizophrenia or substance abuse) may, with greater understanding of their genetic basis, turn out to have at least some genetic causal pathways that can be well understood.

Quite aside from the question whether GWASs will yield insights about genetic causal mechanisms, the method by itself permits researchers to identify only correlations between SNPs and phenotypes; GWASs alone do not guarantee that the identified SNPs cause the phenotype. In most cases, even when SNPs are causally related to observed differences, geneticists do not yet understand those relationships any better than they did when the twin researchers established that genetic differences in general matter for phenotypic differences—although researchers using GWASs do have a more focused idea about where to look in the human genome for such causal information.

To respond to the fact that GWASs were detecting SNPs with small effects that, per the fourth law, could each account for only a very small percentage of the observed variation within a population, geneticists undertook a new approach that could make use of those SNPs. As mentioned above, this new approach, which took off in the 2010s, entails the creation of what we are referring to as "polygenic indexes" ("PGIs") and what others have referred to by other labels, including "polygenic risk scores" ("PRSs") and "polygenic scores" ("PGSs").

Although the technological and statistical tools for creating PGIs are complex, the basic idea is simple. Essentially, to create a PGI, researchers assess the magnitude of the correlation between each SNP and the phenotype of interest, as detected in a GWAS, and then add up the magnitudes associated with each of the SNPs.[48]

Correlations, Causal Effects, and Causal Mechanisms

Before we turn to discussing PGIs and their potential for predicting educational attainment, here we want to clarify the language we use in this report to distinguish cases in which there can be said to be a causal relationship between SNPs and phenotypes from cases in which there can be said to be only a correlation between them. Notice, however, that there is nothing "mere" about the potential use of correlations to make predictions or to hypothesize about causes.

Researchers, including those who work with SNPs, sometimes reserve the term "effect" for when they mean a causal relationship, but sometimes they offhandedly use the term when they really mean a correlational relationship, which is not necessarily causal. To avoid eliding the difference between causal relationships and correlations, we will reserve the term "causal effect" to refer to causal relationships, and we will avoid using the term "effect" when referring to correlations.

When we say that a genetic variant has a "causal effect," we are using the term in a way that is standard across the sciences. By definition, in a given environment, a genetic variant has a causal effect if the phenotype would have been different had the genetic variant been different. Since researchers cannot run experiments that compare two human individuals who are the same with respect to every genetic variant except one, this counterfactual cannot be observed for a given individual, and it is thus impossible to know the causal effect of a genetic variant on a given individual. However, the *average* causal effect across individuals in a population could be estimated by comparing the average phenotype across two groups of individuals when one group randomly inherited the genetic variant and the other

did not. (As we discuss below, within-family, as opposed to population, GWASs approximate such a comparison.) When we refer to a genetic variant's causal effect, we are referring to this average effect.

Notice that the definition of a causal effect is completely silent on the mechanism through which changing the genetic variant would have changed the phenotype. In a famous example of a causal effect that operates through a mechanism that involves genes but is not "genetic" or "biological" in the intuitive sense of those terms, consider a genetic variant that causes a person's hair to be red.[49] In a society in which people with red hair are discriminated against, that genetic variant could have a (negative) causal effect on many phenotypes, including how much education a person gets. In the standard framework used by behavior geneticists, this would be counted as a causal effect, even though it operates largely through a social or environmental, rather than biological, mechanism (the redness of the hair is biological, while the response is social or environmental). Some causal effects of genetic variants likely operate through biological mechanisms in the intuitive sense of "biological," but many others, perhaps especially but not only for SBG phenotypes, likely operate through environmental mechanisms, much as a causal effect is imagined to operate in the red hair example. In other words, environmental mechanisms can be baked into the causal effects of genetic variants (and into the causal effects of PGIs that we discuss later).

Population GWASs versus Within-Family GWASs

To appreciate the distinction between what we are referring to as "causal" and "correlational" relationships between SNPs and phenotypes, it is important to recognize the difference between traditional *population GWASs*, which include people who are not members of the same families, and *within-family GWASs*, which include at least two members of each family.

Until recently, most GWASs have included people who are *not* members of the same families and have focused on identifying *correlations* between genetic variants and phenotypes. This is to say that, as we mentioned above, these studies identify genetic variants that are more common in individuals with higher (or lower) values of some phenotypic measure. This study design, by its nature, does not necessarily identify the genetic variants that *cause* the phenotypic differences, because third factors (including, as we will explain below, human migration patterns) could affect both the frequency of genetic variants and the phenotype. GWAS studies typically use sophisticated approaches to try to "control for" such potential confounds, but the confounds cannot be entirely ruled out.

However, as the cost of measuring genetic variation has continued to fall, it has become possible to conduct GWAS not only on individuals who are not part of the same families but also on individuals who are part of the same families.[50] Since the parents' genetic variants can be directly identified (or inferred) in these within-family GWASs, it is possible to identify genetic variants that are randomly inherited by individuals in the same family. Because the genetic variants are effectively randomly assigned, this "natural experiment" study design holds constant other factors. It therefore makes it possible to infer (average) causal effects of SNPs on the phenotype of interest.

Parenthetically, some researchers today refer to what we are calling "causal effects" by the term "direct effects." We are avoiding the latter term based on our concern that it gives the misleading impression that something is known about *how* these SNPs are involved in the emergence of the phenotype—as opposed to having reason to believe *that* these SNPs are involved in likely staggeringly complex causal chains that give rise to the phenotype of interest. (For example, in the hypothetical example we described above, SNPs that cause reduced educational attainment due to their effects on red hair would be described as having "direct effects" on educational attainment—even though the effects are not at all direct in the usual meaning of that word.) Again, knowing *that* these SNPs are causally related is *not* to know anything about the mechanism that explains how genetic variants have the effects they do. That is not to minimize the fact that there are researchers today who believe that some of the SNPs identified in GWASs have led to, and will continue to lead to, insights regarding mechanisms.

Although the family samples available for GWASs are becoming larger (for example, Laurence J. Howe et al. analyze approximately 100,000 sibling pairs[51]), they are still much smaller than the population samples available for GWASs (for example, the educational attainment study by Aysu Okbay et al. that we will discuss in the next section analyzes approximately three million individuals). Moreover, for relatively uncommon conditions (such as schizophrenia), the family samples available for GWASs remain small. Consequently, now and at least for the next few years, the causal effects of genetic variants are and will be identified much less precisely than the genetic correlations. For this reason, the polygenic indexes that we discuss in the next section, which are derived from population GWASs, are currently based on correlations of genetic variants with the phenotype, rather than causal effects.

PART 3: POLYGENIC INDEXES: *General Principles and Application to Educational Attainment*

As discussed in part 2, for the PGIs that are commonly used today, the "effects" of SNPs that are added up to create PGIs are statistical correlations—and by themselves prove nothing about causation. As we alluded to there, it is equally crucial to notice that, although such effects cannot, by themselves, indicate anything about whether or how a given SNP is causally related to a given phenotype, adding up their (usually) tiny weighted correlations can contribute to making *predictions* about the likelihood that someone with a given genomic profile will develop or exhibit a phenotype of interest.

We need to explain here the term from statistics that researchers use when they talk about prediction: R^2 concerns the proportion of the variation in the dependent (or response) variable of a study (such as an outcome of interest like a phenotype) that is statistically accounted for by the independent variable (such as the weighted sum of SNPs captured by a PGI). The value of R^2 ranges between 0 and 1. It is used ubiquitously in the social (and other) sciences to indicate, for instance, the percentage of the observed variance between adults who do and do not exhibit aggression that can be predicted (or "explained") by an abusive childhood environment. This kind of predictive power is not the kind that crystal balls are said to have: independent variables can be better or worse predictors (indeed, the R^2 indicates how weak or strong the predictive power is). Moreover, the prediction is not even necessarily about the future; it can be simply about a current state of affairs (about an independent sample, for instance).

In this statistical sense, PGIs, like environmental variables, can be "predictive." For medical researchers, PGIs are of interest as potentially useful risk indicators in conjunction with other risk indicators to identify patients for whom certain treatments may be more effective or who need more preventive care.[52] In part 6, we will discuss at greater length why some social scientists are interested in PGIs. In this section, however, we are seeking to explain what it means to speak of a PGI for educational attainment.

Beyond the fact that, in general, PGIs today are almost always based on genetic correlations (rather than known causal genetic effects), it is also important to recognize that PGIs created by studying one "genetic ancestral population" cannot be generalized or applied to another genetic ancestral population to make predictions about that population.[53] Understanding this is important for discussions in parts 5 and 6 about the potential risks and benefits of creating PGIs. To explain the generalizability or portability problem, we need to briefly wade into the complex question concerning what geneticists mean when they refer to genetic ancestral populations, a concept that can be correlated with, but is distinct from and often confused with, race or ethnicity.

Genetic Similarity, Genetic Ancestry, and the Problem of Misleading Associations

The scientific term "genetic ancestry" describes how genetic variants are passed down through generations from ancestors to their offspring (notably, if someone traces their lineage back far enough, they find that they have many ancestors from whom they have not inherited any chunks of chromosomes). It is meant to reflect how genetically similar individuals within a group are, and how genetically dissimilar members of that group are to those outside of it, with respect to specific genetic variants, or alleles. Genetic ancestry is usually inferred, often based on individuals' having similar allele frequencies and linkage disequilibrium patterns compared to some reference population, such as the 1000 Genomes populations.[54] (Linkage disequilibrium is the nonrandom association of alleles at different loci in a given population; different evolutionary histories produce different patterns of those nonrandom associations.[55]) People who are genetically similar to that reference population are grouped together into a population, typically called a "genetic ancestry." Although human genetic similarities and dissimilarities are facts of nature, populations—including genetic ancestries—also entail an element of social construction: although population groupings are informed by the statistical aim of balancing sample size and power, researchers also choose the reference populations based on contemporary geographic classifications and who can be conveniently sampled, and researchers make choices about how similar is "similar enough" to comprise a "population." Indeed, the lack of clear and distinct breaks between the frequency of the appearance of gene variants is at least part of why it has been possible to carve up the world's populations in such radically different ways over the last seventy years.[56] The population geneticist Molly Przeworski, recalling the adage that "[a]n academic discipline is a set of individuals who agree not to question the same assumption," has noted that for population genetics, at least, "that assumption is a population."[57]

Geneticists ubiquitously group people into relatively genetically similar populations, or genetic ancestries, whether they are studying heart disease, schizophrenia, or educational attainment. They do so, despite the imprecision of such grouping, to try to reduce the number of associations they might otherwise find (and incorporate into their

GWASs and PGIs) that are misleading with respect to the trait they are studying.

This potential risk of detecting misleading associations when studying people with relatively different allele frequencies and linkage disequilibrium patterns results from the fact that genetic variants can appear with different frequencies in groups of people who have been separated by time and geographical barriers. That is, allele frequencies can change over time due to the random process of genetic drift that occurs when genes are passed from one generation to the next.[58] (Natural selection, too, could play a substantial role in linkage disequilibrium and, for less polygenic phenotypes, allele frequency differences. The evidence to date, however, suggests that natural selection plays a limited role for the polygenic phenotypes that researchers have studied.[59]) When groups are geographically or socially separated with limited opportunities for gene flow, these chance changes in allele frequencies are not shared, and more distinct variations in allele frequencies between groups emerge over time. Of course, such groups may also differ in terms of their environments or cultures. So, if geneticists do not control for these differences, such as by excluding relatively genetically dissimilar participants from their samples, they are at risk of identifying associations between genetic variants (whose allele frequencies randomly drifted apart) and complex traits that are actually due to environmental or cultural differences. In more technical terms, they are at risk of being misled by the fact of "population stratification." That is, if they do not control for population stratification, they are at risk of detecting SNPs that are systematically associated with a given population, but not with the trait within a population.[60]

Today, as we have said, the "populations" that result from this sample inclusion-and-exclusion practice are often called "genetic ancestries."[61] As others have noted,[62] continental-level genetic ancestry, especially, is imperfect as a proxy for genetic similarity. It is also problematic because these populations do not reliably map onto, yet they are easily and erroneously conflated with, social groupings such as race and ethnicity. We call on the genomics community to continue to work toward better human-population descriptors.

The Need to Specify Genetic Populations Cuts in Two Ethical Directions

We should note that this need to specify the genetic ancestral population under study when creating a PGI can both produce and allay ethical concerns. On the one hand, to the extent that PGIs are population-specific and not portable from one genetic ancestral group to another, the creation of PGIs creates the ethical concern that any benefits of such research will not be equitably distributed among all people. This concern arises because creating PGIs depends upon analyzing DNA samples from biobanks, and the largest biobanks today include overrepresentation of people of "European genetic ancestry"—or, more precisely, "European genetic ancestr*ies*."[63] This means, as those who are hopeful about the future of precision medicine point out, that if benefits accrue from PGIs, those benefits will disproportionately benefit such people.[64] That is why there are such strong calls to develop repositories with more diverse and more representative DNA samples.[65]

On the other hand, the fact that PGIs are not portable or generalizable can be of some help in allaying the ethical concern that PGIs will be used for the sake of trying to create invidious comparisons between some groups. At least today, absent strong (and extremely difficult-to-test) assumptions, it is simply not scientifically legitimate to take a PGI created by studying members of one genetic ancestral group and use it to make a comparison between that group and a different genetic ancestral group. (We will return to this crucial issue in part 6.)

And, despite any solace one might find in the scientific illegitimacy of using PGIs to compare groups, it remains true that the creation of PGIs for complex traits that are as valorized as educational attainment raises concerns based on the history recounted in part 1 of this report. Thus, in the next section, we will take a closer look at research into PGIs for educational attainment, which does not entail comparisons between genetic ancestral groups but does entail comparisons of individuals within a continental-level genetic ancestry group.

PGIs for Educational Attainment

Of the PGIs that are available today for anthropometric, medical, behavioral, and social traits, one of the most predictive (one having the highest R^2) is that for educational attainment.[66] The main reason the PGI for educational attainment—typically measured as the number of years of school someone has completed—is among the most predictive PGIs created so far is that the GWASs of educational attainment have had among the largest sample sizes to analyze. The simple reason for those large sample sizes is that the people who contribute their DNA samples to biobanks are far more likely to include information about how many years they went to school than they are to contribute information about virtually any other phenotype. Moreover, education is strongly correlated with many other phenotypes that are of interest to a wide range of researchers, and therefore information about it is ubiquitously collected in research, including biobank studies.

Obviously, in the contemporary United States (and in most societies), educational attainment is associated with a huge range of environmental differences, including living

For the PGIs that are commonly used today, the "effects" of SNPs that are added up to create PGIs are statistical correlations—and by themselves prove nothing about causation.

in poverty or in privilege, going to poorly or well-resourced schools, and being able or unable to stay in school when a family crisis arises. Equally obviously, success in school is correlated with a large range of abilities—including the ability to sit still and focus, to tolerate frustration, and to be agreeable with peers and teachers—that are also influenced by both genetic factors and nongenetic factors (and their interactions). As we noted in part 2, the strength of an association between particular genetic variants and a complex phenotype—and hence the extent of a PGI's predictive power—depends on these and other environmental factors. For instance, some genetic variants are likely associated with educational attainment due to their association with numerous individual and societal factors, from personality traits to education laws. Changes to any of those pathways—changes in the length of compulsory school, for example, or a pedagogical shift from didactic to experiential learning that is more suited to some personality types—may affect which genetic variants are associated with how much education someone receives, the effect size of each variant, and the overall heritability of a complex phenotype like educational attainment.

Finally, although we are emphasizing that educational attainment reflects myriad social and emotional skills and is not at all identical to intelligence, it is correlated with it. By the term "intelligence"—or "general intelligence" or "cognitive ability"—psychologists refer specifically to the ability "to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience."[67] With that term they are *not* referring to many other abilities that are valued in our society, including what some call "interpersonal intelligence" or "emotional intelligence."[68] Nor should it need to be said that being high in "general intelligence" is very different from being thoughtful, empathic, or wise. Nor does being high in intelligence have anything to do with the fundamental assumption that all human beings have equal moral worth. According to a comprehensive review of the literature, by scholars fiercely critical of putting the genetics of intelligence to regressive political purposes, general intelligence is one of the most valid and reliable constructs that psychologists use[69]—that is, they can say what "it" is with more clarity and can measure it more accurately than they can define or measure virtually any other trait they study, from emotional intelligence to extroversion. Thus, in contrast to those who would seek to resist regressive political actors by disputing the validity of the intelligence construct, we seek to resist regressive political actors by pointing out the unscientific and unethical ways in which they deploy that construct.

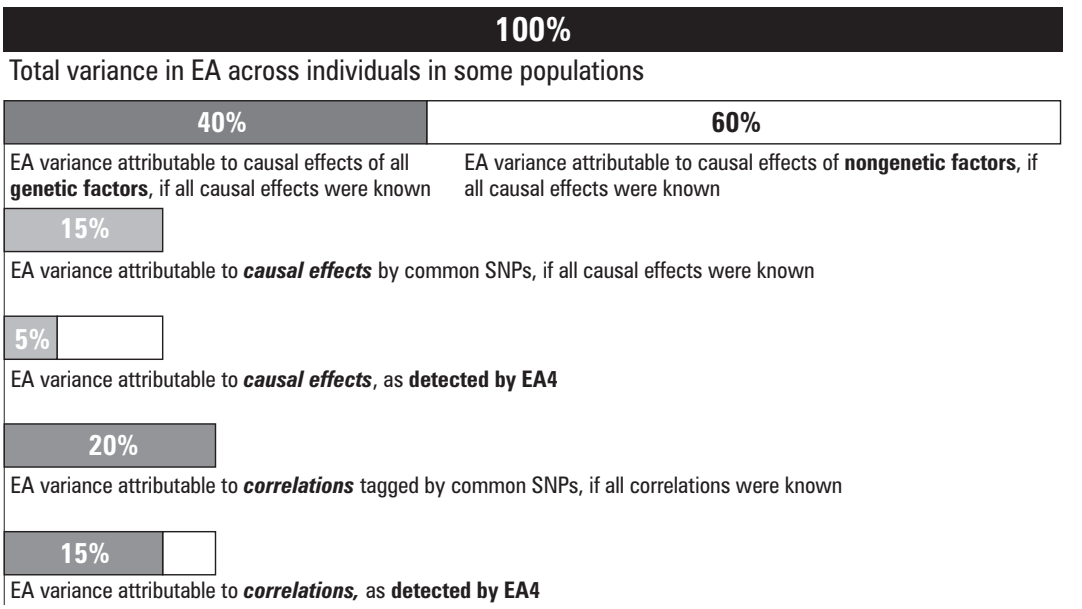
So, again, the most predictive PGIs that have been constructed thus far pertain to educational attainment, which is correlated with, but distinct from, intelligence. Before we say what the PGIs for educational attainment have shown, we should reiterate that these PGIs have been created by analyzing the DNA of people of European genetic ancestries—and therefore would, per above, have substantially reduced predictive power to explain variation among people of other genetic ancestries. At least for today, PGIs created in one genetic ancestral group cannot be used to make scientifically valid comparisons between that group and another one.

Considering Educational Attainment PGIs from Two Different Perspectives

With all the preceding caveats in mind, it can be useful to consider, from at least two perspectives, the most recent of the four major educational attainment GWASs from which PGIs were created.[70] We will refer to this fourth educational attainment study as "EA4." Figure 1 provides a highly simplified way of thinking about what EA4 did and did not detect about the relationship between genetic differences and educational attainment.

Suppose geneticists could identify all the genetic variants (common SNPs, rare SNPs, copy number variants, and so on) that have causal effects on educational attainment and could know their effect sizes with full accuracy. Then twin studies predict that, taken together, these genetic variants would have an R^2 of approximately .40, on average, across populations.[71] (Here and below, in our discussion of twin heritability, "correlational SNP heritability," and "causal-effect heritability," we are offering specific numbers, rounded to increments of .05. Also, to avoid additional complexity in the exposition, we are ignoring adjustments for factors such as assortative mating, gene-gene interactions, and gene-environment correlation (including "genetic nurture"[72]) that would affect these numbers.) Another way of saying that twin studies predict an R^2 of .40 is to say that the causal effects of these genetic variants would account for approximately 40 percent of the total variation in educational attainment among individuals within a population. (The remainder—approximately 60

Figure 1.
Share of Predictive Power of EA4 PGI That Is Attributable to Correlations versus Causal Effects



percent—would be accounted for by nongenetic factors, including environmental variation that is not itself affected by an individual's genes, as well as noise in the measurement of educational attainment.)

If the geneticists identified every common SNP (the sort currently used to create PGIs, which does not include rare SNPs) and could know their causal effects with full accuracy, the R^2 would be approximately .15 on average across populations,[73] predicting approximately 15 percent of the variance; this is called the "causal-effect SNP heritability."[74] The drop from .40 to .15 is, to a large extent, due to the fact that twin studies pick up the effects not just of common SNPs but of all genetic variants (including, for example, rare SNPs and CNVs).[75]

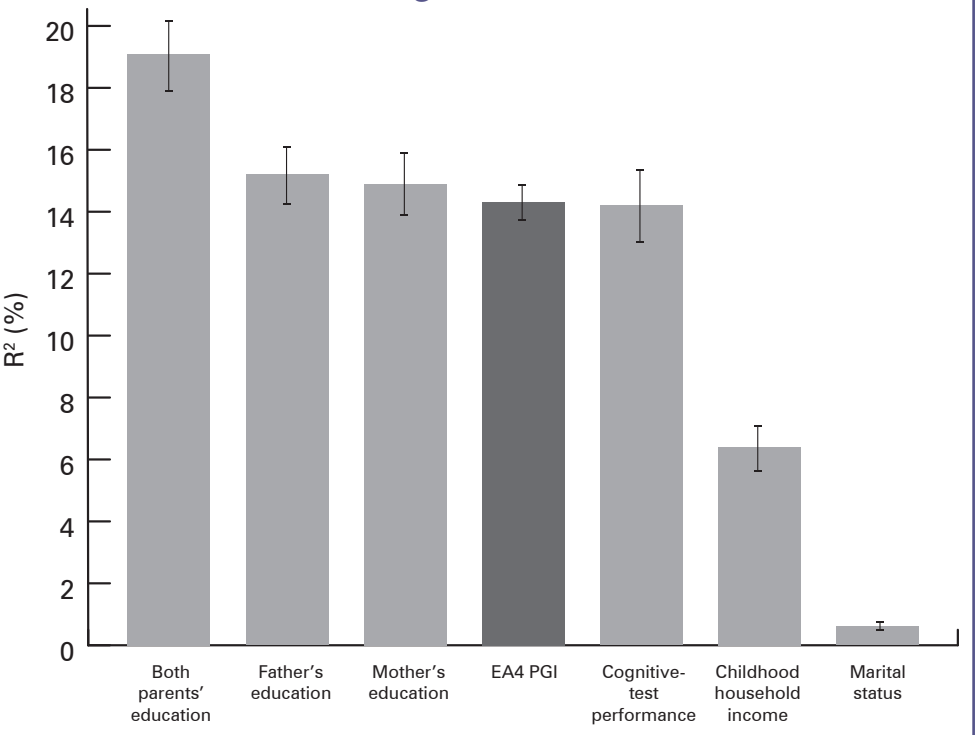
If the geneticists identified every common SNP and could know their correlations with the educational attainment phenotype with full accuracy, the R^2 would be approximately .20,[76] predicting approximately 20 percent of the variance; we are calling this quantity the "correlational SNP heritability." The difference between this .20 and the .15 is, to a large extent, attributable to the fact that correlational SNP heritability includes not only the causal effects of the SNPs but also their associations that are due to gene-environment correlation (including "genetic nurture"). Assortative mating is another factor that makes the correlational SNP heritability larger than the causal effect SNP heritability.

The educational attainment PGI that can account for the highest percent of the total variance (or the PGI with the highest R^2) so far has in fact accounted for approximately 15 percent of the total variation among individuals in education attainment, with only about a third of that, or 5 percentage points, associated with causal effects. (These are the effects that can be detected within sibling pairs and therefore are plausibly causal.) The remaining approximately 10 percentage points are due to an unspecified mix of environmental confounds, including population stratification, various types of gene-environment correlation (including "genetic nurture"), and assortative mating. The correlational SNP heritability detected by EA4—15 percent—is smaller than the predicted ceiling of 20 percent, partly because GWAS samples are not yet large enough to estimate the SNP correlations with educational attainment with sufficient accuracy.[77] Relatedly, the R^2 of the PGI that is due to causal effects is expected to grow beyond 5 percent as GWAS samples get larger and especially if large GWAS are conducted in family samples (eventually getting close to the ceiling causal-effect SNP heritability of 15 percent). Moreover, the ceilings of 20 percent for correlational SNP heritability and 15 percent for causal-effect SNP heritability are expected by some members of our working group to grow larger in the future, when other types of genetic variants that are currently excluded from PGIs are incorporated into the data used to create PGIs.

Being able to account for 5 percent of the total variance among siblings raised in the same family in terms of causal effects of genetic variants can seem to some observers so small as to be a waste of time to discuss. From their perspective, even accounting for 15 percent of the variance in terms of SNPs can seem too little to warrant attention.

But from another perspective, accounting for 15 percent of the variance among individuals raised in different families is of significant interest, even if only some of this predictive power is based on causal effects. (Figure 2 indicates that the PGI is a stronger predictor than income and marital status, about the same as either parent's EA or performance on cognitive tests, and not as strong as using both parents' EA.) In the world of social science research—which day in and day out seeks to identify meaningful correlations—identifying a variable that might account for up to 15 percent of the observed variation is indeed worthy of attention. For a causal variable, even 5 percent of the variance is worthy of attention in the world of social science research. For example, it would be considered moderately large in experimental social psychology, and it is very large compared with most causal effects of environmental interventions that have been identified by field experiments in economics.

Figure 2.
Predictive Value of EA4 PGI in Comparison to Predictive Value of Nongenetic Controls



The EA PGI is from EA4 (A. Okbay et al., "Polygenic Prediction of Educational Attainment within and between Families from Genome-Wide Association Analyses in 3 Million Individuals," *Nature Genetics* 54, no. 4 [2022]: 437-49). The nongenetic predictors are from EA3 (J. J. Lee et al., "Gene Discovery and Polygenic Prediction from a Genome-Wide Association Study of Educational Attainment in 1.1 Million Individuals," *Nature Genetics* 50, no. 8 [2018]: 1112-21). (The R^2s depicted are technically incremental R^2s, which reflect the gain in the R^2 when the score is added as a covariate to a regression of the phenotype on a set of baseline controls [sex, birth year, their interaction, and ten principal components of the genetic-relatedness matrix].)

Moving from the Educational Attainment PGIs to the Risks and Benefits of PGIs in General

SBG research is about many more phenotypes than educational attainment and about more methods than GWASs and PGIs. We restricted our discussion to the case of PGIs for educational attainment because that is the case that, for good reason, exerted a gravitational pull on our working group's attention. As we have noted, there is reason to fear that, despite the best intentions of the researchers in our working group, results from SBG research will be used to reinforce the status quo.[78] Yet results from SBG research can also be used to advance social science research, with a view to producing benefits in basic understanding and, as some of the social scientists in our working group hope, benefits in the form of more effective social policies.

In the next two parts of the report, we will turn to a more in-depth consideration of what our working group sees as the potential risks and benefits associated with SBG research. In the end, we cannot offer an algorithm for weighing those risks and benefits, but we believe that an honest accounting of where our working group agreed and disagreed can be helpful to others who, in the future, face questions concerning the funding, conduct, and communication of SBG research.

PART 4: THE RISKS OF SOCIAL AND BEHAVIORAL GENOMICS (SBG) RESEARCH

SBG research has the potential to harm individuals, groups, and societies in a variety of ways. People can be harmed by research even if they are not research participants.[79] Risks can arise at any stage in the conduct of SBG research, from how—and from whom—data are collected, to the research questions that are asked of the data, to how outcomes are characterized or measured, to how the data are analyzed. Harms can also arise from how SBG research is disseminated or interpreted or from downstream applications of its results (see figure 3). SBG research that is poorly conducted, poorly communicated by researchers or media, or misinterpreted by others carries clear and present dangers. But even well-conducted, well-communicated, well-understood SBG research has the potential to do harm. Here, we do not provide a comprehensive discussion of all possible risks of SBG research but, rather, briefly describe some of the primary forms that harms from SBG research (and other research) might take.

Many—though not all[80]—of what we are calling the "risks of SBG research" are in fact risks that people will erroneously ascribe to genes properties that they do not in fact possess, especially with respect to social and behavioral phenotypes—for instance, that people will describe or regard genes as determining immutable outcomes or defining races or human worth. To the extent that SBG research suggests such things directly or fails to do enough to disabuse readers of these long-standing myths while emphasizing the importance of genetics, SBG research is partly to blame. As we note in the next part, however, *responsible* SBG research not only can avoid causing some of these harms but is in some cases essential to refuting existing harmful stereotypes based on theories of the relationship between genes and social and behavioral outcomes.

Racism and Reification of Race as a Biological Category

As part 2 mentioned, one of the foundational methods of much contemporary genomics research—including but not limited to SBG research—is including and excluding participants based on their genetic ancestry or attempting to control for genetic ancestry during analysis. As we discuss in more detail in part 6, genetic ancestry is not identical to race or ethnicity. Nevertheless, there is a risk

Figure 3.
Risks of Sociobehavioral Genomics Research

Self-fatalism: An individual's belief that their actual or predicted phenotype is determined, fixed, or inevitable.

Self-stigmatization: An individual's belief that they are less capable or worthy.

Stigmatization of others: The belief that other individuals are less capable or worthy.

Discrimination against individuals: Adverse actions taken against individuals in various domains (e.g., in education, employment, health, insurance, immigration, reproduction, or criminal justice).

Group stigmatization: The belief that certain groups are less capable or worthy.

Race or ethnicity biologizing and racism: Reification of race and ethnicity as biological concepts and—in combination with one of the other risks listed here—possible resulting racism.

Harmful or inequitably distributed policy applications: Scientifically invalid or ethically inappropriate applications of SBG research, or beneficial applications to which not all have access.

Genetic distractionism: Devoting resources to SBG research or applications could distract from more effective ways of understanding or addressing the same phenotype or phenomena.

Policy fatalism: Results from SBG research—even despite excellent science communication—could be understood through the lens of genetic determinism and used to justify the status quo as inevitable and environmental interventions as futile.

Note: Each of the above should be understood as occurring on the basis of genetic information.

Several companies sell individual PGI reports, including for social and behavioral phenotypes and phenotypes for which PGIs are only very weakly predictive, without any meaningful attempt to explain the very substantial limitations of these PGI reports to consumers.

that continental-level genetic ancestries, in particular, will be mistaken for racial or ethnic groups, and therefore help reify the erroneous, dangerous notion of race and ethnicity as biological concepts. This result is much more likely than it might otherwise be because current standards for labeling these groups overlap considerably with labels used to self-identify race and ethnicity.

For instance, the 1000 Genomes Project's "superpopulations"—African ancestry (AFR), East Asian ancestry (EAS), South Asian ancestry (SAS), European ancestry (EUR), and (admixed) American ancestry (AMR), which is sometimes reported in research papers as "Latin American"[81] or "Native American"[82]—are very easily conflated with U.S. race and ethnicity census categories, such as "Black or African American," "Asian," "American Indian or Alaska Native," and "Hispanic or Latino."[83] Historically, shifting from notions of biological "race" to biological "populations" did little to combat the mistaken view of race as biological, and there is little reason to believe that appending words like "ancestry," "ancestries," or "population" after "EUR," "AFR," and the like will prevent the conflation of these concepts with races or ethnicities today.[84] This means that SBG research findings, in turn, have the potential to be racialized and used in racist ways—whether this takes the form of stigmatization, discrimination, or fatalism.

Stigmatization, Discrimination, and Fatalism

SBG research that associates certain outcomes or traits with particular kinds of people or that enables individuals or groups to be "scored" according to their likelihood of experiencing a particular outcome or exhibiting a particular trait has the potential to contribute to stigmatization, discrimination, and fatalism. Genetic stigmatization involves regarding an individual or a group as less valuable, less capable, or less favorable in some other way, compared to others, because of their genes. Genetic discrimination entails taking an adverse action against someone on the basis of their genetic or genomic information. Finally, genetic fatalism involves the erroneous belief that some social or behavioral outcome that is associated with genes is unavoidable; as we explain in the introduction and parts 2 and 7 of this report, genes alone do not determine complex social or behavioral outcomes.

Stigma and fatalism can each be externalized toward others or internalized and directed toward the self. Exposure to some SBG research or its products could cause or contribute to damaging self-conceptions by individuals or members of some groups or to damaging conceptions by others, perhaps leading to negative self-fulfilling prophecies. For example, someone who learns that they have a "low" PGI for educational attainment might mistakenly decide on that basis that there is no point in their attempting to pursue additional education, in which case they may indeed turn out to have relatively low educational attainment. A dystopian system in which PGIs are used to exclude those with "low" PGIs from educational opportunities could have a similarly self-fulfilling effect.

Policy Fatalism and Distractionism

SBG research also has the potential to harm societies. By definition, SBG research investigates outcomes that are powerfully shaped by environments as well as being influenced by genes. One risk of investigating genomic contributions is *policy fatalism*: results from SBG research could be miscommunicated or misunderstood by policy-makers or other decision-makers through the familiar and erroneous lens of genetic determinism and then used to justify the status quo—that is, to argue that observed individual or group differences are biologically caused and immutable and, hence, that environmental interventions would be futile. For example, policy-makers could misinterpret research on genomic contributions to educational attainment in such a way as to conclude—as Arthur Jensen hypothesized—that environmental interventions to improve educational attainment were not worth investing in.

Even in the absence of a fatalistic attitude to change, giving attention and allocating resources to SBG research could lead to *genetic distractionism*. Material and human resources that otherwise might have been devoted to other kinds of research might be diverted instead to SBG research, to a harmful degree. This could happen, for instance, if genomics is viewed as a "harder" or more cutting-edge science. SBG research might similarly distract policy attention from more effective or cost-effective interventions to address social problems.

Box 3. Some Potential Uses of Sociobehavioral PGIs in Medicine and Education

A. Predicting disease risk in patients. Many large research studies are currently investigating the clinical use of disease PGIs. Adding sociobehavioral PGIs to these composite scores could be used in clinical practice to improve disease prediction and, potentially, morbidity and mortality. Those at high risk could be recommended for more frequent screenings or preventative behaviors.

B. Polygenic embryo screening. A few companies recently began offering polygenic scoring of in vitro fertilization embryos, and in addition to screening for polygenic disease, at least one company has explored using an educational attainment PGI to screen for intellectual disability, defined as having an IQ of 70 or below, or screening for the full range of educational attainment and other "experimental" sociobehavioral phenotypes, like income, cognitive ability, and subjective well-being.[1] Sociobehavioral PGIs could also be used, as in example A, to more accurately screen for disease in embryos.

C. Better prognoses of developmental disorders. If educational attainment or IQ PGIs turn out to more precisely predict how specific genetic variants will be expressed, parents of newly diagnosed children could be given a more precise prognosis instead of merely being told, for instance, that 30 percent of people with 22q11.2 deletion syndrome have intellectual disability. Similarly, if research shows that IQ and/or schizophrenia PGIs help predict the odds that a patient with 22q deletion syndrome will also have schizophrenia, this could be used for early intervention.

D. Disability benefits. Under section 504 of the U.S. Rehabilitation Act, students with disabilities can receive accommodations, modifications, and learning aids. Eligibility is determined on the basis of a school evaluation that takes into account "any medical diagnosis from a physician, aptitude and achievement tests, teacher recommendations, physical condition, social and cultural background, and adaptive behavior." A high PGI for a qualifying condition like attention deficit-hyperactivity disorder could serve as an additional piece of information helping to secure access to these benefits.

E. Pupil premiums. In the United Kingdom, schools receive a so-called pupil premium of £1345 per year for each student who is of low socioeconomic status, which is associated with poorer educational outcomes. The premium reflects a recognition that disadvantaged children generally face additional challenges in reaching their potential at school, and its purpose is to improve educational outcomes for disadvantaged students. Schools could receive additional premiums for each student with a sufficiently low educational attainment PGI.

[1] P. Turley et al., "Problems with Using Polygenic Scores to Select Embryos," *New England Journal of Medicine* 385, no. 1 (2021): 78-86.

Scientifically Invalid and Unethical Applications in Policy or Practice

Finally, SBG research results or their products (such as PGIs) might be used to support harmful policies or practices—for instance, ones that deprive people of liberty, property, or educational, professional, or other opportunities or that subject them to other adverse actions on the basis of genomic information. In the United States, the Genetic Information Non-Discrimination Act[85] prohibits discrimination on the basis of an individual's genetic information in health insurance and in most employment, but does not cover life, disability, long-term care, or other insurance, much less many other important domains, such as education, employment, health care, immigration, reproduction, or criminal justice. Although it is possible, and some have argued, that genomic information can be ethically used in some of these contexts[86]—a debate that is beyond the scope of our report—the lack of legal protection over the use of genomic information in at least some of these contexts raises ethical issues. And, of course, whatever legal protections exist in one jurisdiction do not necessarily extend to other jurisdictions. Yet the basic deliverable of SBG research (like most research) is information, which knows no geopolitical bounds. Even well-intentioned policies that aim to help individuals or groups—say, offering an environmental intervention as compensation for genes that predispose one to a disadvantageous phenotype—could have the potential to do more harm than good. The intervention could be ineffective (because, say, it is based on poor or misinterpreted SBG research), could cause people to stigmatize themselves or others (or otherwise harmful), or both ineffective and harmful.

Industry research and commercial applications of SBG research raise concerns (as do commercial applications of many other kinds of research) for many, including members of our working group's "community sounding board."[87] For instance, to the extent that SBG research can contribute to beneficial policies or practices, there is a risk that, in the hands of commercial entities, these benefits will not be equally accessible to all or that SBG research

will enrich companies without compensating research participants. For many, commercial entities are, all else equal, more worrisome than other actors because they are driven by the profit motive and typically lack meaningful oversight or other checks and balances.[88] For instance, several companies sell individual PGI reports, including for social and behavioral phenotypes and phenotypes for which PGIs are only very weakly predictive, without any meaningful attempt to explain the very substantial limitations of these PGI reports to consumers, and with no evidence that they have any concern that these reports may have negative psychosocial effects.[89] Companies now also offer polygenic scores for embryos.[90]

Although the working group agrees that SBG research has the potential to be used in scientifically invalid and morally unacceptable ways in policy or practice, we do not always agree on whether particular potential uses (for examples, see box 3) would meet either or both of these criteria.

PART 5: THE POTENTIAL BENEFITS OF SBG RESEARCH

We now turn to what we see as the potential benefits of SBG research—again, largely focusing on GWASs and PGIs. Because both environmental and genetic factors are important for human variation in virtually any human trait, SBG research is in principle relevant for understanding humans. Indeed, failing to consider the role of genes in *some* way means that the scientific record on important phenomena is likely to be incomplete at best and possibly wrong. That said, in studying the causes of social and behavioral outcomes, understanding how nature and nurture combine is a formidable—sometimes impossible—task. Nevertheless, not all beneficial research seeks to understand causation—many value its ability to predict outcomes—and some argue that SBG research has potential even there.

Better Understanding Environmental Causes and the Limits of Genomic Influence

SBG research can also help identify and illuminate environmental causes of social and behavioral outcomes. For instance, research has found strong associations between the PGIs of parents and the outcomes of their children.[91] However, these associations are not necessarily causal. Recent work using parental siblings to estimate the causal effect of the parental PGI on children's outcomes has found only insignificant and imprecise results.[92] Nevertheless, with more data, in the future, researchers may be able to evaluate whether and in what ways parental genes influence their children.

Similarly, SBG research can help demonstrate the limits of genetic influences. For instance, some researchers had hypothesized that genetic influences on the X chromosome are an important source of differences in the variance in cognitive phenotypes across the sexes. But SBG research found no such evidence for this version of the sex-differences claim.[93] Other discriminatory claims might similarly be refuted with the aid of GWASs.

SBG research can help identify heterogeneous effects of environmental exposures on different people. For instance, even if it is the case that, on average, education has a positive effect on income, it is likely that the effect is different for different people. And it might be important, for policy reasons, to understand that heterogeneity: for instance, it would be important to know if education's effect on income on average is largely driven by its effect in high socioeconomic groups who are already advantaged. Work that investigates gene-environment interactions is likely the most active area of SBG research. For instance, in general, losing one's job is associated with a decrease in body mass index (BMI), but researchers have found that people with low BMI variance PGIs[94] experienced greater BMI reductions following job loss than those with high BMI PGIs.[95] However, although most current work uses population-level data, again, much of the most interesting work of this type will require within-family data.

Finally, genetics can help when phenotypes are confounded with one another. Consider the paradoxical associations of different levels of alcohol consumption with health outcomes. Many readers will have been introduced through the popular media to research showing that both abstainers and heavy drinkers are at higher risk for stroke compared to moderate drinkers. However, this U-shaped curve for the effect of alcohol on stroke using self-reported alcohol consumption might be confounded; for instance, those who know they are at increased risk for stroke for other reasons might eliminate alcohol, helping to explain why abstinence is associated on average with higher stroke risk. There are two genetic variants that are common in people of East Asian genetic ancestries (but rare in people of European genetic ancestries) that cause individuals with those variants to become very ill with any alcohol use and

can therefore be used to predict a phenotype of alcohol abstinence. When researchers investigated the extent to which genotype-predicted mean alcohol consumption predicts stroke incidence, they found that the apparent protective effect of alcohol on stroke is largely noncausal and *any* alcohol consumption is associated with at least some increased risk of stroke.[96] Although this is a promising example of using genetic epidemiology to overturn incorrect assumptions based on traditional epidemiology, this kind of work requires strong assumptions that may not be met in many cases.[97]

Improving Randomized Controlled Trials

One of the highest-potential uses of PGIs is as a simple variable in other research—which may have nothing to do with genetics—to control for the influence of genes on the outcome being studied.[98] In a randomized controlled trial or natural experiment, statistical power is increased by including control variables with significant predictive power. Boosting power, in turn, means that fewer participants are needed, which can save time and money, leading to faster discoveries. The precise value of using PGIs as control variables will depend on several factors, including the availability and comparative predictive

Box 4. How PGIs Could Improve Trials of Environmental Interventions

One potential use of PGIs is to increase statistical power in randomized controlled trials and natural experiments to determine the effectiveness of interventions—including, as in the examples we present here, environmental interventions. By accounting for uninformative noise in data, PGIs can allow researchers to perform well-powered studies in smaller samples, saving researcher dollars and potentially reducing recruitment time. This would lead to faster, more-efficient social (and health) science. Here we present a back-of-the-envelope calculation showing how much money could be saved in two well-known early-life education interventions if those studies were run today, using PGIs as a control variable and reducing the sample size such that the study was equally powered. The two interventions considered are the Perry preschool study[1] and the Abecedarian program study.[2] The Perry preschool program was a trial that randomly assigned some disadvantaged preschool children to high-quality preschool education. The program cost roughly $27,000 per participant (here and below in 2022 USD). The Abecedarian study was a larger comprehensive Intervention from infancy through age five that provided a high-quality educational program over several years to disadvantaged children. The program cost approximately $116,000 per participant.

The gains in statistical power by including a PGI as a control variable are a function of the predictive power (measured in R^2) of the nongenetic control variables available versus the predictive power of the control variables with the PGI included. For this example, we assume that the predictive power of a set of relevant nongenetic controls (IQ, parental education, and household income) is 19 percent and that the predictive power of the controls after adding the PGI for EA increases to 23.6 percent. These values come from the 2018 study, EA3. The predictive power of the educational attainment PGI has improved since that study was published, but we use these lower, outdated values in this calculation to produce conservative estimates of the value of controlling for PGIs. We further assume that genotyping costs $50 per participant, which is comparable to the price of genotyping today, but this cost has fallen quickly over the past several years due to improving technology and competitive forces, and we anticipate that it will continue to fall in the future.

Based on the predictive power of the control variables and the PGI, we calculate that, if the Perry preschool or Abecedarian study were launched today but also included a PGI as a control variable, it could recruit a sample that was 6.7 percent smaller than the original samples and still be equally powered. For the Perry preschool study, this would amount to a savings of $1483 per participant. For Abecedarian, it would amount to $6564 per participant. Part of the reason for the large savings is the high cost of the interventions. However, we calculate that there are expected cost savings for any study that costs more than $881 per participant. This threshold will fall as the predictive power of PGIs improves and as the cost of genotyping falls.

[1] "Perry Preschool Project," Social Programs That Work, accessed January 17, 2023, https://evidencebasedprograms.org/programs/perry-preschool-project/.

[2] "Abecedarian Project," Social Programs That Work, accessed January 17, 2023, https://evidencebasedprograms.org/programs/abecedarian-project/; J. J. Lee et al., "Gene Discovery and Polygenic Prediction from a Genome-Wide Association Study of Educational Attainment in 1.1 Million Individuals," *Nature Genetics* 50, no. 8 (2018): 1112-21.

Although the PGIs created by most SBG research are very weak predictors of individual outcomes, some PGIs have predictive power that rivals or exceeds nongenetic predictors that are already routinely used in research.

power of other, nongenetic predictors and the marginal costs of including additional research participants (see box 4). Sociobehavioral PGIs can also be used to increase the comparability of the people who do and do not receive an intervention in an experiment, thus similarly enabling trials to be powered with fewer participants.

Although the PGIs created by most SBG research are very weak predictors of individual outcomes, some PGIs (as explained in part 3) have predictive power that rivals or exceeds nongenetic predictors that are already routinely used in research. For instance, researchers conducting a field experiment of an innovative curriculum might want to control for heterogeneity in participants' propensity to succeed in school even apart from the experimental intervention. These researchers might use the father's educational attainment, mother's educational attainment, or household income, since these all correlate with an individual's eventual educational attainment. The current best PGI for educational attainment (EA4,[99] discussed in part 3) has more predictive power (with an R^2 of about 15 percent) than household income (6 percent), mother's education (15 percent), or father's education (15 percent)—though not as much as mother's and father's educational attainment combined (19 percent).[100]

The first GWAS of educational attainment (EA1), published in 2013, explained only 2 percent of the variance among individuals. As predicted, the R^2 has grown with sample size. By 2018, the best PGI for educational attainment (EA3) had explained 11 percent of the variance among individuals. And, as mentioned, the most recent PGI (EA4), in 2022, predicts about 15 percent. Notably, the jump in four years from 11 percent to 15 percent was achieved only by tripling the GWAS sample size.[101] This shows that sociobehavioral PGIs with relatively low R^2s can increase in meaningfulness to researchers as sample sizes grow, as well as that substantially larger sample sizes can become available relatively quickly. But it also makes clear that growth is not infinite. On average, the SNP heritability of educational attainment in a single, relatively genetically homogenous sample is estimated to be about 20 percent[102] (see figure 1). Each sociobehavioral PGI should be evaluated on its own terms, and in the present day, to determine whether it has the potential to improve social science. Although larger GWASs have produced and in some cases will continue to produce substantially more powerful PGIs, for some sociobehavioral phenotypes (such as those for which the SNP heritability is substantially lower than 20 percent), PGIs will serve as substantially less powerful control variables (although a PGI with an R^2 of only 1 percent that is created for "free" from already-genotyped samples may nevertheless be marginally useful).

Only about one-third of the current PGI's explanatory power regarding educational attainment (an R^2 of approximately 5 percent) is due to the causal effect of genes that differ within sibling pairs. Depending on the purpose to which the PGI is being put, however, that may be irrelevant. For instance, if the PGI is being used only to predict an outcome, no matter the cause, the R^2 of approximately 15 percent will be the appropriate measure of the PGI's predictive power. Moreover, it is unclear what portion of the predictive power of nongenetic predictors like parental education is causal—and that is, in fact, another advantage of genomic predictors: because they are randomly assigned among siblings, it is possible, at least in within-family studies, to estimate the causal effect.

As discussed in part 3, neither demographic variables (like income) nor PGIs have a necessary relationship with years of education. Rather, both kinds of predictors reflect the contingencies and policy choices of the society where the correlations were measured—such as the facts that postsecondary education is often expensive, that any education beyond what is compulsory carries opportunity costs that not all can bear equally, and that wealth is inherited—and so may be meaningful only in the time and place they were measured. Nevertheless, if one wants to develop policy interventions that produce fairer outcomes in the future, it is important to be able to understand the factors that gave rise to current outcomes, notwithstanding the unfairness of some of the factors that produced them. On the one hand, one difference between demographic and genomic index predictors is that the latter are likely perceived as innate, while the former are not. On the other hand, as explained in box 5, genes have several advantages compared to nongenetic predictors of the same phenotype—or to simply measuring the actual phenotype directly.

Although we cannot currently point to examples of sociobehavioral PGIs being incorporated into trials, existing PGIs with the predictive power to serve in this role (such as

Box 5. Why Use Genes, and Not Other Variables, to Study or Predict Phenotypes?

In practice, PGIs are typically used in combination with other predictors, not in lieu of them. But a common question is, what is the advantage of incorporating genomic information in any way, compared to relying solely on nongenetic predictors of the same phenotype, or simply measuring the actual phenotype directly? First, as mentioned in part 1 and discussed in more detail below, some PGIs have predictive power on par with that of nongenetic predictors and, when combined with those nongenetic predictors, can sometimes improve predictions beyond the latter alone.

Second, because genes are randomly assigned (conditional on the biological parents' genes), studies of genetic variation within families are especially useful in research for illuminating what in parts 2 and 3 we called "causal effects." And because genetic-based study designs allow researchers to observe the genetic differences among family members, they have an advantage over traditional twin studies, in which genetic differences are not directly observed. Genetic data also have the advantage of having greater external validity. Not everyone has a twin (and those that do are a special population), but everyone has genes and parents, which is sufficient to estimate causal effects.

Third, sometimes the phenotype of interest is either not observable or unavailable. Even when the phenotype is both observable and available, it can be more expensive or more difficult in other ways to measure than genetic data. For instance, participants who are not sampled, who die, or who are lost to follow-up—substantial problems in most human subjects research—cannot be measured, and these participants may differ from those who are included in research samples in ways that bias research results. (Because biobank samples are not representative of the general population, if biobanks were the source of the genomic information, incorporating genomic information might introduce its own form of bias. However, biobanks are not the only source of genomic information. For instance, if PGIs were used as a covariate in a clinical trial, presumably researchers would compute PGIs for all participants directly.) As for using nongenetic phenotype predictors, parental demographics are often strong predictors of the child's eventual phenotype, but not everyone knows this information about their biological parents. And although genotyping may seem like an expensive method compared to surveys, it can now be done for less than $50 per person (and costs will continue to fall), and that genotyping can serve a potentially very large number of studies.

Fourth, phenotypes can be confounded by other phenotypes. For instance, economists have long been interested in the effects of smoking or drinking on social or labor market outcomes, but in our society, these outcomes are correlated for myriad reasons, and so even direct measurements of one phenotype will always be confounded by the other. As discussed further below, researchers can use genes to help untangle the causal portion of phenotypic relationships using so-called Mendelian randomization (under strong assumptions and in specific cases).

Fifth, a person's DNA sequence doesn't change over time, which means that it can be used to travel back in time. Researchers can collect genetic information on people who participated in a study (such as a randomized controlled trial of a preschool curriculum) or who were exposed to an environmental change of interest (such as a policy reform) years or even decades ago. In contrast, most forms of phenotypic data collected after the fact are subject to biases and errors due to retrospective recall or could have been affected by the environmental exposures of interest.

Finally, sometimes phenotypic relationships are confounded by genes. For example, a researcher who observes an effect of parenting style on child outcomes cannot be sure that the outcomes are directly influenced by the parenting style as opposed to by genes that are influencing both the parenting style and the outcome. Ignoring the latter possibility risks getting the wrong empirical answer.

Each of the above reasons does not necessarily apply to every choice to incorporate genes into a study. Moreover, as research tools, genes have their own drawbacks (for example, PGIs are themselves confounded in various ways). But collectively, they suggest several advantages to incorporating genes into research and triangulating among them and the environment, compared to not including them.

Learning which SNPs are jointly associated with a sociobehavioral phenotype and a disease and which are uniquely associated with the disease can help advance understanding of both the biology of the disease and the effect of the sociobehavioral phenotype.

PGIs for educational attainment and for body mass index [BMI]) are relatively few in number. Incorporating genomics into traditional social science studies is also relatively novel and therefore unfamiliar to most social science researchers. Some clinical geneticists are focusing on the potential of clinical-phenotype PGIs to expedite clinical trials in this way—say, by using PGIs to select participants in preventive drug trials who are at substantially higher risk for a particular health outcome—and this exploration, too, is just beginning.[103]

Understanding Heterogeneity in Treatment Effects

Most trials are powered to identify, and report, average treatment effects. But often, a treatment that is ineffective on average is effective for a minority of participants—or differentially effective across participants. For example, a brief mindset intervention—an intervention aimed at increasing participants' belief that intellectual ability can grow over time—has been shown to increase enrollment in higher math courses, on average, but the intervention is stronger in students from schools with lower test scores. Implementing the intervention would therefore promote equity, which many policy-makers would consider important to know when allocating resources.[104] One could imagine an extension where a PGI that predicts enrollment in math class, the target outcome of the intervention, is tested as a moderator: does the mindset intervention effect everyone equally, or is there heterogeneity across the PGI? (Researchers might also ask, conversely, whether changing students' mindsets attenuates the association between the PGI and enrollment in high math courses, which would tell something about a psychological mechanism through which the PGI is associated with enrollment in higher math courses.)

Advancing Health Research

Not surprisingly, many social scientists who develop PGIs are primarily motivated by the prospect that these tools will be helpful in their own disciplines, such as sociology,[105] psychology,[106] economics,[107] and demography.[108] But because social and behavioral phenotypes often correlate with other outcomes, such as health, SBG research and the PGIs they sometimes produce have the potential to accelerate progress in those areas as well.

Since sociobehavioral phenotypes and health outcomes are often highly correlated and because some sociobehavioral PGIs (such as for educational attainment) are substantially more predictive than some disease PGIs, sociobehavioral PGIs have the potential to improve the prediction of disease or disease phenotype. The educational attainment PGI, for instance, is correlated with hypertension, ischemic heart disease, myocardial infarction, hypercholesterolemia, type 2 diabetes, asthma, osteoporosis, rheumatoid arthritis, migraine, and major depression. Although PGIs have been created for each of these diseases, adding the educational attainment PGI to them predicts these diseases substantially better than using either PGI alone.[109] This makes sociobehavioral GWASs, PGIs, and other SBG research potentially useful not only in social science research but also in *medical* research (we address below the *clinical* use of sociobehavioral PGIs). For instance, researchers recently used the educational attainment PGI to investigate risk factors for autism spectrum disorder.[110]

Moreover, it is not only that sociobehavioral and health phenotypes are correlated; learning which SNPs are jointly associated with a sociobehavioral phenotype and a disease and which are uniquely associated with the disease can help advance understanding of both the biology of the disease and the effect of the sociobehavioral phenotype. For instance, the SNP profile of educational attainment looks somewhat like—but is far from identical to—that of schizophrenia and Alzheimer's disease; some but not all SNPs that are associated with educational attainment are also associated with one or both of these diseases. And so, for example, researchers have used SNPs identified in an earlier GWAS as being associated with educational attainment to propose subtypes of schizophrenia.[111] In principle, comparing SNPs that are not shared by educational attainment and these diseases might help illuminate biological pathways of the diseases. Some SBG researchers are most excited about their work's potential to advance medical research (including medical genetics).

PGIs can also be used to provide more precise predictions of how, if at all, a disease caused by a monogenic variant will express itself in an individual. For instance, some developmental disorders caused by copy-number variation (CNV, a circumstance in which the number of copies of

a specific segment of DNA varies among individuals, for instance, due to insertions, deletions, or duplications) produce phenotypes that vary widely across individuals—from intellectual disability (defined as having an IQ two or more standard deviations below average) requiring considerable interventions to IQ in the standard range. Prior research has found that parental mean IQ predicts the phenotype of a child with a CNV disorder, suggesting that although the same CNV "shifts" IQ down by approximately the same amount in all children, those whose polygenic indexes for intelligence are higher than average can end up with average intelligence despite the CNV.[112] Research is now investigating whether the child's PGI for intelligence or educational attainment can even more precisely predict the variable expression of these CNVs.[113] Similarly, 22q deletion syndrome, which occurs when someone is missing a small part of chromosome 22, is associated with a 25 percent risk of schizophrenia—twenty-five times the risk for the average person. Because schizophrenia and IQ are positively correlated, researchers might use IQ and schizophrenia PGIs to better understand why some 22q patients have schizophrenia and others do not.

Uses of SBG Research and Its Products (Such As PGIs) in Practice and Policy

Because social science research can inform policy and practice, improving the precision, speed, and cost of social science research could *indirectly* lead to any number of policy and practice benefits.[114] Some of the most vocal proponents of SBG research, however, tout its potential to be *directly* incorporated into policy or practice.[115] As noted in the last part, working-group members do not agree about whether various applications of SBG research in practice or policy (see box 3) would be appropriate and hence count as potential downstream benefits of basic SBG research. For some members, at least some of these applications would be inappropriate and should be considered downstream *risks* of SBG research.

A few inflection points underlie this lack of consensus. With respect to scenario A, "Predicting disease risk in patients," in box 3, for instance, some working-group members believe that it is ethically inappropriate to make predictions without any understanding of mechanisms—or without at least being able to rule out indirect effects that reflect social factors—and these working-group members also believe that these requirements will never be met for social or behavioral phenotypes. In response, some other working-group members say they are optimistic about the potential of SBG research to illuminate mechanisms (in psychiatric conditions, for example). Yet others point out that many medical-risk predictions are informed by correlations whose underlying mechanisms are either opaque or known to reflect social injustices and that offering an accurate, actionable prediction can be ethically appropriate, sometimes more so than withholding that prediction while awaiting knowledge of mechanisms.[116] In the domain of education (consider scenarios D, "Disability benefits," and E, "Pupil premiums"), some argue that, since resources are allocated to students and schools whose individual or mean environments are associated with disadvantage, consistency and fairness suggest that resources also be allocated to students and schools whose individual or mean PGIs are associated with disadvantage. Others are skeptical that any benefits would be worth the price of the stigmatization that is likely to result.

Despite these disagreements, we agree that, for multiple reasons, it would rarely be straightforward to translate SBG research and the PGIs they produce into scientifically and ethically acceptable practices or policies. We urge proponents of such applications to calibrate their enthusiasm accordingly. Most SBG research results would need to be translated into plausible interventions. For instance, the GWASs that have identified SNPs associated with educational attainment do not themselves explain *why* those SNPs are associated with more or fewer years of education. As a result, no one knows from these GWASs alone how to intervene to help those with a low educational-attainment PGI; just because a PGI predicts educational attainment does not necessarily imply that it will also predict the effectiveness of an intervention aimed to increase educational attainment. Moreover, most scientifically and ethically acceptable uses of PGIs in practice or policy would first require research to validate these uses and to measure any negative psychosocial or other effects of using PGIs. For instance, responsible innovation might call for a pragmatic trial of PGI use in some practice context under a research protocol, with both the primary outcome (such as increased educational attainment) and psychosocial outcomes compared to a control group. Such translational research would allow careful, evidence-based consideration of risks and expected benefits prior to implementation at scale.

PART 6: JUSTIFIABLE AND UNJUSTIFIABLE SBG RESEARCH

Before describing two categories of SBG research that deserve greater scrutiny—which we call "SBG research of heightened concern" and "SBG research of greatest concern"—we describe some of the working group's general conclusions. First, research that advances understanding of the world and humans' place in it is valuable, even if that knowledge never translates into concrete improvements in anyone's welfare. If that were not the case, large quantities of research in the arts and sciences—and perhaps some entire fields—would have to be deemed a waste of time and other resources.

Second, and relatedly, a majority of working-group members agree that, as a general matter, researchers—including SBG researchers—should not be expected to identify in advance specific potential benefits of their work for human welfare before beginning it. Even if research should, in theory, be justified primarily in terms of the benefits it might yield rather than by its intrinsic value—a position that a majority of working-group members reject—in practice, it is often impossible to accurately anticipate the specific benefits of research, especially of basic science.

Third, we note that there is considerable value in allowing scientists and other scholars "to pursue lines of inquiry and the communication of knowledge and ideas without fear of repression or censorship."[117] The freedom to ask questions, use the scientific method to try to answer them, and then share one's findings is critical to a healthy civil society.

Fourth, we agree that SBG research on a wide range of phenotypes can be worth conducting, funding, and publishing.[118]

Fifth, notwithstanding the above, much SBG research entails risks, and researchers have an ethical obligation to anticipate and take reasonable measures to avoid preventable harms or mitigate risks that may arise from their work or its communication.[119] We reiterate that research can impose risks not only on research participants but also on individuals and groups who were not in a study, or on society at large.[120] These obligations increase in proportion to the riskiness of the research. The greater the risks posed by a line of research or a particular study, the more confident researchers should be in their results before publishing or disseminating them.[121] Similarly, the more vigilant they should be about heading off foreseeable misinterpretations when they communicate them. We elaborate some of these obligations in part 7 below.

These considerations lead us to articulate two levels of concern, which we more fully describe in the next two subsections of the report. First, we consider SBG research involving sensitive phenotypes to be SBG research of heightened concern. At a minimum, heightened obligations of responsible conduct and communication of this research apply; we articulate these in part 7.

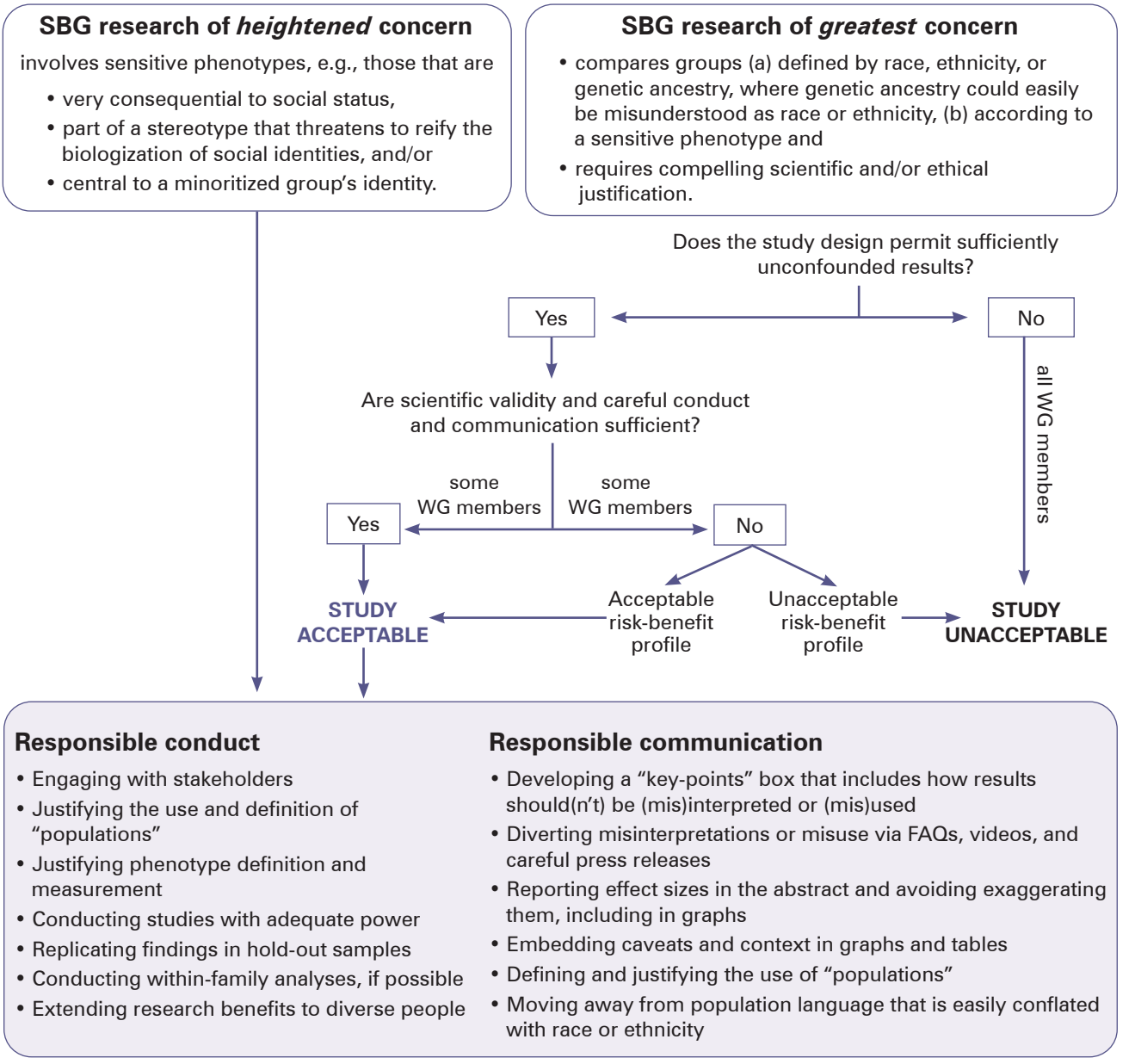
Second, we consider SBG research of the greatest concern to be research on sensitive phenotypes that compares groups defined by race, ethnicity, or genetic ancestry where—due to similarities in how races, ethnicities, and genetic ancestral populations are typically identified—genetic ancestry could easily be misunderstood as race or ethnicity ("group-comparison research," for short). As depicted in figure 4, we all agree that such research requires a compelling justification of the study's scientific validity. While some of us believe that researchers should be free to pursue any scientifically valid research, others of us would additionally require a compelling justification of the study's risk-benefit profile. We all hold that, absent a compelling justification—a criterion that some of us think will never be met—researchers should not conduct, funders should not fund, and journals should not publish such research.

SBG Research of Heightened Concern: Sensitive Phenotypes

Some kinds of SBG research are more ethically fraught than others. In particular, we have more concern about studying (and creating PGIs) for some phenotypes than for others. Nearly any social or behavioral phenotype has some potential to be sensitive. For instance, a PGI that predicts television consumption could be viewed as predicting people who are more or less likely to enjoy low-brow pleasures. But there are at least some social and behavioral phenotypes—for instance, introversion or religiosity—for which that risk seems lower. It is worth distinguishing phenotypes that pose lesser versus greater risks, of the sort described in part 4. Because we recognize not only the potential benefits of SBG research (see part 5) but also its potential risks (part 4), heightened obligations of responsible conduct and communication of this science apply to conducting, funding, and publishing this research.

What makes a phenotype of greater concern? We have in mind several criteria, and some phenotypes will fit more than one criterion that would heighten concern about research investigating it. First, phenotypes that are viewed in a society (rightly or wrongly) as being very consequential to social status are of heightened concern. This criterion is likely met if (but not only if) the phenotype can affect reputation, employability, financial standing, educational advancement, or legal risk.[122] Examples in contemporary U.S. society include obesity, substance-use disorders, intelligence-test scores, educational attainment, income, and

Figure 4.
Responsible Behavior in the Context of Sociobehavioral Genomics (SBG) Research



criminalized behaviors, all of which are associated with, or have implications for, socioeconomic status, social mobility, or both. As discussed in part 1, this category of phenotypes was also of keen interest to eugenicists and continues to be a focus of racists today.

Second, even if a phenotype is not strongly tied to social status, and in isolation is only moderately—or not at all—favored or disfavored, phenotypes that are or have historically been part of harmful stereotypes about minoritized groups[123] are of increased concern because they threaten to reify the biologization of social identities. Examples include athleticism, musical beat synchronization,[124] hypersexuality, hysteria, and financial prowess.[125] Even when research on such phenotypes is conducted in a nonminoritized population, it may carry this risk. For instance, even if genetic research on athleticism is, as is typical, conducted only in people of European genetic ancestries, it has the clear potential to be misinterpreted in ways that help reify both the erroneous notion of race as a biological category and the

Because we recognize not only the potential benefits of SBG research but also its potential risks, heightened obligations of responsible conduct and communication of this science apply to conducting, funding, and publishing this research.

stereotype that those who identify as Black are genetically predisposed to athleticism.[126]

Third, phenotypes that are central to a minoritized group's identity are of heightened concern. Examples include sexual orientation, sexual behavior, and gender identity.

Notably, the above criteria are socially contingent. In particular, they have both a geopolitical and a temporal dimension. Some phenotypes that were central to social status or lack thereof in the Progressive Era—for instance, "promiscuity"—remain disfavored in some circles but no longer play that central role today. Similarly, some groups are minoritized in some times and places but not (or at least less so) in others. For instance, although gay, lesbian, and bisexual people remain minoritized in many ways, both the social status and legal protections afforded these individuals in the United States are dramatically better than they were a relatively short time ago. Yet these protections were denied to sexual minorities in the United States in living memory (and are once again under threat[127]), and they remain denied today in other locations around the world, where being a member of a sexual minority is criminalized. Researchers and other decision-makers should therefore bear in mind geopolitical diversity both within their own country and across the world when assessing the risks of SBG research: studies that are not especially fraught in the researchers' or participants' region may be highly fraught in another region, where research results will inevitably be disseminated and potentially used in the creation of dangerous policies or practices.

Researchers and other decision-makers should also bear in mind the possibility of temporal changes: groups not minoritized today might be so in the future, stereotypes that don't exist or are not prevalent today could become so in the future, and phenotypes that are not of fundamental value today could be in the future as societies evolve. Indeed, SBG research itself has the potential to help usher in any of these future developments. For instance, a phenotype can become newly sensitive when groups are compared according to it, and genomic research can also create newly defined groups for favor or disfavor. Although, in principle, any phenotype could become problematic in some possible future, we do not suggest that all phenotypes therefore be treated as of equal concern today. Rather, we recommend that those assessing SBG research phenotypes do their best to attend to current and likely near-term future factors affecting the sensitivity of phenotypes, although we acknowledge that the difficulty of anticipating future developments is a limitation of the effectiveness of this approach.

Before describing what we mean by research of greatest concern, we need to acknowledge that the line between research of heightened and greatest concern is not bright and that it thus will be up to human beings to decide whether a given study should be subject to our recommendations. An example of a difficult border case is a study done in 2018 that investigated correlations between educational attainment PGIs and levels of socioeconomic success.[128] The researchers analyzed those correlations after they had separated the participants in the educational attainment study (which we discussed in part 3) into three groups: those who started out with low, middle, and high socioeconomic status (SES). Because this study was about sensitive phenotypes (educational attainment and social mobility), it is clearly of heightened concern. The fact that it compares groups with respect to SES heightens the concern. This study does not, however, reach the level of greatest concern because it compares only members of SES groups within the larger group of people of European genetic ancestries; although class is certainly a fraught social division and genetics has historically been misused in class oppression, the study does not attempt to compare people in the most fraught social divisions of the contemporary United States—race or ethnicity—nor does it compare genetic ancestral groups that could be easily conflated with racial or ethnic groups. We turn now to that most fraught kind of SBG research.

SBG Research of Greatest Concern

As noted above, we consider SBG research of the greatest concern to be research on sensitive phenotypes that compares two or more groups defined by race, ethnicity, or genetic ancestry, where—due to similarities in how races, ethnicities, and genetic ancestral populations are typically identified—genetic ancestry could easily be misunderstood as race or ethnicity. Examples of group-comparison research would include an attempt to compare participants who identify as Black with those who identify as White (a race-based comparison), an attempt to compare participants who identify as Ashkenazi Jews with those who iden-

tify as Sephardi Jews (an ethnicity-based comparison), and an attempt to compare participants of European genetic ancestries to those of Asian genetic ancestries (a comparison of genetic ancestral populations that could easily be misunderstood as a racial comparison). Although we are referring to such research as "group-comparison research," we emphasize that, in this subsection, we are talking about particular kinds of group comparisons, with respect to genes and "sensitive" phenotypes.

As depicted in figure 4, we all agree that such research requires a compelling justification of the study's scientific validity. While some of us believe that researchers should be free to pursue any scientifically valid research, others of us would additionally require a compelling justification of the study's risk-benefit profile. We all recommend that, absent a compelling justification—a criterion that some of us think will never be met—researchers not conduct, funders not fund, and journals not publish such research. We make this recommendation because of the following problems regarding the science and ethics of such comparisons.

As we explain below, we do not entirely agree about what kind of justification would be sufficiently compelling to support such "group-comparison" research (see figure 4). In brief, we do all agree that group-comparison research must employ methods that allow researchers to arrive at scientifically valid conclusions and that such methods are currently unavailable for complex SBG phenotypes. Assuming that valid methods will be available in the future (a point about which working-group members have different intuitions), we disagree about whether additional, nonscientific justifications (beyond assurances of responsible conduct and communication) will be needed for such research to proceed.

Again, whereas genetic ancestry is a biological concept, neither race nor ethnicity is. The reason we include some genetic ancestral group comparisons in our default rule is that, in practice, "race," "ethnicity," and "ancestry" are so frequently used interchangeably. This conflation is due in part to the fact that the labels used to describe people of different genetic ancestries—for instance, the 1000 Genomes Project's five superpopulations (EUR, AMR, SAS, EAS, and AFR) and twenty-six populations (such as the Han Chinese in Beijing, China, or CHB)—are very similar to labels that people use to describe (and governments use to track) racial or ethnic identity. As a result, genomics research that attempts to compare people of many different genetic ancestries will very likely be misunderstood as entailing comparisons of racial or ethnic groups, and such attempted comparisons will also face the same profound challenges to scientific validity.

Many identities besides race and ethnicity are minoritized, of course, and in the previous section, we call out genetic research that implicates these identities as of heightened concern. In the contemporary United States, however, no type of identity is more fraught in genetics research than race and ethnicity. We emphasize, however, that although we recommend against *comparing* groups defined by race, ethnicity, or genetic ancestries that can easily be mistaken for races or ethnicities, we join others in urging that racially, ethnically, and genetically diverse people be *included* in genomic studies; to the extent that SBG research has benefits, it is imperative that such studies be available to everyone.

Next, we note some common cases of sociobehavioral and population genetics research that do *not* reach the level of research of greatest concern. For instance, researchers may want to investigate whether there is heterogeneity in schizophrenia symptoms among different groups, such as "negative" symptoms like becoming socially withdrawn or nonresponsive versus "positive" symptoms like hallucinations. Schizophrenia is a sensitive phenotype, and so a study attempting to compare the prevalence of schizophrenia in groups defined by race, ethnicity, or genetic ancestries that can easily be mistaken for races or ethnicities should be justified before it is conducted, funded, or published. But the kind of schizophrenia symptoms that people are more likely to experience may not be a sensitive phenotype. As a result, comparing symptom profiles between racial, ethnic, or genetic ancestral groups falls outside our category of research of greatest concern (and, if valid, the potential clinical benefits of such a study's results are clear).

Another example of research that does not reach the level of greatest concern is the investigation of R^2 "shrinkage" in cross-ancestry use of PGIs. As discussed above, the greater the genetic distance between people in the GWAS training set from whom a PGI is created, on the one hand, and the individuals to whom a PGI is applied, on the other hand, the lower the predictive power the PGI will have. Alicia R. Martin et al. demonstrated and quantified the extent of this "portability problem," and thus the urgency of including more diverse participants in biobanks and GWASs, by comparing the R^2 of PGIs for seventeen phenotypes among the five 1000 Genome Project superpopulations.[129] Even when the phenotypes included in such a study are sensitive, the analysis does not involve comparing mean PGIs for sensitive phenotypes among groups—only the varying predictive power.

We now turn to the reasoning behind our recommendation against conducting some other kinds of studies. Today, typically, predicting phenotypic differences across genetic ancestral groups (whether or not those groups reside in the same geographic region) is not scientifically meaningful because such predictions are confounded due to population stratification, differences in linkage disequilibrium, and the effects of different environments, gene-environment correlation (including genetic nurture), gene-environment in-

While some of us believe that researchers should be free to pursue any scientifically valid research, others of us would additionally require a compelling justification of the study's risk-benefit profile.

teractions, gene-gene interactions, assortative mating, and other methodological problems. Because it is a foundational principle of human subjects research ethics that studies should be conducted only when they are validly designed to answer a meaningful question,[130] when this criterion is not met, there is no need for additional scientific and ethical analysis. All working-group members agree that, especially in light of the fraught nature of such work, SBG research on sensitive phenotypes that seeks to compare groups defined by race, ethnicity, or genetic ancestry, where genetic ancestry could easily be misunderstood as race or ethnicity, but whose study designs prevent valid conclusions, should not be conducted, funded, or published.

However, today's scientific infeasibility should not be a source of comfort to those who are concerned about genetics and pernicious group comparisons. Although we differ on how likely we believe that this is, none of us can guarantee that future research methodologies will never address enough of these methodological problems to render research of this sort as valid and robust as many other kinds of research. Therefore, although it is tempting to avoid the thorny question whether scientifically valid research should be avoided because of its social risks, it is too easy to simply conclude that between-group analyses will forever be scientifically meaningless; societies must confront the tradeoff between social risk and the value of pursuing scientific understanding. And here, members of the working group part ways (see figure 4).

For a first group of members, scientific conclusions about the world—including human-group differences—that result from valid, robust study designs and that are responsibly communicated are ethically permissible, without further justification. The "compelling justification" these members seek, therefore, is justification that a study design could or has overcome the very substantial methodological obstacles associated with comparing different racial, ethnic, or some genetic ancestral groups. Even among the working-group members in this first group, there is disagreement. For instance, some view the pursuit of scientific knowledge as an absolute freedom; these members might acknowledge legitimate limits on the free pursuit of science in certain areas—say, gain-of-function research that could be weaponized to kill countless humans—but do not recognize a reasonable analogy to such exceptionally dangerous research in social and behavioral genomics. Others are open in theory to the idea that some social science research is too dangerous to justify (perhaps especially with the considerable benefit of hindsight) but believe that fair, reasonable methods of implementing such limits are not possible in practice and that attempts to do so are likely to cause more overall harm than good to human welfare and even justice.[131]

A second group of working-group members follows a different path. While acknowledging the importance of scientific knowledge and freedom, they emphasize that these are not absolute values but must be balanced with others, including welfare and justice. Assuming that a study could yield scientifically valid conclusions, these members would therefore require an additional "compelling justification"—namely, that the researchers demonstrate that the study has a sufficiently favorable risk-benefit profile. These members think that, in almost all cases, the social risks of the research would outweigh the potential benefits (whether in terms of basic or translational research). However, they leave open the possibility that, in rare cases, the ethical analysis could turn out differently.

Given the highly contextual nature of research risks and potential benefits, a case-by-case assessment is required, especially in light of the fact that research risk is generally assessed not in isolation but, rather, in comparison to existing risk,[132] which is not static. Here, we offer a nonexhaustive list of factors that some working-group members believe might—alone or in combination—tend to make the risk-benefit profile of group-comparison research more favorable than it otherwise would be and that these members think may be sufficiently favorable to justify the research.

For instance, if ethically or scientifically problematic group-comparison research were already being conducted and taken seriously, that could be one factor in allowing responsible, better-skilled scientists to engage in and show the limitations of similar research; arguably, that would not entail unacceptable incremental risk and might indeed mitigate harm from others' work.[133]

Similarly, in some cases where observed phenotypic differences between groups are already well-known, a genomic explanation might be more likely to be destigmatizing than stigmatizing—say, if the prevailing explanation overemphasizes personal responsibility. For instance, although obesity is not disfavored in all social groups, in many, it is often assumed to be the result of weak character (for ex-

ample, laziness, lack of impulse control, or lower intelligence). Somewhat as sexual minorities who have embraced that they were "born this way," rather than making a "lifestyle choice," and have found this new understanding destigmatizing, a finding that genes make it harder for some individuals—or, potentially, for some genetic ancestral groups—than others to maintain a healthy weight could reduce body-size stigmatization by both those who struggle with their weight and others. There are still myriad ways for such a study to go awry: for example, its results could lead to fatalism or to stigmatization. As with research on any sensitive phenotype, the results of such a study would therefore need to be carefully communicated. Researchers should also engage people with obesity in each group to be compared before proceeding (we discuss such engagement more in part 7).

Other factors that might tend to render a group-comparison study's risk-benefit profile more favorable include comparison of only nonminoritized groups, comparison of minoritized but comparable-status groups, comparison regarding an only marginally sensitive phenotype, strong potential for concrete benefits to the groups under study arising from the research, and support for the research among group members.

Notably, all members of the working group treat group-comparison SBG research differently, in one way or another, from how we treat within-group SBG research. Some working-group members view the risk-benefit profile of between-group SBG research—but not within-group research—as presumptively insufficient to justify the research. And all working-group members view between-group SBG research—but not within-group SBG research—as presumptively scientifically invalid. In resting on such assumptions, we want to acknowledge a challenge: some people might question the coherence—conceptually, and in terms of both scientific validity and risk-benefit profiles—of distinguishing between-group and within-group analyses. We recognize that both within-group analyses (in which members of only one group are compared) and between-group analyses depend on the concept of a group. And we recognize that, although genetic ancestry is a scientific concept, it is, as we emphasized above, constructed by human beings, so it is by no means "written in nature" or necessarily unproblematic.[134] Moreover, we recognize that even within-group analyses have methodological challenges. Nevertheless, the nature and magnitude of the ethical, social, conceptual, and methodological difficulties associated with group-comparison research demand even greater scrutiny than research that investigates within-group differences. We note, again, that although we agree about this overall conclusion, some working-group members emphasize the greater methodological difficulties inherent in analyses between groups (as compared to within groups), while others emphasize the increased social risks. We discuss each of these in turn.

First, we address the comparative scientific validity and value of between- versus within-group SBG research. The value of any research is limited by the extent to which it can yield scientifically reliable conclusions.[135] Although within-group comparisons might always be confounded, the magnitude of the confounding will be vastly greater (but not necessarily forever insurmountable) in between-group comparisons. The confounding will be so great that it is hard to imagine between-group comparisons having substantial scientific value. Working-group members disagree both about how much causal or mechanistic knowledge can result from within-group SBG research, and about how valuable prediction without causal or mechanistic knowledge can be, but we agree that, in principle, within-group analyses are more likely than between-group analyses to generate causal and/or mechanistic knowledge and to generate predictions that are less confounded by influences other than genetic effects are. For instance, when conducting a within-group analysis, comparing the DNA of individuals in the same family (conducting within-family analyses) can allow researchers to attempt to explain the effects of genetic variants. In families, each sibling, during gamete formation and conception, is randomly assigned to receive one each of two parental alleles at each place in the genome. Approximating this method in between-group analyses is not feasible. Although, in families where parents have different genetic ancestries, the proportion of admixture is randomly assigned, generalizing from the results of such admixed family studies to draw conclusions about genetic ancestral groups as a whole would raise concerns about external validity, gene-environment interactions, and genetic effects operating through environmental mechanisms. In particular, any finding of an association between genes and observed differences between groups defined by race, ethnicity, or (in most cases) genetic ancestry would continue to be confounded by, and reflect, postrandomization environmental mechanisms like racial and ethnic discrimination.

Second, we address the comparative potential harms and benefits of between- versus within-group SBG research. We do not deny that some will find between-group comparisons with respect to sensitive traits of inherent interest, nor that there is some intrinsic value in affirming the free pursuit of any research. However, although we recognize rare cases in which such research might entail specific benefits, we think the likelihood that any potential benefits will outweigh the scientific weakness and social risks is exceedingly low (yet, as noted above, we disagree about whether a favorable risk-benefit is required to justify scientifically valid, responsibly conducted and communicated research). If it were possible, someday, to conduct scientifi-

In some cases where observed phenotypic differences between groups are already well known, a genomic explanation might be more likely to be destigmatizing than stigmatizing—say, if the prevailing explanation overemphasizes personal responsibility.

cally valid between-group comparisons, the social costs of anything other than a null result—and, in particular, the costs of a result that confirms existing stereotypes against already-minoritized groups—would be vast. And in the meantime (which may be forever), the social costs of exaggerating what we can learn from SBG research are different in within- versus between-group analyses.

Balancing these considerations of scientific validity and potential benefit and individual and social risk, we conclude that, today, the combination of the dubious (at best) scientific validity and the acute social risks of between-group SBG research means that—one way or the other—there should be a very strong presumption against its being conducted, funded, or published, absent the overriding justifications discussed above.

As much as we worry about group-comparison research, however, we are all highly averse to the idea of policing the production of knowledge. Yet researchers, funders, journal editors and reviewers, and journalists—most of whom we imagine share these competing worries (even if, like us, some of them worry a bit more about one than the other) —do not have the commentator's luxury of remaining paralyzed between them; agents in all these categories must make decisions about group-comparison research. We acknowledge that our attempt here to provide actionable guidance to these actors—by recommending a default rule against group-comparison research involving sensitive phenotypes that is defeasible with a "compelling justification"—is highly imperfect. As we have noted, critical factors in our default rule, such as the sensitivity of phenotypes and whether a group is minoritized, are contingent on time and place, including on future circumstances that are not always easily anticipated and in any case cannot be known with certainty. Nor can we typically know in advance what research will find (for example, substantial average group differences—or not) or what the societal effects of those results might be (for instance, reducing—or increasing—stigma). Like most genomic information itself, these factors are all highly probabilistic. We are not the only ones to struggle with these issues.[136] A 2022 *Nature Human Behaviour* editorial[137] announcing a new Springer Nature ethics policy about research involving human groups[138] received considerable backlash,[139] and editors soon published a second editorial clarifying the first.[140] We do not imagine that our own attempt at addressing these complex issues has remotely solved them, but we hope it contributes helpfully to the conversation.

PART 7: RESPONSIBLE BEHAVIOR IN THE CONTEXT OF SBG RESEARCH

Social and behavioral genomics research is very likely to remain a part of mainstream genetics and social science research, and it is sure to remain an active area of research by some at the edges of mainstream science. As a result, we believe that efforts should be redoubled to promote the responsible conduct of research and to communicate SBG research in ways that proactively mitigate the risks of its misinterpretation and misuse.

In this section, although we cannot provide an exhaustive list of all the actions that responsible conduct and communication of SBG research entails, we offer some recommendations. In so doing, we consider the entire time line of research, from the decision whether to conduct a study; to study design, data collection, and analysis; to dissemination.[141] Because researchers themselves are central to each of these decision points, we focus on this group of agents. However, other parties, including funders, journal reviewers and editors, and the media, have an important hand in how this research is conducted and communicated. We encourage all these stakeholders to consider how a particular study was conceived, conducted, and (as applicable) communicated when considering whether and how to fund, publish, or report on it.

Responsible Behavior before and during Research[142]

First, researchers should reflect on the considerations noted in this report (and by other commentators on SBG research) and consider whether they want to conduct a particular SBG study or embark on a particular line of SBG research at all. By this statement, we do not mean to put a thumb on the scale against SBG research. Rather, we simply mean to encourage reflection on questions about, for example what risks the study poses, what its potential benefits might be, and whether it constitutes a good use of resources in light of the risk-benefit profile and other research opportunities the same researchers might pursue instead.[143] Indeed, researchers in all disciplines would benefit from such reflection and deliberation.

Second, assuming that researchers believe the study to be worth conducting, they should engage the people with the most at stake in the study, at least if it is a novel line of research about which the relevant communities have not been engaged by other researchers. Engagement can take a variety of forms, from informal conversations to participation on formal advisory boards or in mixed-methods research about population attitudes toward the research and from time-limited, pre-research consultation to an ongoing relationship as advisors to, or even investigators in, the research.[144] For instance, before publishing the largest GWAS of same-sex sexual behavior to date,[145] the research team (which included one working-group member, Benjamin Neale) engaged LGBTQ outreach and advocacy groups and dozens of LGBTQ rights advocates and community members and subsequently held workshops where representatives of the public, activists, and researchers discussed the results of the study.[146] The study team leads also included multiple members of the LGBTQ community.[147] There are benefits to diverse study teams; for example, it is less likely (though not impossible) that a study team with relevant forms of diversity will design or communicate research in ways that are culturally insensitive. If the research team does not include members of the groups under study, the researchers should take care to "be reflective of their authorial perspective."[148]

We recognize that many forms of engagement (especially if done well) can be costly and time-consuming, and we do not insist on any one form of engagement over another. We also emphasize that members of the same community will often have different views, that most people are members of multiple communities, and that there is rarely someone with the political, moral, or other authority to speak on behalf of an entire community. Deliberate efforts should therefore be made to engage diverse members of relevant communities, and the fact that community engagement may have gone well should not be viewed as moral license to cease considering the viewpoints and interests of the full community. These limitations aside, however, individual members of relevant communities can offer helpful input throughout the research life cycle, including the formulation of research questions, data collection and analysis, and dissemination. Of course, researchers should also be open to the possibility that community members may recommend that the study not be conducted at all, and indeed researchers should explicitly solicit their views about this question.

Third, as some genetics researchers themselves have recently argued, "geneticists should think more carefully about what data [they] select for [their] analysis—and how [their] choices could lead to misappropriation of [their] work."[149] In particular, if researchers use membership in a group or population as an inclusion or exclusion criterion (or otherwise) in their studies, they should be clear in their own minds about why they are doing so. The social constructs of race and ethnicity should never be used as proxies for genetic similarity or diversity unless absolutely necessary (and then, only with exquisitely careful explanation during dissemination).[150] If, instead of social identities like race and ethnicity, it is the degree of genetic similarity

In genomics research, it is important to explicitly distinguish among "race," "ethnicity," "genetic ancestry," and other group or population terms.

among individuals that actually matters to study inclusion and exclusion (or to other aspects of the study), researchers should make sampling choices that reflect (and, as discussed below, use language that reflects) that intention, as discussed in part 3. This should entail thinking in advance about how what degree of genetic similarity is sufficient for their purposes (for instance, is similarity to a 1000 Genomes Project reference superpopulation sufficient, or should the study use more granular, subcontinental populations, reflecting relatively more genetic similarity?) and how they will measure genetic similarity. Researchers often receive data from large biobanks that are precategorized and prelabeled; we call on data repositories to do their part to more carefully characterize their data subjects and how they were assigned to groups.

Fourth, researchers should be able to justify how they define and measure the phenotypes under study (and, as we discuss below, describe the limitations of this measurement).

Fifth, researchers should conduct only adequately powered studies and replicate their findings in hold-out samples.

Sixth, researchers should, whenever possible and assuming doing so would be adequately powered, conduct within-family analyses (and, as discussed below, report the corresponding effect size).

Seventh, the research community should work to ensure that any benefits of SBG research and PGIs extend to all and do not exacerbate existing disparities.[151] Among other things, this will require diversifying biobanks through equitable global partnerships[152] and developing new statistical methods of analyzing genomic data.[153]

Dissemination via Scientific Papers

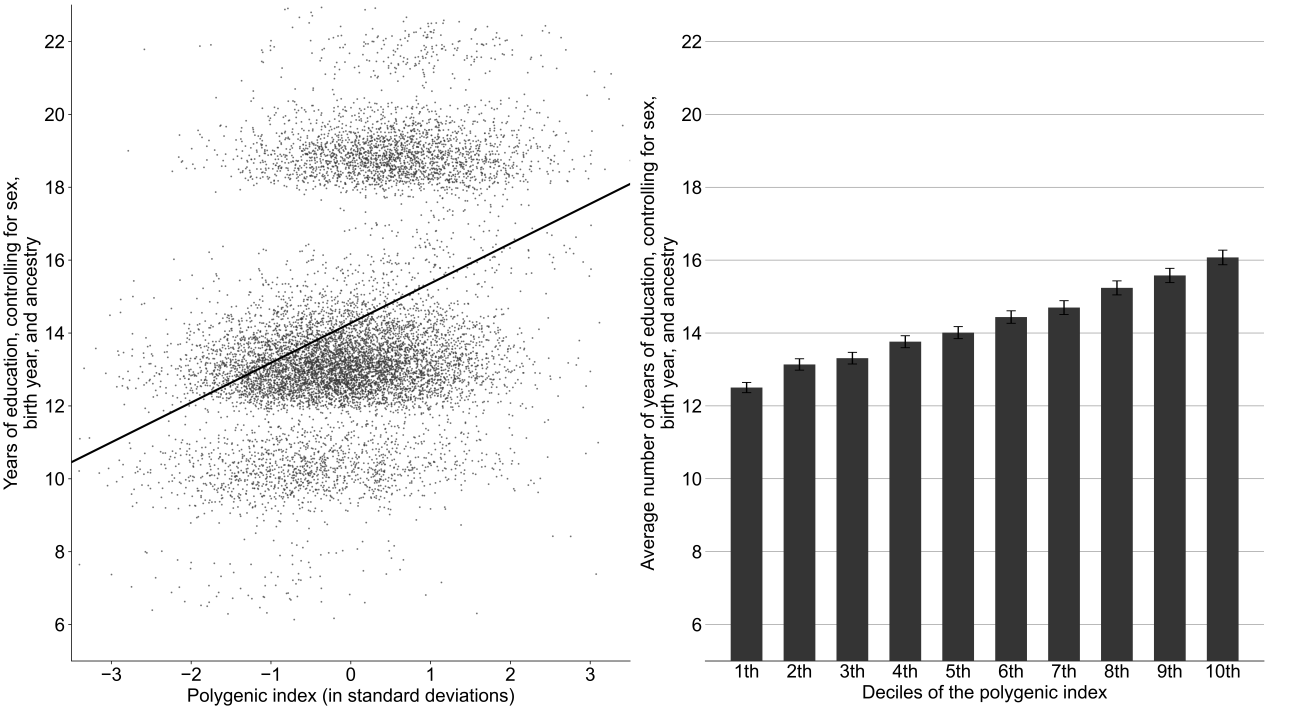
A study conducted responsibly can nevertheless be irresponsibly communicated. This, of course, is true of any study, not only of SBG research. The goals of responsibly communicating any scientific study are to be clear, accurate, and sensitive to the relevant stakeholders (including research participants and other individuals and groups who may be affected); to be appropriately sober about and to contextualize the research results; and to proactively mitigate the risks of misinterpretation and misuse of research. Such communication occurs through not only traditional research dissemination via peer-reviewed scientific articles and conference presentations but also via press releases and, increasingly often, websites, videos, and social media. We first discuss aspects of responsible communication of SBG research in scientific papers and then turn to these other forms of dissemination. However, recommendations that we make about the content that should appear in one form of communication (for example, content disabusing readers of any belief in genetic determinism) will often apply to other forms. Which messages belong in which modes of communication will depend on the space constraints and intended audience of each mode.

First, assuming that groups or populations are used in any way in the study, we agree with the research ethics policy of the portfolio of *Nature* journals that, in scientific publications, researchers should (in either the main text or, if necessary, the supplement) "[e]xplicitly describe their methods of categorizing human populations"; "[d]efine categories in as much detail as the study protocol allows"; "[j]ustify their choices of definitions and categories, including . . . whether any rules of categorization were required by their funding agency"; and "[e]xplain whether (and if so, how) they controlled for confounding variables in their analyses."[154] In particular, in genomics research, it is important to explicitly distinguish among "race," "ethnicity," "genetic ancestry," and other group or population terms.

When using a genetic ancestry concept, it is important both to say that this is neither race nor ethnicity and to affirmatively specify what genetic similarity is and how it is being operationalized in the study. A recent investigation found that scientists use "genetic ancestry" in a variety of ways, including in ways they themselves cannot fully articulate or explain; while some of them view ancestry as closely related to genetics, others view it as only tangentially related.[155] For instance, researchers might explain that their high-level goal (for scientific reasons they explain) was to include people in their study who were relatively genetically similar to one another and to exclude those who exceeded some threshold of dissimilarity, that one way of thinking about genetic similarity is the time since two people shared a common genetic ancestor, and that they operationalized that by using participants' report that all four biological grandparents hailed from the same continent (or did not).

Because continental-level genetic ancestry labels especially encourage the conflation of genetic ancestry with

Figure 5.
Two Different Ways of Displaying PGI Data



race and ethnicity and because, as part 2 mentions, genetic variation is continuous[156] and highly diverse within continents (especially Africa), the genetics community should move as quickly as possible away from continental-level genetic ancestry groupings and labels and toward practices and corresponding labels that reflect this continuum and are less easily conflated with race or ethnicity.[157] We recognize that no ideal alternative to traditional categories such as 1000 Genomes Project "populations" and "superpopulations" exists today, and we call on the field to work together to make progress. Empirical research should explore which ways of describing human groups are least likely to be conflated with race and ethnicity.

Second, social and behavioral genomics researchers, like all researchers, should not exaggerate their effect sizes, either in the text of a paper or by how they display results in graphs. For instance, empirical studies have found that the same results can be interpreted very differently by both laypeople and relevant professionals, depending on whether researchers display the full distribution of individual outcomes, on the one hand, or provide only summary statistics or means, on the other hand.[158] In the case of PGIs, as K. Paige Harden and Daniel Belsky have suggested,[159] quintile and decile plots, which show the *mean* phenotype in each quintile or decile, may suggest more genetic determinism than scatter plots, which show the often-wide *range* of individual outcomes associated with each quintile or decile. Consider figure 5, which shows the same EA4 data as a decile plot and as a scatter plot. Although studies should be conducted to confirm this, it is plausible that providing only a decile plot, as opposed to both kinds of plots, may exaggerate a PGI's predictive power.

Moreover, effect sizes—including not only those of population analyses but also of any within-family analyses—should be reported in a paper's abstract. If within-family analyses were infeasible or impossible to conduct—as they currently often will be, due to insufficient family samples to adequately power such analyses—this should be explained in the manuscript, along with the fact that within-family effect sizes are typically smaller than population-level effect sizes. Any speculation about what effect sizes might be in the future (with larger sample sizes, for example), such as those based on statistical models, should be clearly acknowledged as resting on assumptions. Finally, researchers should be clear that between-family (population-level) associations cannot be interpreted as causal because they are confounded by gene-environment correlation and assortative mating.

Third, researchers should think carefully about graphs and tables, which are (along with abstracts) among the most salient and disseminated portions of scientific articles. For instance, researchers should anticipate that readers will

Many of the concerns about SBG research pertain to how the results might be misinterpreted or misused by nonscientists. It will often be important for SBG researchers to engage in some nontraditional forms of dissemination.

capture and share (potentially virally) snapshots of such display items that are therefore removed from the context of the manuscript.[160] To the extent possible, therefore, caveats and context should be embedded in graphs and tables—at minimum, in the figure captions, but at best, in the figures themselves—so that true manipulation of the image, not mere cropping (which might not be ill intended), would be required to share the display item without important context. Again, empirical research should investigate optimal ways of presenting genomic research results that give genes their due, but no more than their due.

Fourth, at many journals, scientific articles contain not only abstracts but also something like key points, which typically appear on the first page of the article, set off in a box or in some other salient way. We encourage SBG researchers to use at least some of their allotted key-point space to warn readers against misinterpretations and misuse. A similar approach is to devote one display item to a text box noting the ethical, legal, and social implications of the study.[161] The most important potential misinterpretations and misuses to warn readers away from, of course, will depend on the research, including the phenotype and the population(s) under study, but our point is that, given SBG research's propensity to be misinterpreted, the limitations of such research should not be pointed out or discussed merely in the standard paragraph near the end of the manuscript. We also encourage journals to expand the space researchers are allotted for key points, display items, and sections discussing limitations, as necessary. Researchers should not face a trade-off between highlighting their scientific contributions (important for citations and career advancement) and doing their part to discourage misinterpretation and misuse, and decades of behavioral science shows that making it easy to do the right thing substantially increases the odds that the right thing will be done. To that end, we also encourage journals that do not already use key-points sections to adopt them, at least for easily misinterpreted research.

Dissemination beyond Scientific Papers: Press Releases, FAQs, and So Forth

Although the core of research dissemination consists of scientific articles and presentations, scientific results can be, and are, disseminated in other ways and to other audiences, including via press releases, frequently-asked-questions (FAQs) documents, websites designed to showcase a significant study, videos, and social media.[162] Many of the concerns about SBG research pertain to how the results might be misinterpreted or misused by nonscientists, including the media, policy-makers, practitioners, and members of the public. As a result, it will often be important for SBG researchers to engage in at least some of these nontraditional forms of dissemination—and extremely important to do so responsibly. Because these additional materials about how a study should and should not be interpreted or used are helpful only to the extent that they are accessible, the associated scientific paper should, as prominently as possible, tell readers where they can find such materials. Similarly, if a researcher, lab, or research consortium maintains a website where articles in scientific journals are cross-posted or linked to, these other materials should also be cross-posted or linked to, much as open data, code, and materials are linked to papers.

Researchers who are engaged in easily misunderstood, socially risky research like SBG research should think carefully about whether they want to issue a press release. On the one hand, the traditional purpose of issuing a press release for an academic paper is to increase press coverage of the paper, and, given how bad some media coverage of genetics research has been, it is worth asking whether inviting media coverage of easily misunderstood, easily misappropriated research is a good idea, especially if the results are preliminary. On the other hand, another purpose of issuing a press release is to help shape whatever media coverage of the paper may emerge anyway. To the extent that an SBG research paper is likely to receive media attention (say, because it is published in a high-profile journal), issuing a very careful press release (along with FAQs or other products we discuss below) may, on balance, be wise. Like all press releases, any SBG research press release should studiously avoid hyping or exaggerating research results. But SBG press releases should go beyond that by affirmatively seeking to anticipate and bust common myths about social and behavioral genomics and to contextualize the findings.

FAQs can play a similar role in myth busting and in contextualizing SBG research. To our knowledge, FAQs accompanying SBG research publications were pioneered

between 2011 and 2013 by the Social Science Genetic Association Consortium, including working-group members Daniel Benjamin and Patrick Turley, and their advisors, including working-group co-PI Meyer.[163] FAQs (and similar materials) should be distributed to journalists consistent with journal embargo practices and whenever a journalist asks to interview a researcher, as they can help ensure responsible media coverage. For instance, the SSGAC's FAQs on their educational-attainment studies have been explicitly discussed or linked to in coverage of the research by *NBC News*,[164] PBS,[165] *Wired*,[166] and the *Atlantic*,[167] and many more media outlets paraphrased key messages in the FAQs, such as that environment plays a more important role in outcomes like educational attainment than do genes and that most sociobehavioral PGIs are poor predictors of individual outcomes.[168] These FAQs have been praised as a best practice that other researchers should emulate by the editors of *Nature*[169] and by bioethicists.[170] Since the emergence of FAQs in the field of SBG research in 2013, a number of other research teams have joined the SSGAC in writing them to give the context, scope, and limitations of SBG research. FAQs are quickly becoming standard in the field of SBG research and a recognized-enough bioethics tool that they are collected in a repository hosted by The Hastings Center and curated by bioethicists (including working-group member Daphne Oluwaseun Martschenko).[171] Because SBG research can be quite complex to explain to nonspecialists, FAQs must balance comprehensiveness and length, but tables of contents and digital FAQs with collapsing and expanding content for those want to learn more (or less) can help. Especially relevant portions of FAQs can be screenshot and embedded in social media, where study questions, concerns, or misinterpretations frequently arise.

Short videos (animated or otherwise) can be a more engaging way to convey important information, especially to lay audiences, and they can be easily embedded in social media. For instance, researchers who conducted the recent large GWAS of same-sex behavior (including working-group member Neale) created a website devoted to explaining and contextualizing the study. In addition to FAQs about who the researchers and funders are, why the researchers conducted the study, where the data came from, what the researchers did, what they found, and what the study's limitations were, the researchers created an animated video that, in less than three minutes, covers many of these points (and likely encouraged some website visitors to take a deeper dive by reading the more detailed FAQs).[172] The Broad Institute, academic home to several of the lead authors of this GWAS (as well as some of the study's critics), hosted several blog posts from affiliates that were either supportive or critical of the study, as well as five additional brief videos featuring senior author Neale explaining what the study found, why the researchers conducted the study, how they engaged relevant communities, and his views about the potential for the results to be misinterpreted and about the study's central messages.[173]

Myth Busting and Contextualization

We have said that responsible dissemination of SBG research in scientific papers, press releases, FAQs, videos, and the like should bust myths and contextualize, but what, precisely, does that entail? As noted above, the most important potential misinterpretations and misuses to warn readers away from will depend on the study. We encourage interested readers to peruse the FAQs in the Hastings Center repository for examples of messages that a range of SBG researchers felt were important to emphasize as part of responsibly communicating their science. With that said, here we describe a handful of important points that are relevant to most or all SBG research studies.

First, many laypeople are puzzled by the very idea of a study of genomic influences on a social or behavioral phenotype because they believe that genes have nothing to do with such phenotypes. They may therefore benefit from an introduction to the first law of behavioral genetics: all human behavioral traits are heritable.[174] Similarly, many will be skeptical of the motivations of researchers who study sensitive phenotypes, especially in research that yields PGIs for these phenotypes, and will want to know the purpose of the research.

Most messages, however, should be devoted to trying to ensure that genes are not overemphasized. For instance, although many laypeople may not be genetic determinists,[175] it is important to say explicitly that genes contribute to but do not determine social and behavioral phenotypes. Instead, for these and other complex phenotypes (including many medical phenotypes), the environment—including one's family, friends, and community; the physical spaces where one lives and works; and one's income, education, and more—generally has more influence. Moreover, genes and environments interact with one another in myriad complex ways that scientists are far from having sorted out.

Nor, audiences should be reminded, are genetic influences on behavioral traits immutable. Genes that predispose someone to a disadvantage can be overcome through environmental interventions (such as eyeglasses for nearsightedness), while, conversely, genes that predispose someone to an advantage may be undermined or completely mooted by environmental interference (such as malnutrition and height).

The mechanisms through which genetic variants are associated with social and behavioral outcomes are poorly understood and cannot be divorced from environmental or social processes.

Standard scientific jargon used in the peer-reviewed paper reporting the results of a study may mislead nonscientists and benefit from explanation. For instance, PGIs for social and behavioral traits are not "predictors" in the fortune-teller sense that laypeople may understand that term. As scientists use that term, it is perfectly consistent to say that a PGI "predicts" an outcome and also that the prediction is only marginally better at predicting a binary outcome than is a coin flip.

Lay audiences will also need help understanding SBG research effect sizes. On average, laypeople estimate that PGIs, including social and behavioral PGIs, have much more predictive power than they do.[176] For most social and behavioral traits, PGIs alone are not currently useful predictors of individual outcomes, and for many such traits, they will never be good individual predictors. PGIs can increase predictive accuracy for individuals when combined with other predictors,[177] and they can be useful in research.

Finally, even when an SBG study concerns only people of European genetic ancestries, researchers should anticipate that nonspecialists—both well-intentioned and not—may believe that these results speak both to race or ethnicity and to racial or ethnic differences. They should be reminded, therefore, that PGIs created from one group of genetically similar people cannot be meaningfully extrapolated to other groups, for reasons that are not yet fully understood. Alleles associated with an outcome in one genetic ancestral group may not be as strongly associated with that outcome in another group, may not be associated with that outcome at all, or may be associated with the opposite outcome. It is therefore scientifically invalid to identify variants as associated with an outcome in one genetic ancestral group and then compare the frequency with which these alleles appear in different such groups (much less different races or ethnicities), as White supremacists and others have often done.

The Hard Problem beyond Scientific Literacy and Clear Communication

As difficult as the problems of science literacy and clear communication about complex science are, another at-least-as-difficult problem is that different people can and do bring different values to the same set of facts.[178] In the case of genetics, as suggested in part 4, some people may interpret the fact that genetic differences can make a difference for social and behavioral phenotypes as evidence that certain policies are futile or that societies should not invest in some people because of their genomic profile. But others argue that the same facts are evidence that society should redouble its efforts in enriching the environment of those people. Similarly, around the world, White and other supremacists and separatists interpret the fact of genetic diversity to mean that each group is *essentially* different and that these diverse groups should be kept "pure" so that each group's essential attributes—and, in particular, those of the "superior" (in the United States, White) group—are preserved. The rest of us reject not only the scientifically inaccurate assumptions that discrete groups of genetically similar people exist and that groups have essential natures but also the idea that some groups are morally or culturally superior to others.[179] To the extent that what separates these dramatically different perspectives is values rather than science, the many well-intentioned calls by geneticists, journal editors, and others (including our own working group) for more genetics education and clearer communication about genetics research—with a view to resisting its misunderstanding, misuse, or misrepresentation[180]—will only partially address this problem.

Much as we cannot offer a simple algorithm for weighing the potential harms and benefits of any given SBG protocol, we cannot offer a simple solution to the hard problem regarding invidious values. But recognizing why and how SBG research raises questions that demand to be wrestled with is, we think, in itself an important step in the right direction. Meanwhile, we hope that our description of the historic, scientific, and ethical terrain and our recommendations for the responsible conduct and communication of SBG research will be useful to others as they wrestle with social and behavioral genomics research.

Statement of Authorship

Michelle N. Meyer and Erik Parens contributed equally to this report.

Acknowledgments

We would all like to thank Hastings Center president Mildred Solomon for her unfailing support; Hastings Center former and current research assistant-project managers Isabel Bolo, Ben Wills, Aashna Lal, and Margaret Matthews and Geisinger research assistant Tami Gjorgjieva for excellent research assistance; and Vicki Peyton for excellent administrative support. Finally, we thank professors Vanessa Northington Gamble and Bridget Goosby for helpful conversations early in the project.

This special report is a major outcome of a project funded by the Robert Wood Johnson Foundation, the Russell Sage Foundation, and the JPB Foundation. Additional support for this special report comes from The Gil Omenn and Martha Darling Fund for Trusted and Trustworthy Scientific Innovation, a Hastings Center fund. Individual coauthors thank the following additional funders for supporting their time on this project: the National Institute on Aging of the National Institutes of Health (through grants R24-AG065184 and R01-AG042568, supporting Daniel J. Benjamin and Michelle N. Meyer, and grant R00 AG062787, support-

ing Patrick Turley) and Open Philanthropy (through grant 010623-00001, supporting Benjamin, Turley, and Meyer).

Of course, the usual caveat applies: none of these individuals or funding organizations necessarily shares the perspectives reported here.

Notes

1. See R. Plomin et al., eds., *Behavioral Genetics in the Postgenomic Era* (Washington, DC: American Psychological Association, 2002).

2. Roughly the same field has been described elsewhere as "genoeconomics" (D. J. Benjamin et al., "The Promises and Pitfalls of Genoeconomics," *Annual Review of Economics* 4, no. 1 [September 1, 2012]: 627-62), "social genomics" (D. Conley and J. Fletcher, *The Genome Factor: What the Social Genomics Revolution Reveals about Ourselves, Our History, and the Future* [Princeton: Princeton University Press, 2017]), "sociogenomics" (C. Bliss, *Social by Nature: The Promise and Peril of Sociogenomics* [Stanford: Stanford University Press, 2018]), and "social science genomics" (A. Angers et al., *Genome-Wide Association Studies, Polygenic Scores, and Social Science Genetics: Overview and Policy Implications* [Luxembourg: Publications Office of the European Union, 2019]). In our multidisciplinary project, we preferred to use a term that is not limited to a single discipline (such as "genoeconomics") and that does not imply a limitation to either social or behavioral phenotypes.

3. The earliest paper constructing a PGI was S. M. Purcell et al., "Common Polygenic Variation Contributes to Risk of Schizophrenia and Bipolar Disorder," *Nature* 460 (2009): 748-52. But the idea (as applied to humans) was described earlier in P. D. Pharoah et al., "Polygenic Susceptibility to Breast Cancer and Implications for Prevention," *Nature Genetics* 31, no. 1 (2002): 33-36, and in N. R. Wray, M. E. Goddard, and P. M. Visscher, "Prediction of Individual Genetic Risk to Disease from Genome-Wide Association Studies," *Genome Research* 17 (2007): 1520-28.

4. We are using "PGI," following an observation from legal scholar Martha Minow that "score" may connote a judgment that is not necessarily intended in social and behavioral phenotypes. See the box in J. Becker et al., "Resource Profile and User Guide of the Polygenic Index Repository," *Nature Human Behavior* 5, no. 12 (2021): 1744-58.

5. For earlier efforts at a similar sort of wrestling, see E. Parens, A. Chapman, and N. Press, eds., *Wrestling with Behavioral Genetics: Science, Ethics, and Public Conversation* (Baltimore, MD: Johns Hopkins University Press, 2006), and E. Parens, "Genetic Differences and Human Identities: Why Talking about Behavioral Genetics Is Important and Difficult," *Hastings Center Report* 34, no. 1 (2004): S1-S36.

6. On Galton's life, see C. P. Blacker, *Eugenics: Galton and After* (London: Duckworth, 1952); R. S. Cowan, "Francis Galton's Statistical Ideas: The Influence of Eugenics," *Isis* 63, no. 219 (1972): 509-28; P. Froggatt and N. C. Nevin, "Galton's 'Law of Ancestral Heredity': Its Influence on the Early Development of Human Genetics," *History of Science* 10, no. 1 (1971): 1-27; R. S. Cowan, *Sir Francis Galton and the Study of Heredity in the Nineteenth Century* (New York: Garland Publishing, 1985); D. W. Forrest, *Francis Galton: The Life and Work of a Victorian Genius* (New York: Taplinger Publishing Company, 1974); M. Brookes, *Extreme Measures: The Dark Visions and Bright Ideas of Francis Galton* (New York: Bloomsbury, 2004); N. W. Gillham, "Sir Francis Galton and the Birth of Eugenics," *Annual Review of Genetics* 35 (2001): 83-101; K. Pearson, *The Life, Letters and Labours of Francis Galton* (Cambridge: Cambridge University Press, 2011); N. W. Gillham, *A Life of Sir Francis Galton: From African Exploration to the Birth of Eugenics* (New York: Oxford University Press, 2001); and M. Bulmer, "The Development of Francis Galton's Ideas on the Mechanism of Heredity," *Journal of the History of Biology* 32, no. 2 (1999): 263-92. On Galton's revival and his remade reputation, see R. E. Fancher, "Francis Galton's African Ethnography and Its Role in the Development of His Psychology," *British Journal for the History of Science* 16, no. 1 (1983): 67-79, and D. J. Galton and C. J. Galton, "Francis Galton: His Approach to Polygenic Disease," *Journal of the Royal College of Physicians of London* 31, no. 5 (1997): 570-73.

7. On "stirp," see F. Galton, "A Theory of Heredity," *Journal of the Anthropological Institute* 5 (1876): 329-48. For a modern take on Galton's heredity, see M. Bulmer, "The Development of Francis Galton's Ideas on the Mechanism of Heredity," *Journal of the History of Biology* 32, no. 2 (1999): 263-92.

8. F. Galton, *Inquiries into Human Faculty and Its Development* (London: J. M. Dent, 1883), 17.

9. Both quotations are from F. Galton, "Hereditary Talent and Character," *MacMillan's Magazine* 12 (1865): 165.

10. Cf. K. P. Harden, "The Science of Terrible Men," *Aeon*, March 11, 2021, https://aeon.co/essays/what-do-we-do-with-the-science-of-terrible-men.

11. The historical literature on eugenics is vast. A brief survey of the eugenics literature would include E. A. Carlson, *The Unfit: A History of a Bad Idea* (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 2001); D. J. Kevles, *In the Name of Eugenics: Genetics and the Uses of Human Heredity*, 2nd ed. (Cambridge, MA: Harvard University Press, 1995); D. B. Paul, *The Politics of Heredity: Essays on Eugenics, Biomedicine, and the Nature-Nurture Debate* (Albany, NY: State University of New York Press, 1998); W. Kline, *Building a Better Race: Gender, Sexuality, and Eugenics from the Turn of the Century to the Baby Boom* (Berkeley, CA: University of California Press, 2001); A. Stern, *Eugenic Nation: Faults and Frontiers of Better Breeding in Modern America* (Berkeley, CA: University of California Press, 2005); N. Comfort, *The Science of Human Perfection: How Genes Became the Heart of American Medicine* (New Haven, CT: Yale University Press, 2012); T. Keel, *Divine Variations: How Christian Thought Became Racial Science* (Stanford, CA: Stanford University Press, 2018); K. Tallbear, *Native American DNA: Tribal Belonging and the False Promise of Genetic Science* (Minneapolis: University of Minnesota Press, 2013); and A. Nuriddin, "Engineering Uplift: Black Eugenics as Black Liberation," in *Nature Remade: Engineering Life, Envisioning Worlds*, ed. L. A. Campos et al. (Chicago: University of Chicago Press, 2021): 186-203. For the suggestion that, if Galtonian statistical genetics had prevailed over Mendelism in the early twentieth century, much genetic misconception and eugenic mischief might have been circumvented, see G. Radick, *Disputed Inheritance: The Battle over Mendel and the Future of Biology* (Chicago: University of Chicago Press, 2023).

12. W. Bateson, "Letter to Adam Sedgwick," April 18, 1905, available through the Cambridge University Library, https://exhibitions.lib.cam.ac.uk/linesofthought/artifacts/naming-genetics/; W. Bateson and B. D. Bateson, *William Bateson, Naturalist: His Essays & Addresses, Together with a Short Account of His Life* (Cambridge: Cambridge University Press, 1928).

13. W. Bateson, *Biological Fact and the Structure of Society. The Herbert Spencer Lecture Delivered at the Examination Schools on Wednesday, February 28, 1912* (Oxford: Clarendon Press, 1912): 6.

14. On American nativism and its relation to immigration and disease, see J. Higham, *Strangers in the Land: Patterns of American Nativism, 1860-1925* (New Brunswick, NJ: Rutgers University Press, 1955); A. M. Kraut, *Silent Travelers: Germs, Genes, and the "Immigrant Menace"* (New York: Basic Books, 1994).

15. A. E. Garland "The Eugenics Record Office at Cold Spring Harbor, 1910-1940," *Osiris* 2 (1986): 225-64; D. J. Kevles, "Charles Davenport and the Worship of Great Concepts" and "The Gospel Becomes Popular," chaps. 3-4 in *In the Name of Eugenics: Genetics and the Uses of Human Heredity* (Cambridge, MA: Harvard University Press, 1995).

16. W. E. B. Du Bois, *The Talented Tenth* (New York: James Pott and Company, 1903). On Roosevelt and eugenics, see the primary-source documents at The Eugenics Archives ("Roosevelt, Theodore," The Eugenics Archive, accessed September 18, 2022, http://eugenicsarchive.ca/discover/connections/5233dd265c2ec500000000c9), and G. K. Chesterton, *Eugenics and Other Evils: An Argument against the Scientifically Organized State* (Seattle: Inkling Books, 2000), as well as D. Haraway, "Teddy Bear Patriarchy: Taxidermy in the Garden of Eden, New York City, 1908-1936," in *Grasping the World: The Idea of the Museum*, ed. D. Preziosi and C. J. Farago (New York: Routledge, 2018). On Sanger, see M. Sanger, *The Pivot of Civilization* (New York: Brentano's, 1922); J. H. Baker, *Margaret Sanger: A Life of Passion* (New York: Farrar, Straus and Giroux, 2011); A. Franks, *Margaret Sanger's Eugenic Legacy: The Control of Female Fertility* (Jefferson, NC: McFarland, 2014); and M. Ziegler, "Eugenic Feminism: Mental Hygiene, the Women's Movement, and the Campaign for Eugenic Legal Reform, 1900-1935," *Harvard Journal of Law and Gender* 31 (2008): 211-35. For an overview of left-wing eugenic agenda in the early twentieth century, see D. Paul, "Eugenics and the Left," *Journal of the History of Ideas* 45 (1984): 567-90.

17. H. H. Goddard, *Feeble-Mindedness, Its Causes and Consequences* (New York: Macmillan Company, 1914). See also L. Zenderland, *Measuring Minds: Henry Herbert Goddard and the Origins of American Intelligence Testing* (New York: Cambridge University Press, 1998).

18. On *Buck v. Bell*, see P. A. Lombardo, *Three Generations, No Imbeciles: Eugenics, the Supreme Court, and Buck v. Bell* (Baltimore, MD: Johns Hopkins University Press, 2008). In 1942, in *Skinner v. Oklahoma*, despite striking down an Oklahoma law permitting the state to compel the surgical sterilization of anyone who had been convicted three or more times of certain felonies involving "moral turpitude," the Court expressly declined to reconsider the central holding of *Buck*, holding only that sterilization laws must be applied equally to, for example, white-collar and blue-collar criminals, unless the state successfully argues that crimes by the latter were more heritable than those by the former. See V. F. Nourse, *In Reckless Hands:* Skinner v. Oklahoma *and the Near Triumph of American Eugenics* (New York: W. W. Norton, 2008). *Buck* was cited approvingly in *Roe v. Wade* (1973) and has never been explicitly overturned, though it is of dubious precedential value today.

19. P. Reilly, *The Surgical Solution: A History of Involuntary Sterilization in the United States* (Baltimore, MD: Johns Hopkins University Press, 1991).

20. J. Schoen, *Choice & Coercion: Birth Control, Sterilization, and Abortion in Public Health and Welfare* (Chapel Hill: University of North Carolina Press, 2005); J. Schoen, "Reassessing Eugenic Sterilization: The Case of North Carolina," in *A Century of Eugenics in America*, ed P. Lombardo (Bloomington & Indianapolis: Indiana University Press, 2011): 141-60; A. M. Stern, *Eugenic Nation: Faults and Frontiers of Better Breeding in Modern America* (Oakland, CA: University of California Press, 2016); N. C. Comfort, *The Science of Human Perfection: How Genes Became the Heart of American Medicine* (New Haven: Yale University Press, 2014): 97-129.

21. Comfort, *The Science of Human Perfection*, 97-239.

22. A. Jensen, "How Much Can We Boost Scholastic Achievement?," *Harvard Educational Review* 39 (1969): 1-123, 82 (see also 81-88).

23. R. J. Herrnstein, "I.Q.," *Atlantic* 228, no. 3 (1971): 43-64.

24. R. J. Herrnstein and C. A. Murray, *The Bell Curve: Intelligence and Class Structure in American Life* (New York: Free Press, 1994).

25. A. Panofsky, K. Dasgupta, and N. Iturriaga, "How White Nationalists Mobilize Genetics: From Genetic Ancestry and Human Biodiversity to Counterscience and Metapolitics," *American Journal of Biological Anthropology* 175, no. 2 (2021): 387-98.

26. K. P. Harden, *The Genetic Lottery: Why DNA Matters for Social Equality* (Princeton: Princeton University Press, 2021); J. Tabery, *Beyond Versus: The Struggle to Understand the Interaction of Nature and Nurture,* Life and Mind: Philosophical Issues in Biology and Psychology (Cambridge: MIT Press, 2014); J. Tabery, "Why Is Studying the Genetics of Intelligence So Controversial?," *Hastings Center Report* 45, no. S1 (2015): S9-S14.

27. L. L. Heston, "Psychiatric Disorders in Foster Home Reared Children of Schizophrenic Mothers," *British Journal of Psychiatry* 12 (1966): 819-25.

28. T. J. C. Polderman et al., "Meta-analysis of the Heritability of Human Traits Based on Fifty Years of Twin Studies," *Nature Genetics* 47, no. 7 (2015): 702-9.

29. J. Joseph, *The Trouble with Twin Studies: A Reassessment of Twin Research in the Social and Behavioral Sciences* (London: Routledge, 2016).

30. E. Turkheimer, "Three Laws of Behavior Genetics and What They Mean," *Current Directions in Psychological Science* 9, no. 5 (2000): 160-64. See also D. Conley et al., "Heritability and the Equal Environments Assumption: Evidence from Multiple Samples of Misclassified Twins," *Behavior Genetics* 43, no. 5 (2013): 415-26; S. Purcell, "Variance Components Models for Gene-Environment Interaction in Twin Analysis," *Twin Research* 5, no. 6 (2002): 554-71; and D. Conley and J. Fletcher, *The Genome Factor: What the Social Genomics Revolution Reveals about Ourselves, Our History, and the Future* (Princeton, NJ: Princeton University Press, 2017).

31. R. A. Fisher, "The Correlation between Relatives on the Supposition of Mendelian Inheritance," *Transactions of the Royal Society of Edinburgh* 53 (1918): 399-433.

32. J. L Lush, *Animal Breeding Plans* (Ames, IA: Collegiate Press, 1937).

33. S. M. Downes and E. Turkheimer, "An Early History of the Heritability Coefficient Applied to Humans (1918-1960)," *Biological Theory* 17, no. 2 (2022): 126-37.

34. Turkheimer, "Three Laws of Behavior Genetics and What They Mean."

35. R. E. Nickel et al., "Velo-Cardio-Facial Syndrome and DiGeorge Sequence with Meningomyelocele and Deletions of the 22q11 Region," *American Journal of Medical Genetics* 52, no. 4 (1994): 445-49.

36. C. Suplee, "Study Provides New Evidence of 'Gay Gene,'" *Washington Post*, October 31, 1995.

37. N. Wade, "First Gene to Be Linked with High Intelligence Is Reported Found," *New York Times*, May 14, 1998.

38. "'Warrior Gene' Blamed for Maori Violence," *Sydney Morning Herald*, August 9, 2006.

39. M. Duenwald, "Gene Is Linked to Susceptibility to Depression," *New York Times*, July 18, 2003.

40. J. N. Hirschhorn et al., "A Comprehensive Review of Genetic Association Studies," *Genetics in Medicine* 4, no. 2 (2002): 45-61.

41. J. N. Hirschhorn and M. J. Daly, "Genome-Wide Association Studies for Common Diseases and Complex Traits," *Nature Reviews Genetics* 6, no. 2 (2005): 95-108.

42. But one of the first GWASs, of macular degeneration, did find replicable associations. R. J. Klein et al., "Complement Factor H Polymorphism in Age-Related Macular Degeneration," *Science* 308, no. 5720 (2005): 385-89; S. Zareparsi et al., "Strong Association

of the Y402H Variant in Complement Factor H at 1q32 with Susceptibility to Age-Related Macular Degeneration," *American Journal of Human Genetics* 77, no. 1 (2005): 149-53. GWASs of schizophrenia have also motivated further exploration of the genetics of schizophrenia. C. A. Dennison et al., "Genome-Wide Association Studies in Schizophrenia: Recent Advances, Challenges and Future Perspective," *Schizophrenia Research* 217 (2020): 4-12.

43. See, for example, N. R. Wray et al., "Genome-Wide Association Study of Major Depressive Disorder: New Results, Meta-analysis, and Lessons Learned," *Molecular Psychiatry* 17, no. 1 (2012): 36-48, and P. Sullivan, on behalf of 96 Psychiatric Genetics Investigators, "Don't Give up on GWAS," *Molecular Psychiatry* 17, no. 1 (2012): 2-3.

44. C. F. Chabris et al., "The Fourth Law of Behavior Genetics," *Current Directions in Psychological Science* 24, no. 4 (2015): 304-12, at 305. The second and third laws, articulated by Turkheimer, concern what are referred to as the "shared" (or "family") and "nonshared environments." Turkheimer, "Three Laws of Behavior Genetics and What They Mean."

45. M. R. Nelson et al., "The Support of Human Genetic Evidence for Approved Drug Indications," *Nature Genetics* 47, no. 8 (2015): 856-60. GWAS has yet to deliver a lot of treatments—drug development takes a long time—but has certainly radically altered the landscape of what the mechanistic hypotheses might be. For instance, Alzheimer's research is now focused on microglial biology because of genetic links rather than neuronal biology, as these seem to be relevant to the disease processes that lead to the neuronal death.

46. W. R. Reay and M. J. Cairns, "Advancing the Use of Genome-Wide Association Studies for Drug Repurposing," *Nature Reviews Genetics* 22, no. 10 (2021): 658-71.

47. To date, the best case for having learned about causal mechanisms for a psychiatric phenotype comes from a GWAS of schizophrenia. V. Trubetskoy et al., "Mapping Genomic Loci Implicates Genes and Synaptic Biology in Schizophrenia," *Nature* 604 (2022): 502-8. For nonpsychiatric behavioral phenotypes, the best case may come from a GWAS of risk tolerance: R. K. Linnér et al., "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, no. 2 (2019): 245-57. Also see this study on nicotinic receptor genes and smoking or nicotine dependence: The Tobacco and Genetics Consortium, "Genome-Wide Meta-Analyses Identify Multiple Loci Associated with Smoking Behavior," *Nature Genetics* 42, no. 5 (2010): 441-47.

48. "Polygenic Risk Scores," National Human Genome Research Institute, accessed January 12, 2023, https://www.genome.gov/Health/Genomics-and-Medicine/Polygenic-risk-scores.

49. C. Jencks, "Heredity, Environment, and Public Policy Reconsidered," *American Sociological Review* 45, no. 5 (1980): 723-36.

50. For example, L. J. Howe et al., "Within-Sibship Genome-Wide Association Analyses Decrease Bias in Estimates of Direct Genetic Effects," *Nature Genetics* 54, no. 5 (2022): 581-92.

51. Ibid.

52. J. W. O'Sullivan et al., "Polygenic Risk Scores for Cardiovascular Disease: A Scientific Statement from the American Heart Association," *Circulation* 146, no. 8 (2022): e93-e118; N. Zeinomar and W. K. Chung, "Cases in Precision Medicine: The Role of Polygenic Risk Scores in Breast Cancer Risk Assessment," *Annals of Internal Medicine* 174, no. 3 (2021): 408-12.

53. A. R. Martin et al., "Clinical Use of Current Polygenic Risk Scores May Exacerbate Health Disparities," *Nature Genetics* 51, no. 4 (2019): 584-91. For concerns about portability within populations, see also H. Mostafavi et al., "Variable Prediction Accuracy of Polygenic Scores within an Ancestry Group," *ELife* 9 (2020): e48376.

54. The 1000 Genomes Project Consortium, "A Global Reference for Human Genetic Variation," *Nature* 526, no. 7571 (2015): 68-74.

55. M. Slatkin, "Linkage Disequilibrium—Understanding the Evolutionary Past and Mapping the Medical Future," *Nature Reviews Genetics* 9, no. 6 (2008): 477-85.

56. A. Panofsky and C. Bliss, "Ambiguity and Scientific Authority: Population Classification in Genomic Science," *American Sociological Review* 82, no. 1 (2017): 59-87; A. C. F. Lewis et al., "Getting Genetic Ancestry Right for Science and Society," *Science* 376, no. 6590 (2022): 250-52; Y. Ding et al., "Polygenic Scoring Accuracy Varies across the Genetic Ancestry Continuum in All Human Populations," bioRxiv, preprint, posted September 29, 2022, https://www.biorxiv.org/content/10.1101/2022.09.28.509988v1.

57. M. Przeworski, Twitter, September 14, 2021, https://twitter.com/molly_przew/status/1437863695200264197.

58. D. Nelson et al., "Inferring Transmission Histories of Rare Alleles in Population-Scale Genealogies," *American Journal of Human Genetics* 103, no. 6 (2018): 893-906.

59. M. Sohail et al., "Polygenic Adaptation on Height Is Overestimated Due to Uncorrected Stratification in Genome-Wide Association Studies," *Elife* (2019): e39702; J. J. Berg et al., "Reduced Signal for Polygenic Adaptation of Height in UK Biobank," *Elife* (2019): e39725.

60. Although today some work aims at creating "cross-ancestry PGIs" (see, for example, P. Tshiaba et al., "Cross-Ancestry Polygenic Risk Score for Breast Cancer Risk Assessment," supplement, *Journal of Clinical Oncology* 40, no. 16 [2022]: 10540), for at least the foreseeable future, that work does not obviate the need to use genetically similar samples in creating PGIs. That is because, to do such work today, researchers first split their samples into multiple genetically similar groups; then they perform a GWAS or PGI analysis on each group; and finally, they meta-analyze or jointly report the results. Current tools cannot handle doing all the analysis in one step if the sample is genetically diverse.

61. Genetics researchers frequently write simply of "ancestry" (rather than "genetic ancestry"), which is even more easily conflated with genealogy, as well as with race and ethnicity. B. Dauda et al., "Ancestry: How Researchers Use It, and What They Mean by It," *Frontiers in Genetics* 14 (2023): doi:10.3389/fgene.2023.1044555.

62. G. Coop, "Genetic Similarity versus Genetic Ancestry Groups as Sample Descriptors in Human Genetics," arXiv, preprint, revised version, submitted January 7, 2023, https://arxiv.org/abs/2207.11595#; Lewis et al., "Getting Genetic Ancestry Right for Science and Society."

63. A. B. Popejoy and S. M. Fullerton, "Genomics Is Failing on Diversity," *Nature* 538, no. 7624 (2016): 161-64.

64. Martin et al., "Clinical Use of Current Polygenic Risk Scores May Exacerbate Health Disparities."

65. A. R. Bentley, S. Callier, and C. N. Rotimi, "Diversity and Inclusion in Genomic Research: Why the Uneven Progress?," *Journal of Community Genetics* 8, no. 4 (2017): 255-66.

66. Recent PGIs for height and BMI explain about 40 percent and 6 percent, respectively, of the variation among people of European genetic ancestries. L. Yengo et al., "A Saturated Map of Common Genetic Variants Associated with Human Height," *Nature* 610 (2022): 704-12; L. Yengo et al., "Meta-analysis of Genome-Wide Association Studies for Height and Body Mass Index in ≈ 700000 Individuals of European Ancestry," *Human Molecular Genetics* 27 (2018): 3641-49.

67. L. S. Gottfredson, "Mainstream Science on Intelligence: An Editorial with 52 Signatories, History, and Bibliography," *Intelligence*

24, no. 1 (1997): 13-23, cited in R. E. Nisbett et al., "Intelligence: New Findings and Theoretical Developments," *American Psychologist* 67 (2012): 130-59, at 131.

68. P. Salovey, J. Mayer, and D. Caruso, "Emotional Intelligence: Theory, Findings, and Implications," *Psychological Inquiry* 15, no. 3 (2004) 197-215.

69. Nisbett et al., "Intelligence."

70. A. Okbay et al., "Polygenic Prediction of Educational Attainment within and between Families from Genome-Wide Association Analyses in 3 Million Individuals," *Nature Genetics* 54, no. 4 (2022): 437-49.

71. This is approximately an average of the estimates from twin studies. The estimated heritability varies by population; see K. Silventoinen et al., "Genetic and Environmental Variation in Educational Attainment: An Individual-Based Analysis of 28 Twin Cohorts," *Scientific Reports* 10, no. 1 (2020): doi:10.1038/s41598-020-69526-6.

72. A. Kong et al., "The Nature of Nurture: Effects of Parental Genotypes," *Science* 359, no. 6374 (2018): 424-28. Also included as part of the confound of gene-environment correlation are what are called "indirect genetic effects" (or "genetic nurture"), which are the causal effects of parents' genetic variants on parental phenotypes (such as a parent's educational attainment, a parent's income, and parenting behavior), which in turn affects the individual. These indirect effects are reflected in the associations of an individual's own genetic variants because the genetic variants are shared between the individual and their parents. Similarly, indirect genetic effects that operate through siblings or other relatives are also included among the environmental confounds.

73. A. I. Young et al., "Relatedness Disequilibrium Regression Estimates Heritability without Environmental Bias," *Nature Genetics* 50, no. 9 (2018): 1304-10.

74. The terms we use here, "causal-effect" SNP heritability and the "correlational" SNP heritability, are not standard. The academic literature refers to both by the ambiguous term "SNP heritability." We use our terminology here because we believe that it is important to make the distinction.

75. Other factors also contribute to the difference. These include interaction effects between genetic variants (whose effects are captured by heritability estimates from twin studies to a much greater extent than in causal-effect SNP heritability estimates), violations of the assumptions of the methods used to estimate heritabilities (which could bias those estimates), and differences in samples used in the studies, since heritabilities vary by sample.

76. J. J. Lee et al., "Gene Discovery and Polygenic Prediction from a Genome-Wide Association Study of Educational Attainment in 1.1 Million Individuals," *Nature Genetics* 50, no. 8 (2018): 1112-21.

77. The other main—probably even more important—reason that the PGI R^2 falls short of the correlational SNP heritability is that the GWAS used to construct the PGIs mixes different samples that have different SNP associations. See R. de Vlaming et al., "Meta-GWAS Accuracy and Power (MetaGAP) Calculator Shows That Hiding Heritability Is Partially Due to Imperfect Genetic Correlations across Studies," *PLOS Genetics* 13, no. 1 (2017): e1006495, and F. C. Tropf et al., "Mega-analysis of 31,396 Individuals from 6 Countries Uncovers Strong Gene-Environment Interaction for Human Fertility," *Nature Human Behaviour* 1 (2017): 757-65. In the future, if very large GWAS are conducted in highly homogeneous samples, then the PGI R^2 is expected to approach the correlational SNP heritability.

78. Bliss, *Social by Nature*.

79. "Research Must Do No Harm: New Guidance Addresses All Studies Relating to People," editorial, *Nature* 606, no. 434 (2022): doi:10.1038/d41586-022-01607-0.

80. For instance, downstream uses—for some, misuses—of SBG research (such as PGT-P for social and behavioral phenotypes), by definition, would not exist without the underlying SBG research. That is not, of course, to say that SBG research should therefore not be conducted—and, as we have noted, if SBG research were not conducted by mainstream scientists, it would almost certainly continue to be conducted by those on the margins of the academy; our point here is simply to recognize that there is an inescapable connection between the research and applications.

81. See, for example, A. E. Justice et al., "Genome-Wide Association Study of Body Fat Distribution Traits in Hispanics/Latinos from the HCHS/SOL," *Human Molecular Genetics* 30 (2021): 2190-2204.

82. See, for instance, D. Casares-Marfil et al., "Admixture Mapping Analysis Reveals Differential Genetic Ancestry Associated with Chagas Disease Susceptibility in the Colombian Population," *Human Molecular Genetics* 30 (2021): 2503-12.

83. See Panofsky and Bliss, "Ambiguity and Scientific Authority."

84. L. Gannett, "Biogeographical Ancestry and Race," *Studies in the History and Philosophy of Science, Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 47, Part AS (2014): 173-84; L. Gannett, "Racism and Human Genome Diversity Research: The Ethical Limits of 'Population Thinking,'" *Philosophy of Science* 68, no. S3 (2001): S479-S92.

85. Genetic Information Nondiscrimination Act of 2008 (GINA), Pub. L. 110-233, 122 Stat. 881.

86. D. Conley, "From Fraternities to DNA: The Challenge Genetic Prediction Poses to Insurance Markets," *Millbank Quarterly* 97, no. 1 (2023): 40-43.

87. See, in this special report, D. Martschenko et al., "Wrestling with Public Input on an Ethical Analysis of Scientific Research," in *The Ethical Implications of Social and Behavioral Genomics*, ed. E. Parens and M. N. Meyer, special report, *Hastings Center Report* 53, no. 2 (2023): S50-S65.

88. M. N. Meyer, "There Oughta Be a Law: When Does(n't) the U.S. Common Rule Apply?," supplement, *Journal of Law, Medicine, & Ethics* 48, no. 1 (2020): 60-73.

89. L. J. Matthews et al., "Pygmalion in the Genes? On the Potentially Negative Impacts of Polygenic Scores for Educational Attainment," *Social Psychology of Education* 24, no. 3 (2021): 789-808. But see M. N. Meyer, T. Gjorgjieva, and C. F. Chabris, "Laypeople Overestimate the Predictive Power of Polygenic Scores but Do Not View Them as Any More Anxiety-Producing Than Other Scores for the Same Trait," *Behavior Genetics* 52, no. 6 (2022): 378.

90. P. Turley et al., "Problems with Using Polygenic Scores to Select Embryos," *New England Journal of Medicine* 385, no. 1 (2021): 78-86.

91. Kong et al., "The Nature of Nurture."

92. M. Nivard et al., "Neither Nature nor Nurture: Using Extended Pedigree Data to Elucidate the Origins of Indirect Genetic Effects on Offspring Educational Outcomes," PsyArXiv, preprint, last edited October 6, 2022, DOI:10.31234/osf.io/bhpm5.

93. Lee et al., "Gene Discovery and Polygenic Prediction from a Genome-Wide Association Study of Educational Attainment in 1.1 Million Individuals." To be clear, EA3 tested only one mechanism of genetic explanation for sex differences—X-chromosome compensation. However, there are many other possible genetic mechanisms to explain sex differences such as transacting interactions with Y-chromosome genes. Our point, however, is that SBG research can help refute harmful stereotypes. Indeed, arguably only genetics re-

search can refute stereotypes that are already based on beliefs about genetics, as many are.

94. Generally, PGIs are meant to identify groups of people who, on average, have higher levels of a phenotype or have a greater likelihood of having some phenotype. A "variance PGI" (or "vPGI") is a PGI that identifies a group of people who have a higher variance in the phenotype (due to genetic regulation of phenotypic plasticity) than the group of people with a low-variance PGI. R. Johnson, R. Sotoudeh, and D. Conley, "Polygenic Scores for Plasticity: A New Tool for Studying Gene-Environment Interplay," *Demography* 59, no. 3 (2022): 1045-70.

95. L. L. Schmitz et al., "The Impact of Late-Career Job Loss and Genetic Risk on Body Mass Index: Evidence from Variance Polygenic Scores," *Scientific Reports* 11, no. 1 (2021): doi:10.1038/s41598-021-86716-y.

96. I. Y. Millwood et al., "Conventional and Genetic Evidence on Alcohol and Vascular Disease Aetiology: A Prospective Study of 500,000 Men and Women in China," *Lancet* 393 (2019): 1831-42.

97. First, the genetic variant must be sufficiently strongly associated with the modifiable treatment. Second, the genetic variant cannot affect the outcome through any channel other than the modifiable treatment. The latter "exclusion restriction" rules out using PGIs and most individual genetic variants (due to pleiotropy, linkage disequilibrium, and gene-environment correlation). The former rules out using genetic variants that don't have large effects. Indeed, Emily Oster, in her newsletter ParentData (see https://www.parentdata.org/), has criticized the Millwood et al. study we discuss on this ground. That said, there are a few good examples of Mendelian randomization studies, which use SNPs that are known to be related to a modifiable risk factor to help determine the causal influence of the risk factor. For early reviews, see G. D. Smith and S. Ebrahim, "'Mendelian Randomization': Can Genetic Epidemiology Contribute to Understanding Environmental Determinants of Disease?," *International Journal of Epidemiology* 32, no. 1 (2003): 1-22, and C. A. Emdin, A. V. Khera, and S. Kathiresan, "Mendelian Randomization," *Journal of the American Medical Association* 318 (2017): 1925-26. And in some areas (for example, epidemiology) where randomization is often infeasible, Mendelian randomization, while imperfect, may often be better than observational studies. Some have argued that null results of Mendelian randomization studies are particularly illuminating when the observational correlation is non-null. This result suggests that the observed correlation is not causal and that researchers should not waste time trying to understand a causal relationship. M. Van de Weijer et al., "Disentangling Potential Causal Effects of Educational Duration on Well-Being, and Mental and Physical Health Outcomes," PsyArXiv, preprint, submitted September 28, 2022, https://psyarxiv.com/4uqyg/.

98. See EA1 SI: C. A. Rietveld et al., "GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment," *Science* 340 (2013): 1467-71.

99. Okbay et al., "Polygenic Prediction of Educational Attainment within and between Families from Genome-Wide Association Analyses in 3 Million Individuals."

100. Lee et al., "Gene Discovery and Polygenic Prediction from a Genome-Wide Association Study of Educational Attainment in 1.1 Million Individuals." These numbers come from a combined analysis of the Health and Retirement Study and the National Longitudinal Study of Adolescent to Adult Health.

101. Ibid.

102. We say "average" because the SNP heritability depends on the sample. For example, for educational attainment, it appears to be higher than 20 percent in Add Health (a large, U.S. federally funded longitudinal survey of adolescents) and lower than 20 percent in the Health and Retirement Study. However, in EA4, researchers meta-analyzed across different samples with an average genetic correlation with one other of about 0.70 (according to EA3 estimates). This reduces the SNP heritability of the meta-analysis sample relative to what it would be in the constituent samples (de Vlaming et al., "Meta-GWAS Accuracy and Power [MetaGAP] Calculator"), probably to something more like 15 percent. See EA4 SI (Okbay et al., "Polygenic Prediction of Educational Attainment within and between Families from Genome-Wide Association Analyses in 3 Million Individuals") and EA3 estimates (Lee et al., "Gene Discovery and Polygenic Prediction from a Genome-Wide Association Study of Educational Attainment in 1.1 Million Individuals"). However, in the future, as larger GWAS samples become available in homogeneous samples, it should be possible to conduct large GWAS in the homogeneous samples, which have higher SNP heritability.

103. A. C. Fahed, A. A. Philippakis, and A. V. Khera, "The Potential of Polygenic Scores to Improve Cost and Efficiency of Clinical Trials," *Nature Communications* 13 (2022): doi:10.1038/s41467-022-30675-z.

104. D. S. Yeager et al., "A National Experiment Reveals Where a Growth Mindset Improves Achievement," *Nature* 573 (2019): 364-69.

105. R. Sotoudeh, K. M. Harris, and D. Conley, "Effects of the Peer Metagenomic Environment on Smoking Behavior," *PNAS* 116, no. 33 (2019): 16302-7; M. C. Mills and F. C. Tropf, "Sociology, Genetics, and the Coming of Age of Sociogenomics," *Annual Review of Sociology* 46, no. 1 (2020): 553-81.

106. For instance, in developmental psychology, the stability of genetic variants over time and the randomization of DNA sequence variation within families make PGIs particularly helpful in addressing questions about specific mechanisms of parent-to-child transmission of behavior, psychopathology, and human capital; who is being served by interventions and policies and who is being left behind; and developmental precursors of adult lifespan outcomes (such as the "long arm of childhood"). L. Raffington, T. Mallard, and K. P. Harden, "Polygenic Scores in Developmental Psychology: Invite Genetics in, Leave Biodeterminism Behind," *Annual Review of Developmental Psychology* 2, no. 1 (2020): 389-411.

107. S. H. Barcellos, L. S. Carvalho, and P. Turley, "Education Can Reduce Health Differences Related to Genetic Risk of Obesity," *PNAS* 115, no. 42 (2018): E9765-72; N. W. Papageorge and K. Thom, "Genes, Education, and Labor Market Outcomes: Evidence from the Health and Retirement Study," *Journal of the European Economic Association* 18, no. 3 (2020): 1351-99; D. Barth, N. W. Papageorge, and K. Thom, "Genetic Endowments and Wealth Inequality," *Journal of Political Economy* 128, no. 4 (2020): 1474-1522; M. A. Houmark, V. Ronda, and M. Rosholm, "The Nurture of Nature and the Nature of Nurture: How Genes and Investments Interact in the Formation of Skills," IZA Discussion Papers 13780, Institute of Labor Economics, 2020; A. Sanz-de-Galdeano and A. Terskaya, "Sibling Differences in Educational Polygenic Scores: How Do Parents React?," IZA Discussion Papers 12375, Institute of Labor Economics, 2019.

108. D. W. Belsky et al., "Genetic Analysis of Social-Class Mobility in Five Longitudinal Studies," *PNAS* 115 (2018): E7275-84; A. Abdellaoui et al., "Genetic Correlates of Social Stratification in Great Britain," *Nature Human Behaviour* 3 (2019): 1332-42.

109. Okbay et al., "Polygenic Prediction of Educational Attainment within and between Families from Genome-Wide Association Analyses in 3 Million Individuals."

110. D. Antaki et al., "A Phenotypic Spectrum of Autism Is Attributable to the Combined Effects of Rare Variants, Polygenic Risk and Sex," *Nature Genetics* 54 (2022): 1284-92.

111. V. Bansal et al., "Genome-Wide Association Study Results for Educational Attainment Aid in Identifying Genetic Heterogeneity of Schizophrenia," *Nature Communications* 9, no. 1 (2018): doi: 10.1038/s41467-018-05510-z.

112. B. Finucane et al., "Shift Happens: Family Background Influences Clinical Variability in Genetic Neurodevelopmental Disorders," *Genetics in Medicine* 18, no. 4 (2016): 302-4.

113. M. T. Oetjens et al., "Quantifying the Polygenic Contribution to Variable Expressivity in Eleven Rare Genetic Disorders," *Nature Communications* 10, no. 1 (2019): doi:10.1038/s41467-019-12869-0; C. M. Taylor et al., "Phenotypic Shift in Copy Number Variants: Evidence in 16p11.2 Duplication Syndrome," *Genetics in Medicine* 25, no. 1 (2023): 151-54.

114. By "policy," we very broadly mean a law, regulation, procedure, administrative action, rule, guideline, or incentive program of a governmental or other institution. By "practice," we have in mind customary ways of doing things—for instance, a practice of risk stratification in medicine/healthcare delivery.

115. K. Asbury and R. Plomin, *G Is for Genes: The Impact of Genetics on Education and Achievement* (Chichester, West Sussex: John Wiley & Sons, 2013).

116. Compare A. J. London, "Artificial Intelligence and Black-Box Medical Decisions: Accuracy versus Explainability," *Hastings Center Report* 49, no. 1 (2019): 15-21.

117. "Research Ethics," Nature Portfolio, accessed November 19, 2022, https://www.nature.com/nature-portfolio/editorial-policies/ethics-and-biosecurity.

118. We are speaking here in absolute, not relative, terms. We do not here address, for instance, complex macrolevel questions of whether any line of SBG research is more worthy of funding dollars or journal resources than other research.

119. This conclusion is in agreement with the ethics policy of Nature Portfolio, "Research Ethics," cited above.

120. Ibid.

121. Of course, researchers should always stand by the validity and reproducibility of whatever they publish, as far as it goes. But there is a difference between, for example, exploratory and hypothesis-confirming research and between a single hypothesis-confirming study and a line of direct or conceptual replications. Here, we are suggesting that the more socially risky an area of research is, the more confident the researchers should be that the current study will be confirmed in the long term.

122. See 45 C.F.R. 46.104(d)(2), requiring limited institutional review board review of data security procedures for studies that would otherwise be fully exempt but collect data that, if identified, would subject the participant to any of these risks. HHS's list of research activities that are eligible for expedited review has a similar limitation: "The expedited review procedure may not be used where identification of the subjects and/or their responses would reasonably place them at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, insurability, reputation, or be stigmatizing, unless reasonable and appropriate protections will be implemented so that risks related to invasion of privacy and breach of confidentiality are no greater than minimal." 63 Fed. Reg. 60,364-67 (Nov. 9, 1998).

123. Whereas a "minority" group may refer, or be read to refer, to a group of people who are smaller than the majority group, in using "minoritized" groups, we intend to refer to groups—of whatever size—that are on the "wrong side" of a power imbalance in a society.

124. M. Niarchou et al., "Genome-Wide Association Study of Musical Beat Synchronization Demonstrates High Polygenicity," *Nature Human Behaviour* 6, no. 9 (2022): 1292-1309.

125. Barth, Papageorge, and Thom, "Genetic Endowments and Wealth Inequality."

126. B. Carrington, *Race, Sport and Politics: The Sporting Black Diaspora* (London: SAGE Publications, 2010) (describing the myth of the "black athlete"); J. Schultz, "Racialized Osteology and Athletic Aptitude, or 'Black' Bones as Red Herrings," *Journal of Sport History* 46, no. 3 (2019): 325-46.

127. R. S. Maril, "Queer Rights after Dobbs v. Jackson Women's Health Organization," *San Diego Law Review* (forthcoming). For an October 6, 2022, preprint version, see doi:10.2139/ssrn.4178187.

128. Belsky et al., "Genetic Analysis of Social-Class Mobility in Five Longitudinal Studies."

129. Martin et al., "Clinical Use of Current Polygenic Risk Scores May Exacerbate Health Disparities"; L. Duncan et al., "Analysis of Polygenic Risk Score Usage and Performance in Diverse Human Populations," *Nature Communications* 10, no. 1 (2019): doi:10.1038/s41467-019-11112-0.

130. E. J. Emanuel, D. Wendler, and C. Grady, "What Makes Clinical Research Ethical?," *Journal of the American Medical Association* 283, no. 20 (2000): 2701-11.

131. See Speech First v. Cartwright, 11th Cir. No. 21-12583 (April 21, 2022) (Marcus, J., concurring). The concurrence states, "By depriving itself of academic institutions that pursue truth over any other concern, a society risks falling into the abyss of ignorance. Humans are not smart enough to have ideas that lie beyond challenge and debate."

132. 45 C.F.R. 46.111(a)(2).

133. K. A. Bird, "No Support for the Hereditarian Hypothesis of the Black-White Achievement Gap Using Polygenic Scores and Tests for Divergent Selection," *American Journal of Physical Anthropology* 175, no. 2 (2021): 465-76.

134. Lewis et al., "Getting Genetic Ancestry Right for Science and Society."

135. Emanuel, Wendler, and Grady, "What Makes Clinical Research Ethical?"

136. D. Fox, "Subversive Science," *Penn State Law Review* 124, no. 1 (2019): 153-91.

137. "Science Must Respect the Dignity and Rights of All Humans," editorial, *Nature Human Behaviour* 6, no. 8 (2022): 1029-31.

138. Nature Portfolio, "Research Ethics."

139. J. Rauch, "Nature Human Misbehavior: Politicized Science Is Neither Science nor Progress," Foundation for Individual Rights and Expression, September 14, 2022, https://www.thefire.org/news/nature-human-misbehavior-politicized-science-neither-science-nor-progress; S. Pinker, Twitter, August 26, 2022, https://twitter.com/sapinker/status/1563179979667476482.

140. "Why and How Science Should Respect the Dignity and Rights of All Humans," editorial, *Nature Human Behaviour* 6, no. 10 (2022): 1321-23.

141. Responsibility for SBG research does not end with dissemination. Long after a study has been communicated and researchers have moved on to new projects, studies can reemerge in academic and popular discourse, where they may be misinterpreted or misused. The field has a collective duty to publicly criticize work that uses SBG research data or results in scientifically invalid or harmful ways. We acknowledge that rapidly responding to every misinterpretation and misuse of an SBG research study would be a full-time job and that scientists generally are not incentivized for such work. As with service generally, much of this work should fall to senior, tenured members of the field. We also acknowledge that the skills necessary to do good science and to engage in effective science communication are not identical and not necessarily shared by the same people. Finally, we

acknowledge that when misinterpretation or misuse occurs in a marginal outlet, there is a risk that if reputable scientists engage with such work, it will lend legitimacy to it and have a net negative effect. There have been similar debates about whether and how biologists should engage with creationists. M. Greener, "Taking on Creationism: Which Arguments and Evidence Counter Pseudoscience?," *EMBO Reports* 8, no. 12 (2007): 1107-9.

142. We take for granted that institutional review board review, data access committee review, and other standard research oversight processes will be complied with, as applicable, and we do not describe those processes here.

143. See, for example, J. Carlson et al., "Counter the Weaponization of Genetics Research by Extremists," *Nature* 610, no. 7932 (2022): 444-47 (asking scientists to "consider the potential harmful impacts of their work," 447).

144. For more on some of the forms that public engagement in research can take, see the companion piece in this special report, Martschenko et al., "Wrestling with Public Input on an Ethical Analysis of Scientific Research."

145. A. Ganna et al., "Large-Scale GWAS Reveals Insights into the Genetic Architecture of Same-Sex Sexual Behavior," *Science* 365, no. 6456 (2019): eaat7693.

146. "Genetics of Sexual Behavior: A Website to Communicate and Share the Results from the Largest Study on the Genetics of Sexual Behavior," Genetics of Sexual Behavior, accessed November 18, 2022, https://geneticsexbehavior.info.

147. B. Neale, "Opinion: Community Engagement Strengthens Science," Broad Institute (blog), August 29, 2019, https://www.broadinstitute.org/blog/opinion-community-engagement-strengthens-science.

148. "Science Must Respect the Dignity and Rights of All Humans," editorial, 1029.

149. Carlson et al., "Counter the Weaponization of Genetics Research by Extremists," 146-47.

150. P. Sankar and M. K. Cho, "Toward a New Vocabulary of Human Genetic Variation," *Science* 298, no. 5597 (2002): 1337-38; J. H. Fujimura and R. Rajagopalan, "Different Differences: The Use of 'Genetic Ancestry' versus Race in Biomedical Human Genetic Research," *Social Studies of Science* 41, no. 1 (2011): 5-30.

151. Popejoy and Fullerton, "Genomics Is Failing on Diversity"; Bentley, Callier, and Rotimi, "Diversity and Inclusion in Genomic Research"; Martin et al., "Clinical Use of Current Polygenic Risk Scores May Exacerbate Health Disparities."

152. A. R. Martin et al., "Increasing Diversity in Genomics Requires Investment in Equitable Partnerships and Capacity Building," *Nature Genetics* 54, no. 6 (2022): 740-45.

153. Y. Wang et al., "Challenges and Opportunities for Developing More Generalizable Polygenic Risk Scores," *Annual Review of Biomedical Data Science* 5 (2022): 293-320; P. Turley et al., "Multi-Ancestry Meta-Analysis Yields Novel Genetic Discoveries and Ancestry-Specific Associations," bioRxiv, preprint, submitted April 24, 2021, https://www.biorxiv.org/content/10.1101/2021.04.23.441003v1.

154. Nature Portfolio, "Research Ethics."

155. Dauda et al., "Ancestry."

156. Y. Ding et al., "Polygenic Scoring Accuracy Varies across the Genetic Ancestry Continuum in All Human Populations," bioRxiv, preprint, September 29, 2022, https://www.biorxiv.org/content/10.1101/2022.09.28.509988v1.

157. Lewis et al., "Getting Genetic Ancestry Right for Science and Society."

158. J. M. Hofman, D. G. Goldstein, and J. Hullman, "How Visualizing Inferential Uncertainty Can Mislead Readers about Treatment Effects in Scientific Results," *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020): 1-12; S. Zhang et al., "An Illusion of Predictability in Scientific Results," SocArXiv, preprint, submitted September 29, 2022, https://osf.io/preprints/socarxiv/5tcgs/; E. Soyer and R. M. Hogarth, "The Illusion of Predictability: How Regression Statistics Mislead Experts," *International Journal of Forecasting* 28, no. 3 (2012): 695-711.

159. K. P. Harden and D. W. Belsky, "Predicting Education from DNA?," BOLD, July 27, 2018, https://bold.expert/predicting-education-from-dna/; K. P. Harden, "Genetic Determinism, Essentialism and Reductionism: Semantic Clarity for Contested Science," *Nature Reviews Genetics* 24 (2022): 197-204.

160. Carlson et al., "Counter the Weaponization of Genetics Research by Extremists."

161. For example, in addition to developing FAQs, to accompany a recent GWAS of "musical beat synchronization," the authors devoted box 1 of their discussion to ELSI considerations, warning against deterministic interpretations of the study and misuses of the study to make individual predictions about musicality or allocate musical opportunities and emphasizing the importance of historical context and including participants of diverse ancestries in future studies. Niarchou et al., "Genome-Wide Association Study of Musical Beat Synchronization Demonstrates High Polygenicity."

162. J. Carlson and K. Harris, "Quantifying and Contextualizing the Impact of bioRxiv Preprints through Automated Social Media Audience Segmentation," *PLOS Biology* 18, no. 9 (2020): e3000860.

163. SSGAC principal investigators Daniel Benjamin, David Cesarini, and Philipp Koellinger, along with other social scientists and geneticists—David Laibson, Christopher Chabris, and Peter Visscher—began working with one of us, ethicist Meyer, and consulted Mary Carmichael, a former journalist and then an independent communications consultant. Carmichael suggested that the SSGAC develop and distribute FAQs about their studies to help journalists avoid misinterpreting and misreporting findings.

164. D. Yuhas, "A New Way of Predicting Which Kids Will Succeed in School: Look at Their Genes," *NBC News*, October 14, 2020.

165. A. Nathan, "How Genetic Differences Could Make Schools Better," *NOVA*, PBS, August 3, 2018.

166. M. Molteni, "Are Diplomas in Your DNA?," *Wired*, July 31, 2018.

167. E. Yong, "An Enormous Study of the Genes Related to Staying in School," *Atlantic*, July 23, 2018.

168. Jedidiah Carlson et al. criticize FAQs as "post-hoc communications" that "presuppose[] that non-specialist audiences might 'frequently' draw racist conclusions from the data," and in concluding that "a different approach is warranted," argue that "if we state that contemporary genetics research as a whole is incompatible with racist interpretations, yet acknowledge that human-genetics studies might foster racist interpretations, then something is awry with how we are conducting those studies and communicating the results." Carlson et al., "Counter the Weaponization of Genetics Research by Extremists." First, ideally (and in our experience, often in practice), FAQs are not posthoc but, rather, are planned from the outset of a study and written and revised as the associated scientific article is drafted and revised. Second, leaving aside the fact that FAQs can and do cover much more than warnings against racist misappropriations and therefore serve many other purposes, the history of racist misappropriations of genetics research is such that we think there is room—and, indeed, a critical need—both to improve the way genetics research is communicated in scientific papers and to communicate these important points about scientific papers outside the papers themselves. This is especially the case because academic journals nec-

essarily limit the extent to which authors can speak to nonspecialist audiences in that forum.

169. "Dangerous Work," editorial, *Nature* 502, no. 7469 (2013): 5-6.

170. D. O. Martschenko and M. Smith, "Genes Do Not Operate in a Vacuum, and Neither Should Our Research," *Nature Genetics* 53, no. 3 (2021): 255-56; E. Parens and P. S. Appelbaum, "An Introduction to Thinking about Trustworthy Research into the Genetics of Intelligence," in *The Genetics of Intelligence*, ed. E. Parens and P. Appelbaum, special report, *Hastings Center Report* 45, no. 5 (2015): S2-S8.

171. D. O. Martschenko and L. J. Matthews, "Genomic Findings on Human Behavior and Social Outcomes," in "FAQs on Human Genomics Studies," The Hastings Center, accessed December 28, 2022, https://www.thehastingscenter.org/genomics-research-index; D. O. Martschenko et al., "FoGS Provides a Public FAQ Repository for Social and Behavioral Genomic Discoveries," *Nature Genetics* 53, no. 9 (2021): 1272-74.

172. "Genetics of Sexual Behavior."

173. Broad Communications, "Perspectives on the Complex Genetics of Same-Sex Sexual Behavior," Broad Institute, August 29, 2019, https://www.broadinstitute.org/news/perspectives-complex-genetics-same-sex-sexual-behavior.

174. Turkheimer, "Three Laws of Behavior Genetics and What They Mean."

175. C. M. Condit, "Laypeople Are Strategic Essentialists, Not Genetic Essentialists," in *Looking for the Psychosocial Impacts of Genomic Information*, ed. E. Parens and P. Appelbaum, special report, *Hastings Center Report* 49, no. 1 (2019): S27-37.

176. M. N. Meyer, T. Gjorgjieva, and C. F. Chabris, "Laypeople Overestimate the Predictive Power of Polygenic Scores but Do Not View Them as Any More Anxiety-Producing Than Other Scores for the Same Trait," paper presented at the 2022 Behavior Genetics Association Annual Meeting, Los Angeles, CA, June 25, 2022.

177. Okbay et al., "Polygenic Prediction of Educational Attainment within and between Families from Genome-Wide Association Analyses in 3 Million Individuals."

178. "E. Parens," Aeon, accessed January 12, 2023, https://aeon.co/users/erik-parens.

179. A. Angers et al., *Genome-Wide Association Studies, Polygenic Scores, and Social Science Genetics: Overview and Policy Implications* (Luxembourg: Publications Office of the European Union, 2019).

180. C. Rotimi, "ASHG Statement regarding the Warping of Genetic Knowledge to Feed Racist Ideology," American Society of Human Genetics, May 27, 2022, https://contentsharing.net/actions/email_web_version.cfm?recipient_id=4111175288&message_id=21822314&user_id=ASHG&group_id=6364825&jobid=56100554; "IBG Statement in Response to the Buffalo Shooting," Institute for Behavioral Genetics Research and Innovation Office at the University of Colorado, May 19, 2022, https://www.colorado.edu/ibg/2022/05/19/ibg-statement-response-buffalo-shooting; Carlson et al., "Counter the Weaponization of Genetics Research by Extremists."