Philosophy of Medicine
Original Research

# Polygene Risk Scores
# A Philosophical Exploration

James Woodward[1*] and Kenneth Kendler[2]

1 Emeritus, Department of History and Philosophy of Science, University of Pittsburgh, Pittsburgh, PA, USA.
2 Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA.

*Correspondence: jfw@pitt.edu

## Abstract

This paper explores the interpretation and use of polygenic risk scores (PRSs). We argue that PRSs generally do not directly embody causal information. Nonetheless, they can assist us in tracking other causal relationships concerning genetic effects. Although their purely predictive/correlational use is important, it is this tracking feature that contributes to their potential usefulness in other applications, such as genetic dissection, and their use as controls, which allow us, indirectly, to "see" more clearly the role of environmental variables.

## 1. Introduction

This paper explores the interpretation and use of polygenic risk scores (PRSs). These make use of associations between individual single nucleotide polymorphisms (SNPs) obtained from genome-wide association studies (GWAS), and traits of interest, such as height, liability to a disease, or educational attainment (EA). PRSs are increasingly widely used in medicine, psychiatry, genetics, and social science. They raise a number of interesting issues, many of which have not yet been explored in the philosophical literature. PRSs are most obviously used for predictive purposes but some scientists, such as James W. Madole and K. Paige Harden (2022), suggest that PRSs (or the SNP/trait correlations that go into them) can be interpreted causally in some circumstances via an analogy between randomized experiments and meiosis in the case of within-siblings designs, as discussed below. Other writers (Bourrat 2020) do not explicitly discuss PRSs but do suggest that the SNP/trait correlations that go into them should be understood as causal, at least in many cases.

This raises one (but not the only) set of questions we explore in this paper. What criteria need to be met for such correlations to count as "causal"? Do SNP/trait correlations or PRSs typically satisfy such criteria? In keeping with a good deal of recent philosophical discussion (for example, Bourrat 2020), we argue that the appropriate criteria for causality are



ISSN 2692-3963 | Date Received 09-02-2023| Date Accepted 12-06-2023| Date First Published 05-07-2023
Philosophy of Medicine | DOI 10.5195/pom.2023.156 | Volume 4 | Issue 1 | pp.1–21

interventionist in the sense of James Woodward (2003). However, the particular criterion for which we argue in the context of GWAS is different from the one usually assumed in the literature—it makes use of the notion of average causal effect, which is a population-level notion, rather than a notion of what it is for a genetic factor to be causal in a particular individual.[1] Our conclusion is that for the most part SNP/outcome correlations are not causal according to this criterion and hence that the correlations described by PRSs are not causal. However, PRSs can in some cases (when there is appropriate control for confounding) *track* or *indicate* the presence of genetic factors that *are* causally relevant to outcomes, without the PRS correlations themselves having a straightforward causal interpretation. We say more below about how we understand the notion of "tracking" but the basic idea is that it involves a correlational but noncausal relationship that arises as a consequence of (and hence "tracks") an underlying causal structure that is at best only partially identified. In the case of PRSs, this underlying structure is, of course, the causally relevant factors in the genotype and it is understood that the correlations in the PRS arise as a result of aspects of the genotype, such as linkage equilibrium, even though the details of how this works are unknown in many cases. Roughly speaking, a good PRS is not just predictively successful but must also satisfy the stronger condition that it is predictively successful *because* it tracks underlying genetic influences that are causal. Thus, to the extent that a PRS is predictively successful only because it reflects the operation of environmental, nongenetic influences (as noted below, a relatively common occurrence if there are inadequate controls), it will not achieve the tracking function just described.

Elucidating the tracking function of PRSs is important for the following reason, among others: One usually thinks of a confounder *Z* for the relationship between *X* and *Y* as a factor that is itself a cause—for example, it might be that *Z* causes *Y* and is correlated with *X*. However, even if *Z* is unknown, if we can find a factor *W* that appropriately tracks *Z,* we can control for the influence of *Z* by controlling for *W,* even if *W* itself is not causal for *Y*. This is what happens when PRSs are used as controls for unknown and potentially confounding genetic factors that *are* causal for some outcome. This point is illustrated by several examples later in this paper. To our knowledge, this is not an idea that has been explicitly discussed (in the way formulated above) in the philosophical literature.

PRSs are, to some extent, like traditional heritability estimates from twin studies, which can indicate that genetic factors are present that causally contribute to differences in outcomes in a population, without further identifying the nature of such factors, with the difference that PRSs are explicitly built from GWAS correlations. Of course, it is not news that heritability coefficients reflect the operation of genetic factors that are causal without explicitly identifying these, but what is novel about PRSs is that the particular information they incorporate can, in turn, support various inferences that would not otherwise be warranted. This includes, in some cases, inferences that would not be warranted by traditional heritability studies, as illustrated by the "nature of nurture" study (Kong et al. 2018) discussed below.

[1] In other words, rather than a notion of causation that applies to manipulation of an SNP in a single individual, or a collection of identical individuals, in asking about the extent to which GWAS correlations are causal, we need a notion of cause that applies to a population of causally heterogeneous individuals (this is the notion of average causal effect) and an accompanying notion of intervention that characterizes what it means to intervene on such a population.

Examples of various uses of PRSs, discussed below, include the following. First (section 7), PRSs can be used for genetic dissection—that is, to uncover the extent to which apparently different diseases share a common genetic basis, as revealed by overlapping PRSs, and the extent to which they can be discriminated genetically. Second, although we acknowledge that this is more controversial, if the appropriate assumptions are satisfied, one can sometimes leverage the associational information in PRSs to reach causal conclusions involving *other* variables, different from those incorporated in the PRSs—that is, PRSs can sometimes be used as controls in reaching causal conclusions, even if the scores are not themselves causal. In particular, it is possible to sometimes use PRSs to provide insight into the causal role of "environmental" variables in contributing to outcomes. Finally, the predictive successes of PRSs can be interpreted as telling us something about overall genetic architecture—in particular as supporting a picture according to which many traits are influenced by a large number of different genes, each of which has a small effect. The idea that *PRSs* can be used in the ways just described and the logic of such uses has, we believe, not been widely discussed in the philosophical literature. One consequence is that for some purposes it may not matter so much whether SNP/trait or PRS correlations are causal—they can be legitimately used in the ways described above even if they are not causal.[2]

Despite these potentially useful features, PRSs (in at least some of their applications) have come under a good deal of criticism recently. One criticism is that because they are not interpretable as causal and because in many cases the predictive successes of PRSs are (at least, so far) modest, the scores do not provide an important improvement on heritability estimates from traditional twin and family studies. In addition, it is claimed that because of their noncausal character, PRSs do not provide information that can be used to guide the discovery of biological mechanisms, [3] or pathways linking genes and outcomes—information that we would like to have in designing interventions, especially in medical contexts. We agree that PRSs do not provide mechanistic information,[4] but think that they provide information beyond traditional family studies of heritability in the ways described in the preceding paragraph.

A second line of criticism (see, for example, Burt 2022) advanced particularly against the development of PRSs for social variables like EA (but by no means only for such variables) is that such scores are inevitably "confounded" by various environmental factors in a way that makes it impossible to use them to disentangle the role of environmental versus genetic influences. We agree that PRSs for social (and a number of other) variables involving unrelated individuals are often confounded but the recent literature has developed ways of detecting some of the sources and the extent of the resulting confounding and correcting for it. One leading approach has been to use PRSs with within-sib designs

[2] In addition PRSs can be estimated for any population of individuals and do not require the special circumstances—twins separated at birth, and so on—needed for more traditional studies of heritability.
[3] Without wishing to wade into the voluminous philosophical literature on mechanisms, we take it that in the scientifically relevant sense a mechanism by which a cause produces an effect must (at least) specify intermediate causal factors or causal pathways that are "between" the cause and effect. Within an interventionist framework, establishing that a causal relationship exists and identifying a mechanism are very different matters: one can establish that *C* causes *E* without knowing or identifying the mechanism linking *C* to *E*. For example, it was widely known that aspirin ingestion causes headache relief before the mechanism by which aspirin produces this effect was known. Establishing that *C* causes *E* is just a matter of establishing that some intervention on *C* is associated with changes in *E;* establishing the mechanism linking *C* to *E* requires much more.
[4] This is widely recognized—see, for example, Plomin and Von Stumm (2022).

that can be compared with results from unrelated individuals (see, for example, Young et al. 2019; Howe et al. 2022). When this is done, the role of "environmental" factors in the causation of social variables appears to be quite substantial, so that the upshot is very far from vindicating strong forms of genetic determinism. In general, our goal is not to defend all uses of PRSs in social genomics but rather to explore the logic of their use in various applications and when they can and cannot be informative.

We also emphasize that our goal is to discuss PRSs rather than to systematically survey the recent philosophical literature on genetic causation, although we will refer to the latter at several points. Nonetheless, it is worth briefly noting some respects in which our treatment extends and departs from this literature. First, the philosophical literature has focused largely on issues having to do with causation by individual SNPS rather than the interpretation of PRSs. Second, we emphasize that the relevant causal notion for interpreting GWAS and PRSs is a population-level notion of average causal effect, rather than SNP causation in individuals, the latter often being the focus of the philosophical literature. Third, we emphasize the importance of the contrast between total or net causation and causation along a path, which is important in the interpretation of PRSs. Fourth, because we hold that SNP/trait correlations are mostly noncausal, we take the additional step of extending the notions of stability and specificity (as characterized in Woodward 2010), previously defined only with respect to causal relations, to predictive relations. Fifth, we emphasize the extent to which the recent scientific literature shows that GWAS correlations and PRSs are often confounded, not just as a consequence of linkage disequilibrium (which is commonly recognized) but in other ways as well. Finally, the scientific literature on PRSs is developing rapidly. Our goal is not to make original scientific contributions to this literature but rather to provide philosophical readers with an introduction to PRSs and to discuss some interpretive and methodological problems they raise.

The remainder of this paper is organized as follows: Section 2 provides some historical and scientific background to the use of PRSs. Section 3 discusses conflicting claims about the causal status of SNPs and PRSs. Section 4 takes up the issue of what it is for SNPs to have a causal interpretation. Sections 5–6 and 8 argue, in agreement with others (for example, Bourrat 2020), that even when causal, the relationships in which SNPs are involved will typically be unstable and nonspecific, notions we characterize below. Previously, stability and specificity have been applied only to causal relationships. Section 6 extends these notions to noncausal relationships like PRSs, arguing that even qua predictive, PRSs have limited stability and tend to be nonspecific. Section 7 describes the use of PRSs in genetic dissection. Section 8 provides some additional illustrations of how PRSs, although lacking a straightforwardly causal interpretation, can nonetheless be used fruitfully in causal inferences. Section 9 returns to issues about confounding. Section 10 explores in more detail the analogies that have been proposed between meiosis and randomized experiments.

## 2. Background

PRSs are constructed by regressing a trait measurement onto different variants at relevant SNPs (typically with controls for various demographic variables, the selection of a trimmed set of SNPs to reduce their correlation via linkage disequilibrium, and sometimes examining

only SNPs that are significant at a particular *p* value threshold) to obtain a regression coefficient reflecting the magnitude of the association of these variables. In the simplest case, these coefficients are then summed to obtain the PRS for the trait. The score thus has the form:

$$S = \Sigma Bj$$

where the *Bj* are the regression weights associated with various SNPs. The scores are first estimated on a sample drawn from a population and then "tested" by applying them to another sample, ideally from a similar ethnic group. Scores that successfully predict on the new sample are generally found to replicate in further applications on similar samples. Successful prediction is understood in terms of percent of variance explained on new samples, which is a population-level measure.

The increasing use of PRSs reflects recent empirical discoveries about the human genome and the genetic contributions to many complex traits. Since 1900, two major paradigms have dominated human genetics, one based on correlations and predictions—initially called biometrical genetics and now genetic epidemiology—and the other (now molecular genetics) based on Mendelian models that sought to identify specific genes and, through them, to understand biological mechanisms. These two fields were shown to be theoretically consistent by Ronald A. Fisher (1919). If you postulated a very large number of genes of very small effect that interact additively, Mendelian theory predicts exactly the sort of distributions of liability that the biometricians assumed.

Before the discovery of DNA technologies, Mendelians had limited tools but the scientific landscape has recently greatly expanded. The early generation of tools, especially linkage analysis, worked well for finding the genes underlying Mendelian disorders—cystic fibrosis and Huntington's disease. But these methods failed to find the genetic basis for common heritable disorders like hypertension and schizophrenia. Enter the GWAS. The first major report on PRSs came from a 2009 effort to identify individual risk variants for schizophrenia in the International Schizophrenia Consortium (ISC). The authors (Purcell et al. 2009) were disappointed in detecting only a single significant SNP. There had been a variety of approaches to trying to "count up" PRS measures "in the air" for a year or two previously. But the lead author of this paper, Shaun M. Purcell, developed the first detailed approach. After solving some technical problems (for example, intercorrelations of SNPs due to linkage disequilibrium), he generated a PRS for schizophrenia in his original ISC sample and tested it in an independent sample from the Molecular Genetics of Schizophrenia (MGS) consortium. It discriminated the cases and controls in the MGS with a *p* value of $2 \times 10^{-28}$, dwarfing the significance of their one genome-wide significant SNP, which was $4.19 \times 10^{-8}$! However, we should note that this very impressive *p* value for PRS arises from the prediction of only 3% of the variance in liability to schizophrenia in the MGS sample. Because PRS often works with very large samples, it is possible to attain very low *p* values with modest predictions in variance.

Long before the development of modern, high-throughput genotyping, twin and adoption designs established that many traits and disorders were substantially heritable. With the advent of modern molecular techniques, it seemed reasonable to look for "the genes" underlying this heritability. The initial expectation was that most of these genes would be in coding regions of the genome, that some of these would have relatively large

effects, and that much of the genetic variance associated with various traits would be captured by a small number of genes. This led to employing linkage analysis in the 1980s and 1990s. This was not successful. It became clear that nearly all complex traits and diseases resulted from a very large number of variants that typically had very small effects. Because of the smallness of the effects involved, very large samples are needed to detect them. Unlike linkage studies, which replicated poorly, PRSs do successfully replicate (again in the sense of percent of variance explained) when studies with adequately large samples are conducted on the same population—a central part of their appeal.

The fact that so many genetic variants of very small effects contribute to many complex traits makes the task of achieving a full "mechanistic "understanding of the genetic causation of such traits difficult. The use of PRSs can be seen as a reaction (or adaptation) to these facts. It is a strategy for dealing with what would otherwise be overwhelming complexity. Rather than attempting to directly identify individual genes or other variants that are causally responsible for various traits and to elucidate their mechanisms of action, polygenic scores bypass this in favor of a single aggregate measure that is predictively useful.

Given that SNP/trait associations in a PRS are combined according to a simple additive formula that does not include information about SNP or gene interactions, nonlinearities, or patterns of gene expression, it may seem surprising that the scores are so predictively successful. However, a long tradition of twin studies, which typically assume additivity of gene effects, has shown substantial levels of heritability for many traits. Thus, the predictive power of PRSs based on additivity assumptions (which we can think of as incorporating a good deal of the same information produced by twin studies) is not surprising.[5]

Given the weak signal of individual SNPs, distinguishing true from false positive associations is difficult with small sample sizes. Thus, as expected, the predictive power of the PRS has increased rather dramatically with larger sample sizes. This is well illustrated with schizophrenia, where initial studies with cases and controls of ~3,000 predicted around 3% of disease variance while more recent studies with sample sizes of nearly 80,000 cases and 250,000 controls predict over 7% of the variance (Ripke, Walters, and O'Donovan 2020). Within social science, an even more impressive result is seen in the very recent PRS for EA from a sample of over 3 million individuals, which predicts 13% of the variance in EA in other ethnically matched samples (Okbay et al. 2022).[6]

## 3. Are SNP/Trait Correlations and PRSs Causal?

It is widely accepted among users of PRSs that the individual SNPs that go into the scores are rarely in the coding region of genes and in most cases are not causes of the traits they predict. Instead, even when other sorts of confounding are absent (see below), the SNPs are correlated, in virtue of linkage disequilibrium, with variants that *are* causal, and it is this that accounts for their predictive power. For example, two recent large-scale GWAS studies for depression and schizophrenia found that under 10% could be identified as likely to be causal (Levey et al. 2020; Ripke, Walters, and O'Donovan 2020). A recent survey article

[5] The question of how to reconcile the success of the assumption of additivity in heritability estimates or PRSs with the apparent existence of considerable interaction at the molecular level is an interesting one. See Hill, Goddard, and Visscher (2008) for discussion.

[6] However, as noted in section 9, the percent of variance explained declines considerably when scores are estimated with a within-sibs design.

claims flatly that SNPs "typically do not cause the [associated] trait" (Schaid, Chen, and Larson 2018, 491).

However, some philosophers and scientists adopt more straightforwardly causal interpretations of the scores. Pierrick Bourrat (2020) writes: "If a given SNP is more strongly associated with the disease state, one can conclude—with some confidence—that the difference in phenotype is in part due to the nucleotide variant," adding in a footnote that although it is possible that the association in question may be noncausal, this problem may be "mitigated" by increasing sample size, using different populations, and "sophisticated statistical methods" (2020: 1077). He writes that "SNPs can be conceptualized as emulations of ideal interventions" (1080) and that "despite being purely observational studies as opposed to experiments in which interventions are performed, GWAS emulate the minimal condition of causation." (As these quotations suggest, Bourrat takes this minimal condition for interpreting an SNP outcome/association as causal to be the "interventionist" condition described below in section 4.) As noted above, a related but more restricted claim is made by Madole and Harden (2022), who contend that "because meiosis constitutes a naturally occurring manipulation, comparison of within-family genotypes can be conceptualized within the same framework of causal inference as randomized controlled trials (RCTs)" and hence warrants a causal interpretation. We argue below that both Bourrat's and Madole and Harden's claims are unconvincing.

## 4. What Is It for SNPs to Have a Causal Interpretation?

This section takes up the following issues: First, when is it correct to regard SNPs as causes of phenotypic traits with which they are associated? Relatedly, can we infer such causal relationships just from the correlational information in a GWAS? Second, given an appropriate understanding of causation, to what extent, as an empirical matter, are SNPs identified in a typical GWAS causal?

In addressing the first question, we find it useful to think in terms of hypothetical intervention experiments. (Here, we follow a widespread consensus in the recent literature on causal inference, as well as recent philosophical discussion of genetic causation—for example, Bourrat 2020.) Interestingly, the use of such hypothetical experiments (involving individuals at the moment of conception, as described below) to clarify what is meant by genetic causation goes back at least to Fisher (1930: cf. Lee and Chow 2013). Fisher also focuses on a notion of average causal effect similar to the one we describe below. Our aim is to use such experiments to give readers some intuition for what it might mean for a nucleotide at some locus to "cause" some trait, rather than being just correlated with it.

Consider first a single individual, *N,* at the moment of conception and a trait, *P,* and suppose it possible to carry out an intervention—an exogenous, unconfounded experimental manipulation—that replaces a nucleotide (for example, *C*) at one SNP location with an alternative nucleotide (for example, *G*) while at the same time leaving every other feature of *N*'s genome that might affect *P* (except for any features, genetic or otherwise, that are themselves caused by this replacement) unchanged, as well as leaving *N*'s environment unchanged. If, under such an intervention on an SNP, there is an associated change in trait *P,* then this change in nucleotide satisfies what we regard as a *minimal* condition for being a cause of this change in *P*. We regard satisfaction of this condition as sufficient for causation. Hypothetical experiments of this general sort are also employed by Bourrat

(2020) and Lynch and Bourrat (2017), among others, as criteria for causation, although their characterizations differ from ours in various details.

The condition just described has very limited applicability for the purposes of interpreting the results of a GWAS. For one thing, a GWAS concerns a population of individuals and an intervention of the sort described above will have different impacts on the trait of interest for different individuals because of differences in their genetic background, so that it becomes unclear what is meant by "the" effect of such an intervention, unless a particular individual is specified. A natural way to extend the interventionist framework to address this complication is to imagine randomly assigning members of a population to treatment and control groups, and intervening to impose a particular nucleotide—for example, *C*—at an SNP locus in the treatment group, and then comparing the result with a control group in which there is no such intervention (or alternatively one in which a different nucleotide is imposed for every member of that group). If there is a difference in the *distribution* of trait *P* between the treatment and control groups (for example, a difference in the probability or expected value of *P*) , we can think of *C* as having a causal influence on *P* in this population. As suggested by Madole and Harden (2022), we may think of the difference in the expected value of *P* between the treatment and control groups as giving us the average causal effect (ACE) of *C* on *P* within this population, in analogy with the ACE detected in an ordinary randomized experiment. Note that this also fits naturally with the usual assessment of SNPs and PRSs in terms of percent of variance explained, which is also a population-level notion.

Assuming that we adopt this ACE notion as the appropriate condition for the causal interpretation of a GWAS, there is another complication: the condition works for only one notion of cause—a *net* or *total* cause in the sense of Woodward (2003). *C* is a net cause of *E* if it has a net overall effect on *E,* taking account of all the different causal paths by which it affects *E*. This contrasts with the notion of a cause *C* being a *contributing* cause of *E* along a causal path. Roughly speaking, *C* is a contributing cause of *E* along a path if interventions on *C* make a difference to *E* (or the probability of *E*) when what happens on other paths is held fixed (Woodward 2003).[7] It is this contributing cause notion that is captured in path diagrams and directed acyclic graphs that are sometimes used to represent genotype/ phenotype relations. We also need this contributing cause notion to capture cases in which a genetic factor affects a phenotype via two different paths, one more direct and operating entirely "within the skin" and the other operating indirectly through some environmental variable, as in the cases of genotype-environment (G-E) covariance described below. The total cause notion combines the effects of both paths on the phenotype, while the contributing cause notion separates them. By themselves, SNP/trait correlations, even if causal, reflect total or net population-level effects, rather than decompositions along distinct causal paths, and the latter is often of considerable interest. A similar point holds for PRSs.

Turning now to PRSs, recall that in constructing such a score, "weights" for individual SNPs are calculated by regressing a measure of the phenotypic outcome on a variable associated with the individual SNP, based on the magnitude of its effect size. Here, if the result describes a causal relationship, the regression coefficient should correctly describe

[7] For example, if *G* affects *P* via two different paths, one direct and the other via a mediating environmental variable *M,* intervening to fix *M* will give the direct effect alone.

the average or expected change in phenotypic outcome for the population under population-wide interventions that change the variant present at the SNP in analogy with a causal interpretation of regression coefficients in other contexts.

Even an intervention involving an SNP in a single individual of the sort described above will be technically very difficult, although CRSPR and other technologies afford some possibilities. Randomized experiments involving interventions on all the SNPs going into a PRS seem way beyond anything presently possible, as well as ethically problematic. Moreover, the upshot of such an experiment would depend on just which such interventions are carried out. That is, different interventions that change the score by the same amount but involve changing different underlying SNPs would likely have different effects—with the result that "the effect" of intervening to change a PRS by some specified amount is ill defined. This by itself seems to undercut the claim that PRSs are straightforwardly causal (or that they describe causes of phenotypic effects), at least within an interventionist framework.

Suppose that a researcher has access only to nonexperimental ("correlational") data of the sort that comes from a GWAS. Under the framework we are proposing, claims that individual SNPs are "causal" for some outcome are to be interpreted as claims about the outcomes of the hypothetical population-level experiments described above. Causal inference in this context thus requires that we somehow use this correlational information, perhaps in connection with other assumptions, to reliably infer to conclusions about the results of such hypothetical experiments. Reliable causal inference from nonexperimental correlational data is certainly sometimes possible (witness techniques like instrumental variables and, arguably, various machine learning algorithms) but as a general matter (not specific to GWAS) this requires that strong additional conditions obtain. These conditions may be design-based or grounded in other empirical considerations but in any case these will go beyond purely correlational information.[8] Both theoretical considerations and experience show that attempts to correct for confounding through purely statistical means (conditioning on possible confounders, and so on) rarely yield reliable conclusions. Thus, we should not think that causal conclusions emerge just from the application of statistical techniques from a GWAS.

There are a number of possible ways in which noncausal SNP/outcome correlations can arise, where by "noncausal" we mean a relationship that is not even minimally causal in the sense described above. One of the most obvious is through linkage disequilibrium—if *G* is a genuinely causal factor for *P* and *G* is in linkage disequilibrium with a variant *S1* at some SNP (for example, because *G* and *S1* are nearby on the same chromosome), the latter will be correlated with *P,* even though it may have no causal effect on *P*. The noncausal status of *S1* would be revealed in the ideal experiments above—if we were to intervene to change the nucleotide at *S1,* either in an individual or a population, while not affecting *G,* there would be no change in *P*.

A second commonly accepted way in which spurious correlations between SNPs and phenotypic traits can arise is through population stratification or through environmental factors correlated with a subject's genome but not caused by it. This might happen for

[8] By "designed-based", we mean that the design of the study itself rules out certain kinds of confounding—as with the within-sib designs described below. The additional assumptions required for warranted causal inference take various forms—for example, causal inference based on independence relations, as in Spirtes et al. (2001), requires the Causal Markov condition and conditions like Faithfulness.

adventitious cultural or historical reasons, as (in a standard illustration) when genetic variant *V* is associated with Chinese ancestry, Chinese people tend to have dexterity with chopsticks for cultural reasons and *V* is interpreted as a cause of that dexterity. More subtle examples involve parental nurture: an individual's genes are correlated with her parents' genes, and these parental genes may contribute to the nurture the parents provide, influencing *N*'s phenotype, particularly for traits like EA but not limited to these. In such cases the apparent influence of *N*'s genes on the trait (as measured from SNPs and PRSs estimated on unrelated individuals) will be inflated since some of this influence will be due to parental nurture.

Another possibility involves "dynastic" factors: if parental genes are associated with greater economic resources or social status for historical reasons (for example, white people tend to be wealthier than black people) and such resources affect offspring traits (such as EA), this will lead to an association between offspring genes and those traits, even if those genes do not causally contribute to the traits. This is one of many reasons why causal interpretations of SNP/outcome correlations are highly problematic, at least in the absence of additional evidence and why one cannot assume that such correlations "emulate" (Bourrat 2020, 1078) the results of intervention experiments.[9] As we note below, differences in PRSs between siblings is one way of mitigating the impact of such noncausal correlations.

## 5. Causal Interpretation and Gene Environment Correlation

The conditions described in the hypothetical experiments above (in either their individual or population versions) are, as we have said, "minimal" conditions for causation. This appellation is meant to capture the idea, described in Woodward (2010), that there are additional dimensions, besides the minimal conditions just described, relevant to the assessment of causal relationships.[10] Causal relations can vary along these additional dimensions, with those that more fully satisfy criteria associated with these dimensions typically being viewed as more paradigmatically causal, "deeper" or more explanatory. One such dimension is *stability*: suppose that the relationship between *C* and *E* is causal in the sense described above. The stability of the *C*→*E* relationship has to do with whether that

[9] It is certainly possible to sometimes identify "emulations" of interventions in data that are not the result of an experiment. But, in such cases, what is required is the presence of a process that produces the data in a way that is "intervention-like" in the sense of satisfying the conditions for an intervention, even though there is no deliberate experimental manipulation. Identifying an "instrument" (in the sense employed in an instrumental variable analysis) is one of the most obvious cases in which such an intervention-emulating process can be exploited for purposes of causal inference. (The use of Mendelian randomization to make inferences about the causal influence of environmental exposures illustrates such a strategy in a genetics context.) But it is widely agreed in the causal modeling literature that correlational data by itself does not give us information about the presence of such an intervention-emulating process. In general, successful causal inference requires causal assumptions as input, not just "statistical" assumptions. This is reflected in Cartwright's 1979 slogan, "no causes in, no causes out."

[10] This general strategy of distinguishing between minimal conditions for causation and further conditions that a causal relation may or may not satisfy is also adopted by Bourrat (2020), Madole and Harden (2022) and many others. Bourrat argues that even if SNPs are minimally causal, they typically will not satisfy additional conditions having to do with stability and a further condition called proportionality. Madole and Harden reach a similar conclusion, arguing that even if minimally causal, SNP/trait relations typically lack other features of paradigmatic causal relations—they are "shallow," "non-uniform," and fail to provide mechanistic information. We mostly agree with these claims but disagree with Bourrat and Madole and Harden about the extent to which SNP/trait relations are even minimally causal.

causal relationship would continue to hold under changes in other variables or background conditions, *B*—relationships that hold under a wider range of conditions are more stable and, other things being equal, viewed as more paradigmatically causal.[11] For obvious reasons, we should not include in the circumstances *B* anything that is causally between *C* and *E*—as when *C*→ *B*→ *E*. For example, if *G* is minimally causal for outcome *P* in *N*'s genome, the circumstances *B* should not include some protein *X* to which *G* causally contributes and which, in turn, causally contributes to *P* but it should include factors in *N*'s genome and environment that are not on any causal pathway from *G* to *P*. At one extreme, there are SNP→ trait causal relations that will continue to hold regardless of what might be the case elsewhere in a subject's genome or environment. An example not involving an SNP but a variable number of repeats in a gene is provided by the causation of Huntington's chorea. At the other extreme are SNP→ trait relations that, although causal, would not continue to hold if any one of a large number of possible changes were to occur elsewhere in the subject's genome or environment and thus, to this extent, are relatively unstable. Stability thus captures the extent to which a causal relationship holding in one environment or set of background circumstances generalizes or is "portable" to other environments or circumstances. It also captures the extent to which a causal relationship is relatively immutable or inevitable (or difficult or impossible to change) via changes in background circumstances—"difficult" corresponding to cases in which changes in background circumstances that are hard to achieve are required. Given a sufficiently large number of CAG repeats in the Huntington gene, eventual development of Huntington's disease is, unfortunately, inevitable in the sense that whether it occurs is not affected by other genetic or environmental factors. On the other hand, it is entirely possible—in fact, arguably very common—for a gene/trait relation to be causal in the minimal sense described above and yet to be highly unstable in the sense that will fail to hold in other background circumstances. This is so for the cases of reactive G-E correlation discussed below. There is some tendency to assume that whenever genes cause an outcome, the gene→ outcome relation must be relatively stable or immutable—hence that when genes cause outcomes, they make those outcomes largely unavoidable, as in so-called genetic determinism.[12] Separating minimal causation from stability allows us to see that this inference is mistaken.

As one illustration of the notion of stability, recall that the notion of an average effect is population-relative in the sense that it is defined with respect to a particular population. This means that *G* might have a non-null average effect on *P* in some population *X* but not in some other population *X**.[13] To the extent that this is so, the average effect of *G* on *P* will not be stable. Because SNP/trait correlations, even if causal, incorporate information about average effects, they and the PRSs from which they are constructed may be subject to the sort of instability that results from such population-relativity, although the extent to which such instability is present is an empirical matter.

This population-relative source of instability in GWAS and in heritability estimates is widely recognized and discussed by a number of authors (see, for example, Downes and Matthews 2019; Bourrat 2021). A somewhat more interesting and controversial illustration

[11] Note that issues about the stability of a causal relation should be distinguished from whether it is confounded or spurious—the latter implying that it is not even minimally causal.
[12] In other words, genetic determinism is best construed as a claim about the stability of gene→ trait relations, rather than as a claim that genes causally influence traits.
[13] Bourrat (2021) describes this as the locality problem.

of the role of stability considerations, also relevant to PRSs and discussions of genetic causation, is provided by the contrast between "reactive" and "active" or "evocative" G-E correlations. Suppose, following the well-known thought experiment in *Inequality: A Reassessment of the Effect of Family and Schooling in America* (Jencks et al. 1972) that in a certain society there is discrimination against children with red hair, *R,* and only these—redheads receive inadequate nutrition and are denied access to good schools. As a result, there is an association between *R* and low EA. If we were to discover SNPs that were causally implicated in *R,* these would also be associated with low EA. Moreover, this association would be causal in the minimal sense described above: if one were to intervene to replace the nucleotides causing *R* in the affected children with those that cause some other hair color, holding everything else constant, including the presence of discriminatory practices targeting only redheads, the EA of those children would improve.

Some writers claim that the *R*→ EA relationship is not a genuine causal relationship, which implies that the minimal condition for causation described previously is too permissive.[14] Note, however, that the *R*→ EA relationship has a different status from spurious causal relationships involving linkage disequilibrium and population stratification. In these cases, interventions on the genetic variant, holding everything else constant, will not change the outcome. In the redhead scenario, it will. Linkage disequilibrium and population stratification are straightforward sources of *confounding,* producing correlational relationships that look as though they are causal, although they are not. The *R*→ EA relationship is not confounded or spurious in this sense.[15] In this respect, describing this relation as G-E "correlation" is potentially misleading since the relationship is not "merely correlational" in the sense of reflecting a noncausal relation. We thus need some way to distinguish the *R*→ EA relationship (which we agree is a non-paradigmatic and "shallow" causal relationship) from those relationships that are merely correlational. Characterizing the *R*→ EA relationship as causal but only "minimally" so gives us a way of doing this. As argued below, we think that much of the resistance readers may feel to treating the *R*→ EA relationship as causal derives from the fact that it is relatively unstable.

[14] See, for example, Block and Dworkin (1976). Their assessment is based on intuition-like judgments that the *R*→ EA relation is not what "we" ordinarily think of as a causal relation. We think that in scientific contexts it is important to employ a principled notion of causation that is not based on intuitive judgments. The interventionist framework provides this and judges that the *R*→ EA relation is causal. Burt (2022) claims that Jencks et al.'s example should be understood in terms of "downward causation," the idea being that the EA of red-haired children is not caused by their genes but rather by a social practice of discrimination. However, within an interventionist framework, the children's genes, *G,* cause their red hair, *R,* which in turn causes a discriminatory social response, *D,* which causally influences EA (*G*→ *R*→ *D*→ EA—it doesn't matter if we add another variable as another cause of *D*). The *D*→ EA relation might be viewed as downward causation but this does not undercut the claim that *G* causes EA. (Causation is not always transitive but it is in this case.) Lynch and Bourrat (2017) claim that reactive and active G-E have the "same causal structure" and express puzzlement about why they are treated differently. We agree with Lynch and Bourrat that in both cases the gene/trait relation satisfies the minimal interventionist condition for causation. However, our view is that there is a further causally relevant distinction between the two related to stability, so that in this sense the two cases do not share "the same causal structure." Lynch and Bourrat also claim that the differential reactions of some to the active and reactive cases are influenced by judgments about responsibility and agency. We agree but will add that there are both theoretical (Woodward 2006) and empirical reasons (Grinfeld et al. 2020) to think that judgments of responsibility are influenced by stability judgments, so that acknowledging a role for responsibility is not in tension with the idea that stability considerations also apply.

[15] A caveat regarding a potential source of confusion: Suppose that, as in the example described below, genetic factor *G* causes *P,* via both a direct inside the skin path and an indirect outside the skin path. If we attempt to estimate the direct of *G* on *P* via the overall correlation between *G* and *P,* we will "confound" the direct effect with the indirect effect. But this does not show that the indirect causal relation is spurious or not genuine.

The *R*→ EA relationship is described in the literature as a case of reactive or evocative gene/environment correlation, because the subject's genotype "evokes" an environmental response (societal prejudice), which in turn causally contributes to the outcome but where the subject does nothing to create this prejudicial environment. By contrast, in cases involving *active gene/environment correlation* a subject's genotype causes them to select or actively create certain environments, which then causally contribute to outcomes. As an illustration, suppose there are genes that "directly" cause higher EA via inside the skin processes (for example, those affecting brain development) and that also cause higher EA indirectly via an outside the skin route that prompts their possessors to seek out intellectually stimulating environments, which also causes higher EA.

We suggested above that causal relations like *R*→ EA are relatively unstable. This relation will hold only in environments in which there is discrimination against red-haired children. By contrast, some kinds of active/gene environment correlations may be much more stable. In addition to the example involving intellectual curiosity above, consider degree of social support (Furukawa and Shibayama 1997), often thought to depend on factors entirely extrinsic to the individuals involved but which shows a surprisingly high level of stability when those individuals move to new environments, a result consistent with the idea that individuals have somewhat stable tendencies to construct their support networks—a tendency that might have a genetic component, as suggested by twin studies (Kendler 1997).

We suggest that it is such differences in stability that (at least in large part) underlie our inclination to think that reactive G-E correlations are more problematic cases of "genetic causation" than many cases of active G-E correlation. To the extent that, in the latter cases, the outside the skin pathway is operative in many different environments and is present because of general behavioral tendencies that have a genetic component, we are more comfortable labeling such relations as causal. Note, however, that in cases of both reactive and active G-E correlations involving an SNP associated with some trait *P,* a PRS will incorporate such correlations and when there is both an inside the skin and an outside the skin pathway between the SNP and *P,* these paths will be combined into a single "net" measure of association that goes into the PRS. Thus, even if an SNP in a PRS is causal for a trait, the relationship between the SNP and the trait will not (in the absence of other information) distinguish SNP/trait relations that arise from mechanisms that operate entirely inside the skin from those that involve an outside the skin route and will not distinguish cases of reactive and active G-E. Moreover, even if an SNP/trait relation is causal, this by itself tells us little or nothing about how stable that relation is or how stable a PRS incorporating that relation will be in other circumstances. For this reason, even if an SNP/outcome relation is causal (or even if a PRS incorporates causal information) it seems misleading to think of these as telling us anything about "genetic potential" or "propensity," as is sometimes claimed (for example, Plomin and Von Stumm 2022). To the extent that "potential" and "propensity" mean anything clear in this context, they presumably are claims about stability—"potential" suggests genes limit what traits can develop in non-actual circumstances.

Both active and reactive gene environment correlation should be distinguished from "passive gene/environment correlations" in which, for example, features of the parental environment causally influenced by parental genotypes are correlated with their children's genotype but not because they are causally influenced by their children's genotype. For

example, certain parental genotypes may causally contribute to their possessors providing particularly nurturing environments for their children, with the nurturing environment in turn causally contributing to their children's development, but where the child's genotype does not play a causal role in generating the environment provided by the parent. This is illustrated in the study involving non-transmitted heredity for EA by Augustine Kong et al. (2018) described below. Unlike the active/reactive examples discussed above, in this last case, if we were to intervene to change the child's genotype, this would not change the parental genotype's contribution to the environment they provide. Thus, unlike the cases of active and reactive gene/environment correlation, in this last case the relation between the child's genotype (including SNPs) and the parental environment is purely correlational, rather than causal.

## 6. Stability Applied to *Predictive* Relationships

So far, we have focused on stability as a condition that *causal claims* meet to a greater or lesser degree. We can ask parallel questions about *predictive relationships* such as PRSs that do not have a straightforward causal interpretation. That is, we can ask about the extent to which such predictive relationships are stable across different sorts of changes—for example, does the weighted sum of SNP associations that allows for the prediction of height in one population or set of circumstances allow for equally good predictions in other populations or circumstances? Here, since we are focusing just on the stability of PRSs qua predictors, both linkage disequilibrium and population stratification serve as sources of instability, in addition to any instability present in the relations between SNPs and outcomes that are genuinely causal.

In many cases, PRSs qua predictors are not highly stable across different populations. For example, scores for many traits that predict relatively well in European populations are much less predictive in other groups, especially in African populations. This is understood to arise from the differences in the patterns of linkage disequilibrium across human subpopulations because most of the predictive power of the PRS results from linkage disequilibrium relationships between detected and causal SNPs. Indeed, it is a source of concern to those who wish to use PRSs that they are largely calculated just for European populations and comparable results are often not available for other populations. On the other hand, there are exceptions—the PRSs for schizophrenia in East Asian and European populations appear to be highly correlated (Lam et al. 2019).

When a PRS is unstable across different populations or background circumstances, this may suggest (even if it does not establish) that the relationship incorporates a number of noncausal associations or relations that, even if minimally causal, are unstable. Suppose that nucleotide *N1* is associated with trait *P* but noncausally in some population, the association arising because *N1* is correlated in linkage disequilibrium with nucleotide *N2,* which is causal for *P*. Then in a different population in which, because of different patterns of linkage disequilibrium, *N1* and *N2* are not correlated, a PRS that incorporates *N1* will predict *P* less well, assuming that *P* is also caused by *N2* in that population. Thus, to the extent that PRS relations for various traits do not transfer well across different populations this is another piece of evidence that they may incorporate noncausal associations or unstable causal relations. High (predictive) stability in PRS relations, as in the case of PRSs for schizophrenia, on the other hand, may suggest (although not establish) that the PRS is

made up of a substantial proportion of SNPs that are genuinely causal, assuming that stable relations tend to be causal, even if the converse is not always true.

## 7. Specificity and Genetic Dissection

A second dimension along which relations that are minimally causal can vary is the extent to which they are *specific*. For our purposes, a cause is specific to the extent it has just one kind of effect (in some restricted set of possible effects of interest such as diseases), rather than many different kinds of effects. As an illustration, the variants of Huntington gene that cause Huntington's disease are highly specific in the sense that they only cause this disease and not others. On the other hand, the genes implicated in many mental illnesses seem to be relatively nonspecific in the sense that they seem to be involved in many related illnesses. Just as we extended the notion of stability to include the stability of predictive (and not only causal) relations, we can do the same with specificity, asking, for example, whether individual SNPs specifically predict one trait/disorder or many. A recent paper (Cross-Disorder Group of the Psychiatric Genomics Consortium 2019) shows that a number of the same individual SNPs that predispose to least one of eight psychiatric disorders—schizophrenia, bipolar illness, major depression, autism spectrum, ADHD, Tourette's syndrome, anorexia, and OCD—typically predispose to the others. In fact, some SNPs were identified that predisposed to all of the disorders. Similarly, one might expect that there will be considerable nonspecificity in PRSs for these disorders. On the other hand, the SNPs in question (and PRSs built from them) are very likely to be non-predictive for outcomes like height and heart disease—hence specific for mental illnesses as opposed to other sorts of outcomes.

As this suggests, PRSs can be employed in the *genetic dissection* of diseases—that is, they can be used to explore the extent to which apparently different diseases share a common genetic basis, as revealed by overlapping PRSs and the extent to which they can be discriminated genetically. Several recent papers (for example, Ruderfer et al. 2014; Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium 2018) have explored the (predictive) specificity of various SNPs and PRSs for schizophrenia (SCZ) and bipolar disorder (BD). PRSs for these diseases share a significant overlap, which is consistent with other results, suggesting a common genetic basis. However, it is also possible to use PRSs to distinguish between clinical symptoms of the two diseases—for example, BD patients with psychotic symptoms have higher SCZ PRSs than BD patients without such symptoms. Again, note that this does not require that all or many of the SNPs incorporated in these PRS be causal—it is enough if they reliably track or are correlated with underlying genetic differences that *are* causal. At the same time, the ability of PRSs to distinguish between the two diseases suggests that there are underlying genetic differences that might be fruitful to explore.

## 8. The Use of PRSs in Causal Analysis

So far, the picture we have presented of PRSs is that these incorporate a mixture of a few genuinely causal relationships (in the minimal interventionist sense) and many relationships involving noncausal associations. Moreover, the scores do not tell us anything about the mechanisms or intervening variables by which SNPs or genes causally affect traits.

Finally, at least in a number of cases, the scores are not highly stable even in a predictive sense when applied to populations that are different from those on which they were originally estimated. Many PRSs are also relatively nonspecific within classes of disorders (psychiatric, autoimmune, and so on).

These considerations may seem to suggest that PRSs are of rather limited usefulness when it comes to causal analysis. However, we think this assessment is, at least in part, incomplete. In particular, as we shall now argue, the associational information in a PRS, although not straightforwardly causal, can sometimes be used to control for genetic influences of various sorts to reach causal conclusions about the influence of other variables, including those that are "environmental."

As an illustration, Kong et al. (2018) studied the effect of the environment provided by parents on children's EA. In the broad sense of genetic influence under discussion, which incorporates both active and reactive gene/environment correlation, children's genotype clearly influences their EA. However, there are also important influences due to the environment provided by parents, which is influenced by the parents' own genes, some of which are transmitted to their offspring and some of which are not transmitted. Thus, the cumulative effect on children's EA reflects (i) genes transmitted from their parents, including (ii) transmitted genes, the presence of which in the parents affects the environment parents provide, and (iii) genes of the parents that affect the environment they provide but are not transmitted to their child (non-transmitted genes). A standard study that looks only at the relation between parental genotypes transmitted to the child and child's educational attainment will not disentangle these factors. Kong et al. proposed a novel method that could address this question. They obtained a PRS related to EA for children as well as a parental PRS related to EA involving alleles *not* transmitted to their children. They assumed that the direct effect, *d,* of the transmitted alleles on child's EA (that is, the effect not mediated by parental nurture) can be estimated via $d = \Theta(T) - \Theta(NT)$, where $\Theta(T)$ and $\Theta(NT)$ are estimates of the effects of the transmitted and non-transmitted alleles, respectively. As the authors remark, calculating this difference allows one to cancel out or control for genetic nurturing effects, as well as some other potential confounds. Using this methodology, the authors find that "the average estimated effect of the non-transmitted alleles [on EA] is 34.2% of that of the transmitted alleles" (Kong et al. 2018, 425). Thus, there is a substantial genetically based nurture effect on children's EA that is based on non-transmitted alleles as well as on those alleles that are transmitted. By way of contrast, parental alleles associated with height and BMI predict these characteristics in children only insofar as the particular genetic risk factors were transmitted directly to the children (cf. Koellinger and Harden 2018)—a result in accord with commonsense causal expectations. We emphasize that this result about the effects of non-transmitted alleles on children's EA cannot be obtained through traditional heritability studies—it requires the use of PRSs.

This example provides an illustration of the use of PRSs to disentangle, at least in part, different paths by which aggregate effects come about and thus the way in which "predictive" but noncausal information can be used in causal analysis. The example shows how genetic analysis based on PRSs can sometimes be used to identify the role of "environmental" factors in causally affecting outcomes. There are many other examples. PRSs for EA have been found to be better predictors of EA among white men than white women among cohorts from the early twentieth century, but in later cohorts the predictability of white women's scores improves and eventually exceeds the predictability

of the scores for white men (Raffington, Mallard, and Harden 2020). One plausible causal interpretation is that this reflects an improvement in the available educational opportunities for women—in earlier cohorts, environmental factors like gender discrimination function so as to reduce EA for white women, while in later cohorts these play less of a role. PRSs help to highlight the causal role of this environmental change.

## 9. Confounding Again

Having highlighted the possible role of PRSs as control variables, we return in this section to some issues about confounding and its significance. Several recent papers have shown that GWAS associations for a number of phenotypes estimated from within-sib designs (that is, designs that compare PRSs between siblings) are substantially reduced in comparison with estimates for unrelated individuals.[16] For example, James J. Lee et al. (2018) report within-sibling estimates for EA that are 40% lower than those for unrelated individuals. Laurence J. Howe et al. (2022) report smaller within-family estimates for height, educational attainment, age at first birth, number of children, cognitive ability, depressive symptoms, and smoking, although not for many "molecular phenotypes" such as low-density lipoprotein-cholesterol. It is plausibly assumed that the within-sibling estimates control to a significant extent for such factors as assortative mating, population stratification, and environmental factors such as parental nurture shared among close family members (such as siblings). Thus, comparisons with estimates for unrelated individuals may provide a rough estimate of the extent to which the above factors influence the latter.[17] Interestingly, Lee et al. suggest that only a third of the diminishment in EA they observe may be attributable to assortative mating; they suggest that the remaining diminishment may reflect the role of parental nurture, noting that this is roughly consistent with result obtained by Kong et al. (2018). This example suggests that, in some cases, it may be possible to compare different studies involving PRSs to triangulate on an estimation of the magnitude of certain kinds of confounding—particularly confounding due to environmental factors, including genetically based parental nurture. In this way, comparison of within-sibling and unrelated individuals PRSs for various traits again illustrates how these may provide evidence of the role of broadly environmental factors—for example, it is suggestive that the diminishment of the PRS for depression is roughly the

[16] This fact and the implication that PRSs for unrelated individuals are, in a number of cases, confounded by nongenetic factors or genetic effects coming from parents are emphasized in the critical review of Harden (2021) by Coop and Przeworski (2022).

[17] There is some subtlety (deserving of more philosophical attention) around the issue of what it means for largely noncausal relations like PRSs and GWAS to be "confounded." (One usually thinks of a confounded relationship as one that looks causal but is not, while PRSs do not even look causal, even if unconfounded by nongenetic factors.) In this context, confounding has to do with the extent to which the PRSs track genetic factors that are causal, rather than tracking nongenetic environmental factors that are causal. That is, the confounded/non-confounded contrast has to do with a distinction among noncausal factors between what is tracked (genetic versus environmental factors), rather than a distinction having to do with causal relationships. To the extent that PRS correlations track nongenetic environmental causes, their use as controls to reveal the role of environmental factors will underestimate the role of such factors. "Confounding" in this context refers to this possibility.

same as that for EA, suggesting a possibly substantial role for the environment in this condition, a result consistent with other studies.[18]

## 10. Causal Interpretation Redux

We have argued against the claim that GWAS correlations are typically causal. But what about the more restricted claim that, conditional on parental genotype (for example, considering pairs of full sibs), SNP/trait relations will not be confounded, and hence can be interpreted as causal because of the "random" nature of meiosis? As noted above, it is true that conditioning on parental genotype will eliminate or reduce certain forms of confounding, such as those due to population structure and (perhaps) parental nurture. However, the possibility of confounding due to genetic linkage (which operates over considerably larger genomic distances than linkage disequilibrium) remains. That is, because of linkage, even if it is random whether a sib gets A or a, and random whether the sib gets B or b, it does not follow that there will be no correlation between whether the sib gets A and whether she gets B if they are even moderately nearby on a chromosomal arm. Thus, despite the random nature of meiosis if, say, A/a is causal for trait *P,* and B/b is not, B/b can still be correlated with *P* because linkage is present, and hence this relation can look causal. What conditioning on parental genotype does is to make it more likely that a difference in PRSs between sibs is tracking something genuinely genetic (as opposed to, say, the influence of some environmental factor or parental nurture) but this is different from establishing that the SNPs that go into the score are themselves causal.[19]

Another disanalogy is this: in an ordinary randomized controlled trial, values of a single treatment variable are randomly assigned. By contrast, when SNPS for sibs are compared, one in effect has a comparison of the combined upshot of a very large number of *different* treatments (one for each SNP) for pairs of sibs. Because of linkage, one is comparing large collections of different treatments (chunks of chromosomes), rather than treatments corresponding to individual SNPs. This does not undercut the potential role of GWAS correlations and PRSs in tracking genetic factors that *are* causal but it does suggest that we cannot conclude from the random nature of meiosis that SNP/trait correlations that remain, conditional on parental genotype, are causal.

## 11. Conclusion

We have described some interpretive and methodological issues concerning PRSs. We have argued that the correlations described by PRSs are not causal, even in the minimal sense

[18] These observations also illustrate an important contrast between two varieties of "predictive" relationships. Some relationships may be predictive but those using such relationships may have no idea why they are predictively successful, or what factors the relationship tracks that explain its predictive success. A hominid from 100,000 BCE might recognize that autumn predicts winter but might have no understanding of why this predictive relationship holds. PRSs are at least intended to be different—as noted, they are constructed with the goal of tracking specifically genetic causal relationships, so that genetic relationships explain their predictive success. If they are predictively successful only because they track nongenetic, environmental relationships, they will not lead to reliable estimates of environmental effects when used as controls. This is another respect in which thinking of PRSs as merely predictive can be misleading.

[19] Consider Belsky et al. (2018), who found that sibs with higher PRSs for EA achieved more education than their co-sibs with lower PRSs. It is not implausible that this difference indicates the presence of a genetic differences between such co-siblings even if we can't tell from the PRSs which SNPs are causal for EA.

described above. Although their purely predictive/correlational use is important, it is this tracking feature that contributes to their potential usefulness in other applications, such as genetic dissection, and their use as controls that allow us to, indirectly, more clearly "see" the role of environmental variables.

**Acknowledgments**
We would like to acknowledge James Madole and K. Paige Harden for a very useful discussion, and Eric Turkheimer for helpful comments on an earlier draft.

**Disclosure Statement**
No competing interest was reported by the authors.

**References**
Belsky, Daniel W., Benjamin W. Domingue, Robbee Wedow, Louise Arseneault, Jason D. Boardman, Avshalom Caspi, Dalton Conley, et al. 2018. "Genetic Analysis of Social–Class Mobility in Five Longitudinal Studies." *Proceedings of the National Academy of Sciences* 115, no. 31: e7275–7284. https://doi.org/10.1073/pnas.1801238115.

Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. 2018. "Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes." *Cell* 173, no. 7: 1705–1715. https://doi.org/10.1016/j.cell.2018.05.046.

Block, Ned and Gerald Dworkin. 1976. "IQ, Heritability and Inequality." In *The IQ Controversy,* edited by Ned Block and Gerald Dworkin, 410–540. New York: Random House.

Bourrat, Pierrick. 2020. "Causation and Single Nucleotide Polymorphism Heritability." *Philosophy of Science* 87, no. 5: 1073–1083. https://doi.org/10.1086/710517.

———. 2021. "Heritability, Causal Influence and Locality." *Synthese* 198: 6689–6715. https://doi.org/10.1007/s11229-019-02484-3.

Burt, Callie H. 2022. "Challenging the Utility of Polygenic Scores for Social Science: Environmental Confounding, Downward Causation, and Unknown Biology," *Behavioral and Brain Sciences*: 1–36. https://doi.org/10.1017/s0140525x22001145.

Cartwright, Nancy. 1979. "Causal Laws and Effective Strategies." *Noûs* 13, no. 4: 419–437. https://doi.org/10.2307/2215337.

Coop, Graham and Molly Przeworski. 2022. "Lottery, Luck, or Legacy: A Review of *The Genetic Lottery: Why DNA Matters for Social Equality*." *Evolution* 76, no. 4: 846–853. https://doi.org/10.1111/evo.14449.

Cross-Disorder Group of the Psychiatric Genomics Consortium. 2019. "Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders." *Cell* 179, no. 7: 1469–1482. https://doi.org/10.1016/j.cell.2019.11.020.

Downes, Stephen M. and Lucas Matthews. 2019. "Heritability." In *Stanford Encyclopedia of Philosophy,* edited by Edward N. Zalta. https://plato.stanford.edu/entries/heredity/.

Fisher, R.A. 1919. "The Correlation between Relatives on the Supposition of Mendelian Inheritance." *Earth and Environmental Science, Transactions of the Royal Society of Edinburgh* 52, no. 2: 399–433. https://doi.org/10.1017/S0080456800012163.

———. 1930. *The Genetical Theory of Natural Selection*. Oxford: Oxford University Press.

Furukawa, T. and T. Shibayama. 1997. "Intra-individual versus Extra-individual Components of Social Support." *Psychological Medicine* 27, no. 5: 1183–1191. https://doi.org/10.1017/S0033291797005424.

Grinfeld, Guy, David Lagnado, Tobias Gerstenberg, James F. Woodward, and Marius Usher. 2020. "Causal Responsibility and Robust Causation." *Frontiers in Psychology* 11: 1–19. https://doi.org/10.3389/fpsyg.2020.01069.

Harden, Kathryn Paige. 2021. *The Genetic Lottery: Why DNA Matters for Social Equality*. Princeton: Princeton University Press.

Hill, William G., Michael E .Goddard, and Peter M. Visscher. 2008. "Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits." *PLoS Genetics* 4, no. 2: e1000008. https://doi.org/10.1371/journal.pgen.1000008.

Howe, Laurence J., Michel G. Nivard, Tim T. Morris, Ailin F. Hansen, Humaira Rasheed, Yoonsu Cho, Geetha Chittoor, et al. 2022. "Within-Sibship Genome-Wide Association Analyses Decrease Bias in Estimates of Direct Genetic Effects." *Nature Genetics* 54, no. 5: 581–592. https://doi.org/10.1038/s41588-022-01062-7.

Jencks, Christopher, Marshall Smith, Henry Acland, Mary Jo Bane, David Cohen, Herbert Gintis, Barbara Heyns, and Stephanie Michelson. 1972. *Inequality: A Reassessment of the Effect of Family and Schooling in America*. New York: Basic Books.

Kendler, Kenneth S. 1997. "Social Support: A Genetic-Epidemiologic Analysis." *American Journal of Psychiatry* 154: 1398–1404.

Koellinger, Philipp D. and K. Paige Harden. 2018. "Using Nature to Understand Nurture." *Science* 359, no. 6374: 386–387. http://dx.doi.org/10.1126/science.aar6429.

Kong, Augustine, Gudmar Thorleifsson, Michael L. Frigge, Bjarni J. Vilhjalmsson, Alexander I. Young, Thorgeir E. Thorgeirsson, Stefania Benonisdottir, et al. 2018. "The Nature of Nurture: Effects of Parental Genotypes." *Science* 359, no. 6374: 424–428. https://doi.org/10.1126/science.aan6877.

Lam, Max, Chia-Yen Chen, Zhiqiang Li, Alicia R. Martin, Julien Bryois, Xixian Ma, Helena Gasparet, et al. 2019. "Comparative Genetic Architectures of Schizophrenia in East Asian and European Populations." *Nature Genetics* 51, no. 12: 1670–1678. https://doi.org/10.1038/s41588-019-0512-x.

Lee, James J. and Carson C. Chow. 2013. "The Causal Meaning of Fisher's Average Effect." *Genetics Research* 95: 89–109. https://doi.org/10.1017%2FS0016672313000074.

Lee, James J., Robbee Wedow, Aysu Okbay, Edward Kong, Omeed Maghzian, Meghan Zacher, Tuan Anh Nguyen-Viet, et al. 2018. "Gene Discovery and Polygenic Prediction from a Genome-Wide Association Study of Educational Attainment in 1.1 million Individuals." *Nature Genetics* 50, no. 8: 1112–1121. https://doi.org/10.1038/s41588-018-0147-3.

Levey, Daniel F., Murray B. Stein, Frank R. Wendt, Gita A. Pathak, Hang Zhou, Mihaela Aslan, Rachel Quaden, et al. 2020. "GWAS of Depression Phenotypes in the Million Veteran Program and Meta-analysis in more than 1.2 Million Participants Yields 178 Independent Risk Loci." *MedRxiv*. https://doi.org/10.1101/2020.05.18.20100685.

Lynch, Kate E. and Pierrick Bourrat. 2017. "Interpreting Heritability Causally." *Philosophy of Science* 84, no. 1: 14–34. http://dx.doi.org/10.1086/688933.

Madole, James W. and K. Paige Harden. 2022. "Building Causal Knowledge in Behavior Genetics." *Behavioral and Brain Sciences:* 1–76. https://doi.org/10.1017/S0140525X22000681.

Okbay, Aysu, Yeda Wu, Nancy Wang, Hariharan Jayashankar, Michael Bennett, Seyed Moeen Nehzati, Julia Sidorenko, et al. 2022. "Polygenic Prediction of Educational Attainment within and between Families from Genome-Wide Association Analyses in 3 Million Individuals." *Nature Genetics* 54, no. 4: 437–449. https://doi.org/10.1038/s41588-022-01016-z.

Plomin, Robert and Sophie von Stumm. 2022. "Polygenic Scores: Prediction versus Explanation." *Molecular Psychiatry* 27, no. 1: 1–4. http://dx.doi.org/10.1038/s41380-021-01348-y.

Purcell, Shaun M., Naomi R. Wray, Jennifer L. Stone, Peter M. Visscher, Michael C. O'Donovan, Patrick F. Sullivan, and Pamela Sklar. 2009. "Common Polygenic Variation Contributes to Risk of Schizophrenia and Bipolar Disorder." *Nature* 460, no. 7256: 748–752. https://doi.org/10.1038/nature08185.

Raffington, Laurel, Travis Mallard, and K. Paige Harden. 2020. "Polygenic Scores in Developmental Psychology: Invite Genetics in, Leave Biodeterminism behind." *Annual Review of Developmental Psychology* 2: 389–411. https://dx.doi.org/10.1146/annurev-devpsych-051820-123945.

Ripke, Stephan, James T.R. Walters, and Michael C. O'Donovan. 2020. "Mapping Genomic Loci Prioritises Genes and Implicates Synaptic Biology in Schizophrenia." *MedRxiv*. https://doi.org/10.1101/2020.09.12.20192922.

Ruderfer, Douglas M., Ayman H. Fanous, Stephan Ripke, Andrew McQuillin, Richard L. Amdur, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Bipolar Disorder Working Group of the Psychiatric Genomics Consortium, et al. 2014. "Polygenic Dissection of Diagnosis and Clinical Dimensions of Bipolar Disorder and Schizophrenia." *Molecular Psychiatry* 19, no. 9: 1017–1024. https://doi.org/10.1038/mp.2013.138.

Schaid, Daniel J., Wenan Chen, and Nicholas B. Larson. 2018. "From Genome-wide Associations to Candidate Causal Variants by Statistical Fine-mapping." *Nature Reviews Genetics* 19: 491–504. https://doi.org/10.1038/s41576-018-0016-z.

Spirtes, Peter, Clark Glymour, and Richard Scheines. 2001. *Causation, Prediction and Search*. Cambridge, MA: MIT Press.

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. New York: Oxford University Press.

———. 2006. "Sensitive and Insensitive Causation." *Philosophical Review* 115, no. 1: 1–50. http://dx.doi.org/10.1215/00318108-2005-001.

———. 2010. "Causation in Biology." *Biology and Philosophy* 25, no. 3: 287–318. http://dx.doi.org/10.1007/s10539-010-9200-z.

Young, Alexander I., Stefania Benonisdottir, Molly Przeworski, and Augustine Kong. 2019. "Deconstructing the Sources of Genotype-Phenotype Associations in Humans." *Science* 365, no. 6460: 1396–1400. https://doi.org/10.1126/science.aax3710.