Behavioral and Brain Sciences

cambridge.org/bbs

Building causal knowledge in behavior genetics

James W. Madole[a,b] and K. Paige Harden[a]

[a]Department of Psychology, University of Texas at Austin, Austin, TX, USA and [b]VA Puget Sound Health Care System, Seattle, WA, USA
jmadole@utexas.edu
harden@utexas.edu

Target Article

Cite this article: Madole JW, Harden KP. (2023) Building causal knowledge in behavior genetics. *Behavioral and Brain Sciences* **46**, e182: 1–57. doi:10.1017/S0140525X22000681

Target Article Accepted: 12 April 2022
Target Article Manuscript Online: 5 May 2022
Commentaries Accepted: 5 September 2022

Keywords:
behavior genetics; causal inference; counterfactual reasoning; experimental designs; individual differences; philosophy of science

What is Open Peer Commentary? What follows on these pages is known as a Treatment, in which a significant and controversial Target Article is published along with Commentaries (p. 18) and an Authors' Response (p. 50). See bbsonline.org for more information.

Abstract

Behavior genetics is a controversial science. For decades, scholars have sought to understand the role of heredity in human behavior and life-course outcomes. Recently, technological advances and the rapid expansion of genomic databases have facilitated the discovery of genes associated with human phenotypes such as educational attainment and substance use disorders. To maximize the potential of this flourishing science, and to minimize potential harms, careful analysis of what it would mean for genes to be causes of human behavior is needed. In this paper, we advance a framework for identifying instances of genetic causes, interpreting those causal relationships, and applying them to advance causal knowledge more generally in the social sciences. Central to thinking about genes as causes is counterfactual reasoning, the cornerstone of causal thinking in statistics, medicine, and philosophy. We argue that within-family genetic effects represent the product of a counterfactual comparison in the same way as average treatment effects (ATEs) from randomized controlled trials (RCTs). Both ATEs from RCTs and within-family genetic effects are shallow causes: They operate within intricate causal systems (*non-unitary*), produce heterogeneous effects across individuals (*non-uniform*), and are not mechanistically informative (*non-explanatory*). Despite these limitations, shallow causal knowledge can be used to improve understanding of the etiology of human behavior and to explore sources of heterogeneity and fade-out in treatment effects.

1. Introduction

Violent crime is endemic to human society. The second-century poet Juvenal described Ancient Rome as having "no shortage of thieves" and "many opportunities to die" (Juvenal, 1769). Court records from thirteenth-century England show that "murderous brawls and violent death … were everyday occurrences" (Gurr, 1981, p. 305). Now, statistics suggest that 2020 was America's "most violent year in decades," with more than 19,000 people killed in firearm-related incidents (Bates, 2020). Understanding *why* people act in violent and criminal ways remains a societal imperative.

In a 2013 address, then-President Barack Obama offered one possible avenue for reducing crime: improving early childhood education. "Every dollar we invest in high-quality early childhood education," said Obama, "can save more than $7 later on by … reducing crime" (Obama, 2013). At its core, this statement communicates the *causal* hypothesis that high-quality childhood education will reduce crime. This hypothesis about a cause–effect relationship takes the form of a *counterfactual* statement about what could be. *If* the availability of high-quality childhood education were different, Obama predicts, then crime rates would also be different.

As social scientists, one of our primary aims is to produce research that verifies or challenges these sorts of causal claims. We examine evidence as to whether a causal relationship exists between two variables, offer theories for interpreting causal associations, and evaluate whether causal knowledge can be effectively applied to improve public health and well-being. As we will explain in this paper, this process of evaluating causes in social science relies heavily on counterfactual thinking, and it often begins by manipulating a variable in a randomly selected group of people. As we will also explain in this paper, this process of evaluating causes in social science is not limited to environmental exposures such as early childhood education. The same process of evaluating causes applies even when the causes in questions are variables less commonly considered by social scientists: genes.

1.1. Environmental causes in the social sciences: An empirical example

In the early 1960s, disadvantaged children living in Ypsilanti, Michigan were randomly assigned to an intensive two-year preschool education program (High/Scope Perry Preschool Program [HPPP]) that involved over two hours of daily active learning and weekly

© The Author(s), 2022. Published by Cambridge University Press

home visits from teachers (Schweinhart, Barnes, & Weikart, 1993). Children of the same age and socioeconomic background who were not assigned to this program received no preschool education. All participants were assessed throughout the first 40 years of life to determine the effects of the education program on outcomes such as educational attainment, economic earnings, and criminal behavior (Belfield, Nores, Barnett, & Schweinhart, 2006).

This methodological design, known as a randomized controlled trial (RCT), serves as the gold standard for validating the sorts of causal claims advanced by President Obama. By randomly assigning participants to different levels of a manipulated variable (in this case, preschool education), researchers were able to approximate the counterfactual scenario of what would have happened *if* conditions had been different. We can observe both the rate of criminal behavior in those who were given the treatment of better education *and* the rate of criminal behavior among children whose lives proceeded as usual (the control group). Relative to the control group, those children who participated in the high-quality preschool education program received over 50% fewer total arrests and over 80% fewer charges for violent crimes by age 40 (estimates based on data presented in Heckman, Moon, Pinto, Savelyev, & Yavitz, 2010). Because of the experimental design of this study, we can conclude that President Obama was correct: improving early childhood education caused a reduction in adult criminal outcomes.

This conclusion is more meaningful than merely observing a *correlation* between attending preschool and (not) committing crimes. As social scientists, we privilege inferences about causal relationships over such correlational ones and believe that they reveal something unique about the world. As this paper will show, the conclusion that good preschool education causes an average reduction in adult criminal behavior is distinctive, but how we interpret and apply this knowledge depends on the type of causation implied by experimental designs.

To further illustrate both the power and the limitations of this causal knowledge, consider that of the six individuals from the HPPP study who went on to incur the greatest number of lifetime criminal charges, three of them had actually participated in the preschool education program (Heckman et al., 2010). So, despite being exposed to a program that "causes" a reduction in criminal behavior, these individuals nevertheless received a total of 110 criminal charges between them. At the same time, the specific mechanisms underlying the effect of preschool education on crime are opaque. Indeed, researchers were surprised to observe the effects of preschool on adult outcomes, as the benefits of the intervention had appeared to fade out entirely in middle childhood (Heckman, 2006). Whatever intermediary process linked an educational experience at age four with a behavior committed (or not committed) by age 40 is not known (Schneider & Bradford, 2020). Clearly then, the causation implied by the experimental paradigm does not suggest that preschool education is the sole determinant of a person's lifetime criminal behavior, nor that the criminal behavior of any single individual can be attributed to the preschool education they received, nor does it explain anything about the mechanisms generating individual differences in the relationship between preschool education and criminal behavior.

Nevertheless, knowing that preschool education makes an average difference in adult criminal behavior is useful. Most directly, this knowledge has led to calls for policy changes in the United States to develop and disseminate childhood education programs. That is, the most straightforward application of the observation that changing **X** produced an average difference in **Y** is to develop intervention and prevention programs that target **X** on a large scale. In this paper, we refer to this application as *first-generation causal knowledge*. Knowing that **X** caused **Y** in one group of people implies that one could change **Y** in future groups of people by changing **X**.

As many interventionists and policymakers can attest, however, first-generation causal knowledge can be quite limited (Bryan, Tipton, & Yeager, 2021). Treatment effects often fail to sustain over time, to generalize to other samples, or to behave in predicted ways (Bailey, Duncan, Cunha, Foorman, & Yeager, 2020). Further, even when these effects show maintenance and durability across time and place, they often operate through unobserved mechanisms, obfuscating deep understanding of the effect. We may know that improved preschool education caused decreases in crime, but we have limited understanding of for whom this effect will hold, why it holds, for how long it will last, or how portable this effect will be across contexts. Indeed, recent RCTs of preschool programs for children from low-income families have found surprising *adverse* effects of preschool on children's academic achievement, attendance, and disciplinary infractions (Durkin, Lipsey, Farran, & Wiesen, 2022). Clearly, first-generation causal knowledge is not sufficient to anticipate how an effect will play out in a different environmental and historical context.

Overcoming this challenge requires what we refer to in this paper as *second-generation* causal knowledge. By revealing sources of heterogeneity and mechanisms supporting the durability of causal effects, we can better understand when, where, why, for whom, and for how long **X** makes some difference in **Y** – and this knowledge gives us more avenues for effecting change. Knowing that preschool education made an average difference in adult criminal behavior is useful in the near term, because we identify preschool education as a potential intervention target. But we must go beyond that, examining the causal pathway from early education to adult crime to identify *other* intervention targets whose manipulation might yield larger, more enduring, or more generalizable changes in criminal behavior.

1.2. Evaluating genetic causes in the social sciences: An impossible or worthless task?

Let us consider another causal hypothesis: certain genetic variants cause violent and criminal behavior. As evidence for this

James W. Madole is a doctoral student in clinical psychology at the University of Texas at Austin. His research focuses on the development of health-related outcomes across the lifespan, with a particular interest in the philosophical and ethical implications of associations between biology and behavior. His work has been published in *Clinical Psychological Science*, *Biological Psychiatry*, *Molecular Psychiatry*, and *Nature Neuroscience*.

K. Paige Harden, Ph.D., professor of clinical psychology at the University of Texas at Austin, is a behavioral geneticist whose research investigates genetic and environmental contributions to social inequalities across the lifespan, with a particular focus on the development of cognition and psychopathology in childhood and adolescence. She is the author of *The Genetic Lottery: Why DNA Matters for Social Equality* (Princeton, 2021) and the recipient of the American Psychological Association's Distinguished Scientific Award for an Early Career Contribution to Psychology.

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

claim, we might point to behavior genetics research on the heritability of antisocial behavior (see Lahey, Waldman, & McBurnett [1999] and Moffitt [2006] for review) and on specific measured DNA variants that can predict antisocial behavior and involvement with the criminal justice system (Karlsson Linnér et al., 2021; Tielbeek et al., 2017). The suggestion that genetic variants cause criminal behavior likely triggers a stronger intuitive response from many of our readers than the suggestion that being deprived of an education in early childhood causes criminal behavior. Genetic effects tend to be viewed as more essential, more natural, and more immutable than other causes (Dar-Nimrod & Heine, 2011; Lynch, Morandini, Dar-Nimrod, & Griffiths, 2019). Accordingly, claims about genetic causes are more controversial, both to our fellow scientists and to the general public, than claims about environmental ones. Research linking genetics with human behavior (along with some neuroscience research, which we do not address here) has been characterized "subversive science" that has the "power to shake the public's faith" in "cherished ideologies" of responsibility and equality (Fox, 2019, pp. 153, 156). As the biologist Richard Dawkins noted almost four decades ago, genes have acquired a "sinister, juggernaut-like reputation" (Dawkins, 1982/2016, p. 12).

Genetics' sinister reputation has historical roots. In the twentieth century, results from the nascent field of behavioral genetics were used to justify state-sponsored violence against the socioeconomically disadvantaged and people of color, including forcible sterilizations. This history is – and will likely continue to be – a stumbling stone for those asked to consider the idea that genetic differences between people could *cause* the behavioral outcomes that are the province of social science.

Despite these fears that genetics will be misused to justify racist and classist oppression, the search for genetic correlates of human behavior is accelerating. The past decade has witnessed a rapid expansion in the collection and analysis of genomic data. As of June 2021, more than 38 million individuals had contributed DNA to ancestry-testing companies (Janzen, 2021) and over 5,000 genome-wide association studies (GWASs) had been published (Buniello et al., 2019). This includes GWASs of social and behavioral phenotypes, such as educational attainment (Lee et al., 2018), household income (Hill et al., 2019), and criminal activity (Tielbeek et al., 2017). As ancestry-testing companies and national biobanks continue to accrue DNA samples from millions of individuals, and as genetic variants continue to demonstrate associations with more and more biologically distal life outcomes, scientists have an outstanding responsibility to address the implications of genomic research.

The obvious shadow cast by the history of eugenics can make it difficult to see another stumbling block to considering claims about genetic causation: a widespread confusion about the basics of causal inference, about how genetic research in humans could ever establish causation, and about what such causal knowledge would ever be good for, in the absence of the ability to tinker directly with people's genes. The goal of this paper is to resolve this stumbling block by describing how certain genetic research designs map onto what social scientists *already* know about establishing causal relationships and applying causal knowledge. By describing a clear perspective on what it does – and *does not* – mean for genes to be causes, and how that causal knowledge can be ethically applied, we also challenge the genetic determinism and essentialism that have historically characterized the pernicious misapplications of genetics by political extremists.

Let us consider the problem in more detail. Why might the prospect of establishing genetic causes of human behavior seem difficult, perhaps to the point of impossibility? Recall that the first step for testing a cause in an RCT is to manipulate the variable-of-interest in a randomly selected group of people. In many corners of biology, manipulating the genome is not only viable, but widely practiced. Researchers studying rodents, insects, and sea and plant life commonly use gene-modification strategies (e.g., knockout, selective breeding) as a means of gaining experimental control (Nagy, Perrimon, Sandmeyer, & Plasterk, 2003). These techniques allow for a direct assessment of the (counterfactual) causal hypothesis that if the organism's genome had been different, the outcome would have been different too. But when it comes to testing causal hypotheses about the human genome, the very idea of experimental manipulation is provocative at best and contemptible at worst. Gene-editing technologies such as CRISPR have demonstrated that direct alteration of the human genome is possible, but the use of these technologies on any meaningful scale is both scientifically nascent and ethically ambiguous (Gaskell et al., 2017). Regardless of one's moral appraisal of gene modification in humans, the fact remains that at present, manipulating the genomes of a randomly selected group of people is not a practicable option for testing hypotheses about the genetic causes of human behavior.

Moreover, even if we concede that, at a conceptual level, genes *could* cause average differences in human behavior, at a practical level, it is not readily apparent what we would *do* with this knowledge. As Evelyn Fox Keller wrote, "[t]he major practical interest driving the search for the relative importance of different causal factors in producing a given phenomenon is to be found in the wish to effect change in that phenomenon" (Keller, 2010, p. 8). But, for the same reasons that we discussed above, we cannot (and should not) readily apply knowledge of genetic causes to change the genomes of large swathes of the population in the hopes of changing their outcomes. Indeed, many of us cannot even engage in that thought experiment without feeling anxiety or revulsion at the prospect. As a consequence, it might be easy to conclude that establishing genetic causes of human behavior, even if it could be accomplished, is not a worthwhile endeavor. The fruit of that causal knowledge, the idea that we could change behavior by changing people's genes, seems poisonous.

1.3. Goals of the current paper

This skepticism about the feasibility and value of establishing genetic causes, however intuitive and well-meaning it might be, is mistaken. As we discuss in this paper, genetic causes are like nearly all environmental causes investigated in social science: they are *non-uniform*, *non-unitary*, and *non-explanatory*. Indeed, most genetic causes, when appropriately identified, can be interpreted along the same lines as average treatment effects (ATEs) estimated from RCTs or other natural experiments. Genetic causes, such as environmental ones, are not deterministic, explanatory, or homogeneous across place and time, but they *do* make an average difference in social and behavioral outcomes.

We also consider not just the feasibility of causal inference about genes but also the utility of that endeavor. We propose that knowledge about genetic effects on important life outcomes can help us change people's lives for the better, and that these changes may be brought about via social science (i.e., environmental) interventions, not by manipulating genomes. Specifically, we call attention to *second-generation* causal knowledge. Examining

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

the causal pathways from genes to life-course outcomes allows us to improve etiological understanding, uncover sources of heterogeneity in those outcomes, and identify novel targets for intervention.

2. What is a cause and how do we identify them? A brief review of causal inference in the social sciences

For decades, the scholarly community has been polarized by how to interpret findings from behavioral genetics (Fig. 1). For some, the complexity of processes that span genes and behavior and the dynamic interplay between genes and environment preclude researchers from gleaning any sort of meaningful causal knowledge from behavioral genetic research designs (Block, 1995; Lewontin, 1974/2006; Turkheimer, 2011). For others, heritability estimates and correlations with measured genotypes are evidence that genes *determine* life outcomes (Herrnstein & Murray, 1996; Jensen, 1969; Murray, 2020).[1] And for still others, genes are neither non-causal nor supra-causal, but are rather causes of human behavior in a more circumscribed, probabilistic sense (Bourrat, 2020; Dawkins, 1982/2016). How to decide among these competing interpretations?

We think that the seemingly intractable conversation about how to interpret the results of behavioral genetic research can be advanced by first considering a more general, and less controversial question: how do social scientists typically think about (non-genetic) causes and how do they go about finding them?

2.1. "No causes in, no causes out"

Determining that a relationship is causal requires more than plugging data into statistical models. It requires *causal concepts* (Pearl, 2009). Conceptual definitions of causation have historically been expressed in terms of active behavior – a cause "produces" (Locke, 1690/1997), "forces" (Lakoff, 1993), and "changes" (Charlton, 1983). Empirical tests of causation, therefore, involve detecting such activity, and not all statistical associations are up to the task. The familiar adage "correlation does not equal causation" is founded on precisely this principle, that a statistical association between two variables does not inherently demonstrate that one of those variables produced or changed the other. Identifying statistical causes means grounding statistical models in causal concepts and assumptions. In other words, "no causes in, no causes out" (Cartwright, 1995, p. 154).

The predominant causal concept in scientific thinking is the counterfactual (Pearl, 2018). *Counterfactuals* refer broadly to any hypothetical situation that describes what would have happened if conditions had been different. In 1973, David Lewis asserted that the counterfactual was the cornerstone of causal reasoning, arguing that **X** is a cause of **Y** if (a) when **X** occurs, **Y** occurs and (b) in the closest possible alternative world where **X** did not occur, **Y** also would not have occurred (Lewis, 1973a). Boiling water causes a tea kettle to whistle because (a) when water boils in a kettle, it whistles and (b) in a close possible world where water was not boiling in a kettle, it would not have whistled. Causation, in this view, is a matter of counterfactual dependence (Lewis, 1973b).

Counterfactual logic marked a departure from thinking about causation in terms of the regular occurrence of two variables. *Regularity* accounts of causation, which had dominated much of the history of causal reasoning, required that for **X** to cause **Y**, **Y** must invariably follow **X** (Hume, 1748/1999; Mill, 1843/2002). Relying on the constant conjunction of two variables for causation, however, is problematic. Among the problems of regularity accounts is that they evoke the thorny concepts of *necessity* (whether **X** must be present for **Y** to occur) and *sufficiency* (whether **X** alone can bring about **Y**) (Hulswit, 2002; Mackie, 1965). Counterfactual definitions relieve the need for **Y** to be necessarily or sufficiently dependent on **X**. Boiling water causes a tea kettle to whistle, but it is neither necessary (we can create steam in a kettle without boiling water), nor sufficient (if the water is boiling but the spout is open, the kettle will not whistle).



Figure 1. Dimension of causal thinking in human behavioral genetics.

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

Despite these strengths, the counterfactual dependence account offered by Lewis (1973b) has limitations.[2] First, it fares no better than regularity accounts at ruling out third causal variables. Borrowing an example from Woodward (2005), the reading of a barometer and the occurrence of a storm are counterfactually dependent on one another, such that if the barometer reading dropped, a storm would occur and if the barometer reading had not dropped, the storm would not have occurred. Nevertheless, they are not causally related. Both are caused by a third variable, namely, atmospheric pressure (Woodward, 2005). Second, counterfactual dependence does not explain the direction of the causal effect (Brady, 2011). Observing the co-occurring presence and absence of two variables does not reveal which of those variables is causally responsible for the other. Third, and perhaps most critically, the Lewis counterfactual is subject to what Holland (1986) referred to as the *fundamental problem of causal inference*: it is impossible to simultaneously observe **X** and **not-X**. The same kettle of water cannot be boiling and not boiling at the same time.

Manipulationist accounts of causation address some of these limitations. Similar to Lewis' counterfactual, manipulationist thinking relies on hypothetically comparing what would happen to **Y** under different conditions of **X**. Where it deviates is in reserving causal efficacy for those counterfactual situations "that describe how the value of one variable would change under interventions that change the value of another" (Woodward, 2005, p. 15). The critical shift here is from an emphasis on counterfactual *dependence* to counterfactual *control* (Ross, 2015). Manually changing the reading on a barometer will not cause a storm to occur because the barometer lacks causal control over the weather (Woodward, 2005).

This subtle shift from dependence to control has important advantages. First, it ensures that the detected relationship is not an artifact of a common cause. If intervening on **X** changes **Y** (or the probability of **Y**), then holding everything else constant, this rules out the possibility that **X** and **Y** just happen to change together because of **Z** (Ross, 2018). Second, it allows us to determine the direction of the effect. Designating one variable to be manipulated and one to respond establishes temporal precedence and helps to segregate cause from effect (Hill, 1965/2015). That just leaves the *fundamental problem of causal inference* – how can we simultaneously observe the changed and unchanged versions of **X**? For that, we need to create parallel worlds.

2.2. Parallel worlds and potential outcomes

In the United States, more than 256,000 children and adolescents have witnessed or died from school shootings in the past two decades (Cox, Rich, Chiu, Muyskens, & Ulmanu, 2018). The median age of assailants is 16 years old (Cox et al., 2018). While we know that changing preschool education is an effective means of reducing violent crime, if we have already missed the opportunity to improve an individual's preschool experience, we must develop other methods for reducing violent and aggressive behavior during critical developmental windows. Suppose you think that, for gun violence to end, adolescents need to be more compassionate toward one another. Equipped with an understanding of the relevant causal concepts, you know that to demonstrate that compassion causes a reduction in violent behavior, you need to manipulate compassion and see how violent behavior responds. For example, you might design a curriculum for first-year high school students that increases awareness of positive emotions and strengthens empathic communication skills. To test whether this intervention works, you need to create *parallel worlds*, running with the exact same conditions at the exact same time, save for one single difference: the presence of the compassion intervention. Each world then hosts a range of *potential outcomes*, in this case, the prevalence of violent behavior. The difference in the observed outcomes across these worlds represents the *causal effect* of the compassion intervention on violent behavior.

In social science, the simulation of parallel worlds and potential outcomes most often takes the form of a *randomized controlled trial* (RCT; Fisher, 1925). We create parallel worlds by assigning different, but similar, people to different conditions of an intervention (i.e., treatment groups). We consider the response of each treatment group as a representation of potential outcomes, of what would have happened given the opposite condition. We summarize the causal effect by taking the difference of the average effect for each treatment group (ATE; Rubin, 2005). RCTs entitle causal inference because they translate those theoretical causal concepts – *manipulation*, *counterfactual control*, *parallel worlds*, *potential outcomes* – into empirical action. They provide an algebra of the counterfactual (Pearl, 2010).

How well an RCT approximates these causal concepts, however, depends on how well it meets four critical assumptions: *independence*, *sample homogeneity*, *potential exposability*, and *SUTVA* (*stable unit treatment value assumption*). Together, these assumptions build confidence that a study truly tests whether **X** has causal control over **Y**. Fortunately, most of these are satisfied (at least in expectation) by a single methodological tool: *randomization*. By randomizing participants to treatment groups, we neutralize any dependency between treatment assignment and outcome (*independence*; Holland, 1988), and we balance (in expectation) the treatment groups on all variables other than **X** (*sample homogeneity*; Rubin, 1974). Randomization thus forms the basis of our parallel worlds, ensuring that the mechanism splitting our sample into respective worlds operates in a way that maximizes the uniformity of these worlds. Any causal effect is therefore attributable to the control of **X** over **Y**, and not to any artifactual differences between these worlds.

Randomization also helps confirm that all participants can be potentially assigned to any of the treatment conditions (*potential exposability*; Jo & Muthén, 2001). This marks the first step toward preserving the comparison of potential outcomes. If certain participants are unable to receive one of the treatment conditions – that is, if **X** cannot be manipulated for them – then the counterfactual collapses. Holland's (1986) proclamation "No causation without manipulation" is emphasized for exactly this reason (p. 959). If **X** cannot be changed, then the potential outcome of what would have happened had **X** been different does not exist, and no causal comparison can be drawn. Importantly, this proclamation can be extended to cover scenarios in which **X** is only hypothetically manipulatable, but where pragmatic or ethical considerations limit its ability to be manipulated in practice (Holland, 1986; Woodward, 2005).

If randomization sets the counterfactual conditions of a study into motion, *SUTVA* guarantees that they persist as the study unfolds. SUTVA protects the uniformity of parallel worlds and the openness of potential outcomes by stipulating that (a) participants in each treatment group receive identical forms of the treatment and (b) the outcome for each participant is not influenced by the treatment assignment of another participant (Rubin, 1980). Uniting these tenets is the overarching principle that, once parallel worlds have been set to run, no new worlds are created. Consider, for example, if instead of receiving the same compassion

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

curriculum, some students received education focused on building communication skills, while others learned mindful breathing or expressive writing. We could no longer meaningfully compare the potential outcomes of **X** and **not-X** because **X** would represent several divergent conditions. Likewise, if participants from the treatment group share their discoveries with members from the control group, then our parallel worlds have intersected and opened new counterfactual doors. For the difference between potential outcomes to have causal validity, the parallel worlds initiated by randomization must be preserved throughout the study. In theory, "SUTVA is automatically satisfied under the Fisher (1935) null hypothesis of absolutely no treatment effects of any kind" (Rubin, 1986, p. 961), though in practice, meeting SUTVA involves careful methodological design and statistically testing the magnitude of potential interference (Hudgens & Halloran, 2008; Sobel, 2006).

2.3. Conceptualizing causes

We began with causal concepts. Next, we translated those concepts into empirical parameters and assumptions in the form of an RCT. The final step is to export a causal conclusion. Yet drawing an appropriate causal conclusion is not always straightforward. For one, there are many different kinds of causal relationships – some are general rules, others are specific instances; some are direct, whereas others are bridged by a cascade of intermediary forces (Hausman, 2005; Rottman & Hastie, 2014). Moreover, a statistical parameter, by itself, provides little insight into the type of observed causal relationship. An ATE reveals only that there is a mean difference between groups. When it comes to interpreting instances of counterfactual control, however, philosophers have established a set of dimensions along which causal relationships can be conceptualized (see Woodward [2010] on *stability*, *specificity*, and *levels of explanation*). Because RCTs simulate counterfactual conditions, these dimensions can be readily exported and applied to interpreting ATEs (see Deaton & Cartwright, 2018). In most of the social sciences, ATEs are perhaps best understood by describing what they are not: they are not *uniform*, not *unitary*, and not *explanatory*.

Uniform causes produce effects in the same way every time. For example, atmospheric pressure invariably causes a barometer to drop. At least in theory, we often presume that treatment effects will behave uniformly (*unit homogeneity*; Holland, 1986). Despite this expectation, we often observe substantial heterogeneity in treatment effects (Angrist, 2004; Kent, Rothwell, Ioannidis, Altman, & Hayward, 2010). This is an important indication of the type of observed causal relationship – it tells us that the observed relationship is *probabilistic* rather than *deterministic*. Heterogeneity indicates that the cause does not affect the outcome in the exact same way across person, place, or time. And indeed, this is what we find in RCTs: "there is no warrant for the convenient assumption that the ATE estimated in a specific RCT is an invariant parameter, nor that the kinds of interventions and outcomes we measure in typical RCTs participate in general causal relations" (Deaton & Cartwright, 2018, pp. 13–14). This limits the ideographic and external validity of ATEs. They do not tell us about singular causes (i.e., that **X** is the cause of **Y** in a specific instance for a specific person), nor do they tell us about general claims (i.e., that **X** will cause **Y** in all places at all times) (see Cartwright [1988] for a discussion of *singular* vs. *generic* causes).

Unitary causes produce effects entirely on their own. Atmospheric pressure, for example, is singularly capable of dropping the reading on a barometer. Heterogeneity in treatment effect provides another important indication here. It tells us that the causal relationship is dependent on the presence of other factors (i.e., *moderators*). Adolescents with a large emotional vocabulary may show a greater reduction in aggressive behavior after a compassion intervention than those with more limited vocabularies. In this case, compassion is not *causally exclusive*, but rather, its effect on violent and aggressive behavior is embedded within a system of other causes whose collective functioning brings about the outcome. This renders ATEs *local* parameters that reflect causes that are inextricably tied to the demographic composition and environmental context of the measured sample.

Explanatory causes provide a description of how the cause brought about the effect. For example, atmospheric pressure causes a barometer to drop by changing the balance of the weight of mercury and the air pressure inside of the barometer. In contrast, ATEs tell us only that changing one variable will change the other, without explaining how this change comes about (Woodward, 2002). This explanatory, or *causally distal*, gap divorces causes from mechanisms. *Mechanisms* can be conceptualized as complex causal systems whose interrelated parts collectively produce an effect (Glennan, 1996). Identifying mechanisms requires (a) decomposing the effect into the component processes extending from cause to effect and (b) articulating how those processes function together to generate an outcome (Craver & Darden, 2013). These are different concepts than those at work in RCTs, so their empirical validation requires a different set of scientific practices.

2.4. First- and second-generation causal knowledge

In 1949, John Cade reported a series of case studies finding that lithium salts helped to pacify "psychotic excitement" (Cade, 1949). In his initial report, Cade called for "controlled observation of a sufficient number of treated and untreated patients" to test more conclusively whether differences in lithium administration caused differences in manic symptoms (Cade, 1949, p. 518). Seventy years, and dozens of controlled trials later, lithium has been heralded as a "psychiatric success story" (Draaisma, 2019, p. 584). The well-established knowledge that lithium makes an average difference in manic symptoms has been packaged into the first line of treatment for bipolar disorder in clinical practice (Draaisma, 2019; Volkmann, Bschor, & Köhler, 2020). "I don't believe in God," wrote Jaime Lowe, "but I believe in Lithium" (Lowe, 2015, para. 35).

The "controlled observations" upon which the efficacy of lithium was established constitute what we refer to as *first-generation causal knowledge*. This is the knowledge that a variable makes a non-uniform, non-unitary, and non-explanatory (i.e., average) difference in an outcome. As we have demonstrated so far, this is the type of information that is gained from standard counterfactual comparisons under the potential outcomes model. The promise of first-generation causal knowledge has historically been that, despite everything it lacks, it suggests a target that can be manipulated to change the probability of an outcome on a large scale (Gueron & Rolston, 2013). Because we know that lithium treatment causes an average difference in manic symptoms, we can prescribe lithium to bipolar patients, in the hopes of reducing the severity of their manic symptoms, even if we lack a clear sense of who is most likely to benefit from this treatment or how this causal relationship comes about.

And yet, for all the difference that lithium has made, not knowing exactly how or for whom this treatment works has limited its utility. Lithium is effective in fewer than one in three

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

patients and, even after 70 years of research, its mechanisms of action remain largely undefined (Alda, 2015; Harrison et al., 2016). Lithium is far from the only intervention with a positive ATE, high heterogeneity in its effects, and unclear mechanism (s) of action. As a result, scientists have become increasingly vocal about the limitations of first-generation causal knowledge (Bailey et al., 2020). In the behavioral and social sciences, as seminal findings have failed to replicate, generalize, and sustain over time, scholars have criticized "the narrow emphasis on discovering main effects and the common practice of drawing inferences about an intervention's likely effect at a population scale based on findings in haphazard convenience samples that cannot support such generalizations" (Bryan et al., 2021, p. 1). If social science is to advance and reach more people, we need to "revolutionize" our approach to identifying and applying causal knowledge (Bryan et al., 2021, p. 1).

In many corners of science, this revolution has already started. Once again, we can look to lithium treatment for guidance. Knowing that lithium creates an average difference in manic symptoms is useful not only because it identifies an intervention target, but also because it identifies a causal pathway that can be investigated to better understand the pathophysiology of bipolar disorder and sources of heterogeneity in its treatment. For example, recent research has found that variation in properties of neuronal signaling explains differences in response to lithium (Mertens et al., 2015). In particular, studies of lithium responders versus non-responders found that the former show a reduction in the hyperexcitability of hippocampal dentate gyrus neurons, suggesting that this "might be the mechanism that allows [lithium] to improve symptoms in both mania and depression phases" (Stern et al., 2018, p. 1461). With this knowledge, these researchers have been able to predict more accurately who will respond to lithium, to test whether alternate treatments reduce neuronal hyperexcitability in lithium non-responders, and to discover highly specific electrophysiological processes that serve as candidates for pharmacological intervention (Santos et al., 2021; Stern et al., 2018).

All of this followed from the first-generation knowledge that lithium makes an average difference in symptoms of mania. What initially appeared to be a critical flaw in the results from an RCT – that the results are not perfectly portable across all people – turned out to be a boon for scientific discovery. By continuing to investigate the causal pathway, and more specifically, *heterogeneity* in the causal pathway, we have been able to migrate our relatively shallow understanding of this causal effect to a position of greater causal depth. These types of investigations represent a progression toward what we refer to as *second-generation causal knowledge*. This is knowledge that provides a "clear sense of the *mechanisms of change* through which effects (intended and unintended) occur, which specific [causal] components and combinations are likely to be most (and least) effective, and *in what contexts* and with *whom* such effects will potentially be replicable" (Bonell, Fletcher, Morton, Lorenc, & Moore, 2012, p. 10). The promise of second-generation causal knowledge is that, by identifying processes and contexts through which the effect emerges, we will be able to increase uniformity, improve understanding, and isolate steps in the causal path that serve as candidates for intervention.

2.5. Summary

In this section, we discussed one of the primary tools that social scientists use to test causation: RCTs. The *counterfactual* was introduced as the primary causal concept that gives RCTs causal power, with particular emphasis placed on counterfactual situations that involve *manipulation* and *control*. The construction of *parallel worlds* and the comparison of *potential outcomes* across these worlds was discussed as the foundation of the ATE. Guidelines for interpreting ATEs in the context of RCTs were advanced by detailing what these causal relationships are not: they are not the same across all people (*uniform*), they are not isolable causes (*unitary*), and they are not explanations for how a cause changes an effect (*explanatory*). Using the example of lithium administration, we highlighted how the understanding that a cause creates an average difference in an outcome (first-generation causal knowledge) is traditionally used to identify and implement large-scale intervention targets. We reviewed the limitations of this application and highlighted how second-generation approaches can improve our understanding of the mechanisms of action generating an effect and sources of heterogeneity in treatment outcomes. In the next section, we carry forward this experimental and interpretational framework to scaffold our definition of what it means for genes to be causes.

3. Causal inference in genetic research designs

3.1. Overview of behavior genetics

Tracing the causes of human behavior has been of scholarly interest because long before social scientists were using RCTs to manipulate measured variables. In every epoch of documented history, *heredity* has been considered one such source of human action and decision making (see Loehlin [2009] for a complete history of *behavior genetics*). It was only relatively recently, however, that two major breakthroughs transformed this longstanding endeavor from speculation to quantification. The first came in 1869, when Francis Galton redefined the study of heredity as the study of measurable similarities between relatives (Galton, 1869; Kevles, 1995). Then in 2001, researchers successfully sequenced the human genome, making it possible to observe the composition of human DNA (Venter et al., 2001). These empirical milestones have provided critical scientific insight into the etiology of complex human outcomes, and it turns out that the pre-empirical scholars were right: genes do cause human behavior. Arriving at this conclusion, however, requires more than simply obtaining estimates of genetic associations. Once again, "no causes in, no causes out."

These methodological advances have formed the foundation of the two principal methodologies used in behavior genetics: *twin studies* and *genome-wide association studies* (GWASs). In twin studies, pairs of monozygotic twins, sharing 100% of their segregating genetic variance,[3] are contrasted with pairs of dizygotic twins, who share only 50%. The total variance of a measured trait can then be decomposed into three latent sources: *additive genetic variance* (a^2), *shared environmental variance* (c^2), and *nonshared environmental variance* (e^2) (Plomin, DeFries, Knopik, & Neiderhiser, 2013). Of primary interest to behavior geneticists is the proportion of phenotypic variance attributable to additive genetic variance, also known as a trait's *heritability* (h^2). Similar to an R^2 effect size, heritability is useful in that it quantifies the extent to which phenotypic differences are statistically accounted for by genetic differences (writ large), but it fails to specify which genes or, crucially, how those genes are responsible for producing phenotypic differences. Without such mechanistic knowledge, it can be difficult or impossible to predict whether genetic influences will be portable across environmental contexts (Mostafavi et al., 2020; Uchiyama, Spicer, & Muthukrishna, 2021).

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

The breakthrough in genetic sequencing modernized the estimation of genetic associations from a single unobserved variable to millions of observed variables. In GWASs, individual genetic sites (known as *single-nucleotide polymorphisms* [SNPs]) are entered as independent variables in a linear regression predicting a measured phenotype. This hypothesis-free approach tests associations with millions of SNPs in order to glean insight about which specific portions of the genome are associated with the occurrence or degree of a trait (Corvin, Craddock, & Sullivan, 2010; see Box 1 for a technical primer on GWASs).

The estimates from these linear regressions, called *SNP effect sizes*, represent either the probability that cases differ from controls at a particular genetic site or represent the magnitude of the association between a particular genetic site and a continuous outcome. The entire set of SNP effects (collectively referred to as *summary statistics*) is in turn used for a wide array of applications. A popular application in the social sciences is using GWAS summary statistics to create a *polygenic score*, which aggregates information from all SNPs into a single index of each individual's genetic propensity for a trait (Sugrue & Desikan, 2019).

Exactly what h^2 estimates and SNP associations tell us about the relationship between genes and behavior has been the source of much discourse and much disagreement (see Fig. 1). At one end of the spectrum are those that claim – quite extraordinarily – that these coefficients prove that traits such as intelligence are genetically determined and that differences in ability between racial and ethnic groups must be the result of hardwired genetic differences (Herrnstein & Murray, 1996; Jensen, 1969; Murray, 2020). At the other end are those that claim that, not only do these coefficients fail to represent genetic determinism or innate group differences, they fail to represent *anything* meaningful about how genes influence behavior (Block, 1995; Lewontin, 1974/2006).

Both extremist views are mistaken. Heritability estimates and SNP associations are neither supra-causal nor inherently meaningless. They are simply point estimates from statistical models. In the same way that a statistical association between cannabis use and psychotic symptoms would not imply that cannabis use is the ultimate or fixed source of psychosis, nor that a population with a high incidence of psychosis must therefore be using more cannabis, h^2 estimates and SNP effects imply neither deterministic associations nor between-group differences. What can be implied, however, is that using cannabis *potentially* increases risk for developing psychotic symptoms. This is a causal hypothesis that must be evaluated using study designs, such as RCTs, that appropriately instantiate causal concepts. Likewise, most behavior geneticists believe that twin studies and GWASs have utility in identifying genetic factors that potentially predispose for phenotypic differences between people (Visscher, Hill, & Wray, 2008).

But if the conclusion that we aim to defend is that genes cause behavioral and psychological outcomes, then clearly, we need something more than genetic associations alone. For genes to be considered causes, h^2 estimates and SNP effects need to be bolstered by the same causal concepts that privilege some *t*-statistics as ATEs. We need to know what the trait would have looked like if the genotype had been different. Unlike in RCTs, however, we cannot manipulate the treatment to simulate the counterfactual. We cannot randomly assign people to receive a certain genotype – at least not in any agreeable, ethical, or disseminable way. Fortunately, there is no need. The counterfactual has already been simulated for us.

3.2. Natural experiment of genetic inheritance

Consider this portentous lesson from history. In the fall of 1918, an influenza pandemic hit the United States without warning. By January of 1919, the virus had mostly disappeared. This meant that babies born just a few months apart experienced vastly different prenatal conditions. In effect, the virus had manipulated the prenatal environment, randomizing adjacent birth cohorts into those exposed to pandemic conditions – including either the flu itself or related stressors – and those experiencing relatively normal, or control, conditions. These cohorts represented parallel worlds that could be compared at different developmental stages to examine the causal effect of prenatal conditions on economic outcomes. Leveraging this natural randomization to simulate conventional RCT methodology, researchers concluded that *in utero* exposure to pandemic conditions caused lower educational attainment, lower income, and lower socioeconomic status in adulthood (Almond, 2006). The real-world generation of these counterfactual conditions typifies a *natural experiment*, in which treatment and control groups are meted out on the basis of a naturally occurring randomization mechanism.

In the case of genes, counterfactual conditions are created through meiosis, an instance of naturally occurring biological randomization. Meiosis is a process of cell division and DNA recombination that results in the production of unique sex cells (i.e., *gametes*). This process is essentially a natural manipulation of parental DNA. During recombination, segments of DNA from identical (i.e., *homologous*) chromosomes cross over in novel patterns to create new chromosomes to be inherited by offspring. Recombination is a primary source of intergenerational genetic variation (Nachman, 2002; Spencer et al., 2006), and the amount of variation created is vast. Within a single person, recombination results in the production of over 8 million unique chromosomal combinations (Batmanian, Ridge, & Worrall, 2011). When combined with a partner's gametes, there are over 70 trillion genotypes that an offspring could become (Carroll, 2020).

Box 1. Primer on GWAS

Single-nucleotide polymorphisms (*SNPs*) are the sites of DNA that commonly vary in the population (>1%). Each SNP is composed of a pair of allelic variants, or two of four possible genetic "letters" (adenine, thymine, cytosine, and guanine). SNP genotyping identifies each genotyped individual's pair of allelic variants at each polymorphic site (Perkel, 2008). Because of the correlation structure among SNPs (known as *linkage disequilibrium*), it is possible to impute values for hundreds of SNPs not measured during the genotyping process, allowing for the analysis of millions of genetic variants in relation to an outcome. Prior to conducting GWAS, the raw allelic structure of each SNP is converted to an ordinal variable reflecting the number of minor alleles (i.e., the less commonly occurring allele in the population) that an individual possesses. The number of minor alleles at each SNP is what is then associated with the outcome to obtain an *SNP effect size*. SNP effect sizes are used for a growing number of applications, including annotating the biological function of identified SNPs (Watanabe, Taskesen, van Bochoven, & Posthuma, 2017), constructing polygenic scores (Sugrue & Desikan, 2019), modeling genetic associations with other traits (Grotzinger et al., 2019), and estimating SNP-based heritability (Yang, Zeng, Goddard, Wray, & Visscher, 2017).

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

Manipulation is not synonymous with randomization, however. If the variation produced by meiosis creates different treatment conditions, it must also be the case that the inheritance, or assignment, of these conditions is random. Two principles established by nineteenth century geneticist Gregor Mendel reassure us of the validity of randomization: (1) the *Law of Segregation* states that at every point in the genome, offspring randomly inherit one allele from each parent and (2) the *Law of Independent Assortment* states that alleles will segregate to gametes independently of one another (Davies et al., 2019). The astute reader will note that all alleles are not inherited entirely independently from each other because of *linkage* (National Human Genome Research Institute, 2022), the tendency for DNA segments that are positioned close together on a chromosome to be inherited together. We return to defining linkage and considering its implications for causal inference below.

Thus, at each genetic site, you inherit two alleles (one from your mother and one from your father) but *which* alleles you inherit of the possible parental alleles is a completely random event. Even though linkage makes it so that we inherit groups of alleles together (i.e., *haplotype blocks*) (Phillips et al., 2003), Mendel's principles are just as aptly applied to haplotype blocks as to SNPs – haplotype blocks are, at least in part, randomly created (Wang, Akey, Zhang, Chakraborty, & Jin, 2002), and inherited independently of one another (Browning & Browning, 2011). Crucially, this randomness gives genetic inheritance its experimental infrastructure (Davey Smith & Ebrahim, 2003): just as we can compare outcomes between treatment and control groups in the context of an RCT in order to gain insight about the average causal effect of the treatment, we can compare outcomes between family members who inherited different genes in order to gain insight about the causal effect of genotype. Confidence in these counterfactual conditions, however, depends on how well they meet those four critical assumptions of all randomized experiments – *independence*, *sample homogeneity*, *potential exposability*, and *SUTVA*. We consider each one in turn.

3.2.1. Independence

At face value, Mendel's laws satisfy the independence assumption. If genetic variants are randomly and independently assigned, then we should expect no systematic dependency between genotype and outcome (Holland, 1988). In actuality, there exist several possible violations of independence.

First, because of evolutionary factors and non-random mating patterns, different subpopulations, such as those with different ancestral backgrounds, have different frequencies of certain alleles, known as *population stratification* (Cardon & Palmer, 2003). Discrepancies in allele frequency across different groups of people are often systematically associated with environmental differences (*environmental confounding*), non-ancestral-related genetic differences (*genetic confounding*), and mate selection (*assortative-mating confounding*) (Young, Benonisdottir, Przeworski, & Kong, 2019). This means that if we estimate a genetic association in a sample of people who are not close biological relatives, we cannot separate the causal effect of the gene from any of these confounding sources. Conventional GWASs do their best to mitigate these problems. For instance, they are conducted in ancestrally homogeneous samples (Mills & Rahal, 2019), and even within these samples, population stratification is often corrected for by controlling for ancestry-based principal components (Price et al., 2006) or using linear mixed models (Yang, Zaitlen, Goddard, Visscher, & Price, 2014). But, none of these practices guarantees independence (Haworth et al., 2019).

The only way to surmount this problem is to examine genetic associations relative to parental genotypes, for example, by directly comparing an offspring to both of its parents or by comparing siblings from the same family (Brumpton et al., 2020; Young et al., 2018). For any individual, "each of the meiosis and conception events that determined [a person's] DNA is an independent event *conditional on the parental genotypes*" (Davies et al., 2019, p. R174, emphasis added). Here are Mendel's laws in action: the genotype of any individual is a random and independent selection of genes from their parents. Because siblings inherit their genes from the same pool of potential genotypes, the pitfalls of population structure can be avoided if the comparison of siblings is appropriately conditioned on their parental genotypes (Fletcher, Wu, Li, & Lu, 2021; Zaidi & Mathieson, 2020). Novel designs such as within-sibship GWASs and relatedness disequilibrium regression (RDR) exploit the randomization in meiosis that renders treatment assignment and outcome independently (Howe et al., 2021; Young et al., 2018).

Second, alleles are not inherited completely independently from each other. Rather, DNA segments that are positioned closely together on a chromosome are more likely to be inherited together, as there is a lower probability of a recombination event occurring between them. As a loose analogy, if you shuffle a deck of cards and then split the deck, two cards that are right next to each other in the deck before shuffling are more likely to end up in the same half of the deck than cards that are far apart from each other. This co-inheritance results in *linkage disequilibrium* (LD), that is, a correlation between alleles.

The issue of LD raises a more general issue, which we refer to as the *resolution* of genetic effects. The highest resolution for genetic causes is to identify an individual genetic variant. When geneticists talk about identifying a "causal variant," they are using a high resolution for genetic effects: a C1 allele in the *cystic fibrosis transmembrane regulatory (CFTR)* gene, for example, causes cystic fibrosis. The lowest resolution for genetic causes is the entire genome. A method such as RDR can conclude that, if people had inherited different genetic segments from their parents, their phenotypes would be different. This is a causal conclusion but one that is silent regarding *which* genetic variants are causally relevant.

An intermediate resolution for conceptualizing genetic causes, and the resolution most relevant for understanding the results of GWASs, is neither the individual variant nor the entire genome, but instead a set of alleles that are all in high LD with each other (but not in LD with other alleles). A within-sibling GWAS leverages the natural experiment of meiosis, but it does not measure every possible genetic variant. Thus, a "hit" in a within-sibling GWAS, that is, an SNP that is associated with within-sibling differences in phenotypes, might be the causal variant, or it might be in LD with the causal variant. That is, the SNP is best considered a *measure* of an underlying genetic cause, while the specific causal variant often remains unknown.

In order to build an intuition about how an SNP can be a measure of a cause, rather than the cause itself, it might be helpful to consider other types of natural experiments. Consider, for example, the Dutch Hunger Winter studies. In 1944, the Nazis retaliated against Dutch resistance to occupation by imposing an embargo on transport to western Holland, causing a severe famine in large cities. By November, food rations were 450 calories per day, and the famine continued until Holland was liberated by the Allied armies in May 1945.

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

The Dutch famine has become a famous quasi-experiment for studying the effects of prenatal exposure to caloric restriction of adult health and cognition. Because the famine affected cities in a geographically circumscribed area in west Holland, for a circumscribed period of time, exposure to famine can be treated as if random, by comparing individuals conceived during the famine to individuals in the same cities who were conceived before or after the famine, and to individuals conceived at similar times in unaffected cities.

In their landmark 1972 study on the effects of prenatal famine exposure on cognition, males who appeared for military induction at age 18 were asked for their date and place of birth, which researchers used to assign them to "exposed" or "unexposed" groups (Stein, Susser, Saenger, & Marolla, 1972). That is, we can differentiate between the study's *cause* of interest (prenatal exposure to famine) and the study's *measurement* of that cause (participant's self-report of date and place of birth). Obviously, a participant writing down his date of birth is not the cause of his adult health. Rather, his self-report is an indicator used to infer his membership in a group that is as-if randomly exposed to the putative cause. Such a situation, where researchers are relying on potentially imperfect measures of putative causes, is common in natural experiments where researchers are not assigning participants to treatment and control groups, but are rather ascertaining exposures after the fact. Similar to the Dutch Hunger Winter researcher who has not randomly assigned their participants to be exposed to famine or not, a GWAS researcher has not assigned people to genotypes. Nature has randomly assigned offspring to genotypes from their parents, and the GWAS researcher is left trying to ascertain to which genotypes people have been assigned. An SNP array is an imperfect measure of that random assignment.

Putting these lines of reasoning together, the natural experiment of meiosis guarantees that segments of the parental genome are independently and randomly assigned to offspring, but there remains non-independence of specific alleles that are co-inherited and in LD. A within-family GWAS, then, will be able to successfully identify that "genes" have a causal effect on phenotypes, but "genes" are studied at an intermediate level of resolution, encompassing all alleles in LD with the measured SNP. Researchers can then use "fine mapping" techniques to gain higher resolution (LaPierre et al., 2021).

3.2.2. Sample homogeneity

Comparing members of the same family should allow randomization to serve another one of its chief functions: preserving sample homogeneity (Rubin, 1974). Randomization "guarantees, by construction,…that the [difference in means for all other causes] is zero *in expectation*" (Deaton & Cartwright, 2018, p. 4). In practice, sample homogeneity is a function of two factors: (1) the number of participants and (2) the number of trials (Deaton & Cartwright, 2018). GWASs are uniquely suited to address these factors. First, standard GWAS sample sizes tend to tally in the millions (e.g., Evangelou et al., 2018; Karlsson Linnér et al., 2019; Nielsen et al., 2018), orders of magnitude larger than typical RCTs. Although improving the sample sizes of within-family designs remains a critical aim of behavior genetics, recent studies have begun to analyze SNP effects in upward of 40,000 sibling pairs (Karlsson Linnér et al., 2021). Second, meiosis is essentially a series of millions of randomized trials. As the assortment of alleles at each genetic site is a random event, we should have increasing confidence that allele carriers do not differ in systematic ways as we aggregate over the genome. This makes summary indices of genetic effects, such as polygenic scores, particularly powerful tools.

3.2.3. Potential exposability

Potential exposability is directly related to manipulability. If the treatment is something that can be manipulated, or changed, then randomization ensures that every participant is potentially exposable to any condition (Jo & Muthén, 2001). In one sense, the conditions of meiosis easily satisfy the requirement of potential exposability. Meiosis manipulates parental DNA, creating trillions of unique genotypes for an offspring to inherit (Carroll, 2001). The fact that meiosis satisfies genotype-level exposability suggests that, as with sample homogeneity, indices that aggregate across the genome may be particularly suited for causal inference.

3.2.4. SUTVA

Consider a family with two adolescent children, Linda and Maggie. Through meiosis, Linda and Maggie were randomly assigned their genotypes, creating parallel worlds that could be compared to examine whether their genes caused different life outcomes. In particular, Linda inherits variants in the ADH1B gene that affect her metabolism of alcohol, contributing to her refraining from alcohol use (Bierut, 2011). Linda's substance-use choices become part of the environment that she shares with Maggie, a factor that often serves to align substance-use habits among siblings (see Samek, McGue, Keyes, & Iacono [2015] for a review of shared environmental factors in substance use). If observing Linda decline alcohol, return home promptly before curfew, and engage in substance-free recreational activities influences Maggie's alcohol-related behavior, then SUTVA has been violated. Linda's treatment assignment – her genotype – has interfered with Maggie's potential outcome, obfuscating a causal comparison of counterfactual conditions (Rubin, 1980). For SUTVA to be preserved in the natural experiment of genetic inheritance, there can be no *indirect sibling-to-sibling genetic effects* (Eaves, 1976).

The surest way to safeguard against the behavioral transmission of genetic effects between siblings is to analyze data from a single offspring controlling for both of the parental genotypes. Alternatively, one could compare the potential outcomes of siblings who were not raised together (e.g., *adoption studies*; Plomin, DeFries, & Loehlin, 1977). This assures that each sibling's genotype has as little influence on the other sibling's phenotype as possible. To be sure, even adoption studies cannot protect against other sources of indirect genetic effects (see Scarr & McCartney [1983] for a review), but these are more a problem of sample homogeneity than SUTVA. As an analogy, consider an RCT on a pharmacological treatment of depression. If some participants happen to read existential philosophy during their treatment, the threat is that a potential imbalance of philosophy readers across treatment groups will confound depression scores. Reading existential philosophy, however, has nothing to do with whether the depression treatment that one participant receives interacts with another participant's depressive symptoms. Non-sibling indirect genetic effects are like reading existential philosophy – they are sure to affect an offspring's outcome,[4] and they might create a systematic difference in (genetically influenced) environments across allele carriers, but they do not violate SUTVA.

Evidence has begun to suggest, however, that when siblings are raised together, their respective genotypes do in fact influence

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

their siblings' phenotypes (Fletcher, Wu, Zhao, & Lu, 2020). The presence of sibling interference need not undo causal inference entirely (Rosenbaum, 2007). In these cases, addressing SUTVA involves determining (a) the direction of the interference and (b) the magnitude of the effect. Developmental psychologists differentiate between imitation and contrast effects – those patterns of "behavioral acquisition via social learning" that serve to either fuse or drive apart sibling behavior (Carey, 1986, p. 320; see Dolan, de Kort, van Beijsterveldt, Bartels, & Boomsma [2014]; Moscati, Verhulst, McKee, Silberg, & Eaves [2018] for empirical demonstrations). Whether or not Linda refraining from alcohol use causes Maggie to similarly abstain or rebel into greater use depends on factors such as their relative ages (Abramovitch, Corter, & Lando, 1979) and the stage of their dyadic relationship (Carey, 1986).

Empirically examining SUTVA involves quantifying the magnitude of imitation or contrast effects by segregating the direct causal effect from interference effects. This can be achieved through a process of *triangulation* (Lawlor, Tilling, & Davey Smith, 2017), a leveraging of multiple data sources and unique methodological approaches to increase confidence in a causal conclusion. In the case of sibling interference, Kong et al. (2018) provide a paradigmatic example: using genotype data from both siblings and parents and integrating within-sibship comparison with a traditional trio design (see Connolly & Heron [2015] for a review), the researchers were able to triangulate on a direct causal estimate of genotype on outcome. By including the effect of the sibling's genotype and the uninherited portions of the parental genotypes in the model, Kong et al. (2018) estimated the magnitude of the interference and effectively ensured that it, and other confounding sources, were controlled for. Adoption studies may ensure protection against SUTVA violations, but innovative methodological approaches can still rescue causal inference in the face of sibling interference.

3.3. Shallow end of genetic causation

Perhaps no outcome has been more magnetic in contemporary behavior genetics than educational attainment (EA; Martin, 2018). The most recent GWAS of EA, published in 2018, has already been cited over 900 times (Lee et al., 2018). It has also generated a litany of passionate critiques and rebuttals (see, e.g., the blog post titled "Why We Shouldn't Embrace the Genetics of Education"; Warner, 2018). Yet prior to 2013, EA was considered a fairly rudimentary, albeit important, covariate in GWASs (Plomin & von Stumm, 2018). Priority had been given to medical and psychiatric disease states; EA was simply a confound to rule out. As GWAS methodology began to permeate the social sciences, however, the troves of data on EA that had been accrued over the years by large-scale research consortia became invaluable. Suddenly, EA had become the most GWAS-able trait.

The first GWAS of EA detected three SNPs with significant effects in 126,559 individuals, collectively explaining 2% of its variance (Rietveld et al., 2013). Three years later, 74 SNPs were detected in twice as many people, explaining 4% of the variance (Okbay et al., 2016). By 2018, the GWAS of EA included 1.1 million individuals, over 1,000 significant SNP effects, and explained over 10% of the variance (Lee et al., 2018). By social science standards, that is a large and stable effect size (Funder & Ozer, 2019), and one that even outperforms many complex, multivariate approaches to predicting educational outcomes (Salganik et al., 2020). The incremental successes of the EA GWASs are undeniably impressive, but they have not been accompanied by incremental increases in causal inference. Even if the fourth iteration of the EA GWAS detected 5,000 significant SNP effects and explained 50% of the variance in EA, it alone would not move us closer to the conclusion that genes cause educational outcomes.

To be sure, we are currently in a position to conclude that genes cause EA. But this conclusion is only possible because researchers have applied summary statistics from EA GWASs to datasets that allow for counterfactual comparison. By using "within-family genetic design[s]," differences in associations between polygenic scores and educational outcomes allow for "causal inference and explanation" (Selzam et al., 2019, p. 360). So when we find that "children with higher polygenic scores… move up the social ladder in terms of education, occupation, and wealth, even compared with siblings in their own family" (Belsky et al., 2018, p. E7281), the appropriate conclusion is that genes caused these differences in attainment.

For behavior geneticists, this is undoubtedly a triumph. After years of null results and unreplicable false positives, the field can now construct measures of DNA differences that caused important life outcomes. For others, however, this statement rouses ambivalence at best, and outrage at worst. There is a vocal contingent of bloggers, journalists, and scientists who fear that GWAS of social outcomes and its associated applications "will only be fuel for those who think that social inequalities are natural and unchangeable" (Samorodnitsky, 2020, para. 20). Such a picture of genetic causes is unwarranted, however, when we remember what it means for something to be a cause: genetic causes for human behavioral traits are *non-uniform*, *non-unitary*, and *non-explanatory*.

It can be easy to neglect that genetic causes behave just like ATEs from RCTs. Prominent examples from medicine have shaped expectations that genes are of a different class of causes (Ross, 2019). Take cystic fibrosis (CF), for example. CF is an autosomal recessive disorder present in about 70,000 individuals globally. It is caused by two mutated copies of the *CFTR* gene on the seventh chromosome (Cutting, 2015). Unlike most ATEs, this genetic cause is (a) *uniform* – it consistently produces the occurrence of CF across individuals, (b) *unitary* – it alone causes the occurrence of CF, and (c) *explanatory* – it provides an explanation for how CF occurs[5] (Elborn, 2016). Together, these characteristics make CF an instance of *deep genetic causation*[6] (see Turkheimer [1998] on *strong biologisim*; Meehl [1972] on *specific genetic etiology*). Scientists gravitate toward deep causes. They are salient, simplistic, and they provide a coherent framework for the operation of a complex system such as the genome (Engel, 1977; Kendler, 2005). Despite the conceptual attraction to deep causes, almost everything we have learned from GWASs points to genes as *shallow* causes – many variants from across the genome relate to behavioral outcomes, but when they matter and how they matter differs across people, place, and time (Ross, 2019). The appropriate paradigm for genetic causes of human behavior is therefore not the deeply deterministic example of CF, but the local, probabilistic, and distal characteristics of ATEs.

Support for the idea that genes are *non-unitary* causes of behavior is so robust that it has been consecrated as one of the modern laws of behavior genetics (Chabris, Lee, Cesarini, Benjamin, & Laibson, 2015). Indeed, arguably the greatest takeaway from the GWAS era has been that individual genetic variants do not produce behavioral effects on their own. This is not a trivial statement – decades of research were spent hunting for single polymorphisms (i.e., *candidate genes*) that would prove to have

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

causal control in the etiology of behavioral and psychological outcomes (see Munafò [2006] for an overview). Consistent failure of these findings to replicate, however, pushed behavior geneticists to develop more sophisticated models. Most believe now that the genetic architecture of complex traits is polygenic (involving thousands of variants with small effects distributed throughout the genome [Duncan, Ostacher, & Ballon, 2019]) or even omnigenic (involving sundry genome-wide variants that affect behavior by disrupting interconnected gene regulatory networks [Boyle, Li, & Pritchard, 2017]). But if single genes are not unitary causes of behavior, neither is the genome writ large. Even a model that considered every gene in the genome and its higher-order function would fail to be *causally exclusive* because it would fail to account for larger etiological systems such as "history and cohort, the life course, and social structures like gender" through which "genetic influence must be understood" (Herd et al., 2019, p. 1070). Genes might cause EA, but they are certainly not the *only* cause of EA.

The nesting of genetic effects within biological, psychological, and social systems is what makes them *local* parameters. The size and shape of a particular effect will always depend on the size and shape of the other causal factors present in that instance. In theory, this suggests that genetic effects will be *non-uniform*. If context matters, then genetic effects should change across settings. Nevertheless, there persists "the common assumption…that genetic effects are 'universal' across environments" (Tropf et al., 2017, p. 758). This would imply that genetic effects are *deterministic*, that they will produce the same effect in the same way every time, independent of the context. Two takeaways from modern genomics suggest that this assumption is unfounded: (1) genetic effects are heterogeneous across environments and (2) genetic effects show poor generalizability.

That genetic effects vary across environments is a proposition of longstanding tenacity (*gene × environment [G × E] interactions*) (see Jaffee & Price [2007] for a review; Feldman & Lewontin, 1975; Turkheimer & Gottesman, 1996). Reliably identifying such interactions, however, has historically proved difficult (Munafò & Flint, 2009). This is where the substantial increases in the predictive power of GWASs, similar to those seen in EA, have considerable value. Several studies have been able to provide insight into the environments that facilitate the emergence of genetic effects on EA. Genetic effects appear to increase in size when structural barriers such as gender (Herd et al., 2019), class (Rimfeld et al., 2018), and intergenerational mobility (Engzell & Tropf, 2019) are removed. Said differently, the probability that genes matter for EA varies depending on the environmental exposures of the individual. Similar heterogeneity has been observed in genetic effects on reproductive, physical, and psychiatric outcomes (Coleman et al., 2020; Tropf et al., 2017).

The disparity in genetic effects across environments is further corroborated by the fact that GWAS findings have largely failed to be applicable outside of discovery samples. That means that the genes that predict an outcome in one sample fare poorly in predicting the same outcome in a separate sample. Not only is this the case when testing predictive accuracy in diverse populations (Martin et al., 2017), but inconsistent accuracy has also been found when looking within demographic subgroups (i.e., age, gender, socioeconomic status) of the same ancestry (Mostafavi et al., 2020). Moreover, this failure for genetic effects to port was found even when phenotypes were measured consistently across samples. Variation in the measurement or applicability of a phenotype across populations, for example when "educational attainment emphasizes rote memorization or formal writing… [rather than] experiential learning," is likely to be another source of restricted generalizability (Meyer, Turley, & Benjamin, 2020, para. 5). Collectively, this suggests that while genes cause EA, this is neither a singular nor a generic claim (Cartwright, 1988). We know neither that genes are the cause of EA for a specific individual nor that genes are the cause of EA for all people across place and time.

Genes, however, do have generic functions (Dawkins, 1982/2016). Every gene produces biochemical material for cellular encoding, and the specific set of instructions governed by a particular gene is consistent across person, place, and time (Schaefer & Thompson, 2014). This would indicate that the first step in the causal pathway from genes to behavior is uniform. To ultimately arrive at non-uniform effects on behavior, there must therefore be subsequent points along this causal pathway where people diverge. Indeed, we know already that this divergence begins almost immediately after gene function. Even processes such as gene expression and gene regulation show substantial heterogeneity across environments (Bork et al., 1998), and this tends to be more the rule than the exception as the pathway winds through biological (Gough et al., 2017), psychological (Molenaar, 2004), and sociological systems (Scott, 1988). Each point of heterogeneity demarcates a *garden of forking paths* (Borges, 1941/2018; Gelman & Loken, 2014), a splintering of a uniform stream of processes into separable causal pathways. The staggering amount of heterogeneity that exists in the processes that extend from genes to behavior tells us that there are a potentially untraceable number of causal pathways. Identifying a genetic cause provides no insight into which causal pathway ultimately produced the behavior because genetic causes are not mechanisms. Genes might cause EA in the sense that genes made some *distal* difference in the level of attainment, but not in the sense that they provide an explanation for how this difference was made.

This distinction between *causes* and *mechanisms* often gets lost when applied to the relationship between biology and behavior (Thomas & Sharp, 2019). As Gregory Miller writes, "[r]elevant science abounds with demonstrations that…imply causal relationships between psychology and biology…yet we often write as if we know the mechanisms" (Miller, 2010, p. 717). Despite the implicit assumption that biology reveals something inherently mechanistic, there is nothing that necessitates that biological causes need to be mechanistic nor that mechanisms need to be biological. In plainest terms, mechanisms explain an effect. Even putatively biological associations, such as genetic effects on lung cancer, might be largely explained via social processes, such as access to cigarettes (Kendler et al., 2012). Further, whether a biological cause actually provides explanatory insight is a matter of circumstance. Consider the example from Turkheimer (1998) on the origins of vocal muteness for two individuals, one who has suffered a stroke in Broca's area of the brain and the other who has taken a religious vow of silence:

> It seems natural to describe the stroke patient's muteness as biological and the monk's as psychological. What do these attributions mean? It is not simply that the aphasia is 'in' the brain, because the monk's decision presumably resides there also. Instead, the difference involves the nature of the structural relationship between a neurological representation of the condition and a psychological account of it (p. 783).

In other words, the key difference is that only the aphasia patient's identified biological cause is mechanistic: it describes how a

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

localized lesion compromises the neural areas that support linguistic functioning – an outcome that would also happen to be invariant across time and place. Likewise, biology factors into the monk's muteness (see, e.g., research on the neural networks supporting religiosity; Kapogiannis et al., 2009). While this biology may prove useful in forecasting future spiritually-based muteness in other individuals using predictive modeling (Shmueli, 2010; Yarkoni & Westfall, 2017), it would be insufficient to explain this monk's vow of silence. The mechanism behind the monk's silence is psychological – it is his *decision* that generates and explains his outcome.

3.4. Using genetic causes to advance second-generation causal knowledge

In this paper, we have argued for the idea that certain types of genetic effects (i.e., contingent on parental genotypes) constitute first-generation causal knowledge. Similar to ATEs, genes are causal in the sense that "differences in [genotype]…cause phenotypic differences in particular genetic and environmental contexts" (Waters, 2007, p. 558). Unlike with many ATEs, however, this information cannot be used to manipulate the causal variable on a population scale. This would seem to limit the applied value of identifying and conceptualizing genes – or any other immutable cause – as *average difference-makers*. But, as the case of lithium treatment showed us, there is no reason to restrict first-generation causal knowledge to this singular application. Similar to all first-generation causes, genetic effects contingent on parental genotypes represent causal pathways that can be explored to advance second-generation aims.

If we accept the conclusion that genetic variants make an average difference in psychological and behavioral outcomes, then we can begin to embrace the trove of potential scientific discoveries lying along this causally distal pathway. To start, we can improve phenotypic understanding by exploring mediating processes. Traditionally, researchers have used GWAS results to gain deeper insights into the biology of complex behavioral outcomes (Dick et al., 2018). Approaches such as *bioinformatics annotation* make it possible to locate specific cells, tissues, and organs where relevant genetic variants are expressed (Watanabe, Mirkov, de Leeuw, van den Heuvel, & Posthuma, 2019). In *pathway analysis*, genetic variants are clustered by functional relatedness and used to assess whether candidate biological functions are implicated in disease etiology (White et al., 2019). Collectively, these techniques serve to "increase explanatory power" by specifying the "parameters of the nervous system [that] are aberrant as a result close in the causal chain to the gene or genes" (Khatri, Sirota, & Butte, 2012, p. 1; Meehl, 1972, p. 11).

Inspired by this method of biological discovery, researchers have called for an approach that maps genotypes to multifarious aspects of the social environment (*phenotypic annotation*; Belsky & Harden, 2019). By associating polygenic scores for one phenotype with related phenotypes at different stages across the lifespan, we can detail potential behavioral and developmental pathways through which target phenotypes emerge. Already this work has provided considerable insight into how genetic risk for adult outcomes such as body mass index, smoking, educational attainment, and attention-deficit/hyperactivity disorder manifests in childhood and adolescence (Agnew-Blais et al., 2021; Belsky et al., 2012, 2013a, 2013b, 2016). Uncovering more about the biological and behavioral intermediaries bridging genes and behavior improves our ability to develop integrated causal models of complex behavioral phenomena.

As our understanding of the causal structure of psychological and behavioral phenotypes deepens, our discovery of potential prevention and intervention targets improves (Dick, 2018). Indeed, each process that we find mediates cause and effect which represents a candidate for intervention, even if the original cause itself is immutable. Consider again the example of lithium administration: researchers localized the differential pattern of neuronal signaling in lithium responders and non-responders to the expression of a single gene (*LEF1*) (Santos et al., 2021). Rather than structurally alter the gene, these researchers explored its downstream biological consequences (e.g., transcription pathways), thereby identifying "useful phenotypes for drug development" (Santos et al., 2021, p. 12). In these cases, the relevant question is not just whether the cause itself is manipulatable, but (a) which of the mediating processes are manipulatable and (b) which processes' manipulation will generate a meaningful effect on the outcome. Knowing that *LEF1* causes differences in brain signatures characteristic of lithium response allows us to identify mechanistic processes that could be pushed upon to improve treatment responsivity.

The same should be true of behavioral and health-related outcomes. Understanding how genetic factors unfold along biological and behavioral pathways across development allows us to isolate intermediate processes that represent (a) prognostic markers of future outcomes and (b) targets for programmatic manipulation that may serve to close the gap in health disparities (Belsky, Moffitt, & Caspi, 2013b). Behavioral genetics is beginning to turn toward these applications, and research on body mass index (BMI) provides a ready example. Large-scale phenotypic annotation efforts have begun to link genetic variants associated with adult BMI to eating behaviors in childhood and adolescence (Abdulkadir et al., 2020; Herle et al., 2021a, 2021b). These studies have found that, by as early as age 2, a child's eating behavior may demarcate genetic risk for adult BMI. This suggests that eating habits, and possibly related health behaviors, may represent malleable outcomes through which we can mitigate the influence of genetic differences on BMI. Preliminary evidence supports this claim. Correlational research has shown that genetic effects on adult BMI are larger in individuals who live sedentary lifestyles and consume more sweetened beverages (Li et al., 2010; Qi et al., 2012). Early experimental findings point toward physical activity at age 11 as a modifiable behavior for attenuating the association between genes and BMI (Herle, Pickles, & de Stavola, 2021b).

Still, we know that not all 11-year-olds will respond equally to a behavioral intervention. This was one of the main takeaways from the HPPP that we reviewed at the beginning of this paper. Simply being exposed to an intervention does not entail how a given person will respond, for how long the effect will last, or whether it will generalize to related behaviors across development (Bailey et al., 2020; Bryan et al., 2021; Green, 2021). To improve the efficacy and reach of our treatments, we need to understand the sources of individual differences in their outcomes. We need to "be concerned with the otherwise neglected interactions between organismic and treatment variables" (Cronbach, 1957, p. 681).

Genetic causes can help. By integrating genomic data into longitudinal, experimental research designs, we can begin to answer causal questions about heterogeneity in treatment effects and mechanisms generating the fadeout, persistence, and emergence

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

of those effects later in life. A growing body of work in this area has demonstrated that responses to childhood interventions such as HPPP are sensitive to genetic variation (Albert et al., 2015; Brody et al., 2009, 2013; Kuo et al., 2019). Using the framework for genetic causation that we have advanced in this paper, we can develop more robust and comprehensive understanding of how individual differences in constitutional factors influence treatment outcomes. In particular, we can integrate whole-genome measures from family members into *two-shock designs*, which yield an estimate of the interaction of two random sources of variation to provide special insight into the (biological and environmental) contexts in which a particular cause operates (Almond, Currie, & Duque, 2018). These designs may critically advance our understanding of why particular individuals are more or less likely to respond to treatments and why particular treatment effects are more or less enduring or generalizable.

3.5. Summary

In this section, we considered interpretations of the prevailing statistical parameters used in behavior genetics – h^2 and *SNP effects*. We argued that the randomization of offspring to genotype in *meiosis* generates a natural experiment, but that genetic effects on behavior can only be considered causal when other counterfactual conditions are met. The experimental assumptions of *independence*, *sample homogeneity*, *potential exposability*, and *SUTVA* were discussed with respect to genetic causation. The takeaway was that within-family designs that leverage the natural experiment of genetic inheritance are best suited for causal inference. Guidelines for conceptualizing genetic causes were examined with respect to a dimension of causal depth: *deep causes*, that are unitary, uniform, and explanatory, and *shallow causes*, that are local, probabilistic, and causally distal. We discussed how the knowledge of genetic causes as advanced in this paper can be applied to advance second-generation aims: genomic data can improve our understanding of the etiology of complex psychological and behavioral outcomes, can facilitate the discovery of intervention and prevention targets for health-related outcomes, and can provide insight into individual differences in treatment responsivity, fadeout, and emergence.

4. Conclusions

Our motive for writing this paper was to grapple with the conceptual issues that have marked the history of behavioral genetics. To guide our discussion, we turned to philosophical and statistical thinking on the parameters for detecting and interpreting counterfactual causes. We compared the infrastructure of genetic inheritance to that of an RCT, and concluded that genetic effects conditional on the parental genotype are causal in the same sense as ATEs. To conclude, we provided some suggestions for how this knowledge can be used to facilitate scientific inquiry and maximize treatment outcomes.

Doubtless, many will take issue with the conclusions that we have drawn and the solutions that we have offered. Such responses are understandable in a field that is so richly complex and so wildly divisive. We welcome all work that earnestly engages in, challenges, questions, or explores the ideas that we have presented in this paper, insofar as it continues to think cautiously and judiciously about the meaning and applications of genomic research. Knowledge of genetic effects on human behavior will only continue to grow over the next several decades. We must chart the course for how we interpret and use this knowledge. As Dov Fox wrote, "[t]here is nothing especially menacing about knowledge on its own…[a]wareness or understanding of some subject can be troubling only when those facts are sought for bad reasons, or when such data are put to bad effects" (Fox, 2019, p. 155).

Financial support. This work was supported by grants from the Jacobs Foundation and the John Templeton Foundation. K.P.H. is a Faculty Research Associate of the Population Research Center at the University of Texas at Austin which is supported by grant P2CHD042849 from the NICHD.

Competing interest. None.

Notes

1. The American Psychological Association defines genetic determinism as "the doctrine that human … behavior and mental activity are largely (or completely) controlled by the genetic constitution of the individual and that responses to environmental influences are for the most part innately determined" (American Psychological Association, 2023). Determinism thus implies that the presence of certain genetic variants is both *necessary* and *sufficient* to produce a certain phenotype, and that this genotype–phenotype relationship is *immutable* across the range of possible environmental influences.
2. Lewis attempts to address some of these limitations by employing a *similarity metric* (Lewis, 1973a), which indexes the relative similarity between two possible worlds (e.g., a world with oxygen in the atmosphere is a more similar possible world to the current world than a world without oxygen in the atmosphere). While this approach has been lauded for its conceptual logic, its application to the actual analysis of counterfactuals has been questioned (Bowie, 1979).
3. Approximately 99.9% of DNA is identical for all humans (Collins & McKusick, 2001). That leaves only 0.1% of the genome (~2 million genetic variants) to differ across individuals. This sliver of genetic variation within human populations is the object of behavior genetics.
4. It is possible that indirect genetic effects not captured by within-family genetic analyses could in fact be causal. For example, indirect effects in which the genes that are not transmitted from parent to child are nonetheless associated with the child phenotype could be due to causal effects of parental genes on children that are mediated through parental investments: e.g., the parental genotype causes an increase in their education, which in turn changes parental behavior, such as reading aloud to the child, in a way that increases the child's education. Such a path is not explicitly captured by within-family genetic analyses, but might be said to be causal, at a remove of one generation.
5. The presence of two mutated copies of the *CFTR* gene produces an imbalance of sodium chloride to water in epithelial cells in the body. This results in the secretion of abnormally thick and sticky mucous that clogs air passages. Cardinal symptoms of cystic fibrosis include respiratory difficulties (coughing, lung infections, shortness of breath) and unusually salty sweat secretion. Diagnostic testing typically involves assessing sodium chloride levels in sweat (LeGrys, 1996).
6. Importantly, *shallow* and *deep* represent poles of a dimension of causal depth. Even paradigmatically deep causes like CF possess shallow features at some level (e.g., CF is also the product of non-unitary causes, in the sense that it is caused by two mutated copies of the CFTR gene on the seventh chromosome *and* the non-existence of a counteracting mutation *and* the existence of carbon atoms). Two mutated copies of the CFTR gene on the seventh chromosome is a deep cause in the sense that it operates in relative isolation to produce a narrow range of outcomes through an identified mechanism.

References

Abdulkadir, M., Herle, M., De Stavola, B. L., Hübel, C., Santos Ferreira, D. L., Loos, R. J., … Micali, N. (2020). Polygenic score for body mass index is associated with disordered eating in a general population cohort. *Journal of Clinical Medicine*, *9*, 1187. doi: 10.3390/jcm9041187

Abramovitch, R., Corter, C. M., & Lando, B. (1979). Sibling interaction in the home. *Child Development*, *50*, 997–1003. doi: 10.2307/1129325

Agnew-Blais, J. C., Belsky, D. W., Caspi, A., Danese, A., Moffitt, T. E., Polanczyk, G. V., … Arseneault, L. (2021). Polygenic risk and the course of attention-deficit/

hyperactivity disorder from childhood to young adulthood: Findings from a nationally representative cohort. *Journal of the American Academy of Child & Adolescent Psychiatry*, *60*, 1147–1156. doi: 10.1016/j.jaac.2020.12.033

Albert, D., Belsky, D. W., Crowley, D. M., Latendresse, S. J., Aliev, F., Riley, B., … Dodge, K. A. (2015). Can genetics predict response to complex behavioral interventions? Evidence from a genetic analysis of the Fast Track randomized control trial. *Journal of Policy Analysis and Management*, *34*, 497–518. doi: 10.1002/pam.21811

Alda, M. (2015). Lithium in the treatment of bipolar disorder: Pharmacology and pharmacogenetics. *Molecular Psychiatry*, *20*, 661–670. doi: 10.1038/mp.2015.4

Almond, D. (2006). Is the 1918 influenza pandemic over? Long-term effects of *in utero* influenza exposure in the post-1940 U.S. population. *Journal of Political Economy*, *114*, 672–712. doi: 10.1086/507154

Almond, D., Currie, J., & Duque, V. (2018). Childhood circumstances and adult outcomes: Act II. *Journal of Economic Literature*, *56*, 1360–1446. doi: 10.1257/jel.20171164

American Psychological Association. (2023). Genetic determinism. In *APA dictionary of psychology*. Retrieved March 4, 2022, from https://dictionary.apa.org/genetic-determinism

Angrist, J. D. (2004). Treatment effect heterogeneity in theory and practice. *The Economic Journal*, *114*, C52–C83. doi: 10.1111/j.0013-0133.2003.00195.x

Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., & Yeager, D. S. (2020). Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. *Psychological Science in the Public Interest*, *21*, 55–97. doi: 10.1177/1529100620915848

Bates, J. (2020). 2020 ends as one of America's most violent years in decades. *Time Magazine*. Retrieved from https://time.com/5922082/2020-gun-violence-homicides-record-year/

Batmanian, L., Ridge, J., & Worrall, S. (2011). *Biochemistry for health professionals*. Elsevier Australia.

Belfield, C. R., Nores, M., Barnett, S., & Schweinhart, L. (2006). The high/scope Perry preschool program cost–benefit analysis using data from the age-40 followup. *Journal of Human Resources*, *41*, 162–190. doi: 10.3368/jhr.XLI.1.162

Belsky, D. W., Domingue, B. W., Wedow, R., Arseneault, L., Boardman, J. D., Caspi, A., … Harris, K. M. (2018). Genetic analysis of social-class mobility in five longitudinal studies. *Proceedings of the National Academy of Sciences*, *115*, E7275–E7284. doi: 10.1073/pnas.1801238115

Belsky, D. W., & Harden, K. P. (2019). Phenotypic annotation: Using polygenic scores to translate discoveries from genome-wide association studies from the top down. *Current Directions in Psychological Science*, *28*, 82–90. doi: 10.1177/0963721418807729

Belsky, D. W., Moffitt, T. E., Baker, T. B., Biddle, A. K., Evans, J. P., Harrington, H., … Caspi, A. (2013a). Polygenic risk and the developmental progression to heavy, persistent smoking and nicotine dependence: Evidence from a 4-decade longitudinal study. *JAMA Psychiatry*, *70*, 534–542. doi: 10.1001/jamapsychiatry.2013.736

Belsky, D. W., Moffitt, T. E., & Caspi, A. (2013b). Genetics in population health science: Strategies and opportunities. *American Journal of Public Health*, *103*, S73–S83. doi: 10.2105/AJPH.2012.301139

Belsky, D. W., Moffitt, T. E., Corcoran, D. L., Domingue, B., Harrington, H., Hogan, S., … Caspi, A. (2016). The genetics of success: How single-nucleotide polymorphisms associated with educational attainment relate to life-course development. *Psychological Science*, *27*, 957–972. doi: 10.1177/0956797616643070

Belsky, D. W., Moffitt, T. E., Houts, R., Bennett, G. G., Biddle, A. K., Blumenthal, J. A., … Caspi, A. (2012). Polygenic risk, rapid childhood growth, and the development of obesity: Evidence from a 4-decade longitudinal study. *Archives of Pediatrics & Adolescent Medicine*, *166*, 515–521. doi: 10.1001/archpediatrics.2012.131

Bierut, L. J. (2011). Genetic vulnerability and susceptibility to substance dependence. *Neuron*, *69*, 618–627. doi: 10.1016/j.neuron.2011.02.015

Block, N. (1995). How heritability misleads about race. *Cognition*, *56*, 99–128. doi: 10.1016/0010-0277(95)00678-R

Bonell, C., Fletcher, A., Morton, M., Lorenc, T., & Moore, L. (2012). Realist randomised controlled trials: A new approach to evaluating complex public health interventions. *Social Science & Medicine*, *75*, 2299–2306. doi: 10.1016/j.socscimed.2012.08.032

Borges, J. L. (2018). *The garden of forking paths*. Penguin. (Original work published in 1941).

Bork, P., Dandekar, T., Diaz-Lazcoz, Y., Eisenhaber, F., Huynen, M., & Yuan, Y. (1998). Predicting function: From genes to genomes and back. *Journal of Molecular Biology*, *283*, 707–725. doi: 10.1006/jmbi.1998.2144

Bourrat, P. (2020). Causation and single nucleotide polymorphism heritability. *Philosophy of Science*, *87*, 1073–1083. doi: 10.1086/710517

Bowie, G. L. (1979). The similarity approach to counterfactuals: Some problems. *Noûs*, *13*, 477–498. doi: 10.2307/2215340

Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An expanded view of complex traits: From polygenic to omnigenic. *Cell*, *169*, 1177–1186. doi: 10.1016/j.cell.2017.05.038

Brady, H. E. (2011). *Causation and explanation in social science*. Oxford University Press. doi: 10.1093/oxfordhb/9780199604456.013.0049

Brody, G. H., Beach, S. R., Hill, K. G., Howe, G. W., Prado, G., & Fullerton, S. M. (2013). Using genetically informed, randomized prevention trials to test etiological hypotheses about child and adolescent drug use and psychopathology. *American Journal of Public Health*, *103*, S19–S24. doi: 10.2105/AJPH.2012.301080

Brody, G. H., Beach, S. R., Philibert, R. A., Chen, Y. F., Lei, M. K., Murry, V. M., & Brown, A. C. (2009). Parenting moderates a genetic vulnerability factor in longitudinal increases in youths' substance use. *Journal of Consulting and Clinical Psychology*, *77*, 1–11. doi: 10.1037/a0012996

Browning, S. R., & Browning, B. L. (2011). Haplotype phasing: Existing methods and new developments. *Nature Reviews Genetics*, *12*, 703–714. doi: 10.1038/nrg3054

Brumpton, B., Sanderson, E., Heilbron, K., Hartwig, F. P., Harrison, S., Vie, G. Å., … Davies, N. M. (2020). Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. *Nature Communications*, *11*, 3519. doi: 10.1038/s41467-020-17117-4

Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour*, *5*, 980–989.

Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., … Parkinson, H. (2019). The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*, *47*, D1005–D1012. doi: 10.1093/nar/gky1120

Cade, J. F. (1949). Lithium salts in the treatment of psychotic excitement. *Medical Journal of Australia*, *2*, 349–352.

Cardon, L. R., & Palmer, L. J. (2003). Population stratification and spurious allelic association. *The Lancet*, *361*, 598–604. doi: 10.1016/S0140-6736(03)12520-2

Carey, G. (1986). Sibling imitation and contrast effects. *Behavior Genetics*, *16*, 319–341. doi: 10.1007/BF01071314

Carroll, D. (2001). Genetic recombination. In *Encyclopedia of genetics* (pp. 841–845). Elsevier. doi: 10.1006/rwgn.2001.0543

Carroll, S. B. (2020). *A series of fortunate events: Chance and the making of the planet, life, and you*. Princeton University Press.

Cartwright, N. (1988). Regular associations and singular causes. In B. Skyrms & W. L. Harper (Eds.), *Causation, chance and credence* (pp. 79–97). Springer Netherlands. doi: 10.1007/978-94-009-2863-3_5

Cartwright, N. (1995). Precis of *Nature's Capacities and Their Measurement*. *Philosophy and Phenomenological Research*, *55*, 153–156. doi: 10.2307/2108313

Chabris, C. F., Lee, J. J., Cesarini, D., Benjamin, D. J., & Laibson, D. I. (2015). The fourth law of behavior genetics. *Current Directions in Psychological Science*, *24*, 304–312. doi: 10.1177/0963721415580430

Charlton, W. (1983). Causation and change. *Philosophy*, *58*, 143–160. doi: 10.1017/S0031819100068650

Coleman, J. R. I., Peyrot, W. J., Purves, K. L., Davis, K. A. S., Rayner, C., Choi, S. W., … Breen, G. (2020). Genome-wide gene–environment analyses of major depressive disorder and reported lifetime traumatic experiences in UK BioBank. *Molecular Psychiatry*, *25*, 1430–1446. doi: 10.1038/s41380-019-0546-6

Collins, F. S., & McKusick, V. A. (2001). Implications of the Human Genome Project for medical science. *JAMA*, *285*, 540–544. doi: 10.1001/jama.285.5.540

Connolly, S., & Heron, E. A. (2015). Review of statistical methodologies for the detection of parent-of-origin effects in family trio genome-wide association data with binary disease traits. *Briefings in Bioinformatics*, *16*, 429–448. doi: 10.1093/bib/bbu017

Corvin, A., Craddock, N., & Sullivan, P. F. (2010). Genome-wide association studies: A primer. *Psychological Medicine*, *40*, 1063–1077. doi: 10.1017/S0033291709991723

Cox, J. W., Rich, S., Chiu, A., Muyskens, J., & Ulmanu, M. (2018). More than 228,000 students have experienced gun violence at school since Columbine. *Washington Post*. Retrieved from https://www.washingtonpost.com/graphics/2018/local/school-shootings-database/

Craver, C. F., & Darden, L. (2013). *In search of mechanisms: Discoveries across the life sciences*. The University of Chicago Press.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*, 671. doi: 10.1037/h0043943

Cutting, G. R. (2015). Cystic fibrosis genetics: From molecular understanding to clinical application. *Nature Reviews Genetics*, *16*, 45–56. doi: 10.1038/nrg3849

Dar-Nimrod, I., & Heine, S. J. (2011). Genetic essentialism: On the deceptive determinism of DNA. *Psychological Bulletin*, *137*, 800–818. doi: 10.1037/a0021860

Davey Smith, G., & Ebrahim, S. (2003). "Mendelian randomization": Can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology*, *32*, 1–22. doi: 10.1093/ije/dyg070

Davies, N. M., Howe, L. J., Brumpton, B., Havdahl, A., Evans, D. M., & Davey Smith, G. (2019). Within family Mendelian randomization studies. *Human Molecular Genetics*, *28*, R170–R179. doi: 10.1093/hmg/ddz204

Dawkins, R. (2016). *The extended phenotype: The long reach of the gene*. Oxford University Press. (Original work published in 1982).

Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, *210*, 2–21. doi: 10.1016/j.socscimed.2017.12.005

Dick, D. M. (2018). Commentary for special issue of prevention science "Using genetics in prevention: Science fiction or science fact?". *Prevention Science*, *19*, 101–108. doi: 10.1007/s11121-017-0828-7

Dick, D. M., Barr, P. B., Cho, S. B., Cooke, M. E., Kuo, S. I. C., Lewis, T. J., … Su, J. (2018). Post-GWAS in psychiatric genetics: A developmental perspective on the "other" next steps. Genes, *Brain and Behavior*, *17*, e12447. doi: 10.1111/gbb.12447

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

Dolan, C. V., de Kort, J. M., van Beijsterveldt, T. C. E. M., Bartels, M., & Boomsma, D. I. (2014). GE covariance through phenotype to environment transmission: An assessment in longitudinal twin data and application to childhood anxiety. *Behavior Genetics*, *44*, 240–253. doi: 10.1007/s10519-014-9659-5

Draaisma, D. (2019). Lithium: The gripping history of a psychiatric success story. *Nature*, *572*, 584–586.

Duncan, L. E., Ostacher, M., & Ballon, J. (2019). How genome-wide association studies (GWAS) made traditional candidate gene studies obsolete. *Neuropsychopharmacology*, *44*, 1518–1523. doi: 10.1038/s41386-019-0389-5

Durkin, K., Lipsey, M. W., Farran, D. C., & Wiesen, S. E. (2022). Effects of a statewide pre-kindergarten program on children's achievement and behavior through sixth grade. *Developmental Psychology*, *58*, 470–484. doi: 10.1037/dev0001301

Eaves, L. (1976). A model for sibling effects in man. *Heredity*, *36*, 205–214. doi: 10.1038/hdy.1976.25

Elborn, J. S. (2016). Cystic fibrosis. *The Lancet*, *388*, 2519–2531. doi: 10.1016/S0140-6736(16)00576-6

Engel, G. (1977). The need for a new medical model: A challenge for biomedicine. *Science (New York, N.Y.)*, *196*, 129–136. doi: 10.1126/science.847460

Engzell, P., & Tropf, F. C. (2019). Heritability of education rises with intergenerational mobility. *Proceedings of the National Academy of Sciences*, *116*, 25386–25388. doi: 10.1073/pnas.1912998116

Evangelou, E., Warren, H. R., Mosen-Ansorena, D., Mifsud, B., Pazoki, R., Gao, H., … Caulfield, M. J. (2018). Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nature Genetics*, *50*, 1412–1425. doi: 10.1038/s41588-018-0205-x

Feldman, M. W., & Lewontin, R. C. (1975). The heritability hang-up. *Science*, *190*, 1163–1168. doi: 10.1126/science.1198102

Fisher, R. A. (1925). *Statistical methods for research workers* (11th ed. rev.). Oliver and Boyd.

Fisher, R. A. (1935). *The design of experiments*. Oliver and Boyd.

Fletcher, J. M., Wu, Y., Li, T., & Lu, Q. (2021). *Interpreting polygenic score effects in sibling analysis*. bioRxiv. doi: 10.1101/2021.07.16.452740

Fletcher, J. M., Wu, Y., Zhao, Z., & Lu, Q. (2020). *The production of within-family inequality: Insights and implications of integrating genetic data*. bioRxiv. doi: 10.1101/2020.06.06.137778

Fox, D. (2019). Subversive science. *Penn State Law Review*, *124*, 153.

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*, 156–168. doi: 10.1177/2515245919847202

Galton, F. (1869). *Hereditary genius: An inquiry into its laws and consequences*. Macmillan. doi: 10.1037/13474-000

Gaskell, G., Bard, I., Allansdottir, A., Da Cunha, R. V., Eduard, P., Hampel, J., … Zwart, H. (2017). Public views on gene editing and its uses. *Nature Biotechnology*, *35*, 1021–1023. doi: 10.1038/nbt.3958

Gelman, A., & Loken, E. (2014). The statistical crisis in science. *American Scientist*, *102*, 460–465. doi: 10.1511/2014.111.460

Glennan, S. S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, *44*, 49–71. doi: 10.1007/BF00172853

Gough, A., Stern, A. M., Maier, J., Lezon, T., Shun, T.-Y., Chennubhotla, C., … Taylor, D. L. (2017). Biologically relevant heterogeneity: Metrics and practical insights. *SLAS Discovery*, *22*, 213–237. doi: 10.1177/2472555216682725

Green, C. S. (2021). Interventions to do real-world good: Generalization and persistence. *Psychological Science in the Public Interest*, *21*, 43–49. doi: 10.1177/1529100620933847

Grotzinger, A. D., Rhemtulla, M., de Vlaming, R., Ritchie, S. J., Mallard, T. T., Hill, W. D., … Tucker-Drob, E. M. (2019). Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nature Human Behaviour*, *3*, 513–525. doi: 10.1038/s41562-019-0566-x

Gueron, J. M., & Rolston, H. (2013). *Fighting for reliable evidence*. Russell Sage.

Gurr, T. R. (1981). Historical trends in violent crime: A critical review of the evidence. *Crime and Justice*, *3*, 295–353. doi: 10.1086/449082

Harrison, P. J., Cipriani, A., Harmer, C. J., Nobre, A. C., Saunders, K., Goodwin, G. M., & Geddes, J. R. (2016). Innovative approaches to bipolar disorder and its treatment. *Annals of the New York Academy of Sciences*, *1366*, 76. doi: 10.1111/nyas.13048

Hausman, D. M. (2005). Causal relata: Tokens, types, or variables? *Erkenntnis*, *63*, 33–54. doi: 10.1007/s10670-005-0562-6

Haworth, S., Mitchell, R., Corbin, L., Wade, K. H., Dudding, T., Budu-Aggrey, A., … Timpson, N. J. (2019). Apparent latent structure within the UK BioBank sample has implications for epidemiological analysis. *Nature Communications*, *10*, 333. doi: 10.1038/s41467-018-08219-1

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science (New York, N.Y.)*, *312*, 1900–1902. doi: 10.1126/science.1128898

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics*, *94*, 114–128. doi: 10.1016/j.jpubeco.2009.11.001

Herd, P., Freese, J., Sicinski, K., Domingue, B. W., Mullan Harris, K., Wei, C., & Hauser, R. M. (2019). Genes, gender inequality, and educational attainment. *American Sociological Review*, *84*, 1069–1098. doi: 10.1177/0003122419886550

Herle, M., Abdulkadir, M., Hübel, C., Ferreira, D. S., Bryant-Waugh, R., Loos, R. J., … Micali, N. (2021a). The genomics of childhood eating behaviours. *Nature Human Behaviour*, *5*, 625–630. doi: 10.1038/s41562-020-01019-y

Herle, M., Pickles, A., & de Stavola, B. (2021b). *Can interventions mitigate genomic liability for obesity? Using causal inference based mediation analyses in genetically-sensitive studies*. Poster presented at the annual meeting of the Behavior Genetics Association.

Herrnstein, R. J., & Murray, C. A. (1996). *The bell curve: Intelligence and class structure in American life* (1st Free Press pbk. ed.). Simon & Schuster.

Hill, A. B. (2015). The environment and disease: Association or causation? *Journal of the Royal Society of Medicine*, *108*, 32–37. (Original work published in 1965). doi: 10.1177/0141076814562718

Hill, W. D., Davies, N. M., Ritchie, S. J., Skene, N. G., Bryois, J., Bell, S., … Deary, I. J. (2019). Genome-wide analysis identifies molecular systems and 149 genetic loci associated with income. *Nature Communications*, *10*, 5741. doi: 10.1038/s41467-019-13585-5

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*, 945–960. doi: 10.1080/01621459.1986.10478354

Holland, P. W. (1988). Causal inference, path analysis, and recursive structural equations models. In Clogg, C. (Ed.), *Sociological methodology* (pp. 449–484). American Sociological Association. doi: 10.2307/271055

Howe, L. J., Nivard, M. G., Morris, T. T., Hansen, A. F., Rasheed, H., & Cho, Y., … Within Family Consortium (2021). *Within-sibship GWAS improve estimates of direct genetic effects*. bioRxiv. doi: 10.1101/2021.03.05.433935

Hudgens, M. G., & Halloran, M. E. (2008). Toward causal inference with interference. *Journal of the American Statistical Association*, *103*, 832–842. doi: 10.1198/016214508000000292

Hulswit, M. (2002). *From cause to causation: A Peircean perspective*. Springer Netherlands.

Hume, D. (1999). *An enquiry concerning human understanding* (T. L. Beauchamp, Ed.). Oxford University Press. (Original work published in 1748).

Jaffee, S. R., & Price, T. S. (2007). Gene–environment correlations: A review of the evidence and implications for prevention of mental illness. *Molecular Psychiatry*, *12*, 432–442. doi: 10.1038/sj.mp.4001950

Janzen, T. (2021). Autosomal DNA testing comparison chart. In *International Society of Genetic Genealogy Wiki*. Retrieved from https://isogg.org/wiki/Autosomal_DNA_testing_comparison_chart

Jensen, A. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, *39*, 1–123. doi: 10.17763/haer.39.1.l3u15956627424k7

Jo, B., & Muthén, B. (2001). Modeling of intervention effects with noncompliance: A latent variable modeling approach for randomized trials. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 57–87). Erlbaum.

Juvenal (1769). *The satires of Juvenal translated: With explanatory and classical notes, relating to the laws and customs of the Greeks and Romans*. George Faulkner.

Kapogiannis, D., Barbey, A. K., Su, M., Zamboni, G., Krueger, F., & Grafman, J. (2009). Cognitive and neural foundations of religious belief. *Proceedings of the National Academy of Sciences*, *106*, 4876–4881. doi: 10.1073/pnas.0811717106

Karlsson Linnér, R., Biroli, P., Kong, E., Meddens, S. F. W., Wedow, R., Fontana, M. A., … Beauchamp, J. P. (2019). Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature Genetics*, *51*, 245–257. doi: 10.1038/s41588-018-0309-3

Karlsson Linnér, R., Mallard, T. T., Barr, P. B., Sanchez-Roige, S., Madole, J. W., Driver, M. N., … Dick, D. M. (2021). Multivariate analysis of 1.5 million people identifies genetic associations with traits related to self-regulation and addiction. *Nature Neuroscience*, *24*, 1367–1376. doi: 10.1038/s41593-021-00908-3

Keller, E. F. (2010). *The mirage of a space between nature and nurture*. Duke University Press.

Kendler, K. S. (2005). "A gene for…": The nature of gene action in psychiatric disorders. *American Journal of Psychiatry*, *162*, 1243–1252. doi: 10.1176/appi.ajp.162.7.1243

Kendler, K. S., Chen, X., Dick, D., Maes, H., Gillespie, N., Neale, M. C., & Riley, B. (2012). Recent advances in the genetic epidemiology and molecular genetics of substance use disorders. *Nature Neuroscience*, *15*, 181–189. doi: 10.1038/nn.3018

Kent, D. M., Rothwell, P. M., Ioannidis, J. P. A., Altman, D. G., & Hayward, R. A. (2010). Assessing and reporting heterogeneity in treatment effects in clinical trials: A proposal. *Trials*, *11*, 85. doi: 10.1186/1745-6215-11-85

Kevles, D. J. (1995). *In the name of eugenics: Genetics and the uses of human heredity* (1st Harvard University Press pbk. ed.). Harvard University Press.

Khatri, P., Sirota, M., & Butte, A. J. (2012). Ten years of pathway analysis: Current approaches and outstanding challenges. *PLoS Computational Biology*, *8*, e1002375. doi: 10.1371/journal.pcbi.1002375

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., Young, A. I., Thorgeirsson, T. E., … Stefansson, K. (2018). The nature of nurture: Effects of parental genotypes. *Science (New York, N.Y.)*, *359*, 424–428. doi: 10.1126/science.aan6877

Kuo, S. I., Salvatore, J. E., Aliev, F., Ha, T., Dishion, T. J., & Dick, D. M. (2019). The family check-up intervention moderates polygenic influences on long-term alcohol outcomes: Results from a randomized intervention trial. *Prevention Science*, *20*, 975–985. doi: 10.1007/s11121-019-01024-2

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

Lahey, B. B., Waldman, I. D., & McBurnett, K. (1999). Annotation: The development of antisocial behavior: An integrative causal model. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *40*, 669–682.

Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and thought* (2nd ed., pp. 202–251). Cambridge University Press. doi: 10.1017/CBO9781139173865.013

LaPierre, N., Taraszka, K., Huang, H., He, R., Hormozdiari, F., & Eskin, E. (2021). Identifying causal variants by fine mapping across multiple studies. *PLoS Genetics*, *17*, e1009733.

Lawlor, D. A., Tilling, K., & Davey Smith, G. (2017). Triangulation in aetiological epidemiology. *International Journal of Epidemiology*, *45*, 1866–1886. doi: 10.1093/ije/dyw314

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., … Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, *50*, 1112–1121. doi: 10.1038/s41588-018-0147-3

LeGrys, V. A. (1996). Sweat testing for the diagnosis of cystic fibrosis: Practical considerations. *The Journal of Pediatrics*, *129*, 892–897. doi: 10.1016/S0022-3476(96)70034-3

Lewis, D. (1973a). *Counterfactuals* (Rev. ed.). Blackwell.

Lewis, D. (1973b). Causation. *The Journal of Philosophy*, *70*, 556–567. doi: 10.2307/2025310

Lewontin, R. C. (2006). The analysis of variance and the analysis of causes. *International Journal of Epidemiology*, *35*, 520–525. (Original work published in 1974). doi: 10.1093/ije/dyl062

Li, S., Zhao, J. H., Luan, J. A., Ekelund, U., Luben, R. N., Khaw, K. T., … Loos, R. J. (2010). Physical activity attenuates the genetic predisposition to obesity in 20,000 men and women from EPIC-Norfolk prospective population study. *PLoS Medicine*, *7*, e1000332. doi: 10.1371/journal.pmed.1000332

Locke, J. (1997). *An essay concerning human understanding* (R. S. Woolhouse, Ed.). Penguin. (Original work published in 1690).

Loehlin, J. C. (2009). History of behavior genetics. In Kim, Y. K. (Ed.), *Handbook of behavior genetics* (pp. 3–11). Springer. doi: 10.1007/978-0-387-76727-7_1

Lowe, J. (2015). I don't believe in God, but I believe in Lithium. *The New York Times*, 25. Retrieved from https://www.nytimes.com/2015/06/28/magazine/i-dont-believe-in-god-but-i-believe-in-lithium.html

Lynch, K. E., Morandini, J. S., Dar-Nimrod, I., & Griffiths, P. E. (2019). Causal reasoning about human behavior genetics: Synthesis and future directions. *Behavior Genetics*, *49*, 221–234. doi: 10.1007/s10519-018-9909-z

Mackie, J. L. (1965). Causes and conditions. *American Philosophical Quarterly*, *2*, 245–264.

Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., … Kenny, E. E. (2017). Human demographic history impacts genetic risk prediction across diverse populations. *The American Journal of Human Genetics*, *100*, 635–649. doi: 10.1016/j.ajhg.2017.03.004

Martin, N. (2018). Getting to the genetic and environmental roots of educational inequality. *Npj Science of Learning*, *3*, 4. doi: 10.1038/s41539-018-0021-1

Meehl, P. E. (1972). Specific genetic etiology, psychodynamics, and therapeutic nihilism. *International Journal of Mental Health*, *1*, 10–27. doi: 10.1080/00207411.1972.11448562

Mertens, J., Wang, Q. W., Kim, Y., Diana, X. Y., Pham, S., Yang, B., … Yao, J. (2015). Differential responses to lithium in hyperexcitable neurons from patients with bipolar disorder. *Nature*, *527*, 95–99. doi: 10.1038/nature15526

Meyer, M. N., Turley, P., & Benjamin, D. J. (2020). Response to Charles Murray on polygenic scores. *Medium*. Retrieved from https://medium.com/@michellenmeyer/response-to-charles-murray-on-polygenic-scores-e768cf145cc

Mill, J. S. (2002). *A system of logic*. University Press of the Pacific. (Original work published in 1843).

Miller, G. A. (2010). Mistreating psychology in the decades of the brain. *Perspectives on Psychological Science*, *5*, 716–743. doi: 10.1177/1745691610388774

Mills, M. C., & Rahal, C. (2019). A scientometric review of genome-wide association studies. *Communications Biology*, *2*, 9. doi: 10.1038/s42003-018-0261-x

Moffitt, T. E. (2006). Life-course-persistent versus adolescence-limited antisocial behavior. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology: Risk, disorder, and adaptation* (pp. 570–598). Wiley. doi: 10.1002/9780470939406.ch15

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research & Perspective*, *2*, 201–218. doi: 10.1207/s15366359mea0204_1

Moscati, A., Verhulst, B., McKee, K., Silberg, J., & Eaves, L. (2018). Cross-lagged analysis of interplay between differential traits in sibling pairs: Validation and application to parenting behavior and ADHD symptomatology. *Behavior Genetics*, *48*, 22–33. doi: 10.1007/s10519-017-9882-y

Mostafavi, H., Harpak, A., Agarwal, I., Conley, D., Pritchard, J. K., & Przeworski, M. (2020). Variable prediction accuracy of polygenic scores within an ancestry group. *eLife*, *9*, e48376. doi: 10.7554/eLife.48376

Munafò, M. R. (2006). Candidate gene studies in the 21st century: Meta-analysis, mediation, moderation. Genes, *Brain and Behavior*, *5*, 3–8. doi: 10.1111/j.1601-183X.2006.00188.x

Munafò, M. R., & Flint, J. (2009). Replication and heterogeneity in gene×environment interaction studies. *The International Journal of Neuropsychopharmacology*, *12*, 727–729. doi: 10.1017/S1461145709000479

Murray, C. A. (2020). *Human diversity: The biology of gender, race, and class* (1st ed.). Twelve.

Nachman, M. W. (2002). Variation in recombination rate across the genome: Evidence and implications. *Current Opinion in Genetics & Development*, *12*, 657–663. doi: 10.1016/S0959-437X(02)00358-1

Nagy, A., Perrimon, N., Sandmeyer, S., & Plasterk, R. (2003). Tailoring the genome: The power of genetic approaches. *Nature Genetics*, *33*, 276–284. doi: 10.1038/ng1115

National Human Genome Research Institute (NHGRI) (2022). *Talking glossary of genetic terms*. Retrieved from https://www.genome.gov/genetics-glossary

Nielsen, J. B., Thorolfsdottir, R. B., Fritsche, L. G., Zhou, W., Skov, M. W., Graham, S. E., … Willer, C. J. (2018). BioBank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nature Genetics*, *50*, 1234–1239. doi: 10.1038/s41588-018-0171-3

Obama, B. (2013). *State of the union address* 2013 [Transcript]. Retrieved from https://obamawhitehouse.archives.gov/the-press-office/2013/02/12/remarks-president-state-union-address

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., … Benjamin, D. J. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, *533*, 539–542. doi: 10.1038/nature17671

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys*, *3*, 96–146. doi: 10.1214/09-SS057

Pearl, J. (2010). The mathematics of causal relations. In Shrout, P. E. (Ed.), *Causality and psychopathology* (pp. 47–65). Oxford University Press.

Pearl, J. (2018). Theoretical impediments to machine learning with seven sparks from the causal revolution. *arXiv*. Retrieved from https://arxiv.org/abs/1801.04016

Perkel, J. (2008). SNP genotyping: Six technologies that keyed a revolution. *Nature Methods*, *5*, 447–453. doi: 10.1038/nmeth0508-447

Phillips, M. S., Lawrence, R., Sachidanandam, R., Morris, A. P., Balding, D. J., Donaldson, M. A., … Cardon, L. R. (2003). Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots. *Nature Genetics*, *33*, 382–387. doi: 10.1038/ng1100

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2013). *Behavioral genetics: A primer* (6th ed.). Worth.

Plomin, R., DeFries, J. C., & Loehlin, J. C. (1977). Genotype–environment interaction and correlation in the analysis of human behavior. *Psychological Bulletin*, *84*, 309–322. doi: 10.1037/0033-2909.84.2.309

Plomin, R., & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics*, *19*, 148–159. doi: 10.1038/nrg.2017.104

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, *38*, 904–909. doi: 10.1038/ng1847

Qi, Q., Chu, A. Y., Kang, J. H., Jensen, M. K., Curhan, G. C., Pasquale, L. R., … Qi, L. (2012). Sugar-sweetened beverages and genetic risk of obesity. *New England Journal of Medicine*, *367*, 1387–1396. doi: 10.1056/NEJMoa1203039

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., … Koellinger, P. D. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science (New York, N.Y.)*, *340*, 1467–1471. doi: 10.1126/science.1235488

Rimfeld, K., Krapohl, E., Trzaskowski, M., Coleman, J. R. I., Selzam, S., Dale, P. S., … Plomin, R. (2018). Genetic influence on social outcomes during and after the Soviet era in Estonia. *Nature Human Behaviour*, *2*, 269–275. doi: 10.1038/s41562-018-0332-5

Rosenbaum, P. R. (2007). Interference between units in randomized experiments. *Journal of the American Statistical Association*, *102*, 191–200. doi: 10.1198/016214506000001112

Ross, L. N. (2015). *Causal control: A rationale for causal selection*. *PhilSci Archive*. Retrieved from http://philsci-archive.pitt.edu/id/eprint/11743

Ross, L. N. (2018). The doctrine of specific etiology. *Biology & Philosophy*, *33*, 37. doi: 10.1007/s10539-018-9647-x

Ross, L. N. (2019). *Explanation in contexts of causal complexity: Lessons from psychiatric genetics*. *PhilSci Archive*. Retrieved from http://philsci-archive.pitt.edu/id/eprint/15728

Rottman, B. M., & Hastie, R. (2014). Reasoning about causal relationships: Inferences on causal networks. *Psychological Bulletin*, *140*, 109–139. doi: 10.1037/a0031903

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*, 688–701. doi: 10.1037/h0037350

Rubin, D. B. (1980). Comment on: "Randomization analysis of experimental data in the fisher randomization test" by D. Basu. *Journal of the American Statistical Association*, *75*, 591–593. doi: 10.2307/2287653

Rubin, D. B. (1986). Comment: Which ifs have causal answers. *Journal of the American Statistical Association*, *81*, 961–962. doi: 10.1080/01621459.1986.10478355

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, *100*, 322–331. doi: 10.1198/016214504000001880

Salganik, M. J., Lundberg, I., Kindel, A. T., Ahearn, C. E., Al-Ghoneim, K., Almaatouq, A., … McLanahan, S. (2020). Measuring the predictability of life outcomes with a scientific mass collaboration. *Proceedings of the National Academy of Sciences*, *117*, 8398–8403. doi: 10.1073/pnas.1915006117

Samek, D. R., McGue, M., Keyes, M., & Iacono, W. G. (2015). Sibling facilitation mediates the association between older and younger sibling alcohol use in late adolescence. *Journal of Research on Adolescence*, *25*, 638–651. doi: 10.1111/jora.12154

Samorodnitsky, D. (2020). A study tried to use genetics to explain why people are poor. *Vice*. Retrieved from https://www.vice.com/en/article/jgepv8/a-study-tried-to-use-genetics-to-explain-why-people-are-poor

Santos, R., Linker, S. B., Stern, S., Mendes, A. P., Shokhirev, M. N., Erikson, G., … Gage, F. H. (2021). Deficient LEF1 expression is associated with lithium resistance and hyperexcitability in neurons derived from bipolar disorder patients. *Molecular Psychiatry*, *26*, 2440–2456. doi: 10.1038/s41380-020-00981-3

Scarr, S., & McCartney, K. (1983). How people make their own environments: A theory of genotype environment effects. *Child Development*, *54*, 424–435. doi: 10.1111/j.1467-8624.1983.tb03884.x

Schaefer, G. B., & Thompson, J. N. (2014). *Medical genetics: An integrated approach*. McGraw-Hill.

Schneider, B., & Bradford, L. (2020). What we are learning about fade-out of intervention effects: A commentary. *Psychological Science in the Public Interest*, *21*, 50–54. doi: 10.1177/1529100620935793

Schweinhart, L. J., Barnes, H. V., & Weikart, D. P. (1993). *Significant benefits: The High/Scope Perry preschool study through age 27*. Monographs of the High/Scope Educational Research Foundation, No. 10. High/Scope Press.

Scott, J. (1988). Social network analysis. *Sociology*, *22*, 109–127. doi: 10.1177/0038038588022001007

Selzam, S., Ritchie, S. J., Pingault, J.-B., Reynolds, C. A., O'Reilly, P. F., & Plomin, R. (2019). Comparing within- and between-family polygenic score prediction. *The American Journal of Human Genetics*, *105*, 351–363. doi: 10.1016/j.ajhg.2019.06.006

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, *25*, 289–310.

Sobel, M. E. (2006). What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference. *Journal of the American Statistical Association*, *101*, 1398–1407. doi: 10.1198/016214506000000636

Spencer, C. C. A., Deloukas, P., Hunt, S., Mullikin, J., Myers, S., Silverman, B., … McVean, G. (2006). The influence of recombination on human genetic diversity. *PLoS Genetics*, *2*, e148. doi: 10.1371/journal.pgen.0020148

Stein, Z., Susser, M., Saenger, G., & Marolla, F. (1972). Nutrition and mental performance: Prenatal exposure to the Dutch famine of 1944–1945 seems not related to mental performance at age 19. *Science (New York, N.Y.)*, *178*, 708–713.

Stern, S., Santos, R., Marchetto, M. C., Mendes, A. P. D., Rouleau, G. A., Biesmans, S., … Gage, F. H. (2018). Neurons derived from patients with bipolar disorder divide into intrinsically different sub-populations of neurons, predicting the patients' responsiveness to lithium. *Molecular Psychiatry*, *23*, 1453–1465. doi: 10.1038/mp.2016.260

Sugrue, L. P., & Desikan, R. S. (2019). What are polygenic scores and why are they important? *JAMA*, *321*, 1820–1821. doi: 10.1001/jama.2019.3893

Thomas, J. G., & Sharp, P. B. (2019). Mechanistic science: A new approach to comprehensive psychopathology research that relates psychological and biological phenomena. *Clinical Psychological Science*, *7*, 196–215. doi: 10.1177/2167702618810223

Tielbeek, J. J., Johansson, A., Polderman, T. J. C., Rautiainen, M.-R., Jansen, P., Taylor, M., … Posthuma, D. (2017). Genome-wide association studies of a broad spectrum of antisocial behavior. *JAMA Psychiatry*, *74*, 1242–1250. doi: 10.1001/jamapsychiatry.2017.3069

Tropf, F. C., Lee, S. H., Verweij, R. M., Stulp, G., van der Most, P. J., de Vlaming, R., … Mills, M. C. (2017). Hidden heritability due to heterogeneity across seven populations. *Nature Human Behaviour*, *1*, 757–765. doi: 10.1038/s41562-017-0195-1

Turkheimer, E. (1998). Heritability and biological explanation. *Psychological Review*, *105*, 782–791. doi: 10.1037/0033-295X.105.4.782-791

Turkheimer, E. (2011). Commentary: Variation and causation in the environment and genome. *International Journal of Epidemiology*, *40*, 598–601. doi: 10.1093/ije/dyq147

Turkheimer, E., & Gottesman, I. I. (1996). Simulating the dynamics of genes and environment in development. *Development and Psychopathology*, *8*, 667–677. doi: 10.1017/S0954579400007355

Uchiyama, R., Spicer, R., & Muthukrishna, M. (2021). Cultural evolution of genetic heritability. *Behavioral and Brain Sciences*, *45*, e152. doi: 10.1017/S0140525X21000893

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., … Zhu, X. (2001). The sequence of the human genome. *Science*, *291*, 1304–1351. doi: 10.1126/science.1058040

Visscher, P. M., Hill, W. G., & Wray, N. R. (2008). Heritability in the genomics era—concepts and misconceptions. *Nature Reviews Genetics*, *9*, 255–266.

Volkmann, C., Bschor, T., & Köhler, S. (2020). Lithium treatment over the lifespan in bipolar disorders. *Frontiers in Psychiatry*, *11*, 377. doi: 10.3389/fpsyt.2020.00377

Wang, N., Akey, J. M., Zhang, K., Chakraborty, R., & Jin, L. (2002). Distribution of recombination crossovers and the origin of haplotype blocks: The interplay of population history, recombination, and mutation. *The American Journal of Human Genetics*, *71*, 1227–1234. doi: 10.1086/344398

Warner, J. (2018). Why we shouldn't embrace the genetics of education. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/blogs/just-visiting/why-we-shouldnt-embrace-genetics-education

Watanabe, K., Mirkov, M. U., de Leeuw, C. A., van den Heuvel, M. P., & Posthuma, D. (2019). Genetic mapping of cell type specificity for complex traits. *Nature Communications*, *10*, 1–13. doi: 10.1038/s41467-019-11181-1

Watanabe, K., Taskesen, E., van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nature Communications*, *8*, 1826. doi: 10.1038/s41467-017-01261-5

Waters, C. K. (2007). Causes that make a difference. *The Journal of Philosophy*, *104*, 551–579. doi: 10.5840/jphil2007104111

White, M. J., Yaspan, B. L., Veatch, O. J., Goddard, P., Risse-Adams, O. S., & Contreras, M. G. (2019). Strategies for pathway analysis using GWAS and WGS data. *Current Protocols in Human Genetics*, *100*, e79. doi: 10.1002/cphg.79

Woodward, J. (2002). What is a mechanism? A counterfactual account. *Philosophy of Science*, *69*, S366–S377. doi: 10.1086/341859

Woodward, J. (2005). *Making things happen: A theory of causal explanation*. Oxford University Press.

Woodward, J. (2010). Causation in biology: Stability, specificity, and the choice of levels of explanation. *Biology & Philosophy*, *25*, 287–318. doi: 10.1007/s10539-010-9200-z

Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M., & Price, A. L. (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nature Genetics*, *46*, 100–106. doi: 10.1038/ng.2876

Yang, J., Zeng, J., Goddard, M. E., Wray, N. R., & Visscher, P. M. (2017). Concepts, estimation and interpretation of SNP-based heritability. *Nature Genetics*, *49*, 1304–1310. doi: 10.1038/ng.3941

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*, 1100–1122.

Young, A. I., Benonisdottir, S., Przeworski, M., & Kong, A. (2019). Deconstructing the sources of genotype–phenotype associations in humans. *Science*, *365*, 1396–1400. doi: 10.1126/science.aax3710

Young, A. I., Frigge, M. L., Gudbjartsson, D. F., Thorleifsson, G., Bjornsdottir, G., Sulem, P., … Kong, A. (2018). Relatedness disequilibrium regression estimates heritability without environmental bias. *Nature Genetics*, *50*, 1304–1310. doi: 10.1038/s41588-018-0178-9

Zaidi, A. A., & Mathieson, I. (2020). Demographic history mediates the effect of stratification on polygenic scores. *eLife*, *9*, e61548. doi: 10.7554/eLife.61548

Open Peer Commentary

Mechanistic understanding of individual outcomes: Challenges and alternatives to genetic designs

Olesya Bondarenko

Department of History and Philosophy of Science, St John's College, University of Cambridge, Cambridge, UK
ob317@cam.ac.uk
https://www.hps.cam.ac.uk/directory/bondarenko

doi:10.1017/S0140525X22002126, e183

Abstract

I argue that advancing "second-generation" or mechanistic causal knowledge of individual outcomes requires a comprehensive research programme that uses a variety of different methods in addition to the ones described in the paper under discussion. I also highlight that environment-focused approaches can be as instrumental in identifying potential phenotypic causes as gene-focused approaches.

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

Behavioural genetic methods are gaining traction in various areas of social science. It has been argued that some of these methods, particularly genome-wide association studies (GWAS) and polygenic scores based on them, will generate new causal insights into individual outcomes such as educational attainment (e.g., Freese, 2018; Harden, 2021; Liu & Guo, 2016). Madole & Harden (M&H) offer a theoretical framework for these arguments by sketching a two-tiered view of causal investigation in behavioural genetics that mirrors the commonly used distinction between difference-making and mechanistic types of inquiry (see Tabery [2014] for a philosophical discussion of this distinction in the context of behavioural genetics). In what follows, I will not evaluate the arguments concerning the first tier (including the proposed analogy between a within-family study of polygenic scores and a randomised controlled trial) but will focus on the second tier, which the authors term "second-generation causal knowledge" and which may also be called mechanistic knowledge or mechanistic understanding.

Let us assume one has successfully identified some single-nucleotide polymorphisms (SNPs) linked with causal genetic variants. In the authors' view, once that is done, a goldmine of second-generation causal discoveries (by which they mean discoveries of higher-level causal mechanisms) awaits the researcher who is keen to explore the processes through which genetic effects manifest. The mechanisms' supposed components include both phenotypic traits and environmental factors that contribute to individual outcomes and could serve as intervention targets. The identification of causal mechanisms is sometimes seen as the only strong rationale for genetically informed inquiry in the policy-oriented social sciences. According to Cesarini and Visscher (2017), "it is only to the extent that genetic information makes it possible to tailor more effective interventions that genetic data may be a useful supplement to systems already in place" (p. 3). At the same time, mechanism elucidation is not necessarily guaranteed or even significantly facilitated by "first-generation" genetic findings, contrary to what M&H seem to suggest. Establishing a causal link between a phenotypic or environmental variable and an outcome presents a significant challenge of its own, requiring a mix of observational and experimental evidence generated by a variety of methods. Top-down approaches such as "phenotypic annotation" (Belsky & Harden, 2019) that are mentioned in the paper can help identify networks of phenotypic and environmental correlates, but disentangling the causal relations within those networks – and pinpointing suitable intervention targets – is a task that goes beyond a simple mapping of associations. Some techniques for causal inference about phenotypic mediators based on observational genetic data have been described elsewhere (e.g., Briley, Livengood, & Derringer, 2018; Pingault et al., 2018) but face several significant challenges, including widespread pleiotropy. M&H briefly describe "integrating genomic data into longitudinal, experimental research designs" (target article, sect. 3.4, para. 7) as a way of meeting the demands of causal inference at this stage of investigation but do not articulate a well-developed research programme with a clear added value for genetic methods.

That brings us to another question – why should one use genetic information as a starting point to elucidate causally relevant phenotypic and environmental factors for a particular outcome? Genetic data are not the only possible handle on the kind of phenotypic characteristics and processes scientists are interested in (especially given the adage that all traits result from a combination of genetic and environmental factors). Interestingly enough, "first-generation" knowledge about environmental causes can likewise be used (and has been used) as a springboard for mechanistic inquiry, even probing into some of the same causal intermediaries as genetically informed research, including cognitive or psychological characteristics of individuals. For instance, it is known that socioeconomic status explains a large proportion of the variance in educational attainment in different societies (see e.g., Eriksson, Lindvall, Helenius, & Ryve, 2021). Scientists have linked the effects of socioeconomic status on educational attainment with differences in phenotypes such as executive function (Hackman, Gallop, Evans, & Farah, 2015). This demonstrates that once we arrive at a solid understanding that a particular environmental factor is an important difference maker, it can then be used to investigate possible causal pathways; genetic knowledge is not unique in this sense.

In this context, one worry an advocate of genetic methods might have is that the observed effects of environmental factors such as socioeconomic background could also reflect genetic differences between individuals and therefore fall short of the standards for causal inference. However, it is possible to control for these differences in order to arrive at more accurate estimates of environmental influence. For instance, Kendler, Turkheimer, Ohlsson, Sundquist, and Sundquist (2015) have shown in an adoption study of siblings that being adopted into a family with a higher socioeconomic status generated significant advantages in terms of measured IQ after controlling for genetic factors. This suggests a more limited albeit important role for genetic tools, including polygenic scores: as controls in the study of environmental variables. Even though this application is not without its pitfalls (see Akimova, Breen, Brazel, & Mills [2021] on the potential for introducing bias), when applied with care, genetic controls may help address some of the worries that the "first-generation" knowledge of environmental factors does not meet a stringent epistemic standard.

In summary, "second-generation" goals of causal inquiry in the context of human behaviour cannot be achieved by genetic methods alone, nor do genetically informed research designs provide the only possible path towards a mechanistic understanding. Therefore, it would be desirable to clearly situate these designs within the wider disciplinary and methodological terrain, indicating how they relate to the other known ways of generating the epistemic goods that are being sought.

Financial support. This work was supported by funding from Cambridge Commonwealth, European & International Trust and St. John's College, University of Cambridge.

Competing interest. None.

References

Akimova, E. T., Breen, R., Brazel, D. M., & Mills, M. C. (2021). Gene–environment dependencies lead to collider bias in models with polygenic scores. *Scientific Reports*, *11*(1), 1–9.

Belsky, D. W., & Harden, K. P. (2019). Phenotypic annotation: Using polygenic scores to translate discoveries from genome-wide association studies from the top down. *Current Directions in Psychological Science*, *28*(1), 82–90.

Briley, D. A., Livengood, J., & Derringer, J. (2018). Behaviour genetic frameworks of causal reasoning for personality psychology. *European Journal of Personality*, *32*(3), 202–220.

Cesarini, D., & Visscher, P. M. (2017). Genetics and educational attainment. *Npj Science of Learning*, *2*(1), 4.

Eriksson, K., Lindvall, J., Helenius, O., & Ryve, A. (2021). Socioeconomic status as a multidimensional predictor of student achievement in 77 societies. *Frontiers in Education*, *6*(6), 1–10.

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

Freese, J. (2018). The arrival of social science genomics. *Contemporary Sociology*, *47*(5), 524–536.
Hackman, D. A., Gallop, R., Evans, G. W., & Farah, M. J. (2015). Socioeconomic status and executive function: Developmental trajectories and mediation. *Developmental Science*, *18*(5), 686–702.
Harden, K. P. (2021). *The genetic lottery: Why DNA matters for social equality*. Princeton University Press.
Kendler, K. S., Turkheimer, E., Ohlsson, H., Sundquist, J., & Sundquist, K. (2015). Family environment and the malleability of cognitive ability: A Swedish national home-reared and adopted-away cosibling control study. *Proceedings of the National Academy of Sciences*, *112*(15), 4612–4617.
Liu, H., & Guo, G. (2016). Opportunities and challenges of big data for the social sciences: The case of genomic data. *Social Science Research*, *59*, 13–22.
Pingault, J.-B., O'Reilly, P. F., Schoeler, T., Ploubidis, G. B., Rijsdijk, F., & Dudbridge, F. (2018). Using genetic data to strengthen causal inference in observational research. *Nature Reviews Genetics*, *19*(9), 566–580.
Tabery, J. (2014). *Beyond versus: The struggle to understand the interaction of nature and nurture*. MIT Press.

Human genomic data have different statistical properties than the data of randomised controlled trials

Mirjam J. Borger, Franz J. Weissing and Eva Boon

Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands
m.j.borger@rug.nl; f.j.weissing@rug.nl; e.boon@rug.nl
https://www.marmgroup.eu/

doi:10.1017/S0140525X22002229, e184

Abstract

Madole & Harden argue that the Mendelian reshuffling of genes and genomes is analogous to randomised controlled trials. We are not convinced by their arguments. First, their recipe for meeting the demands on randomised experiments is inherently inconsistent. Second, disequilibrium across chromosomes conflicts with their assumption of statistical independence. Third, the genome-wide association study (GWAS) method has many pitfalls, including low repeatability.

Madole & Harden (M&H) attempt to unravel the role of heredity in human behaviour by arguing that the methods of causal analysis can be applied to behavioural genetic data, thus establishing causal links between genes and behaviour. Their key argument is that "within-family genetic effects represent the product of a counterfactual comparison in the same way as average treatment effects from randomised controlled trials" (target article, abstract). Based on this argument, they "advance a framework for identifying, interpreting, and applying causal effects of genes on human behavior" (target article, abstract). While we agree with the authors that human behaviour genetics needs a sound foundation, we see at least three reasons why their proposed framework is not suitable for providing such a foundation.

The first reason is the inherent inconsistency of the proposed approach. M&H discuss whether and when behavioural genetic experiments meet four critical demands on randomised experiments. They argue that the first three demands (independence, sample homogeneity, potential exposability) can be met if the analysis is based on sibling studies, where siblings grow up in a common environment. In contrast, the fourth demand, SUTVA (stable unit treatment value assumption), requires that the siblings do not affect each others' behaviour, that is, grow up in different environments. The fourth demand (growing up in different environments) is contradictory to the first three demands (growing up in a common environment). Thus, at least one of the demands will be violated in *any* genetic data set. Obviously, this undermines M&H's argument that within-family genetic effects are comparable to the outcome of randomised controlled trials.

The second reason is an unfounded extrapolation from single-gene to genome-wide causation. The key argument in the target article is that Mendelian inheritance has similar properties as the randomisation procedure of controlled trials. Mendel's rules, however, apply to single genes or unlinked pairs of genes, while M&H are mainly interested in the causal analysis of genome-wide association studies (GWASs), where thousands of single-nucleotide polymorphisms (SNPs) are considered simultaneously.

M&H are aware of this problem, in which the physical linkage of genes on a chromosome results in the co-inheritance of alleles at linked loci and subsequent correlations across loci. Consequently, they propose to focus on "a set of alleles that are all in high linkage disequilibrium with each other (but not in linkage disequilibrium with other alleles)" (target article, sect. 3.2.1, para. 6). In this approach, it is crucial to identify such sets of alleles. If the physical linkage of gene loci would be the sole (or most important) cause of linkage disequilibrium, the proposed method might be feasible, as the SNPs used in GWASs provide a physical linkage map, allowing to identify chromosomal regions that are closely linked. However, the term "linkage disequilibrium" is misleading. It suggests that "disequilibrium" (as statistical associations across loci) is mainly caused by physical linkage. Yet, factors like natural and sexual selection, non-random mating, genetic drift, or gene flow can create considerable disequilibrium at *unlinked* loci, such as loci on different chromosomes (Hedrick, 2005). Alleles at different loci can, for example, get associated through selection if they produce a high-fitness genotype in combination (but not on their own).

Theoretical considerations suggest that such "epistatic effects" (statistical interactions between genotypes at two or more loci) are common. For example, the evolution of female preferences in sexual selection largely relies on the build-up of disequilibrium between sender and receiver genes (Kuijper, Pen, & Weissing, 2012). Regulatory networks (such as gene-regulatory networks, metabolic networks, or the immune network) are another important class of examples, as a large percentage of human genes are involved in such networks (Chatterjee & Ahituv, 2017). Genes underlying a regulatory network are functionally linked (through selection on the operation of the network) in intricate and unpredictable ways (Van Gestel & Weissing, 2016, 2018), and their epistatic interaction will likely result in linkage disequilibrium (even in the absence of physical linkage).

Controlled crossing experiments in animals indeed confirm ample disequilibrium caused by epistatic effects (Flint & Mackay, 2009; Mackay, 2014). Such experiments cannot be conducted on humans, but likely epistasis is common in our species too. The problem is that epistasis, and its associated disequilibrium, tends to remain hidden in GWASs (Mackay, 2014).

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

This implies that a major source of statistical dependence remains hidden to the researcher, making it almost impossible to correct for linkage disequilibrium in the way suggested by M&H.

The third reason is based on previously documented pitfalls of the GWAS method. M&H have high expectations regarding the GWAS method, while this method is heavily criticised in other branches of genetics because of its low repeatability and its tendency to produce false positives (e.g., Marjoram, Zubair, & Nuzhdin, 2014; Zhou et al., 2020; Zuk, Hechter, Sunyaev, & Lander, 2012). Low repeatability is a major problem, as it either indicates the limited ability of these studies to generalise (i.e., big differences between study populations in how genes cause behaviour) or that most results are actually artefacts of the model (false positives). In the GWAS method it is possible to set the sensitivity of models. Yet, this is a complicated trade-off, especially when using the method to find many genes with weak effects. When the sensitivity is low, only genes with strong effects can be found, which might result in a bias, as possibly important other genes (with weaker effects) cannot be found. On the contrary, setting the sensitivity high will result in many false positives, which might also result in wrong conclusions. Even if the sensitivity is kept constant between studies, low repeatability is found. To increase repeatability of studies, statistical corrections can be added. However, these corrections are generally limited in their success, as artefacts can still appear (e.g., Mills & Mathieson, 2022).

In conclusion, we argue that the causal framework proposed by M&H is not suited to understand the effects of genes on behaviour. While we agree with the authors that human behaviour genetics needs a sound causal foundation, this remains a formidable challenge.

Financial support. MJB is supported by ALW-NWO Grant No. ALWOP.531.

Competing interest. None.

References

Chatterjee, S., & Ahituv, N. (2017). Gene regulatory elements, major drivers of human disease. *Annual Review of Genomics and Human Genetics*, *18*, 45–63.

Flint, J., & Mackay, T. F. C. (2009). Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Research*, *19*, 723–733. doi: 10.1101/gr.086660.108

Hedrick, P. W. (2005). *Genetics of populations* (3rd ed). Jones & Bartlett.

Kuijper, B., Pen, I., & Weissing, F. J. (2012). A guide to sexual selection theory. *Annual Review of Ecology, Evolution, and Systematics*, *43*, 287–311.

Mackay, T. F. C. (2014). Epistasis and quantitative traits: Using model organisms to study gene–gene interactions. *Nature Reviews Genetics*, *15*, 22–33. https://doi.org/10.1038/nrg3627

Marjoram, P., Zubair, A., & Nuzhdin, S. V. (2014). Post-GWAS: Where next? More samples, more SNPs or more biology? *Heredity*, *112*, 79–88.

Mills, M. C., & Mathieson, I. (2022). The challenge of detecting recent natural selection in human populations. *Proceedings of the National Academy of Sciences, USA*, *119*, e2203237119.

Van Gestel, J., & Weissing, F. J. (2016) Regulatory mechanisms link phenotypic plasticity to evolvability. *Scientific Reports*, *6*, 24524. doi: 10.1038/srep2452

Van Gestel, J., & Weissing, F. J. (2018). Is plasticity caused by single genes? *Nature*, *555*, E20. doi: 10.1038/nature25495

Zhou, X. S., Pierre, C. L., Gonzales, N. M., Zou, J., Cheng, R., Chitre, A. S., … Palmer, A. A. (2020). Genome-wide association study in two cohorts from a multi-generational mouse advanced intercross line highlights the difficulty of replication due to study-specific heterogeneity. *Genes, Genomes, Genetics*, *10*, 951–965.

Zuk, O., Hechter, E., Sunyaev, S. R., & Lander, E. S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences, USA*, *109*, 1193–1198.

When local causes are more explanatorily useful

Pierrick Bourrat[a,b]

[a]Department of Philosophy, Macquarie University, North Ryde, NSW, Australia and [b]Department of Philosophy & Charles Perkins Centre, The University of Sydney, Camperdown, NSW, Australia
p.bourrat@gmail.com, pierrickbourrat.com

doi:10.1017/S0140525X22002072, e185

Abstract

Madole & Harden plead for better integration of causal knowledge of different depths to understand complex human traits. Classically, local causes – a particular type of shallow causes – are considered less useful than more generalisable causes, giving a false impression that the latter causes are more useful and desirable. Using a simple example, I show that sometimes the contrary is true.

Madole & Harden (M&H) provide an insightful analysis showing that different types of causes can play different roles in helping us understand the aetiology of complex traits. They make a point often underappreciated – namely, that average treatment effects (ATEs), such as those obtained from randomised controlled trials (RCTs), have some of the same limitations as those often attributed to heritability estimates and single-nucleotide polymorphism associations. Similarly, heritability estimates have often been charged with being only local parameters – when such a charge is rarely made against RCTs. Further, the charge of locality gives the false impression that a less local causal relationship – one that could be observed under a broader range of conditions – is always more useful than a local one. I show here that this conclusion does not follow; in some cases, which I illustrate with a theoretical example, local causal knowledge can be more useful for explanation and intervention than more generalisable knowledge.

Since Lewontin (1974), it is commonly accepted that heritability estimates originating from an analysis of variance suffer from the problem of locality: one estimate obtained in one population, even if unbiased, cannot and should not be extrapolated to other populations. This position, particularly its extreme form, is questionable (see Sesardic, 2005, pp. 75–80); however, generally, locality is considered a detrimental feature for establishing causal relationships. Being able to generalise a result is an important aspect of science, and locality stands in its way.

The problem of locality ties in with the analysis of causation provided by Woodward (2010) in the context of biological science (see also Bourrat [2020, 2021], for discussions in the specific context of heritability). Intervening on a variable (X) permits establishing whether X is a cause of another variable (Y) but not comparing it to different causes. To that effect, several dimensions of causal relationships have been proposed in the literature, among which is stability. Some causal relationships are less stable than others – that is, they break down more easily when the background (i.e., the variables of the system that are not X or Y) changes. Thus, locality and stability are inversely related. The more a cause is local – the less it would generalise beyond the population where it was established – the less stable it is.

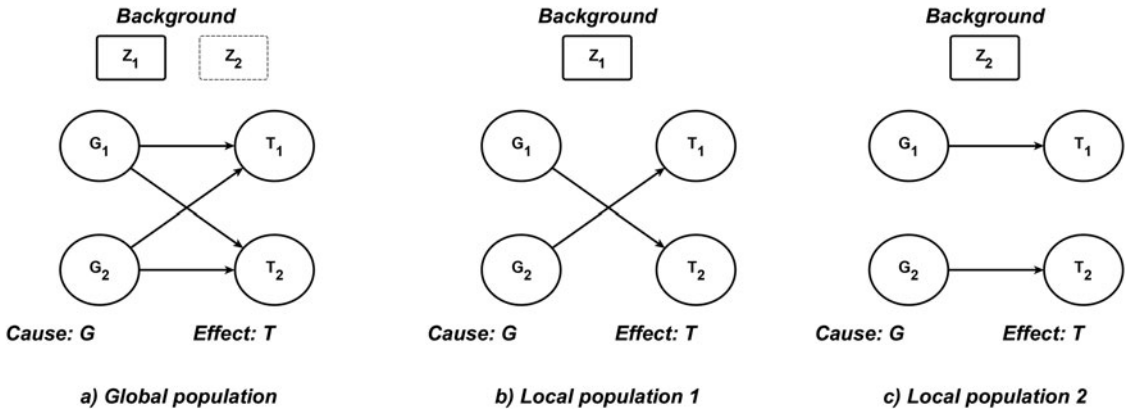


Figure 1 (Bourrat). Causal relationship between *G* and *T* in three populations. (a) In a global population with two randomised backgrounds (Z1 and Z_2), *G* does not appear to be causing *T*: intervening on *G*, on average, does not affect the probability of expressing one of the two values of *T*. (b) In the local population "1," with a constant background Z_1, intervening on *G* leads to a change in *T*, and it is established that this relationship is $G_1 \rightarrow T_2$ and $G_2 \rightarrow T_1$. (c) In the local population "2," with a constant background Z_2, the same is observed as in the local population "1," except that the relationship is reversed so that $G_1 \rightarrow T_1$ and $G_2 \rightarrow T_2$.

An ATE measures whether X makes a difference on Y, while all other variables in the background are randomised. As such, if a difference in Y is observed, one can be confident that the relationship tested is as stable as there is variation in the background. More importantly, however, it tells us nothing about whether this relationship holds under any of the randomised backgrounds, only that X makes, *on average*, a difference on Y (often with a certain magnitude). Conversely, if no causal relationship exists between X and Y, on average, in the range of backgrounds tested, this does not tell us whether it would also be the case in any of the specific backgrounds.

To make the point slightly more concrete (see Fig. 1), suppose a global population of individuals with two possible genotypes (G_1 and G_2) in equal proportions. Each genotype is associated with either two phenotypes with the same probability, T_1 and T_2 (e.g., two levels of anxiety, "low" for T_1 and "high" for T_2), depending on the background with two randomised states in equal proportions, Z_1 and Z_2, that could represent the environment. Intervening on *G* in the global population would lead to the conclusion that the genotype is not a cause of *T*. However, suppose that (unknown to the experimenter) in a local population "1" (part of *G*) containing the same proportion of the two genotypes but where only the background Z_1 exists, intervening on *G* would lead to a deterministic change in *T* (with $G_1 \rightarrow T_2$ and $G_2 \rightarrow T_1$). Further, in another local population "2" (also part of *G*), identical to the first except that only Z_2 exists, the opposite deterministic causal relationship would be established (with $G_1 \rightarrow T_1$ and $G_2 \rightarrow T_2$). In each local population, the conclusion would be that an individual's genotype causes the trait, but that a different genotype causes a different trait's value in the two populations.

This last conclusion would be more adequate than the conclusion reached in the global population that *G* does not cause anxiety. This is so because intervening on the background of some individuals in the population experiencing the "wrong" environment could affect their phenotype and have the benefit of reducing their anxiety.

A similar demonstration using more complex variables than binary variables – although more tedious – could be devised to show that a small global average effect can be the result of two (or more) large effects established in more local backgrounds but going in different directions.

The lesson from this simple case is that local or unstable causal relationships can have more value than more stable ones when the causal relationship is characterised by averages. This flies in the face of the commonly accepted view that more is better when it comes to causal stability and uncovers a well-known trade-off in the philosophy of modelling literature between generality and precision (Levins, 1966). The point sketched here speaks directly to M&H's urge not to dismiss shallow causes once integrated into a more thorough causal analysis.

Acknowledgment. I thank James Madole and Paige Harden for discussions on this topic.

Financial support. This work was supported by an Australian Research Council Discovery Early Career Research Award (Grant ID: DE210100303).

Competing interest. None.

References

Bourrat, P. (2020). Causation and single nucleotide polymorphism heritability. *Philosophy of Science*, *87*, 1073–1083.

Bourrat, P. (2021). Heritability, causal influence and locality. *Synthese*, *198*, 6689–6715.

Levins, R. (1966). The strategy of model building in population biology. *American Scientist*, *54*, 421–431.

Lewontin, R. C. (1974). The analysis of variance and the analysis of causes. *American Journal of Human Genetics*, *26*, 400.

Sesardic, N. (2005). *Making sense of heritability*. Cambridge University Press.

Woodward, J. (2010). Causation in biology: Stability, specificity, and the choice of levels of explanation. *Biology & Philosophy*, *25*, 287–318.

All that glisters is not gold: Genetics and social science

Callie H. Burt

Department of Criminal Justice & Criminology, Andrew Young School of Policy Studies & Center for Research on Interpersonal Violence (CRIV), Georgia State University, Atlanta, GA, USA
cburt@gsu.edu
www.callieburt.org

doi:10.1017/S0140525X22002217, e186

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

Abstract

In their target article, Madole & Harden offer an account of "what it means for genes to be causes" of social outcomes to bolster their claim that genetics should be incorporated into social science with practical implications. Here I object to several key features of their arguments, their representation of the state of science, and claims about the utility of genetics for social science and society.

In their target article, Madole & Harden (M&H) provide a lengthy exposition on genetic causation to undergird their argument that social scientists not only can but should incorporate genetics into their research on human behavior. Pointing to the recent "discovery of genes" associated with social outcomes and the potential utility for social science, the authors argue that careful analysis of what it means for genes to be causes of social behaviors is needed. I agree.

The devil is, as usual, in the details. With much to say and limited space, I first identify and set aside points of agreement between M&H and me while highlighting key disagreements. I then discuss my objections to their arguments. With unlimited space, I would challenge methodological issues related to their model of genetic causation, for example, violation of SUTVA, the non-generalizability of causes of sibling differences to causes of differences between unrelated people in the population, and so on. Here, I focus on what I view as more fundamental challenges.

Shared understandings and points of departure

There is, in my reading, much upon which M&H and I agree about the role of genetics in human behavioral differences. Our disagreement is rooted in what we can know about causes of human behavior, given the nature of development and given limitations of current methods and biological knowledge, and thus whether genetics is useful for social science. In particular:

(1) We agree that genetic differences matter for human social outcomes – achievements, behavior, physical health, personality – in a complex, context-sensitive way. We disagree that complex, social, non-disease achievements or behaviors, like educational attainment, having ever had same-sex sex, or income, are appropriate "traits" for genetic study.
(2) We agree that studying processes that are malleable prognostic markers of human outcomes and thus targets for intervention to reduce disparities and improve health and well-being is valuable for science and society. We disagree that studying putative *genetic* linkages to these intermediate processes is necessary or useful, especially at the current state of the science.
(3) We agree that genetic research has the potential to advance understanding of human health and disease. We disagree that genetic research is gainfully employed to enhance knowledge on the etiology of complex human *social* behavior.
(4) We agree that those previous iterations of social science genetics employed in the service of "scientifically" demonstrating the genetic inferiority of socially subordinate groups should be rejected as flawed and denounced. We disagree that the problems with social science genetics are rooted in political orientation or limited to ethical issues and thus solved by publicly repudiating genetic determinism and recognizing context-dependency and gene–environment interactionism. Crucially, as I outlined in my target article, the problems are not just political but are conceptual (e.g., Burt, 2023; Kaplan & Turkheimer, 2021), methodological (Coop & Przeworski, 2022; Morris, Davies, Hemani, & Smith, 2020; Richardson & Jones, 2019), and biological (Crouch & Bodmer, 2020; McClellan & King, 2010).

I do not doubt that M&H have good intentions. But good intentions – including an explicit "anti-eugenics" approach – are not enough.

Having sketched key sources of disagreement rooted in shared understandings, I turn to four issues that deserve response, challenge, and/or clarification.

Oversimplifications and obfuscations

Throughout M&H's target article, biological complexity is downplayed, and terminology is misleading, which obscures substantial difficulties and biological unknowns. For example, M&H frame their study as being about "genes" building on "the discovery of genes associated with human phenotypes like educational attainment and substance use disorders" (target article, abstract). To be sure, given that the intended audience includes social scientists who lack genomic expertise including familiarity with genetic terminology, the gains in readability by referencing "genes identified" versus a more accurate, "genetic variants that are non-causal markers of some unknown causal SNPs likely in proximity" are substantial. However, this simplifying language is misleading. Many social science readers might reasonably interpret this use of "genes" to indicate identified different versions of genes with defined functions that affect social outcomes through well-characterized biological pathways. The reality is nothing of the sort.[1] As I discuss at length in my target article, because of a host of limitations, current sociogenomics methodologies (i.e., genome-wide association studies [GWASs] and polygenic scores [PGSs]) are ill-suited for identifying specific "genes associated with" complex highly polygenic outcomes (Burt, 2023; Charney, 2022; Kaplan & Turkheimer, 2021).

In glossing over the difficulty of biological interpretation, M&H misrepresent the complexity and uncertainty in moving from risk loci to causal variants acting in genes with defined functions. They note that because "'genes' are studied at an intermediate level of resolution.… Researchers can then use 'fine mapping' techniques to gain higher resolution" as if fine-mapping techniques were straightforward or unproblematic rather than highly sophisticated guesswork based on limited biological knowledge (see discussions in Burt, 2023; Charney, 2022; Crouch & Bodmer, 2020). Were fine mapping and gene identification so easy.

Irreducibly social behaviors resist genetic reductionism

While their specific claims about utility are meager and usually vague, M&H point the value of genetics to enhance social scientific understanding of "how genetic factors unfold along biological and behavioral pathways across development" (target article, sect. 3.4, para. 5). To be sure, the idea of tracing genetic pathways from molecules to behavior sounds compelling. In contrast to genetic diseases, however, tracing genetic variants to complex *social* behavioral differences is impracticable. Because of biological and conceptual limitations – including the fact that complex social traits are defined by social context and thus irreducibly social –

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

mapping genetic variants to social outcomes is infeasible, even leaving aside statistical issues (e.g., environmental confounding, counterfactual model assumption violations) and ignoring the problem of misidentifying downward causation as upward genetic causation (see below).

Given this, the search for specific genetic causes of complex social (non-disease) traits like education and crime remains a misguided endeavor, even if our current methodologies facilitated such precise identification, which they do not.

To be clear, alleviation of this situation will not come from new genomic research tools, sophisticated statistical algorithms, larger sample sizes, or within-family studies. Complex social traits like educational attainment and crime, unlike cystic fibrosis and sickle cell anemia, are social traits not biological ones (Burt, 2023; Dupré, 2012).

Lack of utility

This inability to map specific (miniscule) genetic effects to social outcomes is, however, no great loss. We do not need genetics to "isolate intermediate processes that represent (a) prognostic markers of future outcomes and (b) targets for programmatic intervention" (e.g., early childhood eating habits, health behaviors, education-related behaviors) (target article, sect. 3.4, para. 5).

Using body-mass index (BMI) as a "ready example" to illustrate the potential value of genetics for social science, M&H note that studies linking genetic associations with "phenotype annotation efforts" "have found that, by as early as age two, a child's eating behavior may demarcate genetic risk for adult [high] BMI" (target article, sect. 3.4, para. 5). Given what we know about the importance of early childhood and eating behaviors, do we need genetics to suggest that early childhood eating behaviors are important in shaping adult BMI? I think not. As I have argued elsewhere, the justification for incorporating genetics into social science to reveal well-established social patterns is lacking. This is particularly true because, as we all agree, *genetic does not mean unchangeable*, and however genetically influenced, environments always matter. The targets for interventions and policies are the intervening psychosocial mechanisms (diet, self-control) and/or the environments (e.g., parenting) to reduce adverse health outcomes, reduce social inequalities, and enhance social flourishing.

In sum, from the fact that genetic differences matter for development and behavioral differences, it does not follow that incorporating measures of genetic differences (invariably imprecise and environmentally confounded) will advance social science models or policy.

Misidentifying downward (social) causation as upward (genetic) causation

Finally, even a rigorous counterfactual study of genetic causation cannot distinguish authentic ("upward") genetic causation (from genetic differences to trait differences through biological pathways) from downward social causation (e.g., Burt, 2023). Although not now, we hope one day we can all agree that a model that identifies downward causation as "genetic" (and thus would identify darker skin pigmentation alleles as genetic causes of lower income shaped by sociohistorical processes of racial/ethnic discrimination) is wrong and misleading.

In sum, while M&H extol the utility of these studies for social science and society, given limitations in what we know, can know, and can measure, what genetics offers social science remains vague, misguided, and/or overhyped. Moreover, this is to say nothing of the potential dark side, including reifying an oversimplified, flawed, and catchy notion of a "genetic predisposition" for complex social traits stratified along class, racial/ethnic, and other axes of inequality.

In conclusion, complex social phenotypes like educational attainment are not biological phenotypes; human behaviors are irreducibly social and contingent; and their causes are heterogeneous, intertwined, and unable to be "unbraided" by observational methods with even the best statistical genetic methodologies. Given this, the endeavor to identify "genes" or genetic causes of normal variation in educational attainment given both the current state of science and the nature of the "phenotype" is misguided. Such endeavors also waste time, energy, and skills of talented scientists, such as M&H.

Acknowledgment. The author is grateful to Kara Hannula for helpful feedback.

Financial support. Callie H. Burt is supported by funding from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (5K01HD094999).

Competing interest. None.

Note

1. The authors' use of genetic disease examples with well-characterized biological pathways (cystic fibrosis) surely reinforces such simplified understandings.

References

Burt, C. H. (2023). Challenging the utility of polygenic scores for social science: Environmental confounding, downward causation, and unknown biology. *Behavioral and Brain Sciences*, *46*, doi: 10.1017/S0140525X22001145

Charney, E. (2022). The "Golden Age" of behavior genetics? *Perspectives on Psychological Science*, *17*(4), 1188–1210.

Coop, G., & Przeworski, M. (2022). Lottery, luck, or legacy? *Evolution*, *76*(4), 846–853.

Crouch, D. J., & Bodmer, W. F. (2020). Polygenic inheritance, GWAS, polygenic risk scores, and the search for functional variants. *Proceedings of the National Academy of Sciences*, *117*(32), 18924–18933.

Dupré, J. (2012). *Processes of life: Essays in the philosophy of biology*. Oxford University Press.

Kaplan, J. M., & Turkheimer, E. (2021). Galton's Quincunx: Probabilistic causation in developmental behavior genetics. *Studies in History and Philosophy of Science*, *88*, 60–69.

McClellan, J., & King, M.-C. (2010). Genetic heterogeneity in human disease. *Cell*, *141*(2), 210–217.

Morris, T. T., Davies, N. M., Hemani, G., & Smith, G. D. (2020). Population phenomena inflate genetic associations of complex social traits. *Science Advances*, *6*(16), eaay0328.

Richardson, K., & Jones, M. C. (2019). Why genome-wide associations with cognitive ability measures are probably spurious. *New Ideas in Psychology*, *55*, 35–41. doi: 10.1016/j.newideapsych.2019.04.005

Addressing genetic essentialism: Sharpening context in behavior genetics

Brian Byrne[a] and Richard K. Olson[b]

[a]University of New England, Armidale, NSW, Australia and [b]Institute for Behavioral Genetics, University of Colorado, Boulder, CO, USA
bbyrne@une.edu.au
www.une.edu.au/staff-profiles/hass/bbyrne
richard.olson@colorado.edu
www.colorado.edu/ibg/richard-k-olson

doi:10.1017/S0140525X22002266, e187

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

Abstract

Evidence of a causal role for genes in human behavior underpins genetic essentialism, the scientifically flawed and socially hazardous idea that heritable characteristics are immutable. Behavior geneticists can challenge this idea by designing research that brings the contextual dependence of heritability estimates into sharper focus, and by incorporating a relevant statement into research reports and public outreach.

In their well-argued paper, Madole & Harden take aim at concerns of some scientists that genome-wide association study (GWAS) research will reinforce the idea that inequalities in society are fixed and proper.

They declare that "such a picture…is unwarranted" on the grounds that "genetic causes for human behavioral traits are *non-uniform*, *non-unitary* and *non-explanatory*" (target article, sect. 3.3, para. 4). We agree that the concerns have no scientific bases, but consider that the concerns, and more generally worries about *genetic essentialism* (Dar-Nimrod & Heine, 2011), are grounded in reality. In this commentary, we explore that reality and suggest ways in which behavioral genetics might address its influence on public discourse and policy.

Genetic essentialism is a family of beliefs centered on the ideas that if genes are known to play a role in shaping human differences then those differences are, to a considerable extent, natural and entrenched. The differences cover not only society-wide structures that license privilege, so-called social Darwinism, but ones expressed in individuals, often enough focused on less favorable outcomes in health and behavior, which are also seen as hard to alter. Admittedly, this pessimism can go hand-in-hand with a sense of relief in that a genetically influenced difficulty can be seen as no-one's "fault." This is the mixed-blessing model developed by Haslam and Kvaale (2015). But genetic essentialism can, and does, undermine the acceptance by individuals of efficacious interventions and hamper their implementation by professionals (Haslam & Kvaale, 2015; Lebowitz, Ahn, & Nolen-Hoeksema, 2013).

We propose two ways that behavior-genetic research community could contribute to combatting genetic essentialism, whose adverse effects are now well documented (Dar-Nimrod & Heine, 2011). One is in the design of the research agenda, the other in public communication of research results. Regarding research: The aphorism that "genes are not destiny" can be, and if possible should be, brought into clear focus by demonstrating its truth *within the study of single phenotype*. There exist useful demonstrations of different patterns of heritability across groups defined by location (e.g., lower heritability of early reading in Scandinavia compared to the United States and Australia, Samuelsson et al., 2008) and across groups defined by time (e.g., an increase in heritability of educational attainment in Norway for males born after 1940, Heath et al., 1985). In addition, demonstrations of gene–environment interplay, where it exists (see Grasby, Coventry, Byrne, & Olson, 2019), underline the dependence of some heritability estimates on environmental factors such as socioeconomic level. And in an additional and informative level of complexity, illustrations of different cross-national patterns of gene–environment interaction can deepen insights into environmental influences on phenotypes (Grasby et al., 2019; Tucker-Drob & Bates, 2016).

Researchers can, and should, also maximize opportunities afforded by analytic maneuvers to expose contextual influences on heritability estimates. For example, in a study using the classical twin design when samples are mixed in terms of location, time, or participant type, the choice to standardize scores within groups can generate higher estimates of genetic influence than the choice of not to standardize. In our cross-national study of early reading (Byrne, Olson, & Samuelsson, 2013), Scandinavian children in grade 1 read at a lower average level than those in the United States and Australia, because of the later start of reading instruction in Sweden and Norway. Employing within-country standardization, the genetic influence on reading comprehension was estimated at 0.62; in the absence of within-country standardization, the estimate dropped to 0.38 with a corresponding rise in the shared environment effect. As a thought experiment, imagine a whole-world twin study without country standardization; doubtless the biggest influence on variation in reading ability at any particular age will not be genes but curriculum policies and the resources available for education, showing up in analyses as shared environment. Such a picture would reinforce the value of projects like that of Lyytinen and colleagues (Ojanen et al., 2015), aimed at bringing the benefits of basic reading research to educational practice in less developed nations. Unthinking acceptance of the conclusion, derived from twin studies conducted largely in the United States, Europe, and Australia, that half or more of the variance in reading skill is heritable would tend to undermine efforts like Lyytinen's. Thus, where it is appropriate to privilege a *humans in general* perspective over *humans within particular contexts*, analytic choices can be made to do exactly that.

Thus, in designing behavior-genetic research, features that enhance the visibility of context on heritability estimates should be front and center. International cooperation would be one way to promote this goal, in cases at least where there might be reasons to believe that country differences could influence levels and expression of a phenotype. Geneticists working on devising polygenic risk scores for human conditions, especially diseases, may be ahead of behavior geneticists in that there are calls to increase the ancestral diversity of participant groups (Surigo, Williams, & Tishkoff, 2019) in response to (a) the predominance of populations with European ancestry in studies to date, and (b) accumulating evidence that scores derived from European samples do not always hold up in predicting disease status in African, Asian, Latino, and other ancestry groups (Belsky et al., 2013; Grinde et al., 2019). Behavior genetics would probably replace ancestral groups with sociocultural ones on the grounds that it is those groups that afford the most likely source of contextual influence on patterns of heritability, not ancestry-driven differences in linkage disequilibrium, and in the other sources of variability in polygenic risk score analyses (Surigo et al., 2019).

Armed with a substantial corpus of data on contextual dependence of heritability estimates, behavior genetics could develop a core statement to allay the concerns of "end-users" such as educators, clinical psychologists, criminologists, and lawyers that genes are indeed destiny. A group such as the Behavior Genetics Association could be tasked with crafting such a statement, which could then be incorporated, suitably modified for particular circumstances, into research publications and, more generally, into public outreach campaigns.

Acknowledgments. Many colleagues have contributed to the research and ideas incorporated into our commentary; here we acknowledge in particular Stefan Samuelsson, Linköping University, who initiated and led the Scandinavian component of our twin studies.

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

Financial support. Our research was supported by the Australian Research Council (DP 150102441, DP 0663498, DP 0770805), the National Institute for Child Health and Human Development (P50 HD 27802, R01 HD 38526), the Swedish Research Council (2011-1905), and the Swedish Council for Working Life and Social Research (2011-0177).

Competing interest. None.

References

Belsky, D. W., Moffitt, T. E., Sugden, K., Williams, B., Houts, R., McCarthy, J., & Caspi, A. (2013). Development and evaluation of a genetic risk score for obesity. *Biodemography and Social Biology*, *59*, 85–100. doi: doi.org/10.1080/19485565.2013.774628

Byrne, B., Olson, R. K., & Samuelsson, S. (2013). Subsample standardization in twin studies of academic achievement. In *Annual conference of the behavior genetics association*, July, Marseilles, France.

Dar-Nimrod, I., & Heine, S. J. (2011). Genetic essentialism: On the deceptive determinism of DNA. *Psychological Bulletin*, *137*, 800–818. https://doi.org/10.1037/a0021860.

Grasby, K. L., Coventry, W. L., Byrne, B., & Olson, R. K. (2019). Little evidence that socioeconomic status modifies heritability of literacy and numeracy in Australia. *Child Development*, *90*, 623–637. doi: 10.111/cdev.12920

Grinde, K. E., Qibin, Q., Thornton, T. A., Simin, L., Shadyab, A. H., Chan, K. H. K., … Sofer, T. (2019). Generalizing polygenic risk scores from Europeans to Hispanics/Latinos. *Genetic Epidemiology*, *43*, 50–62. doi: doi.org/10.1002/gepi.2216

Haslam, N., & Kvaale, E. P. (2015). Biogenetic explanations of mental disorder: The mixed-blessings model. *Current Directions in Psychological Science*, *24*, 399–404. https://doi.org/10.1177/096372141558808

Heath, A. C., Berg, K., Eaves, L. J., Solaas, M. H., Corey, L. A., Sundet, J., … Nance, W. E. (1985). Education policy and the heritability of educational attainment. *Nature*, *314*, 734–736.

Lebowitz, M. S., Ahn, W.-K., & Nolen-Hoeksema, S. (2013). Fixable or fate? Perceptions of the biology of depression. *Journal of Consulting and Clinical Psychology*, *81*, 518–527. https://doi.org/10.1037/a00311730

Ojanen, E., Ronimus, M., Ahonen, T., Chansa-Kabali, T., February, P., Jere-Folotiya, J., … Lyytinen, H. (2015). GraphoGame – A catalyst for multi-level promotion of literacy in diverse contexts. *Frontiers in Psychology*. *6*(671), 1–13. doi: 10.3389/fpsyg.2015.00671

Samuelsson, S., Byrne, B., Olson, R. K., Hulslander, J., Wadsworth, S., Corley, R., … DeFries, J. C. (2008). Response to early literacy instruction in the United States, Australia, and Scandinavia: A behavioral-genetic analysis. *Learning and Individual Differences*, *18*, 289–295. doi: doi.org/10.1016/j.lindif.2008.03.004

Surigo, G., Williams, S. M., & Tishkoff, S. A. (2019). The missing diversity in human genetic studies. *Cell*, *177*, 26–31. doi: doi.org/10.1016/j.cell.2019.02.048

Tucker-Drob, E. M., & Bates, T. C. (2016). Large cross-national differences in gene × socioeconomic status interaction on intelligence. *Psychological Science*, *27*, 138–149. doi: doi.org/10.1177/0956797615612727

Causal dispositionalism in behaviour genetics

María Cerezo

Department of Logic and Theoretical Philosophy, Complutense University of Madrid, Madrid, Spain
macere03@ucm.es; https://philpeople.org/profiles/maria-cerezo

doi:10.1017/S0140525X22002060, e188

Abstract

Causal dispositionalism developed in metaphysics of science offers a useful tool to conceptualize shallow causes in behaviour genetics, in a way such that (a) it accounts for complex aetiology and heterogeneity of effects, and (b) genetic causal contribution can be considered to be explanatory. Genes are thus causal powers that make a difference.

One of the virtues of Madole & Harden's (M&H's) approach to causation in behaviour genetics is their ability to combine their honesty in acknowledging the limits and gaps of genome-wide association study (GWAS) methodology and, more generally, in behaviour genetics with their determination to point out the way in which first-generation causal knowledge can open the door to ways in which second-generation causal knowledge can be pursued. As a consequence of the first attitude, their analysis represents a step further to overcome genetic determinism and essentialism (target article, sect. 1.2, para. 4). The second attitude leads them into the sort of pluralism that is commonly accepted in philosophy of science and medicine (see, e.g., Rocca & Anjum, 2020). In their approach, M&H have recourse to some work in the metaphysics of science to deploy the theoretical framework of their investigation. The interventionist theory of causation developed by Woodward (2005), which has been applied to genetics by Waters (2007) and by Woodward (2010) himself, provides them with a useful tool to conceptualize the notion of cause behind first-generation causal knowledge in genetics. In addition, recourse to mechanisms (Craver & Darden, 2013; Glennan, 1996) accounts for the kind of explanation required in second-generation causal knowledge, crucial for understanding the aetiology of complex phenomena. In this brief commentary, I intend to bring to light another metaphysical tool that can help M&H to overcome some limitations and offer a richer picture, namely, the concept of *capacity* or *disposition* (Cartwright, 1989; Mumford, 1998). Mumford and Anjum have developed a dispositionalist theory of causation (Mumford & Anjum, 2011), and have applied it to science (Anjum & Mumford, 2018), medicine (Rocca & Anjum, 2020), and genetics (Mumford & Anjum, 2011, Ch. 10), conceiving of genes as powers or bundles of powers "coded" into the structural complexity of DNA strands.

Capacities, powers, or dispositions (I use these terms as equivalent) are "properties or potentials of things or systems, that can become manifested under certain conditions" (Rocca & Anjum, 2020). Clinical randomized controlled trials (RCTs), for instance, can give rise to claims about the capacity or power of certain substances to induce specific effects in particular contexts. Such association of RCTs to capacity claims "provide a conduit from RCTs to effectiveness" (Cartwright, 2011). Medicine seeks to predict in order to intervene by means of treatments, and behavioural genetics applied to social sciences seeks to predict in order to intervene too. But then, it is necessary that the sort of prediction reached is not of the kind "it works somewhere," but rather of the kind "it will work for us" (Cartwright, 2011). This requires that the treatment reliably promotes the outcome, and this demands a capacity claim (magnets power to attract metals, for instance, grounds the step from "it will attract some metal somewhere" to "it will attract some metal for us"). Capacities thus allows for extrapolation of the knowledge provided by RCTs insofar as they offer the grounding or explanation for the association reached by the RCTs (Cartwright, 2009).

Dispositions have some characteristics that make them suitable for their application in genetics. (a) Given their dispositional nature, powers might not be manifested if the triggering conditions do not occur. This feature allows for a sort of modality that matches very well with claims in genetics, because genes sometimes predispose for something, but do not necessitate it. (b) Triggering conditions make a power manifest and it is then when causation occurs; causation is thus conceived as a process rather than as a relation. Analogously, genes are causes only insofar as genetic expression is triggered. (c) Most powers need mutual manifestation powers, that is, powers that need to be met in order for them to manifest, so that causation occurs as a consequence of the joint manifestation of such powers.

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

This characteristic accounts for phenomena such as polygeny or, in the case of behaviour genetics, heterogeneity across environments. (d) There are interfering and preventing powers, so that a power makes the effect occur differently by influencing the timing, chance, or extent to which the effect occurs (interfering powers) or prevents the effect (preventing powers). Gene silencing and mutation are paradigmatic examples in this case, and many epigenetic phenomena (histone modification and cytosine methylation) instantiate interfering powers that determine in some way gene effects. Demographic composition and environmental context are typical examples of this feature in the case of behaviour genetics. These four features account for the complexity of causation that is so pervasive in genetics. A nice illustration of some of these features is implicit in M&H's example of causal depth in cystic fibrosis, which presents shallow features (see endnote 6).

M&H characterize shallow causes as non-unitary (complex causality), non-uniform (heterogeneity of effects), and non-explanatory (cause and effect are associated without explaining why and how the effect takes place). It is quite immediate the way in which the dispositional framework captures the non-unitary and non-uniform character of shallow causes: Most dispositions manifest jointly and interfering and preventing powers account for heterogeneity of effects in different contexts. Crucially, however, powers are explanatory. It is *because* genes are powers that they produce the effects they do when triggering conditions occur. The picture is then one in which different powers (different genes, other biological components, environmental and cultural factors) contribute to different extent and in various ways to the effect, accounting in this way for the shallowness observed in genetic causes, but insofar as they are explanatory, powers offer a suitable road from shallowness to effective or real causation. Powers offer the way to go from *whether* and *how often* something happens to *why* and *how* something might or might not happen (Rocca & Anjum, 2020). Or, in other words, "there are causes out because there are causes in."

Financial support. This research received financial support from the Spanish Ministry of Science and Innovation (Ministerio de Ciencia e Innovación): Research Projects "Metaphysics of Biology: Framing the Interactions between Metaphysics and Molecular, Developmental and Evolutionary Biology" (Ref: FFI2017-87193-P) and "Metaphysics of Biology: Processes and Dispositions" (Ref: PID2021-127184NB-I00).

Competing interest. None.

References

Anjum, R. L., & Mumford, S. (2018). *Causation in science and the methods of scientific discovery*. Oxford University Press.

Cartwright, N. (1989). *Nature's capacities and their measurements*. Oxford University Press.

Cartwright, N. (2009). Causal laws, policy predictions and the need for genuine powers. In T. Handfield (Ed.), *Dispositions and causes* (pp. 127–157). Oxford University Press.

Cartwright, N. (2011). A philosopher's view of the long road from RCTs to effectiveness. *The Lancet*, *377*, 1400–1401.

Craver, C. F., & Darden, L. (2013). *In search of mechanisms: Discoveries across the life sciences*. The University of Chicago Press.

Glennan, S. S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, *44*, 49–71.

Mumford, S. (1998). *Dispositions*. Oxford University Press.

Mumford, S., & Anjum, R. L. (2011). *Getting causes from powers*. Oxford University Press.

Rocca, E., & Anjum, R. L. (2020). Causal evidence and dispositions in medicine and public health. *International Journal of Environmental Research and Public Health*, *17*(6), 1813.

Waters, C. K. (2007). Causes that make a difference. *The Journal of Philosophy*, *104*, 551–579.

Woodward, J. (2005). *Making things happen: A theory of causal explanation*. Oxford University Press.

Woodward, J. (2010). Causation in biology: Stability, specificity, and the choice of levels of explanation. *Biology & Philosophy*, *25*, 287–318.

Theory matters for identifying a causal role for genetic factors in socioeconomic outcomes

Steven N. Durlauf[a] and Aldo Rustichini[b]

[a]Harris School of Public Policy, University of Chicago, Chicago, IL, USA and [b]Department of Economics, University of Minnesota, Minneapolis, MN, USA
durlauf@gmail.com
sdurlauf@uchicago.edu
aldo.rustichini@gmail.com

doi:10.1017/S0140525X22002084, e189

Abstract

Any empirical claim about the role of genes in socioeconomic outcomes involves successfully addressing the identification problem. This commentary argues that socioeconomic outcomes such as education are sufficiently complex, involving so many mechanisms, that understanding the role genes requires the use of formal theoretical structures.

Madole & Harden (M&H) provide an interesting and wide-ranging argument on the senses in which genetically informed studies may be said to produce causal inference. We focus on their discussion of genome-wide association studies (GWASs) of educational attainment (EA). The authors summarize this evidence as:

> Genes might cause EA in the sense that genes made some distal difference in level of attainment, but not in the sense that they provide an explanation for how this difference was made.

The paper's claims amount to arguing that (1) GWAS type studies produce shallow causal evidence and (2) such evidence can help direct research that seeks mechanisms, that is, explanations. Our argument is that both objectives need theory to be successfully met.

From the perspective of statistical/econometric theory, EA is an example of an outcome determined via a system of interactions, that is, it is a dependent variable in a simultaneous equations system. This is evident from the basic logic of economic models of EA. Suppose that individuals come in types T_i determined by genotype g_i, an unobservable η_i capturing *in utero* effects, and so on. Suppose that individuals experience family influences F_i, social influences S_i, which are determined by one another, the types and associated unobservables ν_i and ξ_i. Together, background factors obey a simultaneous system:

$$T_i = a(g_i, \eta_i) \tag{1}$$

$$F_i = b(T_i, S_i, \nu_i) \tag{2}$$

$$S_i = c(T_i, F_i, \xi_i) \tag{3}$$

Suppose, in turn, that type, family, and social factors combine with unobservables δ_i to produce choices C_i (e.g., effort)

$$C_i = d(T_i, F_i, S_i, \gamma_i) \tag{4}$$

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

and that EA_i is determined by background factors, choices, and unobservables ψ_i

$$EA_i = e(T_i, F_i, S_i, C_i, \psi_i) \quad (5)$$

M&H argue for the informational value of the conditional probability of EA given genotype:

$$\Pr(EA_i|g_i) \quad (6)$$

Variation in (6) for different g_i value demonstrates, in the authors' sense, a causal relationship between genes and education.

We focus, instead, on what is learned about (1)–(5) from (6). From the perspective of economic theory, (1)–(5) is a simultaneous equations system, while (6) is a reduced form quantity that summarizes aspects of the data. The classic identification problem of simultaneous equations systems asks what features of the structural relations (1)–(5) are revealed by reduced form evidence such as (6). The simultaneous equations perspective, as has been long understood, reveals the need for *a priori* assumptions to make credible empirical claims about structural relationships from reduced form ones. How does that logic apply to EA?

First, the determination of mechanisms that produce EA requires *a priori* assumptions on the structure producing between the joint density of all observables. Examples of such assumptions include exclusion restrictions that represent ways to delimit the paths that link different endogenous and predetermined variables. This is the first sense in which social science theory is needed: Deep causal claims require credible *a priori* assumptions and economic theory provides precisely that. Any search for explanations needs to be theoretically informed.

Second, claims that (6) reveals statistical causality implicitly depend on the structure (1)–(5) that produces (2). M&H draw analogies between randomized controlled trials and genomic analyses, arguing that the genetic lottery acts as a randomization device. But the information involved in (6) does not translate into interpretable objects of any type unless one has made background assumptions about (1)–(5). Conditions for causal inference, such as the single unit value treatment assumption and strong ignorability are statements about the properties of a system. Randomized controlled trials, for example, succeed because the assignment mechanism rules out certain pathways linking a treatment to outcomes.

The necessity of theory is illustrated by comparing GWAS evidence on EA with the M&H example on the causal effect of lithium on depression. They defend empirical claims of causal links between lithium and reduced depression even though the biological pathway from lithium to mental state is not understood. We see essential differences with EA. The lithium evidence is compelling, despite the absence of clear biological pathways, because the randomization can balance family and social factors.

In contrast, computation of polygenic scores for EA do not allow one to conclude that changing the polygenic score of a given person, would (in the probabilistic sense of (6)) change their distribution of EA, unless, one has taken a stance on family and social factor processes in (1) that are induced by genotypes. As often noted, genomic correlations with EA could reflect discrimination as opposed to some intrinsic academic ability. Without a theory of these pathways, we do not see how (6) answers substantive questions.

M&H may answer that we are eliding the shallow statistical causality concept with the deeper explanation-based causality concept that they acknowledge is not revealed by (6). We see the issues differently: While the lithium experiments produced useful knowledge in the sense of Marschak (1953) we do not see the same applying to EA. First, while the lithium studies were policy-relevant, that is, led to recommendations on treatment, the same is not true for (6). We see no way of mapping (6), in isolation, to any policy implications if one wishes to rectify inequalities, promote fair equality of opportunity, and so on, without knowledge of mechanisms. The same claim with respect to whether (6) can lead to useful knowledge about mechanisms *per se* – this is the classic failure of identification in simultaneous equations systems without *a priori* assumptions. An individual's genotype, as it is associated with different family and social pathways, does not admit a reduction of the set of potential mechanisms determining EA, let alone their magnitudes, based on (6) alone. While we are inexpert in how heterogeneity in lithium effects facilitated the search for biological pathways, we suspect that prior biological knowledge was required to do this.

We applaud M&H for beginning the process of integrating genomic research with the existing literatures in statistics and econometrics. We endorse their call to use statistical causal findings on genotypes to help guide the search for explanations. Where we differ is that we believe social science areas such as education require social science behavioral models. Genomic data may help with the identification, as it provides observables that help reveal unobserved individual types (in the sense of Eq. (1)), but it cannot succeed alone.

Acknowledgment. A.R. acknowledges the U.S. Department of Defense, contract W911NF2010242.

Financial support. This research received no specific grant from any funding agency, commercial. or not-for-profit sectors.

Competing interest. None.

Reference

Marschak, J. (1953). Economic measurements for policy and prediction. In W. Hood & T. Koopmans (Eds.), *Studies in econometric method* (pp. 293–322). Wiley.

Where *not* to look for targets of social reforms and interventions, according to behavioral genetics

Nikolai Haahjem Eftedal[a,b] and Lotte Thomsen[a,c,d]

[a]Department of Psychology, University of Oslo, Forskningsveien 3A, Oslo, Norway; [b]PROMENTA Research Center, University of Oslo, Forskningsveien 3A, Oslo, Norway; [c]Department of Political Science, Aarhus University, Bartholins Allé 7, Aarhus, Denmark and [d]Center for the Experimental Philosophical Investigation of Discrimination, Aarhus University, Bartholins Allé 7, Aarhus, Denmark
n.h.eftedal@psykologi.uio.no
lotte.thomsen@psykologi.uio.no

doi:10.1017/S0140525X22002059, e190

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

Abstract

Behavioral genetics typically finds that the so-called shared environment contributes little or nothing to explaining within-population variation on most traits. If true, this has important implications for where *not* to look for good targets of interventions: Namely all things that are within the normal range of variation from one rearing environment to the next in that population.

If a socially valued trait was found to be "hardwired" and immutable, this would imply that all individual differences on it are because of the inherent superiority of some over others, and that all efforts toward improvement or equality would be futile. We applaud Madole & Harden's attempt to correct the widespread misunderstanding that several findings of such genetic determinism have already been made, and their clarification that even if genes have causal effects on phenotypes, these effects are typically probabilistic and context dependent with multiple mediating processes. Behavioral genetics is presented as a boon, rather than impediment, to social reform and progression, as it can help identify these mediating processes between genotypes and phenotypes so that we may intervene for the greater common good.

Relevant to this, here we direct attention to the number nine finding by Plomin, DeFries, Knopik, and Neiderhiser (2016), namely that most phenotypes are not influenced much by the so-called shared environment. This finding is quite helpful when searching for targets of interventions, but also a bit discouraging: It tells us where *not* to look. Specifically, it tells us not to look in most of the places people are currently looking. The shared environment represents all things except genes that cause pairs of relatives to become more similar to each other. It thus encompasses not just variation in the family environments where children grow up (parenting, nutrition, etc.), but also all variation between families in broader geographical, economic, and cultural settings. If no variance is explained by the shared environment, this either means that it makes no difference to the relevant phenotype, or that these environmental effects are so unpredictable that they are uncorrelated even between close relatives. Either way, they are bad targets for interventions.

Intelligence, mental health, and personality are among the many important phenotypes largely unaffected by the shared environment, and so are sociopolitical traits such as social dominance orientation (SDO; Kleppestø et al., 2019) and justice sensitivity (Eftedal et al., 2022). Notable exceptions include educational attainment, and political conservatism (Willoughby et al., 2021) and authoritarianism (Eftedal et al., 2020), which often have shared-environmental variance components of at least 20%.

A lack of shared-environmental influences on a trait does not mean that it is necessarily "hardwired" or immutable (e.g., mean IQ has been rising in recent decades). But it does imply that an environmental intervention must be well outside of the normal range of variation to work. For example, if you hope to increase intelligence by encouraging children to read and learn, you would need to provide more than what some parents already give: If it is true that the large amount of variation that exists on parenting, or social expectations, or number of books on the shelf, explains little or no variation in a trait, why, then, should such things suddenly make a difference if we intervene on them? Interventions that only represent nudges on factors that already vary substantially across contexts within a relevant sample are unpromising, *regardless of which part of the causal chain from genotype to phenotype they purport to address*.

A critical caveat to the conclusion that the shared environment has little influence, however, is the fact that this general pattern of findings primarily stems from samples confined to single cultures, typically white, educated, industrialized, rich, and democratic (WEIRD) ones, at one single point in time. That is, behavioral genetic studies can only conclude that there is no effect of the shared environment within their samples, but these samples mostly fail to measure the substantial variation in environments that we see across human cultures and ecologies in both space and time. For example, while attitudes toward group hegemony, as reflected in SDO, are unaffected by shared-environmental variation within a sample of Norwegian middle-aged twins (Kleppestø et al., 2019), there is nevertheless substantial variation across nations and states of the United States in the SDO levels of members of dominant societal groups, which correlate with ecological levels of macrostructural inequality (Kunst, Fischer, Sidanius, & Thomsen, 2017). And when looking across time within cultures or countries, the increases we have seen in IQ over recent decades, for instance, are far larger than what genetic shifts alone can plausibly account for (Flynn, 2009). Ever more large-scale social psychological studies continue to pinpoint the ecological correlates of psychological phenomena such as these.

A lesson from behavioral genetics to those looking for effective societal interventions on shared-environmental variables is then to broaden one's horizon, and look at the things that vary across cultural ecologies, rather than the things that vary within them. The shared-environmental differences that actually seem to make a difference are those we see between, for example, Boston and Mumbai, or between 2022 and 1942, rather than those we see between the Smiths from west of town and the Harpers in the east. Behavioral genetics should broaden its empirical scope beyond single-culture WEIRD samples to adequately identify critical effects of the shared environment and thus potential societal targets of environmental intervention.

When such data collection is not feasible, investigating the *unique* environment of individuals, containing everything that makes relatives *differ*, also offers lessons for behavioral science. While this unique variance component appears to include mostly idiosyncratic and unsystematic influences (Turkheimer, 2000), there are still important things to learn from looking at how non-shared-environmental versus genetic variance components from different phenotypes are *correlated*. And as Franić et al. (2013) describe, such correlations provide more stringent tests of latent psychological constructs than do regular factor analyses. These techniques revealed, for example, that sensitivities to being the perpetrator, victim, beneficiary, and observer of injustice are undergirded by separate and heritable latent motivations to be morally principled and opportunistic (so that the worse one reacts to injustice to oneself, the less one reacts to injustice to others). Furthermore, the unique environments that increase moral opportunism tend to also increase SDO, while genetic substrates that increase moral opportunism also increase SDO and lower altruism and generalized trust (Eftedal et al., 2022).

As non-shared-environmental correlations control for genetic and shared-environmental confounders, they can also help narrow our search for pairs of variables that are connected causally. For instance, individual educational attainment is associated with lower authoritarianism even when controlling for genetic and shared-environmental confounding (Eftedal et al., 2020). Surely

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

such knowledge about the underlying nature of latent psychological constructs and their potential causal connections are useful stepping stones in the search for good interventions.

Financial support. The research was funded by grants 0602-01839B from Independent Research Fund Denmark, 231157/F10 from the Norwegian Research Council, and 101040978 from the European Research Council (all to L.T.) and by center grants DGF-144 from the Danish National Research Foundation and 288083 from the Norwegian Research Council.

Competing interest. None.

References

Eftedal, N. H., Kleppesto, T. H., Czajkowski, N. O., Eilertsen, E. M., Sheehy-Skeffington, J., Vassend, O., … Thomsen, L. (2020). Causality and confounding between right wing authoritarianism, education, and socio-economic status; a twin study. *PsyArxiv*. https://doi.org/10.31234/osf.io/vws83

Eftedal, N. H., Kleppestø, T. H., Czajkowski, N. O., Sheehy-Skeffington, J., Røysamb, E., Vassend, O., … Thomsen, L. (2022). Justice sensitivity is undergirded by separate heritable motivations to be morally principled and opportunistic. *Scientific Reports*, *12*(1), 1–14.

Flynn, J. R. (2009). Requiem for nutrition as the cause of IQ gains: Raven's gains in Britain 1938–2008. *Economics & Human Biology*, *7*(1), 18–27.

Franić, S., Dolan, C. V., Borsboom, D., Hudziak, J. J., van Beijsterveldt, C. E., & Boomsma, D. I. (2013). Can genetics help psychometrics? Improving dimensionality assessment through genetic factor modeling. *Psychological Methods*, *18*(3), 406.

Kleppestø, T. H., Czajkowski, N. O., Vassend, O., Røysamb, E., Eftedal, N. H., Sheehy-Skeffington, J., … Thomsen, L. (2019). Correlations between social dominance orientation and political attitudes reflect common genetic underpinnings. *Proceedings of the National Academy of Sciences*, *116*(36), 17741–17746.

Kunst, J. R., Fischer, R., Sidanius, J., & Thomsen, L. (2017). Preferences for group dominance track and mediate the effects of macro-level social inequality and violence across societies. *Proceedings of the National Academy of Sciences*, *114*(21), 5407–5412.

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2016). Top 10 replicated findings from behavioral genetics. *Perspectives on Psychological Science*, *11*(1), 3–23.

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

Willoughby, E. A., Giannelis, A., Ludeke, S., Klemmensen, R., Nørgaard, A. S., Iacono, W. G., … McGue, M. (2021). Parent contributions to the development of political attitudes in adoptive and biological families. *Psychological Science*, *32*(12), 2023–2034.

Genetics can inform causation, but the concepts and language we use matters

Sara A. Hart[a,b] and Christopher Schatschneider[a,b]

[a]Department of Psychology, Florida State University, Tallahassee, FL, USA and [b]Florida Center for Reading Research, Florida State University, Tallahassee, FL, USA
sahart@fsu.edu
www.idcdlab.com
schatschneider@psy.fsu.edu

doi:10.1017/S0140525X22002175, e191

Abstract

Madole & Harden describe how genetics can be used in a causal framework. We agree with many of their opinions but argue that comparing within-family designs to experiments is unnecessary and that the proposed influence of genetics on behavior can be better described as inus conditions.

Madole & Harden (M&H) describe how genetics can be used in a causal framework. We agree with the authors that viewing genetic causality as probabilistic, instead of deterministic, is a fruitful way to view the effects of genetics on behavior. Moreover, we agree that using within-family designs will strengthen our understanding of the causal effects of genetics on development (Hart, Taylor, & Schatschneider, 2013; van Dijk, Norris, & Hart, in press). However, there are two points we wish to discuss from Madole & Harden (M&H). First, we believe we do not need to equate within-family designs to experiments, including randomized control trials (RCTs) and natural experiments. Second, we believe that the effects of genetics on behavior are better described as what they are, inus conditions, rather than describing what they are not, non-uniform, non-unitary, and non-explanatory.

Philosophers have been debating the meaning of causality for centuries, including John Stewart Mill, who formalized three conditions for establishing causality. First, the cause must precede the effect, second, the cause must be related to the effect, and third, we can find no plausible explanation for the effect other than the cause (see Shadish, Cook, & Campbell, 2002). The first two conditions are easily established by many designs. The third condition is the hardest to meet. Experiments, defined by the use of random assignment with a manipulation, are common and powerful tools we use to meet the third condition, as they allow us to rule out other plausible explanations. However, ultimately causality is determined by meeting those three conditions, whether you have experimentation or not. M&H suggest that genetic transmission from parent to child can be interpreted the same as an average treatment effect from an RCT. We agree that a within-family design and an RCT carry a lot of strong causal information, but we disagree with describing the two designs as similar. Within-family designs are different than RCTs in an important way. An RCT is a specific type of experiment that introduces a manipulable cause to examine whether that cause increases or decreases a given behavior. Within-family designs investigate non-manipulable causes to examine whether genetic influences have effects on variability across the range of behaviors. We believe there is no need to make the comparison between within-family designs and RCTs as on their own within-family designs do meet the conditions laid out by John Stewart Mill for causal conclusions. We also believe that we do not need to equate within-family designs to natural experiments, which is often done, as natural experiments also have a manipulation. By forcing the language of experiments, no matter the type, on within-family designs, we leave ourselves open to criticisms that within-family designs do not have a manipulation and therefore can never establish causality. However, establishing causality does not need an experimental manipulation, and as a field we do not need to borrow the language of experimental designs, whether it is a natural experiment or RCTs, to show that our results can still inform us about causality. With a within-family design that is estimating the genetic transmission from parent to child, if the assumption of the equal environments is met, there are no other plausible genetic or environmental explanations for the effects of the specific genes on behavior. We can instead use the language of what we do have, which is a unique design with no formal name that we know of, which has randomization of genes because of miosis and non-manipulated causes, and looks at the impact of

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

variation, as opposed to mean differences. We believe this unique and powerful design can meet John Stewart Mill's three conditions for establishing causality.

Second, M&H assert that genetic effects should most likely be viewed as non-uniform, non-unitary, and non-explanatory. We agree with this position. It is highly likely that most of the effects of genes on behavior are not causally deterministic but instead only impact the chances of a behavior occurring. However, this position describes these effects in terms of what they are not. The reason that these effects may be non-uniform, non-unitary, and non-explanatory is because they most likely operate as inus conditions (Mackie, 1974; Shadish et al., 2002). Inus stands for an insufficient but nonredundant part of an unnecessary but sufficient condition. Insufficient means that the existence of this factor by itself is not enough to cause the effect. Nonredundant means that in the constellation of factors that come together to produce an effect, a particular factor provides something unique that the other factors do not. Unnecessary means that the effect could be produced by other factors even in the absence of a particular factor. Sufficient means that in concert with other factors, it is enough to produce an effect. For example, having a particular polymorphism alone is insufficient to produce a particular behavior. Every behavior, for example, needs an environment that is capable of allowing that effect. It is nonredundant in that a particular polymorphism is unique within a person. It is unnecessary in that the behavior could be observed without the polymorphism, but it is sufficient in that in combination with a constellation of other factors it either promotes or inhibits a behavior. It is because of these conditions that the effects of genes may be non-uniform, non-unitary, and non-explanatory. Most behaviors of interest to scientists fall under the category of inus conditions because most behaviors that we study have multiple causes in the sense that there are many factors that either promote or inhibit a behavior. It is likely that genetic effects operate on behaviors in the same way.

We applaud M&H for reminding us that within-family designs can inform causality and extending this discussion by laying out how genetic influences might operate within a causal framework. We believe our commentary will help sharpen the concepts and language around this causal framework.

Financial support. This project was supported by funding from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development grants P50HD052120 and R01HD095193. Views expressed herein are those of the authors and have neither been reviewed nor approved by the granting agencies.

Competing interest. None.

References

Hart, S. A., Taylor, J., & Schatschneider, C. (2013). There is a world outside of experimental designs: Using twins to explore causation. *Assessment for Effective Intervention*, *38* (2), 117–126.

Mackie, J. L. (1974). *The cement of the universe: A study of causation*. Oxford University Press.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

van Dijk, W., Norris, C. U., & Hart, S. A. (in press). Using twins to assess what might have been: The cotwin control design. *Research on Social Work Practice*.

Extensions of the causal framework to Mendelian randomisation and gene–environment interaction

Claire M. A. Haworth[a] and Robyn E. Wootton[a,b]

[a]School of Psychological Science, University of Bristol, Bristol, UK and [b]Nic Waals Institute, Lovisenberg Diaconal Hospital, Oslo, Norway
claire.haworth@bristol.ac.uk
robyn.wootton@bristol.ac.uk
https://www.bristol.ac.uk/people/person/Claire-Haworth-04ed5882-f1f6-4fb5-8960-5581b0cc8bc4/
https://app.cristin.no/persons/show.jsf?id=1265206

doi:10.1017/S0140525X22002205, e192

Abstract

In our commentary we ask whether we should ultimately endeavour to find the deep causes of behaviours? Then we discuss two extensions of the proposed framework: (1) Mendelian randomisation and (2) hypothesis-free gene–environment interaction (leveraging heterogeneity in genetic associations). These complementary methods help move us towards second-generation causal knowledge, ultimately understanding mechanistic pathways and identifying more effective intervention targets.

Reactions to genetic causes often take one of two extremes: either genetic causes are aversive, or they are deterministic and put upon a pedestal. We believe that Madole & Harden's (M&H's) framework and reasoned argument could help to moderate these common reactions. Genetic variants are not uniquely powerful, nor uniquely flawed. The way M&H draw parallels with established social science methods helps to make this point.

We agree with M&H that the mechanism behind a cause can act through a complex network of biological, social, and psychological pathways. For most behavioural traits it is difficult to disentangle the purely biological from the complex interactions between those biological manifestations and environments, which in turn can feed into the causal mechanistic process. Hence, the authors define these traits as having "shallow" causes – which are non-unitary, non-uniform, and non-explanatory. We question whether this is a property of their nature, or rather a limit of our understanding. While we cannot currently fine-map every genetic locus and build a mechanistic model of how each genetic variant acts within a causal network, we shouldn't underestimate the potential for scientific advance. Do the authors think there is potential for genetic causes of complex traits to move from shallow to deep as our understanding improves?

Some traits have genetic causes which are clearly deep, for example, phenylketonuria or cystic fibrosis. Their causative genetic variants explain both mechanism and treatment. However, single-genetic variants with large explanatory effects are rare. Complex traits are usually highly polygenic, with each genetic variant exerting only a small effect. Consequently, M&H suggest that their causes are shallow. But occasionally these single variants can produce deep mechanistic insights, even within a

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

complex causal network. For example, a novel genetic variant for Crohn's disease implicated autophagy as a mechanism (Rioux et al., 2007), and individual genetic variants implicated metabolic functions in the development of anorexia nervosa (Watson et al., 2019). Consequently, novel mechanistic insight can come before we have identifed most of the components in a complex causal network. Therefore, we see the line between shallow and deep to change over time as our knowledge advances, rather than as an inherent property of a complex behavioural trait.

In addition, we propose extending the applications of this causal framework to Mendelian randomisation. If we accept the premise that a genetic variant could be causal for trait A then a natural extension is to test whether trait A causes trait B. Let us take M&H's example: Does education (trait A) reduce crime rates (trait B)? One alternative to a randomised education intervention would be Mendelian randomisation, using genetic variants as a proxy for levels of education. The Mendelian randomisation method operates under similar assumptions (and similar limitations) to those laid out in section 3.2 of M&H, following the principle of "genetic inheritance as a natural experiment" (Davey Smith & Hemani, 2014). Mendelian randomisation studies (given their speed and relative low cost) are often posited to be a useful first step to guide future randomisation/intervention studies.

Mendelian randomisation estimates are average causal effects and consequently only constitute *first-generation causal knowledge*. But extensions of the method can help us towards *second-generation causal knowledge*. For example, multivariable Mendelian randomisation tests possible mediation pathways (Sanderson, Davey Smith, Windmeijer, & Bowden, 2019) and factorial Mendelian randomisation tests interaction effects (Rees, Foley, & Burgess, 2020). However, these methods can only test possible mediator/moderator variables from genome-wide association studies (GWASs). To obtain sufficient sample sizes for GWASs, the quality of phenotyping is often poor. This currently limits our ability to move beyond shallow causes using Mendelian randomisation methods alone. We must triangulate Mendelian randomisation results with other study designs (Lawlor, Tilling, & Davey Smith, 2016), which can serve two purposes to help us towards *second-generation causal knowledge*: (1) replication in another sample/context can help us to understand the durability and consistency of the causal effect, and (2) exploring sources of heterogeneity can improve our understanding of the causal pathway, beyond limited GWAS phenotypes.

M&H also highlight heterogeneity in their framework. They discuss how investigating individual differences in treatment response (i.e., effect heterogeneity) can indicate that the causal relationship is dependent on other factors (moderators). If we can identify these moderators, it can inform us about modifiable intervention targets because it deepens our understanding of the causal mechanism. We strongly support this point, and further argue that the same is true for genetic associations: Understanding heterogeneity here can help to identify modifiable targets in the pathway between genes and behaviour that could inform behavioural (or pharmacological) interventions to draw out genetic strengths and mitigate genetic risks. We wish to highlight this as a crucial future direction for the field, which could be conceptualised as hypothesis-free gene–environment interaction (or indeed gene–gene interaction, but we focus our discussion on environments). Most gene–environment interaction studies focus on a specific environmental variable with a plausible mechanism because this is one way to protect against false positives (Moffitt, Caspi, & Rutter, 2005). But with the increasing availability of large sample sizes and more robust statistical methods we could use the heterogeneity in genetic association as a method for identifying relevant (environmental) moderators. One way to do this is to explore why some people are resilient to genetic risk, that is, they have high genetic risk for an outcome but have not yet developed that outcome, akin to the wealth of research exploring protective factors (Armitage et al., 2021).

These two suggested extensions of the causal framework are complementary and could be used straight away with available data and appropriate statistical care, to highlight potential modifying exposures for follow-up, either for more deep phenotyping or to include in intervention studies. Both extensions have the potential to move us closer to second-generation causal knowledge. We believe the field should aspire to this, as well as to moving from shallow to deep causes by pursuing mechanistic information. We commend the framework proposed by M&H for providing the foundation to push the field to pursue these ambitious aims.

Financial support. CMAH is supported by a Philip Leverhulme Prize. REW is supported by a postdoctoral fellowship from the South-Eastern Norway Regional Health Authority (2020024).

Competing interest. None.

References

Armitage, J. M., Wang, R. A., Davis, O. S. P., Collard, P., & Haworth, C. M. A. (2021). Positive wellbeing and resilience following adolescent victimisation: An exploration into protective factors across development. *JCPP Advances*, *1*(2), e12024. https://doi.org/10.1002/jcv2.12024

Davey Smith, G., & Hemani, G. (2014). Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. *Human Molecular Genetics*, *23*(R1), R89–R98.

Lawlor, D. A., Tilling, K., & Davey Smith, G. (2016). Triangulation in aetiological epidemiology. *International Journal of Epidemiology*, *45*(6), 1866–1886.

Moffitt, T. E., Caspi, A., & Rutter, M. (2005). Strategy for investigating interactions between measured genes and measured environments. *Archives of General Psychiatry*, *62*(5), 473–481.

Rees, J. M. B., Foley, C. N., & Burgess, S. (2020). Factorial Mendelian randomization: Using genetic variants to assess interactions. *International Journal of Epidemiology*, *49*(4), 1147–1158. https://doi.org/10.1093/ije/dyz161

Rioux, J. D., Xavier, R. J., Taylor, K. D., Silverberg, M. S., Goyette, P., Huett, A., … Brant, S. R. (2007). Genome-wide association study identifies five novel susceptibility loci for Crohn's disease and implicates a role for autophagy in disease pathogenesis. *Nature Genetics*, *39*(5), 596–604. doi: 10.1038/ng2032

Sanderson, E., Davey Smith, G., Windmeijer, F., & Bowden, J. (2019). An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *International Journal of Epidemiology*, *48*(3), 713–727.

Watson, H. J., Yilmaz, Z., Thornton, L. M., Hübel, C., Coleman, J. R., Gaspar, H. A., … Mattheisen, M. (2019). Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nature Genetics*, *51*(8), 1207–1214.

Behavior genetics and randomized controlled trials: A misleading analogy

Jonathan Michael Kaplan[a] and Kevin Andrew Bird[b]

[a]Philosophy Program, School of History, Philosophy, and Religion, Oregon State University, Corvallis, OR, USA and [b]Department of Horticulture and Ecology, Evolution, and Behavior Program, Michigan State University, East Lansing, MI, USA
Jonathan.Kaplan@oregonstate.edu
kevinbird93@gmail.com

doi:10.1017/S0140525X22002187, e193

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

Abstract

Madole & Harden argue that just as the results of randomized controlled trials (RCTs) represent gains in causal knowledge and are useful, despite their limitations, so too are the findings of human behavior genetics. We argue that this analogy is misleading. Unlike RCTs, the results of human behavior genetics research cannot suggest efficacious interventions, nor point toward future research.

Madole & Harden (M&H) draw a comparison between the results of randomized controlled trials (RCTs) and finding of behavior genetics; just as RCTs give us "causal knowledge," they argue, so too does behavior genetics. The causes identified in both cases, they argue, are shallow causes – they are "non-uniform," "non-unitary," and "non-explanatory," but they are causes nonetheless (target article). And, they argue, just as in the case of RCTs, the "shallow" results can be leveraged into research leading to the establishment of "second-generation causal knowledge," by identifying "processes and contexts through which the effect emerges" (target article, sect. 2.4, para. 5). We disagree and argue that the comparison to RCTs is misleading for both first- and second-generation causal analysis. The "shallowness" of the causal knowledge gained in RCTs does not prevent them from being useful guides to practice; this is not the case with the associations found in behavior genetics.

The point of RCTs, in general, is to find actionable interventions. Although these interventions rely on "shallow" causal knowledge, and so cannot be expected to port reliably between different populations and environments, they *can* be expected to be effective in populations and environments similar to those in the RCT. When we, for example, change our prescribing practices based on the results of an RCT, we expect that the drug we are now prescribing will, on average, yield better results, assuming that the population is relevantly similar to that used by the RCT. Or, to take M&H's example, if we decided to promote early childhood education as a means of improving socially desirable outcomes, the RCT would give us a reason to expect those improvements, ceteris paribus. These actionable interventions represent "first-generation" causal knowledge, according to M&H, and can later potentially provide an entry into mechanistic understandings of causation and explanation, so called "second-generation" causal knowledge.

The situation in behavior genetics is nothing like this. Unlike in the case of RCTs, we cannot change the genetic variants associated with the phenotypic variation – and even if we could, doing so would be wildly irresponsible. A reliable finding that a particular childhood intervention has downstream effects that are significant both statistically and "clinically" (or socially, etc.) suggests a reasonable course of action, even if the relationship between the intervention and the outcome is shallow (non-unitary, non-uniform, non-explanatory). But what would we gain from even an accurate finding that a particular genetic variant was associated with downstream effects? If such knowledge was not "shallow," we might be able to use information gleaned from such genetic associations to intervene environmentally – but the very nature of the shallowness of the relationship prevents our being able to make those kinds of interventions; we cannot change the genes, and gain no information useful for making environmental interventions that could not be equally or better gained from straightforward analyses of the effects of intervention on the traits themselves.

The "shallowness" of behavior genetics findings is generally more problematic than that typically encountered for RCTs. In behavior genetics, the associations between particular loci and behavioral traits are almost absurdly small, and the only reason that they rise to statistical significance at all is that genome-wide association studies (GWASs) can leverage sample sizes in the hundreds of thousands or millions. Even when accumulated across thousands of loci in polygenic scores, the application of polygenic scores is plagued by problems with portability because of effects of both real biological complexity and methodological artifacts (Kaplan & Fullerton, 2022; Matthews, 2022). These issues, in addition to the generally low predictive power, render polygenic scores of little to no use at the individual level (Fusar-Poli, Rutten, van Os, Aguglia, & Guloksuz, 2022; Morris, Davies, & Smith, 2020) and a large amount of uncertainty of estimates appears to hinder the ability to even accurately and consistently stratify individuals into high-risk groups (Ding et al., 2022; Muslimova et al., 2023; Schultz et al., 2022).

Put bluntly, knowing that one drug is, say, 20% more effective than another is actionable; knowing that one allele is associated with a tiny fraction of a percent of the variance in the trait in that population is far less so. Finding associations with a small amount of variance is not necessarily useless; disease GWASs have sometimes highlighted promising mechanisms or pathways even from single-nucleotide polymorphisms (SNPs) with small effects; however, sociobehavioral GWASs have not provided any such specific and discrete pathways. Adding together many such alleles yield no more actionable information than any one of them. Indeed, this points toward another weakness of M&H – it elides the distinction between the results of GWASs and heritability estimates from classic "twin" studies. The former point toward loci that at least in principle *might* be useful for moving from "shallow" to "second-generation" causal analyses (though the tiny effect sizes make the value of this questionable). The latter point toward nothing actionable at all. While accumulating small effects into polygenic scores has increased values of R^2, there has been little elaboration of this kind of mechanistic knowledge revealed by functional enrichment or genetic correlations. This does not mean that such studies are necessarily useless – just that they do not produce the kinds of causal information about the influence of genes on behavior that anything useful can be done with (Kaplan & Turkheimer, 2021; Turkheimer, 2016).

There are a number of other important disanalogies that one might point toward. The *controlled* aspect of RCTs makes "red hair" effects (Jencks, 1972) much less likely (though not impossible). The *randomizing* element of RCTs avoids the problems that GWASs have with cryptic population structure. More generally, RCTs, by their nature, avoid many problems with the intervention of interest covarying systemically with the environment experienced – problems that plague both heritability estimates and GWASs.

In the end, the comparison between the causal information gleaned from RCTs and the results of behavior genetics highlights the weaknesses of the latter, and reveals that they share more in common with observational studies, including their weaknesses.

Competing interest. None.

References

Ding, Y., Hou, K., Burch, K. S., Lapinska, S., Privé, F., Vilhjálmsson, B., … Pasaniuc, B. (2022). Large uncertainty in individual polygenic risk score estimation impacts PRS-based risk stratification. *Nature Genetics*, *54*(1), 30–39.

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

Fusar-Poli, L., Rutten, B. P., van Os, J., Aguglia, E., & Guloksuz, S. (2022). Polygenic risk scores for predicting outcomes and treatment response in psychiatry: Hope or hype? *International Review of Psychiatry*, *34*(7-8), 663–6751.
Jencks, C., Smith, M., Henry, A., Jo Bane, M., Cohen, D., Gintis, H., … Michelson, S. (1972). *Inequality: A reassessment of the effect of family and schooling in America*. Basic Books.
Kaplan, J. M., & Fullerton, S. M. (2022). Polygenic risk, population structure and ongoing difficulties with race in human genetics. *Philosophical Transactions of the Royal Society B*, *377*(1852), 20200427.
Kaplan, J. M., & Turkheimer, E. (2021). Galton's Quincunx: Probabilistic causation in developmental behavior genetics. *Studies in History and Philosophy of Science*, *88*, 60–69.
Matthews, L. J. (2022). Half a century later and we're back where we started: How the problem of locality turned in to the problem of portability. *Studies in History and Philosophy of Science Part A*, *91*, 1–9.
Morris, T. T., Davies, N. M., & Smith, G. D. (2020). Can education be personalised using pupils' genetic data? *eLife*, *9*, e49962.
Muslimova, D., Dias Pereira, R., von Hinke, S., Van Kippersluis, H., Rietveld, C. A., & Meddens, S. F. W. (2023). Rank concordance of polygenic indices. *Nature Human Behavior*, *7*, 802–811. https://doi.org/10.1038/s41562-023-01544-6.
Schultz, L. M., Merikangas, A. K., Ruparel, K., Jacquemont, S., Glahn, D. C., Gur, R. E., … Almasy, L. (2022). Stability of polygenic scores across discovery genome-wide association studies. *Human Genetics and Genomics Advances*, *3*(2), 100091.
Turkheimer, E. (2016). Weak genetic explanation 20 years later: Reply to Plomin et al. (2016). *Perspectives on Psychological Science*, *11*(1), 24–28.

A disanalogy with RCTs and its implications for second-generation causal knowledge

Kate E. Lynch[a,b], Rachael L. Brown[c], Jeremy Strasser[d] and Shang Long Yeo[e]

[a]Department of Philosophy, University of Sydney, Sydney, Australia; [b]Charles Perkins Centre, University of Sydney, Sydney, Australia; [c]Centre for Philosophy of the Sciences, School of Philosophy, Australian National University, Canberra, Australia; [d]School of Philosophy, Australian National University, Canberra, Australia and [e]Department of Philosophy, National University of Singapore, Singapore, Singapore
kate.lynch@sydney.edu.au
www.katelynch.net
rachael.brown@anu.edu.au
http://rachaelbrown.net
jeremy.strasser@anu.edu.au
phiysl@nus.edu.sg
http://shang.isaphilosopher.com

doi:10.1017/S0140525X22002242, e194

Abstract

We are less optimistic than Madole & Harden that family-based genome-wide association studies (GWASs) will lead to significant second-generation causal knowledge. Despite bearing some similarities, family-based GWASs and randomised controlled trials (RCTs) are not identical. Most RCTs assess a relatively homogenous causal stimulus as a treatment, whereas GWASs assess highly heterogeneous causal stimuli. Thus, GWAS results will not translate so easily into second-generation causal knowledge.

We agree that family-based genome-wide association studies (GWASs) are an improvement on traditional GWASs in their ability to rule out confounding common causes. We are, however, sceptical that family-based GWASs will guide research aimed at identifying interventions on non-genetic second-generation variables that can be put to practical use in a manner akin to randomised controlled trials (RCTs). Our scepticism stems from an overlooked disanalogy between family-based GWASs and RCTs – *the heterogeneity of the causal stimulus* – and its impact on non-uniformity.

In most RCTs, individuals in the treatment group receive the same, or as similar as possible, treatment or causal stimulus, such as a drug or educational intervention (*causal stimulus homogeneity*). The same is true of Mendelian randomisation trials (which perhaps inspired Madole & Harden's [M&H's] arguments) – the causal variable(s) being investigated are relatively homogenous exposures. In contrast, family-based GWASs make claims about the average treatment effects of thousands of genetic variations distributed across a population. This aggregation is assigned a single variable, "genes," which can be demonstrated as causal to an outcome to some degree, but nonetheless shows high *causal stimulus heterogeneity*.

Causal stimulus heterogeneity differs from the *heterogeneity of treatment effects* – a feature of both RCTs and GWASs – which M&H discuss in the article. The heterogeneity of treatment effects concerns the non-uniform effects of causal stimuli because of interactions with background factors like physiology and environment. This type of non-uniformity is seen in both GWAS and RCT studies, but is particularly challenging for GWASs because of causal stimulus heterogeneity. In GWASs, the complex role of the environment in the expression of the phenotype is amplified because the causal stimulus is varied and heterogeneous between individuals (Lynch, 2021). This non-uniformity means that the associations GWASs uncover between phenotypes and large aggregates of gene variants are very difficult to connect to mechanisms and function (Matthews & Turkheimer, 2022). A similar challenge occurs in microbiome research: Significant within-population physiological and environmental variations make it difficult to track pathways between microbes and outcomes, limiting the scope for causal inference (Lynch, Parke, & O'Malley, 2019).

Causal stimulus heterogeneity increases non-uniformity and hampers tracing of mechanisms. This is because of variation in nature of "the same" treatment upon subjects. A simple hypothetical illustrates this well. Consider three different drug treatments: First, a drug with a single active ingredient (e.g., lithium). Second, a drug with thousands of ingredients of small efficacy. Third, a drug with thousands of ingredients of small efficacy, where each pill has one ingredient or an alternative at random according to a defined chance procedure. In all three cases, an RCT can determine whether the treatment drug has an average effect compared to a control, and thereby generate first-generation causal knowledge. This is possible even in the face of non-uniformity because of the heterogeneity of treatment effects. The prospect for these results to advance second-generation causal knowledge diminishes, however, across the three cases. The high causal stimulus heterogeneity in the third case is likely to produce non-uniform causal pathways from the very first steps, thus making it difficult or impossible to trace mechanisms from particular drug ingredients given only associations between treatments and outcomes.

A high degree of causal stimulus heterogeneity is typical for GWASs, including family-based ones. To analogise with our hypothetical drug case: The first drug is akin to a single-gene cause, the second a specific aggregate of genes, and the third an

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

aggregate of many genes, which at the individual level is summarised by a polygenic risk score. Polygenic risk scores are highly heterogeneous causal stimuli with non-uniform effects that make it extremely difficult to trace mechanisms from particular genetic ingredients in the causal stimulus, in all but the simplest cases of gene expression (where GWASs are unnecessary). Even if we could hold environments fixed between individuals (thereby reducing the potential for non-uniformity because of background conditions), in GWASs there is typically too much variation between individuals in how the causal stimulus works at the "lower" biological level for effective "bottom-up" investigations of intermediate causes through biological mechanisms. In our view, this largely precludes these studies from providing the sort of causal knowledge required to identify mechanisms and intermediaries for investigation with second-generation studies.

A work-around might be Harden's proposal of phenotypic annotation, which rests on the statistical investigation of intermediate causes through mediation analysis (Belsky & Harden, 2019; see, e.g., Belsky et al. [2016]). Mediation analysis test variables correlated with the stimulus to determine whether (and to what extent) they mediate the causal paths from stimulus to outcome. Such intermediaries could be possible targets for intervention in second-generation studies. In the simplest case, the genetic stimulus would act as an instrumental variable on the potential intermediary (as in Mendelian randomisation, see Davey Smith & Ebrahim, 2003) allowing for measurement of the intermediary's causal effect. However, mediation analysis is tricky at the best of times (see Pearl [2014] for a principled approach). The possibility of confounding common causes between intermediary and outcome is a serious challenge. Common causes (such as other genetic or environmental causes) may account for the relationship between potential intermediary and outcome. In this case, intervening on the potential intermediary will not cause the outcome. A heterogeneous causal stimulus, such as a polygenic risk score, effectively carries a potential common cause within itself: The different ways that the causal stimulus may be realised. The use of an average causal stimulus (by definition) precludes control of this common cause. To determine if a potential intermediary is indeed a cause of the outcome, one would need to do an RCT or another kind of study.

In conclusion, we agree that family-based GWASs provides one kind of causal information that has been missing from traditional heritability and GWASs (see Lynch [2017] for the limitations of causal heritability claims). Unfortunately, the general heterogeneous nature of the genetic variation studied means that this information will not translate easily into second-generation causal knowledge.

Financial support. KEL was supported under Australian Research Council's Discovery Projects funding scheme (project number FL170100160); RLB was supported by funding through the ANU VC Futures Scheme.

Competing interest. None.

References

Belsky, D. W., & Harden, K. P. (2019). Phenotypic annotation: Using polygenic scores to translate discoveries from genome-wide association studies from the top down. *Current Directions in Psychological Science*, *28*, 82–90. doi: 10.1177/0963721418807729

Belsky, D. W., Moffitt, T. E., Corcoran, D. L., Domingue, B., Harrington, H., Hogan, S., … Caspi, A. (2016). The genetics of success: How single-nucleotide polymorphisms associated with educational attainment relate to life-course development. *Psychological Science*, *27*, 957–972. doi: 10.1177/0956797616643070

Davey Smith, G., & Ebrahim, S. (2003). "Mendelian randomization": Can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology*, *32*(1), 1–22.

Lynch, K. E. (2017). Heritability and causal reasoning. *Biology & Philosophy*, *32*(1), 25–49.

Lynch, K. E. (2021). The meaning of cause in genetics. In G. D. Smith, R. Richmond, & J.-B. Pingault (Eds.), *Cold Spring Harbour Perspectives in Medicine*, *11*(9), A0405019. doi: 10.1101/cshperspect.a040519

Lynch, K. E., Parke, E, & O'Malley M. (2019). How causal are microbiomes? A comparison with the *Helicobacter pylori* explanation of ulcers. *Biology and Philosophy*, *34*, 62.

Matthews, L. J., & Turkheimer, E. (2022). Three legs of the missing heritability problem. *Studies in History and Philosophy of Science*, *93*, 183–191.

Pearl, J. (2014). Interpretation and identification of causal mediation. *Psychological Methods*, *19*(4), 459.

Meeting counterfactual causality criteria is not the problem

Kristian E. Markon

Department of Psychological and Brain Sciences, University of Iowa, Iowa City, IA, USA
kristian-markon@uiowa.edu

doi:10.1017/S0140525X22002138, e195

Abstract

Counterfactual causal interpretations of family genetic effects are appropriate, but neglect an important feature: Provision of unique information about expected outcomes following an independent decision, such as a decision to intervene. Counterfactual causality criteria are unlikely to resolve controversies about behavioral genetic findings; such controversies are likely to continue until counterfactual inferences are translated into interventional hypotheses and designs.

Madole & Harden (M&H) compellingly argue that estimated within-family genetic effects can be interpreted as causes in much the same way as randomized controlled trials. Increased focus on within-family molecular genetic designs is needed for many reasons, and M&H illustrate the power of such designs in understanding human neurophysiology and behavior.

Their argument relies on counterfactual accounts of causality, however. Counterfactual accounts have many strengths, but have been equally criticized, usually because of their dependence on states that are by definition impossible to observe (Dawid, 2000). Counterfactual causal accounts attempt to estimate what would have been the case in a different set of conditions that did not and cannot eventuate, leaving a user of such models in a predicament about how to use them: If the causal inference is about something in the past that did not occur, what about the present that did occur, and the future?

Traditional causal accounts and many contemporary accounts (e.g., Granger causality; Dawid, 2015; Granger, 1969; Janzing, Balduzzi, Grosse-Wentrup, & Schölkopf, 2013; Schreiber, 2000), address a different, actionable question: To what extent does a potential cause provide unique information about expected outcomes following an independent decision, such as a decision to implement a manipulation or intervention? Designs developed within these paradigms, such as randomized controlled trials,

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

directly address this type of question, in that a decision is randomly made so as to instantiate independence from past states, and the outcomes of this decision are then observed.

Counterfactual theory suggests that it also provides this information, by estimating what would have happened if a decision of sorts by nature had been made, assuming it could have been made. However, this relies on a number of assumptions about that alternate possible world that might or might not be true. In an important sense, moreover, it takes an unactionable epistemological stance: Even if a counterfactual account informs about what would have occurred had things been different, it does not inform about what one can do now, given things as they are.

This is a critical distinction given the nature of causal pathways involved in behavior genetics. Usually interest is not actually in the immediate effects of the genotype *per se* – the nucleotide sequences at different loci and their translation – but the affected neurodevelopmental processes, those effects on neurophysiology at a later time, and their effects on experience and behavior. The black box between gene and behavior is in fact the phenomenon often most of interest in terms of causal explanation. In the polygenic risk regime, where each polymorphism might have an almost unmeasurable unique effect on phenotype, effects of a particular polymorphism or haplotype may be still further removed from the aggregate neurodevelopmental endpoint of primary interest.

There are some ways to construe genetic effects in terms of decision outcome information. For instance, it might be argued that genotype provides additional predictive information about the outcomes of particular decisions for particular individuals – that is, in deciding what intervention to provide to whom, above and beyond any nongenetic data. This is a reasonable argument, but again, with numerous intraindividual and extraindividual inputs into behavioral development, it may be that downstream predictors provide more information about behavior, being causally more proximal mediators of any genetic effects (Morris, Davies, & Davey Smith, 2020). If a gene is one of many causes of an easily identifiable condition, to treat the condition isn't it more efficient to identify those with the condition, rather than the gene? Moreover, in such a setting the emphasis is still on how to improve efficacy of an intervention, such as an educational or medical intervention.

Another way to construe genetic effects in terms of decision outcome information is in terms of genetic manipulation. Gene editing is a reality (Anguela & High, 2019; Saha et al., 2021), and randomized controlled trials of genotype manipulation may become salient considerations in the neurobehavioral sciences sooner than is often appreciated. Given developmental cascades (Elam, Lemery-Chalfant, & Chassin, in press), it is likely that if genotypes are not altered early in neurodevelopment, in many cases genetic effects likely will be irreversible. In that case, the options for intervention are again further downstream from genes both in time and causal proximity to experience and behavior. Alternatively, one could implement preventative gene editing, but that would be a form of eugenics with all the attendant ethical challenges it implies.

M&H's argument for counterfactual causality is telling in that it implies behavior genetic designs have so far often *not* been seen as causally compelling – otherwise their argument would be unnecessary. Reluctance to construe behavior genetic effects in terms of causes has likely been because of numerous factors, including limitations of common designs, such as lack of precise genotypic information or lack of within-family controls for variables varying between families. However, the reluctance also arguably reflects a perception that behavior genetic studies have generally not provided information about what is changeable or targetable. Translating behavioral genetic effects into interventional hypotheses and designs, where the information they provide can be leveraged to prevent and treat, will likely increase the reception and perceived relevance of such findings. Sometimes causes have effects that, set into motion in a causal chain, are impossible to reverse. But the actionable, effective relevance of the causal chain only goes back so far.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interest. None.

References

Anguela, X. M., & High, K. A. (2019). Entering the modern era of gene therapy. *Annual Review of Medicine*, *70*(1), 273–288. doi:10.1146/annurev-med-012017-043332

Dawid, A. P. (2000). Causal inference without counterfactuals. *Journal of the American Statistical Association*, *95*, 407–424. doi:10.1080/01621459.2000.10474210

Dawid, A. P. (2015). Statistical causality from a decision-theoretic perspective. *Annual Review of Statistics and Its Application*, *2*(1), 273–303. doi:10.1146/annurev-statistics-010814-020105

Elam, K., Lemery-Chalfant, K., & Chassin, L. (in press). A gene–environment cascade theoretical framework of developmental psychopathology. *Journal of Psychopathology and Clinical Science*.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*, 424–438.

Janzing, D., Balduzzi, D., Grosse-Wentrup, M., & Schölkopf, B. (2013). Quantifying causal influences. *Annals of Statistics*, *41*(5), 2324–2358. doi:10.1214/13-AOS1145

Morris, T. T., Davies, N. M., & Davey Smith, G. (2020). Can education be personalised using pupils' genetic data? *eLife*, *9*, e49962. doi:10.7554/eLife.49962

Saha, K., Sontheimer, E. J., Brooks, P. J., Dwinell, M. R., Gersbach, C. A., Liu, D. R., … Zhou, J. (2021). The NIH somatic cell genome editing program. *Nature*, *592*(7853), 195–204. doi:10.1038/s41586-021-03191-1

Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters*, *85*, 461.

Benefits of hereditarian insights for mate choice and parenting

Geoffrey F. Miller

Psychology Department, University of New Mexico, Albuquerque, NM, USA
gfmiller@unm.edu;
https://psych.unm.edu/people/faculty/profile/geoffrey-miller.html

doi:10.1017/S0140525X22002102, e196

Abstract

Madole & Harden develop some good ideas about how to understand genetic causality more clearly, but they frame the benefits of behavior genetics research at a largely collective level, focused on the pros and cons of different ways to engineer the gene pool or social behavior. This neglects the individual benefits of hereditarian insights for mate choice and parenting.

The rise of genome-wide association studies (GWASs) and polygenic scores has triggered a new moral panic about behavior genetics. In their target article, Madole & Harden (M&H) aim to neutralize some of this panic by clarifying the causal inference

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

logic behind GWAS research. They identify some illuminating similarities between within-family genetic effects (given the sexual randomization of parental genotypes), and average treatment effects in randomized controlled trials.

However, their implicit sociopolitical framework risks handicapping the utility of behavior genetics for ordinary people. M&H write as if there are only two main ways that behavior genetics could influence humanity: collective eugenics (changing the gene pool through government policy, or through new cultural norms around embryo screening or CRISPR gene editing), or collective social engineering (using insights into gene–brain–behavior pathways to identify new biopsychosocial interventions that can nudge behavior in socially valued directions – such as reducing violent crime, mental illness, or obesity).

This collectivist viewpoint asks what behavior genetics can do for society at large – rather than for individuals making decisions about their own families. For example, M&H seek scientific insights about how to "identify novel targets for intervention," how to "isolate steps in the causal path that serve as candidates for intervention," and how to identify "targets for programmatic manipulation that may serve to close the gap in health disparities."

An implicit question behind their paper seems to be: How can we make sure that behavior genetics findings are used by good liberal policymakers to solve the social problems prioritized by good liberals – rather than being turned into coercive, racist, neo-Nazi eugenics? They explicitly wish to "challenge the genetic determinism and essentialism that have historically characterized the pernicious misapplications of genetics by political extremists" on the Right (without mentioning the pernicious misapplications of Blank Slate social-constructivism by political extremists on the Left, such as Stalin, Pol Pot, Mao, etc.). Given their aversion to any kind of organized eugenics, they imply that the only other benefit of behavior genetics could be to promote "second-generation causal knowledge" that can make environmental interventions (ranging from new pharmaceuticals to new educational interventions) more effective in bio-hacking the complex genes-to-behaviors pathways.

Their distinction between "deep causes" and "shallow causes" exemplifies some hidden problems in their sociopolitical framework. "Deep causes" (with unitary, uniform, causally explanatory genetic effects) – for example, the way that homozygous mutations in the CFTR gene lead to cystic fibrosis – are seen as the gold standard. "Shallow causes" (with local, probabilistic, causally distal genetic effects) are seen as little more than stepping stones toward further mechanistic biomedical research that can identify new causal pathways for the collective melioration (or manipulation) of behavior.

The trouble is, the most heritable, stable, predictive, and important behavioral traits, such as general intelligence, the Big Five personality traits, and mental disorders, generally show "shallow" rather than "deep" genetic influences. As M&H emphasize, these massively polygenic traits are shaped by thousands of genetic variants that influence intricate neurodevelopmental systems in ways that aren't very mechanistically informative about which biomedical or sociocultural interventions might work.

So, shallow genetic influences won't help the social engineer very much in finding biopsychosocial interventions to reshape society. Yet, for an individual making important life-choices that can shape the next generation of their own family, the massively polygenic core psychological traits are much easier to influence and/or much more important to understand than severe single-gene mutations. When it comes to real-life issues in mating and parenting, where an accurately hereditarian view could guide better decision making, and where a Blank Slate view could lead to high-cost errors and lifelong regrets, "shallow causes" run pretty deep.

Mate choice has been a form of intuitive eugenics at least 540 million years, ever since the Cambrian explosion, when complex senses and centralized nervous systems started to guide sexual selection through mate choice for good genes. People tend to focus their mate choice on traits that are at least moderately heritable, such as physical health, mental health, general intelligence, personality traits, and moral, political, or religious values (Miller, 2000). Mate choice across millions of animal species exerts a high degree of "counterfactual control" over what kinds of offspring get produced, even though animals have no conscious understanding of the "counterfactual dependence" whereby genes influence traits. However, Blank Slate ideologies (Pinker, 2003) have led many people to mistrust their own mate choice preferences and hereditarian intuitions, with potentially disastrous results when choosing a marriage partner, gamete donor, or child to adopt.

Likewise with parenting. In many countries, Blank Slate ideology leads parents to act as if they can micromanage the outcome of their kids' development through intensive hot-house parenting, overprotective coddling, and tiger-mothering. For example, in China there has been much debate about "involution" (内卷), meaning runaway competition in education and labor markets (Yi et al., 2022). This leads parents to put enormous stress on kids to achieve at any cost – regardless of their innate abilities, personalities, and interests. The less that parents believe genes matter, the more frustrated they get by their kids' mediocrities, and the more they blame themselves for failing to provide the right shared family environment. A humbler, more hereditarian perspective allows parents to relax, trust the mate choices they made, trust their genes, and let their kids follow their own path, without the burden of trying to optimize the shared family environment in every possible way (Caplan, 2012).

M&H seem frustrated that they're stuck at this collective policymaking level: "Even if we concede that, at a conceptual level, genes *could* cause average differences in human behavior, at a practical level, it is not readily apparent what we would *do* with this knowledge." Maybe the most important thing *we* can do with this knowledge, as individuals, is to reject the Blank Slate dogma and accept the importance of heredity when choosing our mates, raising our kids, and getting on with our lives.

Competing interest. None.

References

Caplan, B. (2012). *Selfish reasons to have more kids*. Basic Books.

Miller, G. F. (2000). *The mating mind: How sexual choice shaped the evolution of human nature*. Doubleday.

Pinker, S. (2003). *The Blank Slate: The modern denial of human nature*. Penguin.

Yi, D., Wu, J., Zhang, M., Zeng, Q., Wang, J., Liang, J., & Cai, Y. (2022). Does involution cause anxiety? An empirical study from Chinese universities. *International Journal of Environmental Research and Public Health*, *19*, 9826. doi:10.3390/ijerph19169826

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

The providential randomisation of genotypes

Jean-Baptiste Pingault[a], Pasco Fearon[a], Essi Viding[a], Neil Davies[b,c,d], Marcus R. Munafò[b] and George Davey Smith[b]

[a]Department of Clinical, Educational and Health Psychology, University College London, London, UK; [b]Medical Research Council Integrative Epidemiology Unit at the University of Bristol, Bristol, UK; [c]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK and [d]K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Norwegian University of Science and Technology, Trondheim, Torgarden, Norway
j.pingault@ucl.ac.uk
p.fearon@ucl.ac.uk
e.viding@ucl.ac.uk
Neil.Davies@bristol.ac.uk
marcus.munafo@bristol.ac.uk
KZ.Davey-Smith@bristol.ac.uk
www.jeanbaptistepingault.com
https://www.cfr.cam.ac.uk/staff/professor-pasco-fearon
https://www.ucl.ac.uk/pals/people/essi-viding
https://research-information.bris.ac.uk/en/persons/neil-m-davies
https://research-information.bris.ac.uk/en/persons/marcus-r-munafo
https://www.bristol.ac.uk/people/person/George-Davey%20Smith-285dce3f-4498-4e97-82de-250a865b4483/

doi:10.1017/S0140525X2200214X, e197

Abstract

When building causal knowledge in behavioural genetics, the natural randomisation of genotypes at conception (approximately analogous to the artificial randomisation occurring in randomised controlled trials) facilitates the discovery of genetic causes. More importantly, the randomisation of genetic material within families also enables a better identification of (environmental) risk factors and aetiological pathways to diseases and behaviours.

Madole & Harden draw parallels between randomised controlled trials (RCTs) and within-family genetic association designs to elaborate on the notion of genetic causation, that is, whether genes cause behaviours and how to interpret genetic causes. The article is a thoughtful introduction to these topics. Our comment focuses on the core feature shared by both designs, that is, randomisation. First, we discuss how to best capitalise on natural randomisation to help build causal knowledge. Second, despite randomisation being a core feature of both, we caution against drawing too literal parallels between the two designs.

The parallel drawn by the authors was made explicitly by Fisher who established a direct filiation between the (artificially) randomised design he theorised and the (natural) randomisation of genetic material at conception, in his words: "the factorial method of experimentation, now of lively concern so far afield as the psychologists, or the industrial chemists, derives its structure and its name, from the simultaneous inheritance of Mendelian factors.… Genetics is indeed in a peculiarly favoured condition in that Providence has shielded the geneticist from many of the difficulties of a reliably controlled comparison. The different genotypes possible from the same mating have been beautifully randomised by the meiotic process" (Fisher, 1952). As highlighted by the authors, this randomisation within families can help to establish genetic causation. Importantly, genetic causation is always indirect and happens entirely through (molecular and environmental) phenotypes. As such, rather than focusing on genetic causation, perhaps the greatest opportunity for building causal knowledge in behavioural genetics lies in leveraging the beautifully randomised process mentioned by Fisher to understand phenotypic causation in general. To that end, a method called Mendelian randomisation uses genetic variants as instrumental variables to assess phenotypic causation and identify (potentially modifiable) risk factors (Davey Smith & Ebrahim, 2003; Richmond & Davey Smith, 2022; Sanderson et al., 2022). Many Mendelian randomisation studies have focused on complex traits, including a within-family Mendelian randomisation study of the impact of educational attainment on physical health and mortality (Howe et al., 2022b). In addition, the authors rightly state that identifying a causal genetic variant often does not, *per se*, provide insights into causal pathways leading to behaviours. However, novel methods are rapidly developing that leverage genetic variants to understand causal pathways and mechanisms underlying complex phenotypes. Many of these methods directly harness the randomisation of genetic material at conception. For example, recent methods extend Mendelian randomisation to systematically investigate the aetiological role played by gene expression or DNA methylation (Hannon et al., 2018; Porcu et al., 2021). Genetically informed methods for causal inference aiming to identify (environmental) risk factors and biological pathways have been reviewed extensively elsewhere (Davey Smith, Richmond, & Pingault, 2021; Pingault, Richmond, & Davey Smith, 2022).

Critically, the analogy drawn between within-family genetic association studies and RCTs needs to be used with care. In an RCT, the treatment should be well defined (e.g., a given dose of a drug) and can be administered to individuals. By contrast, generally, a genetic variant cannot be administered or modified during the life course. Thus, while RCTs can provide actionable evidence of a specific intervention's efficacy, a within-family genetic association only indicates the effect of inheriting one variant or another. The difference in timing is also essential as inheriting a genetic variant at conception leads to a lifelong exposure as opposed to a time-bound treatment.

Furthermore, the authors argue that genetic causes relevant to behavioural genetics are analogous to causes uncovered by RCTs in that they are shallow – non-unitary (no single isolable cause), non-uniform (people exposed have heterogeneous outcomes), and non-explanatory (not mechanistically informative). However, this analogy is strained for both monogenic and polygenic disorders. A gene implicated in a monogenic disorder such as cystic fibrosis is close to a well-defined and actionable treatment: Although the gene is not, in itself, a treatment, it can potentially be targeted by gene editing techniques such as CRISPR-Cas9 to correct deleterious variants (note that even for monogenic disorders, many deleterious variants can be involved and lead to different phenotypic manifestations) (Jinek et al., 2012). Successful human clinical trials using gene editing techniques for monogenic disorders are emerging (Frangoul et al., 2021). Many rare developmental disorders share a similar genetic architecture, some with relevance to behaviour. For example, adrenoleucodystrophy is a monogenic developmental disorder, which may first manifest with behavioural and cognitive difficulties and can be lethal (Zhu et al., 2020). Even for polygenic disorders such as schizophrenia, some rare variants considerably

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

increase the risk of disease, with odds ratios up to 50 (Singh et al., 2022). As such, genetic causes relevant to behavioural genetics need not be shallow. Variants underlying monogenic disorders or high-risk rare variants may be better conceived as deep causes, that is, close to unitary, uniform, and explanatory.

Conversely, polygenic influences of small effects underlying complex behaviours can indeed be conceived as shallow causes. However, in this case, the "treatment" is not well defined (in content or timing), cannot be refined or changed to increase mechanistic insight and is not directly actionable. Even if extensions of techniques such as CRISPR-Cas9 could theoretically target hundreds of genetic variants at once, this could never be a treatment strategy given unknown and potentially devastating side effects (like a drug RCT consisting of the simultaneous administration of hundreds of compounds). In sum, causes established by RCTs and genetic causes derived from within-family association studies do not necessarily share many features beyond the core concept of randomisation. Further discussion of the notion of cause in genetics and the parallels between RCTs and genetically informed methods such as Mendelian randomisation are available elsewhere (Lynch, 2021; Nitsch et al., 2006).

In conclusion, we agree that behavioural genetics should look to provide causal knowledge. To that end, perhaps the most useful will be exploiting genetic data to understand phenotypic causation and aetiological pathways. Genetically informed designs for causal inference and, in particular, within-family designs, can play a key role in improving aetiological understanding and, ultimately, prevention and treatment (Howe et al., 2022a; Hwang, Davies, Warrington, & Evans, 2021; Pingault et al., 2018).

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interest. None.

References

Davey Smith, G., Richmond, R., & Pingault, J.-B. (Eds.) (2021). *Combining human genetics and causal inference to understand human disease and development*. Cold Spring Harbor Laboratory Press.

Davey Smith, G. D., & Ebrahim, S. (2003). "Mendelian randomization": Can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology*, *32*(1), 1–22.

Fisher, R. (1952). Statistical methods in genetics. *Heredity*, *6*, 1–12 (reprinted in 2010, Int. J. Epidemiol., 39: 329–335). https://doi.org/10.1093/ije/dyp379

Frangoul, H., Altshuler, D., Cappellini, M. D., Chen, Y.-S., Domm, J., Eustace, B. K., … Corbacioglu, S. (2021). CRISPR-Cas9 gene editing for sickle cell disease and *β*-thalassemia. *New England Journal of Medicine*, *384*(3), 252–260. https://doi.org/10.1056/NEJMoa2031054

Hannon, E., Gorrie-Stone, T. J., Smart, M. C., Burrage, J., Hughes, A., Bao, Y., … Mill, J. (2018). Leveraging DNA-methylation quantitative-trait loci to characterize the relationship between methylomic variation, gene expression, and complex traits. *American Journal of Human Genetics*, *103*(5), 654–665. https://doi.org/10.1016/j.ajhg.2018.09.007

Howe, L. J., Nivard, M. G., Morris, T. T., Hansen, A. F., Rasheed, H., Cho, Y., … Davies, N. M. (2022a). Within-sibship genome-wide association analyses decrease bias in estimates of direct genetic effects. *Nature Genetics*, *54*(5), 581–592. https://doi.org/10.1038/s41588-022-01062-7

Howe, L. J., Rasheed, H., Jones, P. R., Boomsma, D. I., Evans, D. M., Giannelis, A., … Davies, N. M. (2022b). Educational attainment, health outcomes and mortality: A within-sibship Mendelian randomization study. *medRxiv*, 2022.01.11.22268884. https://doi.org/10.1101/2022.01.11.22268884

Hwang, L.-D., Davies, N. M., Warrington, N. M., & Evans, D. M. (2021). Integrating family-based and Mendelian randomization designs. *Cold Spring Harbor Perspectives in Medicine*, *11*(3), a039503. https://doi.org/10.1101/cshperspect.a039503

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)*, *337*(6096), 816–821. https://doi.org/10.1126/science.1225829

Lynch, K. E. (2021). The meaning of "cause" in genetics. *Cold Spring Harbor Perspectives in Medicine*, *11*(9), a040519. https://doi.org/10.1101/cshperspect.a040519

Nitsch, D., Molokhia, M., Smeeth, L., DeStavola, B. L., Whittaker, J. C., & Leon, D. A. (2006). Limits to causal inference based on Mendelian randomization: A comparison with randomized controlled trials. *American Journal of Epidemiology*, *163*(5), 397–403. https://doi.org/10.1093/aje/kwj062

Pingault, J.-B., O'Reilly, P. F., Schoeler, T., Ploubidis, G. B., Rijsdijk, F., & Dudbridge, F. (2018). Using genetic data to strengthen causal inference in observational research. *Nature Reviews Genetics*, *19*, 566–580.

Pingault, J.-B., Richmond, R., & Davey Smith, G. (2022). Causal inference with genetic data: Past, present, and future. *Cold Spring Harbor Perspectives in Medicine*, *12*(3), a041271. https://doi.org/10.1101/cshperspect.a041271

Porcu, E., Sjaarda, J., Lepik, K., Carmeli, C., Darrous, L., Sulc, J., … Kutalik, Z. (2021). Causal inference methods to integrate omics and complex traits. *Cold Spring Harbor Perspectives in Medicine*, *11*(5), a040493. https://doi.org/10.1101/cshperspect.a040493

Richmond, R. C., & Davey Smith, G. (2022). Mendelian randomization: Concepts and scope. *Cold Spring Harbor Perspectives in Medicine*, *12*(1), a040501. https://doi.org/10.1101/cshperspect.a040501

Sanderson, E., Glymour, M. M., Holmes, M. V., Kang, H., Morrison, J., Munafò, M. R., … Davey Smith, G. (2022). Mendelian randomization. *Nature Reviews Methods Primers*, *2*(1), 1–21. https://doi.org/10.1038/s43586-021-00092-5

Singh, T., Poterba, T., Curtis, D., Akil, H., Al Eissa, M., Barchas, J. D., … Daly, M. J. (2022). Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature*, *604*(7906), 509–516. https://doi.org/10.1038/s41586-022-04556-w

Zhu, J., Eichler, F., Biffi, A., Duncan, C. N., Williams, D. A., & Majzoub, J. A. (2020). The changing face of adrenoleukodystrophy. *Endocrine Reviews*, *41*(4), 577. https://doi.org/10.1210/endrev/bnaa013

Polygene risk scores and randomized experiments

Lauren N. Ross[a], Kenneth S. Kendler[b] and James F. Woodward[c]

[a]Logic and Philosophy of Science Department, University of California-Irvine, Irvine, CA, USA; [b]Virginia Institute for Psychiatric and Behavioral Genetics and Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA and [c]Department of History and Philosophy of Science, University of Pittsburgh, Pittsburgh, PA, USA
rossl@uci.edu
kenneth.kendler@vcuhealth.org
jfw@pitt.edu
https://www.lps.uci.edu/~rossl/
https://vipbg.vcu.edu/people/kenneth-kendler/
https://jameswoodward.org

doi:10.1017/S0140525X22002151, e198

Abstract

We explore Madole & Harden's (2022) suggestion that single-nucleotide polymorphism (SNP)/trait correlations are analogous to randomized experiments and thus can be given a causal interpretation.

We commend Madole & Harden (M&H) for their lucid discussion of the sense in which genes or single-nucleotide polymorphisms (SNPs) may legitimately be regarded as causes of behavioral traits. We agree with much of what they say but welcome clarification on some issues.

M&H adopt a broadly "interventionist" treatment of causation – the minimal condition for some factor C to count as a cause for an outcome E is that if, hypothetically, unconfounded manipulations of C were to be performed these would lead to changes in E. In the familiar case of a randomized experiment, this leads

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

to the conclusion that an average causal effect (ACE) is a legitimate causal notion. M&H observe that an ACE can be present even though C does not have a uniform effect, even though a similar ACE may not be present in populations different from the population from which the experimental sample was drawn, and even though the experiment tells us nothing about the mechanism by which Cs cause Es. We agree.

M&H suggest that because of the random nature of meiosis, SNP/trait correlations from genome-wide association studies (GWASs) and/or the polygene risk scores (PRSs) that incorporate these (or more precisely, such correlations among full siblings) can be likened to ACEs and hence given a causal interpretation. We explore this claim.

Consider a set of fertilized eggs immediately after conception drawn in a representative fashion from some population. Suppose this set is divided randomly into two groups, such that at a particular SNP position, one nucleotide is experimentally imposed, say A, while for the other group a different nucleotide, for example, C, is imposed. Also suppose that the environments E are uniform across the two groups. Then, any difference in the incidence of some trait T across the two groups can be regarded as the ACE of having A rather than C in that population and environment.

This is not an experiment that is currently technologically possible or morally acceptable. We introduce it only to provide some intuition for what a randomized experiment involving SNP manipulation that provides information about an ACE would look like. If we consider SNP/trait correlations from a GWAS, there are critical differences with the experiment just described. Even putting aside population stratification, the random nature of meiosis does not ensure that individuals with A at some locus in comparison with those with C at that locus are causally similar in other respects (as a genuine randomized experiment does). This is because of linkage disequilibrium – the A/C difference is very likely correlated with other causally relevant differences (often unobserved) nearby in the subjects' genomes that affect trait T. Indeed, the evidence is that most SNPs reported in a GWAS are not causal for traits of interest but are rather merely correlated with factors that are causal – a point recognized by M&H when they suggest that most SNPs have the status of "indicator" variables, tracking through correlations other factors that are causal.

Moreover, there is another, more subtle disanalogy with the randomized experiment described above. In that experiment, a single treatment – for example, A versus C – is randomly imposed on the population. Assuming the random nature of meiosis, a GWAS corresponds to a huge number of *different* randomized treatments in the population: for example, A versus C at SNP1, G versus T at SNP2, and so on. An analogy would be an experiment in which a large number of different drugs D1…D*n* are simultaneously randomly assigned to subjects with unknown correlations among the assignments. Indeed, matters are even more complex because haplotypes are randomized not SNPs. We might perhaps conceptualize this as the assignment of randomized bottles to subjects, each containing a mixture of different drugs. Neither of these scenarios has the straightforward causal interpretation of a standard randomized experiment.

Are these problems ameliorated if, as M&H suggest, one only compares full siblings? This will help with confounds having to do with population stratification and also help, at least somewhat, with potential environmental confounds (to the extent the sibs are exposed to similar environments). However, the challenges posed by genetic linkage remain – given a correlation between, for example, the presence of A at some SNP and trait T, we still don't know whether A is causal for T or merely correlated with some genetic factor that is causal. M&H acknowledge this, suggesting that we should regard the causal factors as whole haplotype blocks.

One problem with this is that haplotype blocks are overly broad candidates for causes, in the sense that although these will contain causally relevant factors, they will also contain many more factors that are causally irrelevant, with no information about which is which. In this respect, citing a haplotype block as a cause seems analogous to saying that something unknown in my refrigerator causes an odor – not false but not particularly informative. Moreover, we wonder whether such a causal interpretation of SNP/trait correlations is necessary. As M&H suggest, one important role for such information is as a control; allowing us to see the causal role of other non-genetic (environmental) variables. Correlational information not having a straightforward causal interpretation can function as such a control as long as it is correlated with the genuinely causal confounds that need to be controlled for. A binary variable indicating whether a voter lived in the South of the United States was often used as a control variable in investigations of the causal influences on voting in the mid-twentieth century. Residence in the South is not, in any ordinary sense, a causal variable, but because it tracks or indicates genuinely causal factors (e.g., racial attitudes) that influence voting, it can be used as a control to isolate the causal role of other variables such as income. Perhaps we should think of PRSs as functioning similarly (for additional discussion, see Kendler & Woodward, under review).

Financial support. L. N. Ross was supported by National Science Foundation (NSF), Award number: 1945647.

Competing interest. None.

Reference

Kendler, K., & Woodward, J. (under review). Polygene risk scores: A philosophical exploration.

Drowning in shallow causality

Hao Shen[a] and Marcus W. Feldman[b]

[a]Department of Human Genetics, University of Chicago, Chicago, IL, USA and [b]Department of Biology, Stanford University, Stanford, CA, USA
shenhao@stanford.edu
mfeldman@stanford.edu

doi:10.1017/S0140525X22002278, e199

Abstract

It has been known for decades that inference concerning genetic causes of human behavioral phenotypes cannot be legitimately made from correlations among relatives. We claim that these inferential difficulties cannot be overcome by assigning different names to causes inferred from within-family and population-level genome-wide association studies (GWASs). For educational attainment, for example, unraveling gene–environment interactions requires more than new names for causes.

Knowledge that a variable makes non-uniform, non-unitary, and non-explanatory average differences in an outcome is defined by Madole & Harden (M&H) as "first-generation causal knowledge" (target article, sect. 1.1, para. 5). This is the kind of knowledge that can be inferred from average treatment effects (ATEs) in

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

randomized controlled trials (RCTs). RCTs are designed to be counterfactual, producing results that can be compared for different values of an experimental condition. The different outcomes of such experiments may, however, depend on temporal, spatial, or environmental contexts in which the experiments are carried out, which may restrict the generality of the results.

M&H introduce "second-generation causal knowledge" (target article, sect. 1.1, para. 7), which derives from understanding the mechanisms that might explain why knowledge inferred from RCTs is not uniform, unitary, or explanatory. Examples of such mechanisms include the effect of the context in which the experiment was carried out, and the role of unintended bias in the choice of subjects (the lithium case in target article, sect. 2.4). As more mediators and confounders are recognized and a more complete causal chain is established, they hope that those closer to the end of a causal chain might be more uniform, unitary, and explanatory, and the reasons for the ATE better understood.

The juxtaposition of first- and second-generation knowledge is blurred by the introduction (target article, sect. 3.3) of genes as "shallow causes" (target article, sect. 3.5) of behavioral phenotypes relative to deep causes. It is not clear whether the authors believe deep causes to be first-generation causal knowledge. However, shallow causes, which are also non-uniform, non-unitary, and non-explanatory, seem to fall under the rubric of first-generation causal knowledge and depend on when, where, how, and on whom the phenotype is assessed. Confusion arises here because second-generation causes are said to provide "a clear sense of the mechanisms of change" by identifying "in what contexts and with whom" causality can be inferred. Do the authors aim for a more complete genetic causal chain, which we assume would involve second-generation causal knowledge, but intend to base it upon shallow causes, which appear to be first generation? A precise dichotomy of first- and second-generation causes, and where deep and shallow causes fall in such a dichotomy, would have been valuable.

Shallow causality's conceptual legitimacy seems M&H to rely on the fact that its limitations are shared with ATEs from RCTs, which do have the advantage of being counterfactually based. Shared limitations are hardly a strong reason to endorse shallow causality as an analytic paradigm.

Neither population genome-wide association studies (GWASs) nor classical heritability studies have a counterfactual basis, and neither should be construed as revealing anything about causality (Feldman & Lewontin, 1975; Lewontin, 1974; Shen & Feldman, 2020). Emphasizing genes as causes, M&H focus on within-family studies, namely comparison between siblings. Given their parents' genotypes, sibs' genotypes can be regarded as a counterfactual experiment only with respect to that family. From within-family GWASs of educational attainment (EA), they conclude that "genes cause EA." "For behavior geneticists" they regard this as "undoubtedly a triumph" (target article, sect. 3.3, paras. 3–4).

But is it? The largest GWAS of EA (Okbay et al., 2022) included 3 million subjects and 53,000 sib pairs. The polygenic score (PGS) for the general sample explained 10–16% of the variance in EA. From the within-family (sib-pair) GWASs, the estimate was that about 31% of the variance explained by the PGS could be classified as "direct effects," which are roughly equivalent to causal. Burt (2023) goes into great detail about the dangers of making general population inferences from within-family GWASs. Here, we note that Okbay et al. (2022) report on GWASs for EA from nearly 2,500 mate pairs and find strong evidence of assortative mating on phenotypes other than EA itself that are correlated with the PGS for EA. Geographic and environmental factors most likely contributed to this assortative mating. Within the general population, there are likely to be differences among families, which may reflect cryptic population stratification. Besides assortative mating, PGS are affected by gene–environment interactions, gene–environment correlations, and environmental variance (Okbay et al., 2022, p. 440). As pointed out by Coop and Przeworski (2022), "the central challenge to identifying genetic causes of behavioral traits" is "the immense difficulty of disentangling population stratification from biological and social effects." Thus, it is not legitimate to claim that within-family studies of EA lead to the conclusion "that genes caused these differences" (target article, sect. 3.3, para. 3). In fact, it is important to stress that PGSs "cannot be used to predict an individual's EA" (Okbay et al., 2022, p. 440).

M&H do recognize the difficulty of extrapolation from inference of genetic causes based on within-family studies to claims about population GWASs. They state (target article, sect. 3.3, para. 10) that genes make "some distal difference in the level of attainment" or that "while genes cause EA, this is neither a singular nor a generic claim" (target article, sect. 3.3, para. 9.). Their justification for the legitimacy of the concept of genes as a shallow cause of traits like EA seems to be that statistical inference of genetic causality shares the properties of being "local, probabilistic, and distal" (target article, sect. 3.3, para. 5) with ATEs. They conclude that "genetic effects conditional on the parental genotype are causal in the same sense as average treatment effects" (target article, sect. 4, para 1). This is actually a statement about within-family GWASs, and the paper's conflation of causal inference from such studies with those of population-level GWASs could be dangerous and should have been avoided. There is no logical reason to believe that claims about causality based on within-family studies also apply to the general population, whether the causal paradigm is first or second generation, deep or shallow (Coop & Przeworski, 2022, p. 851).

M&H are familiar with the shortcomings of the inferential processes that culminate in claims that genes cause behavior. In section 3.3, para. 6, they state "genes might cause EA but they are certainly not the only cause of EA," and in section 3.3, para. 8 "the probability that genes matter for EA varies depending on the environmental exposures." Such statements seem to indicate a genuflection in the direction of Lewontin's (1974) demonstration that causality cannot be inferred from analysis of variance. A straightforward and explicit statement to this effect would have been preferable to introducing complicated definitions of different kinds or levels of causality.

Financial support. This work was supported in part by the Stanford Center for Computational, Evolutionary, and Human Genomics.

Competing interest. None.

References

Burt, C. H. (2023). Challenging the utility of polygenic scores for social science: Environmental confounding, downward causation, and unknown biology. *Behavioral and Brain Sciences*, 46. doi: 10.1017/S0140525X22001145

Coop, M., & Przeworski, M. (2022). Lottery, luck, or legacy. A review of "The genetic lottery: Why DNA matters for social equality." *Evolution*, *76*, 846–853. doi: 10.1111/evo.14449

Feldman, M. W., & Lewontin, R. C. (1975). The heritability hang-up. *Science (New York, N.Y.)*, *190*, 1163–1168. doi: 10.1126/science.1198102

Lewontin, R. C. (1974). The analysis of variance and the analysis of causes. *American Journal of Human Genetics*, *26*, 400–411. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1762622/pdf/ajhg00442-0122.pdf

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

Okbay, A., Wu, Y., Wang, N., Jayashankar, H., Bennett, M., Nehzati, S. M., … Young, A. I. (2022). Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nature Genetics*, *54*, 437–449. doi: 10.1038/s41588-022-01016-z

Shen, H., & Feldman, M. W. (2020). Genetic nurturing, missing heritability, and causal analysis in genetic statistics. *Proceedings of the National Academy of Sciences, USA*, *117*, 25646–25654. doi: 10.1073/pnas.2015869117

Genome-wide association study and the randomized controlled trial: A false equivalence

Paul Siegel

School of Natural and Social Sciences, Purchase College, State University of New York, Purchase, NY, USA
paul.siegel@purchase.edu
https://www.purchase.edu/live/profiles/671-paul-siegel

doi:10.1017/S0140525X22002230, e200

Abstract

Madole & Harden's assertion that the effects derived from within-family genome-wide association studies (GWASs) and from randomized controlled trials (RCTs) are equivalent is misleading. GWASs are substantially more "non-unitary, non-uniform, and non-explanatory" than RCTs. While the within-family GWAS bring us closer to identifying genetic causes, whether it will change behavioral genetics into a causal science is an open question.

Madole & Harden's (M&H's) argument that genetic effects derived from within-family genome-wide association studies (GWASs) are equivalent to average treatment effects from randomized controlled trials (RCTs) rests on the assertion that both methods are "non-unitary, non-uniform, and non-explanatory." This contention is misleading because these three "non-" dimensions are not binary, but very much a matter of degree. Double-blind RCTs are more likely to prevent the entry of other variables whose effects are confounded with those of the treatment, and thereby isolate the treatment as the most likely cause of outcomes. Thus, well-designed RCTs severely limit non-unitariness. RCTs of treatments that prove to be highly efficacious directly demonstrate their greater uniformity of therapeutic effects. Moreover, RCTs that examine treatment effects on not just clinical outcomes but mediating variables like brain activity and physiological arousal can provide mechanistic explanations of therapeutic effects (Horga, Kaur, & Peterson, 2014), and can select between alternative mechanisms for these effects (e.g., Siegel, Cohen, & Warren, 2022).

Even with more than a million randomized trials of alleles, within-family GWASs are much more non-unitary than a well-designed RCT. Besides the indirect sibling-to-sibling genetic effects that the authors address, there are other, multiple environmental sources of variation in phenotypes that cannot be ruled out and thus render within-family GWASs entirely non-unitary. Taking one of many possibilities, differential parenting of siblings, even when stemming from the siblings' genetic differences, has developmental effects that can amplify phenotypic differences over time. Such unshared environmental influences are very difficult to measure (Turkheimer & Waldron, 2000), and it cannot be assumed that they will wash out with large enough samples (McCarthy et al., 2008).

The non-uniformity of GWASs is demonstrated by the needs for very large samples (thousands of cases) and for replication because of limited statistical power (McCarthy et al., 2008). Both of these limitations stem from the need to correct for approximately 1 million independent tests of allele-outcome regressions in a typical GWAS (Visscher et al., 2017). For this reason, the journal *Behavior Genetics* requires replication to consider any GWAS for publication (Hewitt, 2012).

GWAS non-uniformity also results from the need to control for genetic ancestry and thus potentially confounding genetic variants that differ across populations (McCarthy et al., 2008). As a result, virtually all GWASs have been of white Europeans, the most widely appraised ancestry. Therefore, GWAS findings may not apply to other ethnic groups, a non-uniformity that may exacerbate health disparities (Martin et al., 2019).

Another factor that grants RCTs more unitary causal inference than GWASs is the level of validity of outcome assessment. RCTs require accurate, or relatively "deep" assessments of behavioral outcomes to establish treatments as robust causes of changes in those outcomes. Because GWASs require very large samples that are often gathered from biobanks or by consortia across studies, phenotypes are typically assessed superficially to ensure standardization (Friedman, Banich, & Keller, 2021). For example, assessment of depression may be as simplistic as self-reported ratings. Although simulations indicate that the large samples of GWASs have sufficient power to discern genetic effects despite the large error associated with such minimal assessments (Border et al., 2019), these typical GWASs yield considerably lower heritability rates, and identify single-nucleotide polymorphisms (SNPs) with much less specificity, than GWASs with more valid phenotype assessments (Cai et al., 2020).

The above non-unitary and non-uniform factors may explain why GWASs have repeatedly yielded small genetic effects for traits that twin and family studies previously estimated to be large, what has been termed the problem of "missing heritability" (Maher, 2008). GWAS heritability estimates are typically 40–80% lower than those yielded by twin and family studies (Friedman et al., 2021). For example, twin and family studies estimate genetic effects for schizophrenia to be about 60% (e.g., Lichtenstein et al., 2009). In contrast, a large GWAS ($N = 3{,}322$ cases and 3,587 controls without the illness) conducted during the same time identified ~74,000 genetic variants on a single chromosome that accounted for as little as 3%, or as much as 30%, of the variance in schizophrenia, depending on the analytic approach employed (Purcell et al., 2009).

M&H introduce the within-family GWAS as having the potential to address the limitations of traditional heritability approaches like twin studies, which cannot "specify which genes or, crucially, how those genes are responsible for producing phenotypic differences" (target article; sect. 3.1, para. 2). Yet they conclude that the identified SNP has an "intermediate level of resolution, encompassing all alleles in LD [linkage disequilibrium] with the measured SNP" (target article; sect. 3.2.1, para. 10). In other words, the identified SNP is highly correlated with – a marker of – the causal variant, an implicit acknowledgment that the within-family GWAS has the same, aforementioned limitations as twin studies. Critically, modern arrays of genotyped SNPs miss certain variants that are *not* in linkage disequilibrium (LD) with the imputed SNP. While rare, these missed genetic variants can nonetheless have a large effect on variation in phenotypes (Visscher et al., 2017), which is also thought to underlie the aforementioned "missing heritability" (Friedman et al., 2021).

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

By identifying SNPs highly correlated with behavioral traits, the within-family GWAS brings us closer to identifying genetic causes. Whether it will alter the status of behavior genetics as a causal science, however, is a wide open question.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interest. None.

References

Border, R., Johnson, E. C., Evans, L. M., Smolen, A., Berley, N., Sullivan, P. F., & Keller, M. C. (2019). No support for historical candidate gene or candidate gene-by-interaction hypotheses for major depression across multiple large samples. *American Journal of Psychiatry*, *176*, 376–387.

Cai, N., Revez, J. A., Adams, M. J., Andlauer, T. F., Breen, G., Byrne, E. M., … Flint, J. (2020). Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nature Genetics*, *52*, 437–447.

Friedman, N. P., Banich, M. T., & Keller, M. C. (2021). Twin studies to GWAS: There and back again. *Trends in Cognitive Sciences*, *25*(10), 855–869.

Hewitt, J. K. (2012). Editorial policy on candidate gene association and candidate gene-by-environment interaction studies of complex traits. *Behavior Genetics*, *42*(1), 1.

Horga, G., Kaur, T., & Peterson, B. S. (2014). Annual research review: Current limitations and future directions in MRI studies of child- and adult-onset developmental psychopathologies. *Journal of Child Psychology and Psychiatry*, *55*(6), 659–680.

Lichtenstein, P., Yip, B. H., Bjork, C., Pawitan, Y., Cannon, T. D., Sullivan, P. F., & Hultman, C. M. (2009) Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: A population-based study. *Lancet (London, England)*, *373*, 234–239.

Maher, B. (2008). The case of the missing heritability. *Nature*, *456*(7218), 18–22.

Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., & Daly, M. J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics*, *51*(4), 584–591.

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J., & Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nature Reviews Genetics*, *9*(5), 356–369.

International Schizophrenia Consortium, Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., & Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, *460*, 475–752.

Siegel, P., Cohen, B., & Warren, R. (2022). Nothing to fear but fear itself: A mechanistic test of unconscious exposure. *Biological Psychiatry*, *91*(3), 294–302.

Turkheimer, E., & Waldron, M. (2000). Nonshared environment: A theoretical, methodological, and quantitative review. *Psychological Bulletin*, *126*(1), 78.

Visscher, P.M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 years of GWAS discovery: Biology, function, and translation. *The American Journal of Human Genetics*, *101*, 5–22.

Shallow versus deep genetic causes

Adam C. Smith and Stephen M. Downes

Philosophy Department, University of Utah, Salt Lake City, UT, USA
adam.c.smith@utah.edu; s.downes@utah.edu

doi:10.1017/S0140525X22002114, e201

Abstract

We argue that Madole & Harden's distinction between shallow versus deep genetic causes can bring some clarity to causal claims arising from genome-wide association studies (GWASs). However, the authors argue that GWAS only finds shallow genetic causes, making GWAS commensurate with the environmental studies they hope to supplant. We also assess whether their distinction applies best to explanations or causes.

Madole & Harden (M&H) aim to present "a clear perspective on what it does – and does not – mean for genes to be causes" (target article, sect. 1.2, para. 4). We agree that this is an important project but are unsure whether they succeed in their overall aim. The authors propose that genetic causes are like nearly all environmental causes investigated in the social sciences: They are non-uniform, non-unitary, and non-explanatory (target article, sect. 1.3, para. 1). In other words, the genetic causes pointed to by genome-wide association studies (GWASs) hold true only for specific populations, their effects are probabilistic, not deterministic, and the exact nature of their causal relevance to the production of the trait is unknown. M&H call such causes *shallow causes*. The mitigating effects of lithium on mania, understood as an average treatment effect from a randomized controlled trial, is an example of a shallow cause. They acknowledge that this is not the only one way in which genes play a causal role in producing traits and also identify *deep genetic causes*. Deep genetic causes are uniform, unitary, and explanatory. Cystic fibrosis, which is caused by two mutated copies of the cystic fibrosis transmembrane regulatory (CFTR) gene looks to provide an example of a trait with a deep genetic cause.

While this distinction promises clarity, M&H often conflate *cause*, *genetic cause*, and *shallow (genetic) cause* and, as a result, they fail to separate genetic causal claims from environmental causal claims presented by other social scientists. This makes causes discovered by GWASs commensurate with the environmental studies M&H hope to supplant. Second, the distinction between shallow versus deep genetic causes can be helpful, but only if applied carefully and consistently. For example, the authors make apparently contradictory claims about GWASs for educational attainment (EA): The results of GWASs "alone would not move us closer to the conclusion that genes cause educational outcomes" and "we are currently in a position to conclude that genes cause EA" (target article, sect. 3.3, paras. 2 and 3). Their distinction between shallow and deep genetic causes can help resolve this and other apparently contradictory claims about genetic causation. We understand M&H to be saying that while even the most extensive GWASs cannot support the conclusion that certain genes are *deep* causes of a trait, current GWASs do allow the claim that certain genes are *shallow* causes of a trait. This claim is not contradictory and in fact is a clear statement of their understanding of the promise of GWASs. However, M&H introduce further confusion about the nature of genetic causes.

M&H seem to say that deep genetic causes are in fact not genetic causes because they are not causes at all. For example: "Such a picture of genetic causes is entirely unwarranted when we remember what it means for something to be a cause: *non-uniform*, *non-unitary*, and *non-explanatory*" (target article, sect. 3.3, para. 4). Here they define a *cause* as non-uniform, non-unitary, and non-explanatory, implying that deep genetic causes are not just something GWASs can't detect, but in fact are not causes at all. Then, just two sentences later, M&H use the example of cystic fibrosis as a genetic cause that they say is uniform, unitary, and explanatory, that is, a deep genetic cause (target article, sect. 3.3, para. 5).

We propose that this inconsistency on the status of deep genetic causes arises from the account of causation M&H defend. As noted earlier, their account of causation is meant to cover both *genetic causes* and *shallow genetic causes*, which allows them to claim at one point that GWASs can't possibly be giving fuel to eugenicists because *genetic causes* by definition are non-uniform, non-unitary,

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

and non-explanatory (target article, sect. 1.3, para. 1 & sect. 3.5, para. 1). So, what are we to make of deep genetic causes? M&H clearly want to claim that some genetic causes are shallow and some are deep, and that GWASs can only find shallow genetic causes. In fact, M&H's distinction between shallow genetic causes and deep genetic causes seems to be set up to explicitly acknowledge that there are various kinds of genetic causes. If this is right, then they need a different account of *genetic cause*, and possibly *cause* in general, in order to accommodate both shallow and deep genetic causes.

M&H briefly review a few philosophical accounts of causation, then claim to adopt a probabilistic version of Jim Woodward's (2003) manipulationist account. However, it is unclear how this account fits GWASs, because there is no careful manipulation of specific variables in GWASs, a point M&H acknowledge when arguing for the superiority of within-family studies over regular GWASs (target article, sect. 3.5, para. 1). Furthermore, the distinction between shallow versus deep causes does not straightforwardly map on to much philosophical discussion of genetic causation (see e.g., Gannett, 1999; Lynch & Bourrat, 2017; Noble, 2008; Northcott, 2012; Oftedal, 2005; Schaffner, 2016; Sober, 2000; Waters, 2007). For example, M&H say that-causes need not be mechanistic (target article, sect. 3.3, para. 11), while many philosophers, including those cited here, offer accounts ofgenetic causation as being necessarily mechanistic.

M&H's shallow versus deep distinction better tracks Eric Turkheimer's (1998, 2016) distinction between strong and weak genetic explanations than it does distinctions in the philosophy of science literature on genetic causation. For Turkheimer a *weak genetic explanation* says "one way or another, genetic differences among people wind up correlated with phenotypic differences" (Turkheimer, 2016, p. 24). Neither shallow genetic causes nor weak genetic explanations determine the exact difference that certain single-nucleotide polymorphisms (SNPs) make in a phenotype but this is by design. A *strong genetic explanation* by contrast is "the discovery that an observed phenotypic difference is a manifestation of a specific latent genetic mechanism" (Turkheimer, 2016, p. 25). Both uncovering deep genetic causes and providing strong genetic explanations can reveal specific mechanisms.

While M&H's shallow versus deep distinction may resolve some apparent contradictions in their claims about GWASs and causation, we propose that their distinction better points to a way of separating alternate types of genetic explanations. This allows that GWASs do lead to genetic explanations of traits but only in the same sense in which non-genetic social science leads to environmental explanations of traits.

Acknowledgment. The authors thank Hannah Allen for her comments on the manuscript.

Competing interest. None.

References

Gannett, L. (1999). What's in a cause?: The pragmatic dimensions of genetic explanations. *Biology & Philosophy*, *14*(3), 349–373. https://doi.org/10.1023/A:1006583215835

Lynch, K. E., & Bourrat, P. (2017). Interpreting heritability causally. *Philosophy of Science*, *84*(1), 14–34. https://doi.org/10.1086/688933

Noble, D. (2008). Genes and causation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *366*(1878), 3001–3015. https://doi.org/10.1098/rsta.2008.0086

Northcott, R. (2012). Genetic traits and causal explanation. In K. S. Plaisance & T. A. C. Reydon (Eds.), *Philosophy of behavioral biology* (Vol. 282, pp. 65–82). Springer Netherlands. Retrieved from https://doi.org/10.1007/978-94-007-1951

Oftedal, G. (2005). Heritability and genetic causation. *Philosophy of Science*, *72*(5), 699–709. https://doi.org/10.1086/508126

Schaffner, K. F. (2016). *Behaving: What's genetic, what's not, and why should we care?* (1st ed.). Oxford University Press.

Sober, E. (2000). The meaning of genetic causation. In A. Buchanan, D. W. Brock, D. Wikler & N. Daniels (Eds.), *From chance to choice: Genetics and justice* (pp. 347–370). Cambridge University Press. Retrieved from https://doi.org/10.1017/CBO9780511806940.010

Turkheimer, E. (1998). Heritability and biological explanation. *Psychological Review*, *105*(4), 782–791. https://doi.org/10.1037/0033-295X.105.4.782-791

Turkheimer, E. (2016). Weak genetic explanation 20 years later: Reply to Plomin et al. (2016). *Perspectives on Psychological Science*, *11*(1), 24–28. https://doi.org/10.1177/1745691615617442

Waters, C. K. (2007). Causes that make a difference. *The Journal of Philosophy*, *104*(11), 551–579.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press.

Building causal knowledge in behavior genetics without racial/ethnic diversity will result in weak causal knowledge

Moin Syed

Department of Psychology, University of Minnesota, Minneapolis, MN, USA
moin@umn.edu
https://cla.umn.edu/about/directory/profile/moin

doi:10.1017/S0140525X22002163, e202

Abstract

Behavior genetics often emphasizes methods over the underlying quality of the psychological information to which the methods are applied. A core aspect of this quality is the demographic diversity of the samples. Building causal genetic models based only on European-ancestry samples compromises their generalizability. Behavior genetics researchers must spend additional time and resources diversifying their samples before emphasizing causation.

Madole & Harden (M&H) propose that within-family genetic effects are analogous to randomized controlled trials (RCTs), and therefore can help identify potential genetic causes of psychological phenomena. In doing so, the authors repeat the frequent trope of RCTs as the "gold standard" (target article, sect. 1.1, para. 2) for establishing causal claims. Although it is clear that RCTs are optimally set up for such a task, using the language of "gold standard" has been criticized for obscuring the many limitations of the design (see Deaton & Cartwright [2018] for a thorough treatment of the subject and Jones & Podolsky [2015] for a historical discussion). The uncritical use of RCTs as an analogue for genetic effects opens their arguments to many criticisms, ones that are already prominent in work on behavior genetics. Among others, these include insufficient attention to conceptualization (Nguyen, Syed, & McGue, 2021) and measurement (Pelt, Schwabe, & Bartels, 2022) of the target phenotype and lack of sample diversity when seeking to establish generalizable claims (Holden, Haughbrook, & Hart, 2022). By not incorporating these major concerns in their paper, M&H unfortunately

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

compound, rather than resolve, these issues in their otherwise productive set of arguments.

The more general criticism levied toward both RCTs and behavior genetics is the over-emphasis on method over substance. M&H continue in this tradition, devoting careful attention to establishing and calibrating causal claims. This is clearly a useful and much-needed approach to thinking about causation, especially in psychology, but we must not overlook the quality of the substantive information about which we seek to make causal claims. Here, I highlight just one aspect of such quality, the racial/ethnic diversity of samples included in behavior genetic studies, and why such a consideration must be central to any effort to build generalizable causal knowledge.

It is a fact of the design that RCTs sacrifice external validity for the sake of internal validity, being high in *efficacy*, showing promising results in trials, but low in *effectiveness*, or lack of results when translated to real-life conditions (Flay et al., 2005). RCTs have been further criticized by researchers in multicultural psychology and culturally adapted treatments for their lack of inclusion of racial/ethnic minorities and thus limited generalizability (Bernal & Scharrón-del-Río, 2001; Castro, Barrera, & Holleran Steiker, 2010; Whaley & Davis, 2007). Consistent with the mainstream view in psychology, the arguments made by M&H all assume a universal, replaceable person; it does not matter where the information comes from, so long as the proper modeling is applied (see also Yarkoni, 2022). But, of course, strong design features cannot overcome selection issues that can lead to misspecified models (Bradley et al., 2021).

The research on lithium as a treatment for manic symptoms of bipolar disorder, highlighted by the authors, illustrates their lack of attention to sample diversity. The second-generation studies cited by M&H – those that seek to identify specific causal mechanisms beyond an average treatment effect – relied entirely on data from White men (Mertens et al., 2015; Santos et al., 2021; Stern et al., 2018). It may well be that the identified mechanisms are generalizable beyond this very narrow group, but it seems prudent to investigate this question prior to broadly claiming generalizable causal knowledge. Moreover, the authors frame this kind of second-generation investigation as addressing the problem of lack of generalizability beyond the first-generation average treatment effect, but it does so only with respect to individual differences, and not demographic/population heterogeneity.

An argument could be made to justify sample homogeneity in first-generation studies, when putative causal factors are initially identified, and explore generalizability as part of the second-generation process. This is related to the issue of "portability" of findings from genetic studies (or "effectiveness" with RCTs), which refers to the fact that average effects identified through first-generation behavior genetics studies may not generalize to new contexts or populations. Indeed, the impressive Lee et al. (2018) study of educational attainment of 1.1 million individuals included only participants with European ancestry. When the researchers attempted to "port" the polygenic scores derived from the European-ancestry group to a sample of Black Americans, the 10.6% R^2 attenuated by 85%, a result that they indicated was "typical of what has been reported in other studies" (p. 1115). M&H fail to mention the degree of this problem, let alone what the implications are for building *generalizable* causal knowledge. For example, the omnigenic model highlights the need for diversity in discovery samples to identify and separate both core and peripheral variants (Mathieson, 2021; see also Wojcik et al., 2019). Thus, diversity is central to first-generation studies.

Moreover, in discussions of portability, left unsaid is *from* whom *to* whom the results are being ported, which is nearly always from White/European samples to other ancestry or racial groups. Rarely do we seek to, for example, generalize results from African samples to other groups (Adetula, Forscher, Basnight-Brown, Azouaghe, & IJzerman, 2022). This dynamic sets up a standard in which the White/European results serve as the basis for the first-generation causal knowledge, and any deviations from it are problems to be solved or, more often, swept under the rug. Such a perspective is consistent with the *deficit model* that has for decades dominated psychological research on diversity (Cauce, Coronado, & Watson, 1998), and unreasonably constrains the context of discovery.

The issues mentioned here are reminiscent of the aphorism "garbage in, garbage out" in the context of meta-analyses. That is, no degree of sophisticated analyses will save your substantive conclusions if the studies included therein are weak or uninformative. To be fair, M&H clearly know this, but the issue is treated more in passing rather than as a central concern in their quest to build stronger causal knowledge, a quest which I greatly support. They use a catch phrase in similar structure to "garbage in, garbage out" when discussing causal reasoning, "no causes in, no causes out," which, fittingly, pertains to the reasoning and not the quality of the information within it. Combining these two, I might say, "no diversity in, no causes at all."

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interest. None.

References

Adetula, A., Forscher, P. S., Basnight-Brown, D., Azouaghe, S., & IJzerman, H. (2022). Psychology should generalize from – not just to – Africa. *Nature Reviews Psychology*, *1*, 370–371. https://doi.org/10.1038/s44159-022-00070-y

Bernal, G., & Scharrón-del-Río, M. R. (2001). Are empirically supported treatments valid for ethnic minorities? Toward an alternative approach for treatment research. *Cultural Diversity and Ethnic Minority Psychology*, *7*(4), 328–342. https://psycnet.apa.org/doi/10.1037/1099-9809.7.4.328

Bradley, V. C., Kuriwaki, S., Isakov, M., Sejdinovic, D., Meng, X. L., & Flaxman, S. (2021). Unrepresentative big surveys significantly overestimated US vaccine uptake. *Nature*, *600*(7890), 695–700. https://doi.org/10.1038/s41586-021-04198-4

Castro, F. G., Barrera Jr, M., & Holleran Steiker, L. K. (2010). Issues and challenges in the design of culturally adapted evidence-based interventions. *Annual Review of Clinical Psychology*, *6*, 213–239. https://doi.org/10.1146/annurev-clinpsy-033109-132032

Cauce, A. M., Coronado, N., & Watson, J. (1998). Conceptual, methodological, and statistical issues in culturally competent research. In M. Hernandez & R. Mareasa (Eds.), *Promoting cultural competence in children's mental health services* (pp. 305–329). Paul Brooks.

Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, *210*, 2–21. https://doi.org/10.1016/j.socscimed.2017.12.005

Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., … Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*, *6*(3), 151–175. https://doi.org/10.1007/s11121-005-5553-y

Holden, L. R., Haughbrook, R., & Hart, S. A. (2022). Developmental behavioral genetics research on school achievement is missing vulnerable children, to our detriment. *New Directions for Child and Adolescent Development*, *183-184*, 47–55. https://doi.org/10.1002/cad.20485.

Jones, D. S., & Podolsky, S. H. (2015). The history and fate of the gold standard. *The Lancet*, *385*(9977), 1502–1503. https://doi.org/10.1016/S0140-6736(15)60742-5

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., … Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, *50*(8), 1112–1121. https://doi.org/10.1038/s41588-018-0147-3

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

Mathieson, I. (2021). The omnigenic model and polygenic prediction of complex traits. *The American Journal of Human Genetics*, *108*(9), 1558–1563. https://doi.org/10.1016/j.ajhg.2021.07.003
Mertens, J., Wang, Q. W., Kim, Y., Yu, D. X., Pham, S., Yang, B., … Yao, J. (2015). Differential responses to lithium in hyperexcitable neurons from patients with bipolar disorder. *Nature*, *527*(7576), 95–99. https://doi.org/10.1038/nature15526
Nguyen, P. L. L., Syed, M., & McGue, M. (2021). Behavior genetics research on personality: Moving beyond traits to examine characteristic adaptations. *Social and Personality Psychology Compass*, *15*, e12628. https://doi.org/10.1111/spc3.12628
Pelt, D., Schwabe, I., & Bartels, M. (2022). Using item response theory modeling to address bias in genetic findings for happiness, life satisfaction, and social support. *PsyArXiv*. https://doi.org/10.31234/osf.io/7uz9y
Santos, R., Linker, S. B., Stern, S., Mendes, A. P., Shokhirev, M. N., Erikson, G., … Gage, F. H. (2021). Deficient LEF1 expression is associated with lithium resistance and hyperexcitability in neurons derived from bipolar disorder patients. *Molecular Psychiatry*, *26*(6), 2440–2456. https://doi.org/10.1038/s41380-020-00981-3
Stern, S., Santos, R., Marchetto, M. C., Mendes, A. P. D., Rouleau, G. A., Biesmans, S., … Gage, F. H. (2018). Neurons derived from patients with bipolar disorder divide into intrinsically different sub-populations of neurons, predicting the patients' responsiveness to lithium. *Molecular Psychiatry*, *23*(6), 1453–1465. https://doi.org/10.1038/mp.2016.260
Whaley, A. L., & Davis, K. E. (2007). Cultural competence and evidence-based practice in mental health services: A complementary perspective. *American Psychologist*, *62*(6), 563–574. https://doi.org/10.1037/0003-066X.62.6.563
Wojcik, G. L., Graff, M., Nishimura, K. K., Tao, R., Haessler, J., Gignoux, C. R., … Carlson, C. S. (2019). Genetic analyses of diverse populations improves discovery for complex traits. *Nature*, *570*(7762), 514–518. https://doi.org/10.1038/s41586-019-1310-4
Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, *45*, E1. https://doi.org/10.1017/S0140525X20001685.

Behavior genetics: Causality as a dialectical pursuit

Livio Tarchi, Giuseppe Pierpaolo Merola, Giovanni Castellini and Valdo Ricca

Psychiatry Unit, Department of Health Sciences, University of Florence, Florence, FI, Italy
livio.tarchi@unifi.it
giuseppepierpaolo.merola@unifi.it
giovanni.castellini@unifi.it
valdo.ricca@unifi.it

doi:10.1017/S0140525X22002254, e203

Abstract

The overarching theme of causality in behavioral genetics is discussed on epistemological grounds. Evidence is offered in favor of a continuum spectrum in causality, in contrast to discrimination between causal factors and associations. The risk of invalidating exploratory studies in behavior genetics is discussed, especially for the potential impact on those fields of medicine interested in complex behaviors.

Madole & Harden (M&H) propose a discrimination between knowledge of causality and, on the contrary, correlation in behavior genetic studies, with a particular emphasis on genome-wide association studies (GWASs). Far from being a dichotomous difference, we argue that the boundary between these two types of causal relationships is indeed continuous, and any distinction would be arbitrary. Correlation itself is based on a counterfactual assumption, that is, that there is no relationship between the studied factors (Benesty, Chen, Huang, & Cohen, 2009). The line between counterfactuality and causality seems to be briefly addressed in the manuscript, but only implied in both the introduction and in the rest of the article. We agree with the original authors that counterfactuality is not the only acceptable framework of causality, as counterfactuality itself might not prove to be either specific or sensitive to causality (Baumgartner, 2008). In fact, recent developments in the regularity theory of causation (based on the premises that causes are regularly followed by their effects; Baumgartner, 2008) have allowed for a precise estimation and quantification of confidence in causal relationships (Baumgartner, 2008), while also offering the opportunity to assess causality beyond dichotomous categories (whether "ill" or "healthy," or expressing a certain behavior or not). These developments allow researchers to move toward approaches encompassing the possibility of evaluating spectra of continuum in traits, possibly through fuzzy-logic algorithms (Baumgartner & Ambühl, 2020), granting a more flexible evaluation of the interactions between such traits and specific genes. Specific research questions and study designs might then benefit from a careful operationalization of causality in investigation protocols, defining outcome variables either dichotomous, when methodologically sensible, or continuous, in the majority of cases where complex traits need to be considered.

M&H also argue that genetic analyses might not be causally informative, as the relationship between genotype and phenotype can be complex in behavior research. A critique is moved about the results of genetic studies, which are in some cases misquoted or overinflated. Indeed, no single study can inform our understanding of the complex biology behind the interaction of genes and traits. There is a real and actual risk of adopting "evidence-based" policies on frail epistemological and scientific grounds. This risk, however, is contemplated by most contemporary researchers and most of the general public (Visscher, Brown, McCarthy, & Yang, 2012). Nonetheless, supporting a careful review of evidence before adopting interventions does not invalidate the scientific knowledge gained by conducting behavior genetic studies. Even if GWASs were only to offer "association" knowledge rather than "true" causal understanding of underlying factors, it is relevant to point out that other means are available in genetic research to reach this goal, which may be more oriented toward "deeper" and "mechanistic" causal analysis (e.g., pathway analysis, protein functionality studies, endophenotype studies; Kendler & Neale, 2010); but in order for these tools to be used effectively, because of their high relative costs and technical complexity, candidate genes must be identified previously through GWASs. In fact, GWASs have offered considerable insight into nearly every field of contemporary biological sciences, from pharmacodynamics and pharmacokinetics, to proteomics and transcriptomics (Visscher et al., 2012).

A clarification of the term "causality" may also aid in delimiting the space of the discussion. Any operative definition of "causality" should consider scientific endeavors in an integrative manner, and as the result of multidisciplinary efforts. In fact, "science" can be defined as a continuous dialectical pursuit (Popper, 1940), where "knowledge" is constantly updated with evidence derived from different sources. Therefore, scrutiny over preliminary evidence is indeed warranted before it is allowed to inform clinical or policy interventions. However, a primary goal is also to seek a balance between addressing the neglected needs of an individual and violating the right not to be harmed, which is a statutory principle that has guided the field of medicine since its inception. For these reasons, the risk of invalidating exploratory evidence in neuroscience in

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

general, and behavioral genetics in particular, needs to be discussed. The hazard to implicitly propose new criteria in research, that is to consider "association studies" as secondary or even detrimental, should be critically evaluated, as it might severely impact both patients and researchers. For instance, limited funding and scarcity of resources may favor those capable of conducting large-scale "mechanistic" studies, harming scientific independence, tilting the balance in favor of consortium-led enterprises, with negative consequences on originality, scrutiny, and productivity in research (Wang, Veugelers, & Stephan, 2017). Large-scale "mechanistic" studies may also worsen the over-representation of white Anglo-Saxon, European, or East-Asian individuals in genetic studies (Sirugo, Williams, & Tishkoff, 2019). Additionally, research on several clinical conditions may never reach the volume necessary to conduct a large-scale investigation (Rosenberg & Finn, 2022), and no existing knowledge at present may properly guide causal "mechanistic" studies. Especially in those fields of medicine interested by complex behaviors (e.g., psychology, neurology, psychiatry), low prevalence and clinical heterogeneity burden the ease-of-access to interventional programs, as well as the inclusion in observational studies (Mitchell, Maki, Adson, Ruskin, & Crow, 1997). However, it is possible to adopt mitigating options. For example, longitudinal designs can reach higher statistical power than cross-sectional ones, increasing replicability and confidence in the association between genes and phenotype (Rosenberg & Finn, 2022). Again, GWASs offer a cost-effective opportunity to first assess associations between genes and traits in these populations, and later inform more targeted protocols or interventions. Finally, the same conditions interested by low prevalence or high heterogeneity demand urgency in describing causal relationships, as they are taxed by a high rate of inadequacy in treatment (Bulik, 2021). For these reasons, invalidating behavioral genetic studies solely on concerns of describing causal associations may severely impact those individuals who they may benefit the most.

Financial support. The authors did not receive a specific grant from any funding agency, commercial, or non-profit sector for the production of this article.

Competing interest. None.

References

Baumgartner, M. (2008). Regularity theories reassessed. *Philosophia*, *36*(3), 327–354. https://doi.org/10.1007/s11406-007-9114-4

Baumgartner, M., & Ambühl, M. (2020). Causal modeling with multi-value and fuzzy-set coincidence analysis. *Political Science Research and Methods*, *8*(3), 526–542. https://doi.org/10.1017/psrm.2018.45

Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). Pearson correlation coefficient. In I. Cohen, Y. Huang, J. Chen, & J. Benesty (Eds.), *Noise reduction in speech processing* (pp. 1–4). Springer. https://doi.org/10.1007/978-3-642-00296-0_5

Bulik, C. M. (2021). From awareness to action: An urgent call to address the inadequacy of treatment for anorexia nervosa. *American Journal of Psychiatry*, *178*(9), 786–788. https://doi.org/10.1176/appi.ajp.2021.21070697

Kendler, K. S., & Neale, M. C. (2010). Endophenotype: A conceptual analysis. *Molecular Psychiatry*, *15*(8), 789–797. https://doi.org/10.1038/mp.2010.8

Mitchell, J. E., Maki, D. D., Adson, D. E., Ruskin, B. S., & Crow, S. (1997). The selectivity of inclusion and exclusion criteria in bulimia nervosa treatment studies. *International Journal of Eating Disorders*, *22*(3), 243–252. https://doi.org/10.1002/(SICI)1098-108X(199711)22:3<243::AID-EAT3>3.0.CO;2-N

Popper, K. R. (1940). What is dialectic? *Mind; A Quarterly Review of Psychology and Philosophy*, *XLIX*(194), 403–426. https://doi.org/10.1093/mind/XLIX.194.403

Rosenberg, M. D., & Finn, E. S. (2022). How to establish robust brain–behavior relationships without thousands of individuals. *Nature Neuroscience*, *25*(7), 835–837. https://doi.org/10.1038/s41593-022-01110-9

Sirugo, G., Williams, S. M., & Tishkoff, S. A. (2019). The missing diversity in human genetic studies. *Cell*, *177*(1), 26–31. https://doi.org/10.1016/j.cell.2019.02.048

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *The American Journal of Human Genetics*, *90*(1), 7–24. https://doi.org/10.1016/j.ajhg.2011.11.029

Wang, J., Veugelers, R., & Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy*, *46*(8), 1416–1436. https://doi.org/10.1016/j.respol.2017.06.006

Genes, genomes, and developmental process

Jebediah Taylor[a], Staci Meredith Weiss[b,c] and Peter J. Marshall[d]

[a]University of Michigan Law School, Ann Arbor, MI, USA; [b]Perinatal Imaging Partnership, Rosie Maternity Hospital, Cambridge, UK; [c]Department of Psychology, University of Cambridge, UK and [d]Department of Psychology & Neuroscience, Temple University, Philadelphia, PA, USA
jebt@umich.edu
smw95@cam.ac.uk
peter.marshall@temple.edu
https://www.repro.cam.ac.uk/staff/dr-staci-meredith-weiss
https://sites.temple.edu/peterjmarshall/

doi:10.1017/S0140525X22002199, e204

Abstract

The view advanced by Madole & Harden falls back on the dogma of a gene as a DNA sequence that codes for a fixed product with an invariant function regardless of temporal and spatial contexts. This outdated perspective entrenches the metaphor of genes as static units of information and glosses over developmental complexities.

Population geneticists have historically deployed the concept of genes as statistical, rather than material, entities (Griffiths & Tabery, 2008). Although this approach may have sufficed in the era of traditional twin studies of behavior, the advent of genome-wide association studies (GWASs) seems to require some engagement with thorny questions around *how* variation in DNA sequences might be associated with variations in phenotype. The neoclassical view of the gene as a sequence of DNA encoding a single transcript that uniformly produces a particular protein (Portin & Wilkins, 2017) has been adopted by behavior geneticists to provide a biological foundation for their otherwise purely statistical framework. However, decades of empirical findings have long exposed the deficiencies of the neoclassical view. The "post-genomic era" is replete with findings that the same sequence of DNA can be used to derive a variety of transcripts (e.g., Griffiths & Stotz, 2006; McManus & Graveley, 2011; The ENCODE Project Consortium, 2007, 2012), and that products derived from the same DNA sequence can exhibit divergent structures and functions depending on their cellular context (Piatigorsky, 2007). The cellular signals driving these processes, including epigenetic modifications such as DNA methylation, partly reflect responses to an individual's social and ecological situation (e.g., Meaney, 2010). The activity of the genome develops selectively and responsively to fluctuating physiological conditions, undermining the idea of a prespecified, constant function

inherent to a particular sequence of DNA (Neumann-Held, 2001). These realizations necessitate a shift from an "agentive" role for genes to a "reactive" view of the wider genome embedded within the organism as a developmental system (Keller, 2014). Related lines of evolutionary thought emphasizing the primacy of development suggest that genes should be seen as followers, not leaders, of adaptive plasticity (Newman, 2019; West-Eberhard, 2003).

What is a behavior geneticist to do? Rather than engage with the postgenomic complexities, Madole & Harden (M&H) limit themselves to the neoclassical dogma, with DNA framed as a "specific set of instructions" (target article, sect. 3.3, para. 10) that weathers all the variability that "context" can throw at it (target article, sect. 3.3, para. 8). To an extent, they acknowledge the complexities through a "garden of forking paths" metaphor (target article, sect. 3.3, para. 10), but this path only appears to go forward, from a foundational DNA sequence that is inherited at conception and encodes the same product regardless of spatial and temporal context. However, as noted by Oyama decades ago, the metaphorical "information" in the genome is not static: It develops along with the organism (Oyama, 1985). Further, there is no consideration of the circular causation that runs all the way through organismal functioning (Witherington, 2011) and that is apparent even at the level of DNA transcription. In our view, the neglect of circular causation reflects a neglect of development, long regarded as an afterthought by the field of behavior genetics. In contrast to staid models of behavioral genetics, developmental systems perspectives allow for the multifaceted complexities of ontogeny (Gottlieb, 1995; Overton & Lerner, 2014).

Processes that modify DNA transcription and translation are responsive to temporal and situational changes for the organism. Consider, for example, structural brain anatomy, a phenotype popular for study in the genetics literature. Rather than exhibiting a linear growth trajectory, brain development varies over time across types of growth (e.g., cortical thickness vs. surface area), tissues (i.e., gray vs. white matter), and regions (Fjell et al., 2019; Li et al., 2013). Correspondingly, genomic processes relevant to brain development vary across the lifespan as well. For example, while "clusters" of cortical thickness development (i.e., areas of the cortex showing longitudinal intercorrelation over time) were found to overlap substantially with adult "genetic clusters" (i.e., areas previously associated with shared genetic influence), there was only limited overlap between developmental and genetic clusters for cortical surface area, which suggests divergent patterns of developmental organization (Fjell et al., 2019). Further, these kinds of developmental processes are sensitive to experience, with epigenetic influences modifying gene expression relevant to neurodevelopment in response to exposures ranging from lead poisoning and child maltreatment to maternal mental health and exercise (Fujisawa et al., 2019; Miguel, Pereira, Silveira, & Meaney, 2019; Robakis et al., 2022; Senut et al., 2012). Behavior geneticists, for whom linear-additive models of gene and environment account for variation in phenotypes, might overinterpret individual statistical associations between single-nucleotide polymorphisms (SNPs) and outcome measures (e.g., brain volume at a particular point in time) and leave unexamined the entwined, dynamic nature of structural brain development.

The interpretation by M&H of polygenic scores as reflecting an individual's genetic "propensity" (target article, sect. 3.1, para. 4) or "risk" (target article, sect. 3.4, para. 3) further highlights a neoclassical view of DNA as an unmoved mover. Their deployment of polygenic scores to compare the size of "genetic effects" on educational attainment across contexts (target article, sect. 3.3, paras. 8–9), for example, also assumes that polygenic scores capture polymorphisms "for" relevant traits in a context-general sense. This account fails to leave explanatory room for multifinality, obscuring plausible biological and/or social intermediaries (e.g., neurodevelopment, sociocultural biases) between SNPs and the target outcome (Kaplan & Turkheimer, 2021). In contrast, a developmental systems approach engages with the multilevel complexities of how phenotypic variation is generated (Gawne, McKenna, & Nijhout, 2018) and with notions of inheritance that extend beyond DNA (Jablonka & Lamb, 2005). Such an account sees the genome as one resource (among many) used by the developmental system to grow (Overton, 2010) and recognizes the importance of developmental change and associated variation in psychobiological processes, such as epigenetic influences on homeostatic self-regulation (Cao-Lei et al., 2016). If the neoclassical view of DNA and genes, combined with a neglect of developmental process, remains the foundation of behavior genetics, any amount of methodological and statistical prowess in GWAS approaches will fail to move us forward in terms of understanding the complexities of human behavior.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interest. None.

References

Cao-Lei, L., Veru, F., Elgbeili, G., Szyf, M., Laplante, D. P., & King, S. (2016). DNA methylation mediates the effect of exposure to prenatal maternal stress on cytokine production in children at age 13½ years: Project Ice Storm. *Clinical Epigenetics*, *8*, 54. https://doi.org/10.1186/s13148-016-0219-0

Fjell, A. M., Chen, C., Sederevicius, D., Sneve, M. H., Grydeland, H., Krogsrud, S. K., … Walhovd, K. B. (2019). Continuity and discontinuity in human cortical development and change from embryonic stages to old age. *Cerebral Cortex*, *29*, 3879–3890. https://doi.org/10.1093/cercor/bhy266

Fujisawa, T. X., Nishitani, S., Takiguchi, S., Shimada, K., Smith, A. K., & Tomoda, A. (2019). Oxytocin receptor DNA methylation and alterations of brain volumes in maltreated children. *Neuropsychopharmacology*, *44*, 2045–2053. https://doi.org/10.1038/s41386-019-0414-8

Gawne, R., McKenna, K. Z., & Nijhout, H. F. (2018). Unmodern synthesis: Developmental hierarchies and the origin of phenotypes. *BioEssays*, *40*, 1600265. https://doi.org/10.1002/bies.201600265

Gottlieb, G. (1995). Some conceptual deficiencies in "developmental" behavior genetics. *Human Development*, *38*, 131–141. https://doi.org/10.1159/000278306

Griffiths, P. E., & Stotz, K. (2006). Genes in the postgenomic era. *Theoretical Medicine and Bioethics*, *27*, 499–521. https://doi.org/10.1007/s11017-006-9020-y

Griffiths, P. E., & Tabery, J. (2008). Behavioral genetics and development: Historical and conceptual causes of controversy. *New Ideas in Psychology*, *26*, 332–352. https://doi.org/10.1016/j.newideapsych.2007.07.016

Jablonka, E., & Lamb, M. J. (2005). *Evolution in four dimensions: Genetic, epigenetic, behavioral, and symbolic variation in the history of life*. MIT Press.

Kaplan, J. M., & Turkheimer, E. (2021). Galton's Quincunx: Probabilistic causation in developmental behavior genetics. *Studies in History and Philosophy of Science*, *88*, 60–69. https://doi.org/10.1016/j.shpsa.2021.04.001

Keller, E. F. (2014). From gene action to reactive genomes. *The Journal of Physiology*, *592*, 2423–2429. https://doi.org/10.1113/jphysiol.2014.270991

Li, G., Nie, J., Wang, L., Shi, F., Lin, W., Gilmore, J. H., & Shen, D. (2013). Mapping region-specific longitudinal cortical surface expansion from birth to 2 years of age. *Cerebral Cortex*, *23*, 2724–2733. https://doi.org/10.1093/cercor/bhs265

McManus, C. J., & Graveley, B. R. (2011). RNA structure and the mechanisms of alternative splicing. *Current Opinion in Genetics & Development*, *21*, 373–379. https://doi.org/10.1016/j.gde.2011.04.001

Meaney, M. J. (2010). Epigenetics and the biological definition of gene × environment interactions. *Child Development*, *81*, 41–79. https://doi.org/10.1111/j.1467-8624.2009.01381.x

Miguel, P. M., Pereira, L. O., Silveira, P. P., & Meaney, M. J. (2019). Early environmental influences on the development of children's brain structure and function. *Developmental Medicine & Child Neurology*, *61*, 1127–1133. https://doi.org/10.1111/dmcn.14182

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

Neumann-Held, E. M. (2001). Let's talk about genes: The process molecular gene concept and its context. In S. Oyama, P. E. Griffiths & R. D. Gray (Eds.), *Cycles of contingency: Developmental systems and evolution* (pp. 69–84). MIT Press.
Newman, S. A. (2019). Inherency of form and function in animal development and evolution. *Frontiers in Physiology*, *10*, 702. https://doi.org/10.3389/fphys.2019.00702
Overton, W. F., & Lerner, R. M. (2014). Fundamental concepts and methods in developmental science: A relational perspective. *Research in Human Development*, *11*, 63–73. https://doi.org/10.1080/15427609.2014.881086
Overton, W. F. (2010). Life-span development: Concepts and issues. In R. M. Lerner (Series Ed.) & W. F. Overton (Vol. Ed.), *Handbook of life-span development: Vol. 1. Cognition, biology, and methods across the lifespan* (pp. 1–29). Wiley. Retrieved from https://doi.org/10.1002/9780470880166.hlsd001001
Oyama, S. (1985). *The ontogeny of information*. Cambridge University Press.
Piatigorsky, J. (2007). *Gene sharing and evolution: The diversity of protein functions*. Harvard University Press.
Portin, P., & Wilkins, A. (2017). The evolving definition of the term "gene". *Genetics*, *205*, 1353–1364. https://doi.org/10.1534/genetics.116.196956
Robakis, T. K., Roth, M. C., King, L. S., Humphreys, K. L., Ho, M., Zhang, X., … Gotlib, I. H. (2022). Maternal attachment insecurity, maltreatment history, and depressive symptoms are associated with broad DNA methylation signatures in infants. *Molecular Psychiatry*, *27*, 3306–3315. https://doi.org/10.1038/s41380-022-01592-w
Senut, M. C., Cingolani, P., Sen, A., Kruger, A., Shaik, A., Hirsch, H., … Ruden, D. (2012). Epigenetics of early-life lead exposure and effects on brain development. *Epigenomics*, *4*, 665–674. https://doi.org/10.2217/epi.12.58
The ENCODE Project Consortium (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, *447*, 799–816. https://doi.org/10.1038/nature05874
The ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*, 57–74. https://doi.org/10.1038/nature11247
West-Eberhard, M. J. (2003). *Developmental plasticity and evolution*. Oxford University Press.
Witherington, D. C. (2011). Taking emergence seriously: The centrality of circular causality for dynamic systems approaches to development. *Human Development*, *54*, 66–92. https://doi.org/10.1159/000326814

On the big list of causes

Eric Turkheimer

University of Virginia, Charlottesville, VA, USA
ent3c@virginia.edu; https://uva.theopenscholar.com/eric-turkheimer

doi:10.1017/S0140525X22002096, e205

Abstract

The methodological shift from twin studies to genome-wide association studies (GWASs) diminished estimates of true genetic causation underlying statistical heritability of behavioral differences. The sum total of causal genetic influence on behavior is not zero, but, (a) no one cited in the target article ever thought this was the case, and (b) there is still little known about concrete instances of genetic causation.

The target article is pitched as an endorsement of genetic causation, but is mostly concerned with discounting it. Start at the beginning: What do we actually know, with dead solid no-philosophy certainty, about the relationship between the human genome and complex behavioral phenotypes? We know there is a correlation between genetic and phenotypic similarity. In humans (as opposed to the farm animals for whom the concept was originally designed) the "heritability" of a trait is estimated as the unstandardized slope of the regression of phenotypic on genotypic similarity. I will refer to this as the "G-P correlation" to steer clear of the briar patch surrounding heritability *per se*.

Pairs of identical twins are more similar than pairs of fraternal twins, for more or less every behavioral trait. Genetically more similar pairs in a GREML matrix are more similar than less similar pairs, once again for every phenotype. (The magnitude of the GREML G-P correlation is usually much smaller than it is for twins; more about this below.) Although it depends on how you measure it, the G-P correlation is not small. Especially during the twin study era of the previous scientific generation, G-P correlations for the common objects of investigation – intelligence, personality, mental illness, problematic behavior like criminality – were estimated to be from 0.4 to 0.8.

All correlations are caused by something. Thinking about what might cause the G-P correlation of a behavioral phenotype focuses our attention on what we don't know. In the twin study era we did not know which genes were involved; without knowing the genes we could obviously not know anything about the direction, let alone the mechanism, of the genetic effect. Without the direction or the mechanism of the effect, we could know nothing about its potency or its scope. A G-P correlation could be caused by the hard biological consequences of rare genes of large effect, systems of polygenes operating on endophenotypes, violations of the equal environments assumption, uncontrollable correlations between genetic and environmental effects, gene–environment interactions, and so on. The activity of sorting through this hairball of causes, mostly without the benefit of experimental control, is called social science genomics. Social science has its virtues, and twin studies produced a great deal of social science, but twin studies were decidedly unsuccessful at identifying genetic (or for that matter environmental) causes.

My major difference with the authors of the target article involves their contention that this state of affairs was fundamentally changed by the completion of the human genome project and the development of genome-wide association studies (GWASs) and their attendant methods. Modern DNA-based genomics, to be sure, has provided the ability to conduct genetically informed social science in new ways, but these new methods, as well-documented by the target article, have all served to *diminish* our estimation of genetic causation. First, linkage and association studies showed that there are few big genes down there, at least not for behavioral traits in the normal range. Then GWASs showed there were no non-tiny genes down there, and that G-P correlations estimated among unrelated people were much smaller than twin correlations, closer to 0.2 than to 0.6. Then, in what is undeniably the most interesting social scientific method developed since the multivariate twin study, all of the within-family methods described in the target article reduced the causally relevant part of the G-P correlation for behavioral traits by at least half. The actionable part of the correlation, estimated as a real number in the form of a polygenic score (PGS), is less than that, under 5% for even the most intensely studied traits, with samples in the millions. Might there be some solid gold genetic causes down there somewhere in the remaining 5%? Maybe, but with a few exceptions we still don't know the which, why, how, or where of any of them.

Zuckerman (1987) observed on these pages that "All parents are environmentalists until they have their second child." I have three children, and although their parents happen to be concordant for educational attainment (EA), there remains some within-family variation. It seems reasonable that 5% of those differences (the metric doesn't quite apply, but you get the idea) are causally related to unknown, but certainly complex, genetic differences. If there is a big list somewhere called "causes of human differences in behavior," genetic differences deserve to be on it. The nature side of

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

the old debate, however, sometimes lures opponents into defending the absurd null hypothesis that the net causal genetic effect on human behavioral differences is zero. Does anyone actually believe this? I (third from the skeptical left in Fig. 1 of the target article) certainly don't. Block (1995, second from left) doesn't deny genetic causation; in fact, just like the target article it is *about* genetic causation, and the inadequacy of heritability coefficients for quantifying it. Lewontin (1974, first on the left) was, after all, a geneticist: "The analysis of causes in human genetics is meant to provide us with the basic knowledge we require for correct schemes of environmental modification and intervention" (p. 525).

We can now see the outcome of the nature–nurture debate as regards human behavioral differences, and it is not what anyone expected. Genetic differences among humans do not determine behavioral differences. Although genetic differences aren't irrelevant to behavioral differences either, genetic causation of human behavior is weak, thin, contingent, gloomy, first-generation, call it what you will. That conclusion comes as no surprise. The surprise is that the crucial discoveries about the limitations of genetic causation were made not by environmentalists, left-leaning scientists, or philosophers of causation, but by the geneticists themselves. Once upon a time the so-called genetic revolution produced a paradigm shift in social science by showing that human genetic and behavioral differences are always correlated. Now it has produced a second shift, a counter-revolution, by showing that while G-P correlations have certain methodological consequences for social scientific practice, the direct causal genetic effects that underlie them are so small and indeterminate as to place few constraints on our individual or collective self-determination.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interest. None.

References

Block, N. (1995). How heritability misleads about race. *Cognition*, *56*(2), 99–128.

Lewontin, R. C. (1974). Annotation: The analysis of variance and the analysis of causes. *American Journal of Human Genetics*, *26*(3), 400.

Zuckerman, M. (1987). All parents are environmentalists until they have their second child. *Behavioral and Brain Sciences*, *10*(1), 42–44.

Authors' Response

Causal complexity in human research: On the shared challenges of behavior genetics, medical genetics, and environmentally oriented social science

James W. Madole[a,b] and K. Paige Harden[a]

[a]Department of Psychology, University of Texas at Austin, Austin, TX, USA and [b]VA Puget Sound Health Care System, Seattle, WA, USA
jmadole@utexas.edu
harden@utexas.edu

doi:10.1017/S0140525X23000833, e206

Abstract

We received 23 spirited commentaries on our target article from across the disciplines of philosophy, economics, evolutionary genetics, molecular biology, criminology, epidemiology, and law. We organize our reply around three overarching questions: (1) What is a cause? (2) How are randomized controlled trials (RCTs) and within-family genome-wide association studies (GWASs) alike and unalike? (3) Is behavior genetics a qualitatively different enterprise? Throughout our discussion of these questions, we advocate for the idea that behavior genetics shares many of the same pitfalls and promises as environmentally oriented research, medical genetics, and other arenas of the social and behavioral sciences.

R1. Introduction

> When opposing groups of intelligent, highly educated, competent scientists continue over many years to disagree, and even to wrangle bitterly about an issue they regard as important, it must sooner or later become obvious that the disagreement is not a factual one.… If this is, as I believe, the case, we ought to consider the roles played in this disagreement by semantic difficulties arising from concealed differences in the way different people use the same words, or in the way the same people use the same words at different times; … and by differences in their conception of what is an important problem and what is a trivial one, or rather what is an interesting problem and what is an uninteresting one. (Lehrman, 1970, pp. 18–19)

Behavior genetics is a topic about which scientists and scholars have continued over many years to disagree, sometimes bitterly. One feature that makes this rancorous debate curious is that the overarching conclusion of behavior genetics – namely, that there *are* causal genetic effects on human behavioral differences – is often described as obvious and uninteresting. **Turkheimer**, for instance, wrote off the "absurd null hypothesis" that genetic effects on behavior are zero, questioning: "Does anyone actually believe this?" Even the most negativistic commentator (**Burt**) began with the premise that "genetic differences matter for human social outcomes – achievements, behavior, physical health, personality – in a complex, context-sensitive way." Yet our attempt to describe how one might go about conceptualizing, identifying, and leveraging causal genetic effects on human behavioral differences – the very effects that, we are told, everyone *already* believes exist – inspired 23 widely divergent commentaries.

We consider the diversity of the commentators' opinions, and the intensity of their sentiments, a sign of success: our paper surfaced profound disagreements about what, in the study of human behavior, constitutes an important problem versus a trivial one, an interesting problem versus an uninteresting one, and – to add to Lehrman's list – an in-practice difficult problem versus an in-principle impossible one. Commentators contested nearly every single one of our arguments, but not everyone disagreed with the same arguments, and the commentators also disagreed with each other. Further complicating matters, a few commentators disagreed with themselves, advancing contradictory points within the space of their short replies, and a few expressed agreement-masquerading-as-disagreement – that is, they wrote *as if* they were in sharp conflict with our target article, but they were, in fact, restating positions we also advanced.

Our synthesis of these disparate viewpoints is necessarily imperfect (like behavioral genetics itself!): it flattens multidimensional arguments and glosses over nuance. We encourage readers

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

to (re)read the individual commentaries to reconstitute the details that have been lost. In particular, we recommend the commentaries by **Bourrat**; **Ross, Kendler, & Woodward, (Ross et al.)**; **Lynch, Brown, Strasser, & Yeo (Lynch et al.)**; **Syed**; and **Durlauf & Rustichini**, as we thought these authors brought fresh and incisive perspectives to a topic that can often recycle the same stale points and counterpoints. With these caveats and recommendations in mind, the commentaries on our target article can be read as speaking to three overarching questions, which we will consider in turn:

(1) What is a cause?
(2) How are randomized controlled trials (RCTs) and within-family genome-wide association studies (GWASs) alike and unalike?
(3) Is behavior genetics a qualitatively different enterprise?

R2. What is a cause?

Even before genomic data entered the mix, commentators disagreed about what, exactly, made something a "cause," and what was needed to infer a causal relationship. **Ross et al.** aptly summarized our treatment of causal inference:

> M&H adopt a broadly "interventionist" treatment of causation – the minimal condition for some factor C to count as a cause for an outcome E is that if, hypothetically, unconfounded manipulations of C were to be performed these would lead to changes in E. In the familiar case of a randomized experiment, this leads to the conclusion that an average causal effect (ACE) is a legitimate causal notion. M&H observe that an ACE can be present even though C does not have a uniform effect, even though a similar ACE may not be present in populations different from the population from which the experimental sample was drawn, and even though the experiment tells us nothing about the mechanism by which Cs cause Es. We agree.

Their endorsement of our perspective on causation is not surprising (but is reassuring), as our understanding of causation was strongly informed by Woodward's (2005) previous work, particularly his book *Making Things Happen*. This interventionist perspective on what makes something a cause continues the lineage of Holland's (1986) decree: "No causation without manipulation" – even if that manipulation can happen only hypothetically, that is, in the form of a thought experiment.

It turns out that not everyone is a Holland acolyte. Some commentators emphasized *regularity* accounts of causation (Mill, 1843/2002), which prioritize concepts like temporal precedence and the repeated co-occurrence of **X** and **Y**. **Hart & Schatschneider** defined causation as follows: "the cause must precede the effect, second, the cause must be related to the effect, and third, we can find no plausible explanation for the effect other than the cause." Likewise, **Tarchi, Merola, Castellini, & Ricca (Tarchi et al.)** suggested that "[r]ecent developments in the regularity theory of causation (based on the premises that causes are regularly followed by their effects) have allowed for a precise estimation and quantification of confidence in causal relationships." Other commentators noted or implied that causation could (should?) be conceptualized in terms of prediction (**Shen & Feldman**) or mechanism (**Smith & Downes**).

A complete defense of a broadly interventionist treatment of causation is obviously beyond the scope of this reply. "What is appropriately considered a cause?" and "What is appropriately considered evidence for a cause?" are questions that occupy entire careers. More simply, we have two recommendations to readers who are sifting through the various comments.

First, we ask you to reflect on how you *typically* answer those questions when the causes under consideration are not genetic in nature. Imagine, for instance, that you were reviewing a paper showing that children whose families were randomly selected to receive housing vouchers to move out of low-income neighborhoods were more likely, on average, to attend college (Chetty, Hendren, & Katz, 2016). Would you object if the authors concluded that their results were consistent with a *causal* effect of neighborhood environment on educational attainment? Even if the effect increased college attendance by only 2.5 percentage points? Even if you couldn't perfectly predict who would go to college just from knowing whether they moved? Even if the authors could only speculate about the mechanisms linking neighborhood characteristics with educational attainment? Even if the neighborhoods to which people moved were all different from one other, such that everyone in the "treatment" group experienced quite different treatments?

Your answers to such questions are informative about what you *already* think is, and is not, necessary to infer causation, and what you believe to be the scope of that causal inference. We suggest keeping these priors in mind when considering the question of what it means for genes to be causal. Consistency might be the "hobgoblin of little minds" (Emerson, 1841/1993), but in this case, we think some semantic and conceptual consistency about what makes something a "cause" is important for empirical design, statistical interpretation, scientific theory-building, and policy application. Referring back to the Lehrman quotation that we used as an epigraph to this reply, we urge you to avoid using the word "cause" differently at different times.

Second, we encourage you to remember that the binary judgment of whether or not **X** is a cause of **Y** is not the only judgment on the table. As **Cerezo** reminded us in her comment, there are "metaphysical tools" in our toolbox for describing and differentiating among types of causes (see also Kinney, 2019; Ross, 2021, 2022). As we described in our target article, and as we will continue to describe in this reply, inferring that **X** caused **Y** is one of the *first* steps toward understanding the **X–Y** relationship, not the last.

R3. How are RCTs of environmental interventions and within-family GWASs alike and unalike?

In our target article, we described how humans have two copies of every gene, and offspring inherit, at random, just one of them. Because genotypes are randomly "assigned," conditional on the parental genotypes, average phenotypic differences between family members who have been randomly assigned to different genotypes are conceptually analogous to the average difference between people who have been randomly assigned to treatment and control groups in a randomized controlled trial (RCT). That is, some genetic designs, because they capitalize on the randomness of genetic inheritance within families, can estimate the average causal effect of genotypes on phenotypes.

Many of the commentators focused on what they perceived to be the limits of that analogy. Most commonly, they described desirable features of RCTs that they held to be lacking in a within-family genetic study, claiming that (1) only RCTs involve a manipulation of the causal stimulus, (2) RCTs have greater uniformity of the causal stimulus, and (3) RCTs give more actionable information. A few commentators took a different tack, pointing

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

out the ways in which within-family genetic studies share *un*desirable qualities of RCTs, in particular that (4) RCTs make an unsatisfactory trade-off between internal validity and external validity. Although many of these points were incisive, we believe that the gap between within-family genetic studies and nearly all RCTs in the social and behavioral sciences is smaller than most commenters acknowledged.

R3.1. Which designs involve a randomized manipulation?

Some commentators disagreed with the fundamentals of our analogy. For instance, **Hart & Schatschneider** objected to comparing within-family genetic studies to RCTs on the grounds that the former "do not have a manipulation," whereas **Kaplan & Bird** implied that only RCTs have a "randomizing element." We disagree. Although there is no *artificial* manipulation of the genome in human behavioral genetics, the conception of every human involves a *natural* randomized manipulation of genetic material. Indeed, it's ironic that Hart & Schatschneider warned against "forcing the language of experiments … on within-family designs," because the language of experiments actually comes from genetics! In his foundational work on experimental designs, Fisher named experimental "factors" after "Mendelian factors," and strove to create randomization schemes that mimicked the randomization of genetic inheritance. In their comment, **Pingault, Fearon, Viding, Davies, Munafò, & Davey Smith (Pingault et al.)** quoted Fisher (1952) on exactly this point:

> The parallel drawn by [Madole & Harden] was made explicitly by Fisher who established a direct filiation between the (artificially) randomized design he theorized and the (natural) randomization of genetic material at conception, in his words: "*the factorial method of experimentation, now of lively concern so far afield as the psychologists, or the industrial chemists, derives its structure and its name, from the simultaneous inheritance of Mendelian factors.*" (emphasis added)

R3.2. Which designs are complicated by causal stimulus heterogeneity?

In our target article, we discussed the ways in which causal effects can be *non-uniform*, in that they can produce heterogeneous effects across individuals because of moderation by other causal factors. **Lynch et al.** incisively raised the issue of another source of heterogeneity, *causal stimulus heterogeneity*, when not everyone in the "treatment" group receives the same causal treatment. Causal stimulus heterogeneity is patently a problem when using polygenic scores (PGSs), which aggregate effects across very many single-nucleotide polymorphisms (SNPs): two people might have equivalently high or low PGS values, but arrive there via non-overlapping sets of SNPs. Lynch et al. cleverly analogized a PGS to a "a drug with thousands of ingredients of small efficacy, where each pill has one ingredient or an alternative at random according to a defined chance procedure."

Ross et al. raised a similar concern, noting that it applies not only to PGSs, but also to associations with individual single-nucleotide polymorphisms (SNPs), because each SNP "tags" information about other genetic variants (including unmeasured variants) that are in linkage disequilibrium (LD) with the focal SNP included in a genome-wide association study (GWAS). They provided a similar analogy to **Lynch et al.**:

> Assuming the random nature of meiosis, a GWAS corresponds to a huge number of different randomized treatments in the population: e.g., A versus C at SNP1, G versus T at SNP2 and so on. … Indeed, matters are even more complex since haplotypes are randomized not SNPs. We might perhaps conceptualize this as the assignment of randomized bottles to subjects, each containing a mixture of different drugs.

Related concerns were raised by **Borger, Weissing, & Boon (Borger et al.)** ("… thousands of single nucleotide polymorphisms (SNPs) are considered simultaneously"), and by **Pingault et al.**, who *also* analogized polygenic influences to a drug cocktail, the formulation of which differs across people ("… in this case, the 'treatment' is not well defined (in content or timing) … [it is] like a drug RCT consisting of the simultaneous administration of hundreds of compounds").

Lynch et al. correctly pointed out that one can still conclude, on the basis of an appropriately randomized study, that a heterogeneous causal stimulus "works," in that it has a non-zero average treatment effect. But, figuring out *how* it works is be especially challenging:

> The high causal stimulus heterogeneity is likely to produce non-uniform causal pathways from the very first steps, thus making it difficult or impossible to trace mechanisms from particular drug ingredients [i.e., from particular genetic loci] given only associations between treatments [i.e., PGSs] and outcomes.

We agree that causal stimulus heterogeneity does make mechanistic understanding considerably more difficult. But, as we will further explain, we do not think this is a problem that is unique, or uniquely difficult, to the study of genetic causes.

Most discussions of causal stimulus heterogeneity implicitly or explicitly contrasted the messiness of PGSs and SNP arrays with the supposed homogeneity of the causal stimulus in RCTs. **Lynch et al.**, for instance, wrote: "In most RCTs, individuals in the treatment group receive the same, or as similar as possible, treatment or causal stimulus, such as a drug or educational intervention (causal stimulus homogeneity)." **Siegel** made a similar claim, writing that "RCTs of treatments that prove to be highly efficacious directly demonstrate their greater uniformity of therapeutic effects."

From our perspective as clinical psychologists who have worked to deliver "empirically supported" psychotherapy to patients in clinical practice, who have been therapists on RCTs of novel psychotherapeutic interventions, and who have implemented educational interventions in our own classrooms, these characterizations of the alleged homogeneity of environmental interventions sound disconnected from the reality of social and behavioral science. Environmental interventions are, on the whole, more like a PGS than they are like lithium: they are cocktails "with thousands of ingredients of small efficacy," the formulation of which differs across people.

This is perhaps most obvious in the case of therapeutic and educational interventions delivered one-on-one. In an RCT of cognitive behavioral therapy, for instance, the "underlying treatment … may in a real sense differ for every single unit," because the content of every session is tailored specifically to the individual (Smith, 2022, p. 656). But even interventions that are not psychotherapeutic might be, in practice, implemented in a highly idiosyncratic way (see, e.g., an ethnography of welfare case workers by Watkins-Hayes, 2009). And, many interventions in the social and behavioral sciences package together multiple services, not all of which are taken up (or taken up in the same way) by every participant. Consider the High/Scope Perry Preschool Program (HPPP), which we discussed in our target article. This

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

intensive intervention combined attending preschool for 2.5 hours, 5 days a week, for 2 years, home visits by teachers for 1.5 hours per week, and monthly small groups for parents. Given the complexity of the intervention, "one is still left without actually being able to pinpoint what it was in the Perry Preschool Project that actually influenced later adult outcomes" (Schneider & Bradford, 2020, p. 52).

In this way, a binary variable reflecting the presence or absence of intent-to-treat in an environmental RCT does not always, or even usually, give us granular information about the relevant difference-maker(s) or assure that the relevant difference-maker (s) are experienced homogenously across people. Rather, a binary treatment indicator often represents a gross simplification (Heiler & Knaus, 2021). The situation with oft-cited naturally occurring environmental exposures may be of even lower resolution: "treatments" like being drafted into the military (Angrist, 1990) or living in a region of Holland occupied by the Nazis (Stein, Susser, Saenger, & Marolla, 1972) are hardly homogenous or "well-defined."

Thus, even as philosophers and plant geneticists extol the homogeneity of environmental interventions in the social and behavioral sciences, interventionists themselves paint quite a different picture: "We must expect, study and capitalize on the heterogeneity that characterizes most effects in science" (Bryan, Tipton, & Yeager, 2021, p. 986). In fact, environmental interventionists sound remarkably like behavioral geneticists: "The researcher faces a tough trade-off between interpretability and statistical power or, put differently, between learning about the effects of the underlying heterogeneous treatments and the sample size available for studying each treatment" (Smith, 2022, p. 656). These quotes illustrate that the problem of causal stimulus heterogeneity, while definitely a formidable challenge to mechanistic understanding, is not a challenge that is unique to the study of *genetic* causes, but is rather a difficulty that besets most studies in the social and behavioral sciences.

In light of this shared challenge, we wholeheartedly agree with **Bondarenko** that the "'second-generation' goals of causal inquiry in the context of human behavior cannot be achieved by genetics alone, nor do genetically informed research designs provide the only possible path toward mechanistic understanding." Indeed, at no point in our target article did we suggest that genetics can achieve anything alone, nor did we suggest (as **Smith & Downes** alleged) that we "hope to supplant" environmental studies. Although we are optimistic that deeper measurements of the genome and further advances in fine-mapping and gene prioritization methods will result in a higher resolution understanding of genetic difference-makers, we also think that behavior genetics should incorporate more of the conceptual and methodological tools developed by environmental interventionists who are taking heterogeneity seriously, including methods for causal inference when there are multiple versions of treatment (VanderWeele, 2022; VanderWeele & Hernan, 2013). As **Durlauf & Rustichini** highlight, formal theoretical structures are needed to reveal sources of heterogeneity and generate deeper causal explanations of how biopsychosocial systems produce human behavior. Again, we agree with commentators like **Taylor, Weiss, & Marshall** that the "genome [is] one resource (among many) used by the developmental system to grow," and indeed, our entire discussion of average difference-makers as *non-unitary* causes is based on the idea that genes "operate within intricate causal systems" (target article, abstract). As we will discuss next, we also believe that, by virtue of being a difference-making component of the causal system, genetic data may play a key role in helping identify complex etiological models.

R3.3. Which designs give actionable information?

One word that recurred throughout the commentaries was "actionable." **Turkheimer** intimated that knowledge of genetic causes was not actionable because the effect sizes are too small: "The actionable part of the [genotype-phenotype] correlation, estimated as a real number in the form of a PGS, is … under 5% for even the most studied traits…" The 5% figure refers to the within-family effect size of an educational attainment polygenic score (PGS) on years of education in North American samples who have "European" genetic ancestry. Effects of this magnitude were trivialized as "weak" and "small and indeterminate."

We disagree with this characterization. An R^2 of 5% is approximately equal to a correlation (r) of 0.22, or to (assuming equal-sized groups) a Cohen's d of 0.46. By comparison, a study of Swedish children who were adopted into better socioeconomic circumstances found that adoption increased IQ scores, relative to the children's siblings who were reared by their biological parents (Kendler, Turkheimer, Ohlsson, Sundquist, & Sundquist, 2015) by around 4.5 IQ points, or a Cohen's d of 0.34. This effect of "family environment" (certainly a heterogeneous causal stimulus!) was described as "a significant advantage in IQ" (Kendler et al., 2015, p. 4612). Another study of children randomized to foster care, rather than to severely deprived institutional care, found that foster care increased average IQ scores at age 12 by $d = 0.41$ (Almas, Degnan, Nelson, Zeanah, & Fox, 2016). Yet another study examined the effects of a major educational reform in Sweden, which increased the number of years of compulsory schooling, abolished academic tracking at grade six, and rolled-out a unified national curriculum (Meghir & Palme, 2005). Researchers leveraged the gradual implementation of the reform across different areas in Sweden and concluded that the reform increased years of schooling by about 3.5 months, $d \sim 0.2$. What these examples illustrate is that the estimated causal effect of the educational attainment PGS rivals the effect sizes that we observe when people experience radically sweeping changes to their environmental context.

Most effect sizes for specific environmental interventions are even smaller than 5%, and this is exactly what we would expect given the causal complexity of human behavior. Kraft (2020), for instance, reviewed all of the educational studies funded by the U.S. government's Investing in Innovation fund, and found that the median effect size was $d = 0.03$. They concluded: "effects of 0.15 or even 0.10 *SD* should be considered large and impressive" (p. 248). Yeager and Dweck (2020) similarly summarized: "In the real world, single variables do not have huge effects. Not even relatively large, expensive, and years-long reforms do. If psychological interventions can get a meaningful chunk of a .20 effect size on real-world outcomes in targeted groups, reliably, cost-effectively, and at scale, that is impressive" (p. 1281). We agree. Our conclusion that genetic effect sizes are, for some phenotypes, impressive does not stem from grandiosity about genetics, but rather from humility about the difficulty of specifying *any* cause, artificially manipulated or naturally varying, that accounts for even a few percentage points of the variance in complex behavior.

Others pointed out that, unlike the results of some RCTs, causal genetic effects are not directly actionable because we

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

cannot artificially manipulate the genome on the basis of that information. **Pingault et al**. wrote: "Thus, while RCTs can provide actionable evidence of a specific intervention's efficacy, a within-family genetic association only indicates the effect of inheriting one variant or another." **Markon** similarly pointed out that "counterfactual theory … takes an unactionable epistemological stance: even if a counterfactual account informs about what would have occurred had things been different, it does not inform about what one can do now, given things as they are." **Kaplan & Bird** concurred: "[t]he 'shallowness' of the causal knowledge gained in RCTs does not prevent them from being useful guides to practice … the situation in behavior genetics is nothing like this. Unlike in the case of RCTs, we cannot change the genetic variants associated with the phenotypic variation – and even if we could, doing so would be wildly irresponsible." We made a similar point in our target article, writing that "even if we concede that, at a conceptual level, genes could cause average differences in human behavior, at a practical level, it is not readily apparent what we would do with this knowledge.… [W]e cannot (and should not) readily apply knowledge of genetic causes to change the genomes of large swathes of the population in the hopes of changing their outcomes" (target article, sect. 1.2, para. 6).

Given that we are not planning to change people's genotypes, **Kaplan & Bird** ask, "what would we gain from even an accurate finding that a particular genetic variant was associated with downstream effects?" This is a curious question to pose, as it is precisely the question that we address, at length, in the target article on which they are supposedly commenting. Our short answer is that we agree with Lewontin (1974) (quoted by **Turkheimer**): "The analysis of causes in human genetics is meant to provide us with basic knowledge we require for correct schemes of environmental modification and intervention" (p. 409).

And, so far, the project of devising correct schemes of environmental modification and intervention has been far less successful than commonly imagined (Kraft, 2020). Given that there is clearly room for improvement for successful identification of environmental intervention targets, we agree with **Bondarenko** that one "important role for genetic data" is as "controls in the study of environmental variables … when applied with care, genetic controls may help address some of the worries that our 'first-generation' knowledge of environmental factors does not meet a stringent epistemic standard." We also agree with commentators who pointed out that that genetic results can be usefully exploited in Mendelian randomization studies (**Pingault et al.**) for the study of *phenotypic* causation (Pingault, Richmond, & Davey Smith, 2022), which "given their speed and relative low cost" are a "useful first-step to guide future randomization/intervention studies" (**Haworth & Wootton**).

Finally, simply knowing that something has genetic causes – even if you can't identify them or manipulate them – can and has changed clinical practice. For example, genetic research on substance use disorders (SUDs) contributed to a paradigm shift in conceptualizing addiction as a chronic disease that resides within the body of the individual rather than as a moral deficit that resides within their will (Hall, Carter, & Forlini, 2015; Volkow & Koob, 2015). Now, educating patients about genetic effects on addiction is a standard part of psychoeducation that can reduce stigma and increase motivation for treatment (Hassan et al., 2021; Ray, 2012). Causal knowledge has the power to produce important changes in our intentions and actions as scientists and clinicians, whether those causes can be manipulated or not.

R3.4. Which designs have external validity?

Most commentators who were critical of our target article had the intuition that RCTs in the social and behavioral sciences were valuable research endeavors, and objected to our comparing within-family genetic studies to them. A few commentators, however, brought up that RCTs also have their limitations, most prominently, that they sacrifice external validity and generalizability for the sake of internal validity. **Shen & Feldman**, for instance, pointed out that causal knowledge built from within-family comparisons does not necessarily generalize on a population scale. **Borger et al.** also decried the "limited ability…to generalize," and **Siegel** warned that GWAS results "may not apply to other ethnic groups, a non-uniformity that may exacerbate health disparities." **Syed** offered a particularly trenchant summary of the problem:

> It is a fact of the design that RCTs sacrifice external validity for the sake of internal validity, being high in efficacy, showing promising results in trials, but low in effectiveness, or lack of results when translated to real-life conditions … RCTs have been further criticized … for their lack of inclusion of racial/ethnic minorities and thus limited generalizability.

For those interested in "humans in general" (**Byrne & Olson**), causal knowledge that is not built from culturally and demographically representative samples – and therefore not expected to apply equally across subgroups – is inherently flawed. As **Syed** said, "no diversity in, no causes at all." In contrast, others were less interested in "humans in general" knowledge, and were instead concerned with the "individual making important life-choices" (**Miller**). And, in an interesting counterpoint to Syed's commentary, **Bourrat** suggested that, in some cases, "local causal knowledge can be more useful for explanation and intervention than more generalisable knowledge." Similar to **Cerezo**'s emphasis on triggering conditions, Bourrat pointed out that accounting for context can reveal specificities about a causal relationship that are masked when aggregating more generally.

A medical example of the unique contribution of local causation comes from oncology research, where chemotherapy was found to have no average impact on patient survival in a cohort of individuals with stage IB lung cancer, but rather improved patient survival only in those with a tumor size greater than 0.4 centimeters (Strauss et al., 2008). This is what Bakermans-Kranenburg and Van Ijzendoorn (2015) referred to as the "hidden efficacy of interventions." The effect of chemotherapy on patient survival cannot be expected to generalize across all individuals with stage IB lung cancer, but it makes a significant difference for some people.

This example illustrates that locality is not always a curse and that non-portable causes can still sometimes be useful. Finding that a genetic effect holds within a particular ancestral group but not another, or is manifest under certain social conditions but not others, or applies within families but not between them, can be valuable because it allows us to build causal knowledge *within* family or cultural systems.

The tension between **Syed**'s and **Bourrat**'s commentaries – both of which we think make incisive and valuable points – highlights what Richard Levins described as "the contradictory desiderata of

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

generality, realism, and precision…" (Levins, 1966, p. 431). The more we attempt to generalize, the more we collapse over dimensions of variability that meaningfully influence the causal relationship; the more we specify local variables, the more we restrict the applicability of our findings (see Yarkoni [2022] and related commentaries for a discussion about the *precision of estimation/breadth of generalization trade-off*).

With these contradictory desiderata in mind, we strongly agree with **Syed** that randomization is not enough: who is being randomized across what dimensions of human experience? We endorse their conclusion that "diversity is central to first-generation studies" and that "the racial/ethnic diversity of samples included in behavior genetic studies… must be central to any effort to build generalizable causal knowledge." Similarly, we agree with **Byrne & Olson** that, behavior geneticists have a responsibility to make "analytic choices" that "enhance the visibility of context," in particular, sampling across sociocultural contexts that have heretofore been largely excluded from genetics research. And, we agree with **Eftedal & Thomsen** that "[b]ehavioral genetics should broaden its empirical scope beyond single-culture WEIRD samples."

Accordingly, we applaud recent efforts within the field to conduct GWASs in non-European ancestral groups (Gulsuner et al., 2020; Pereira, Mutesa, Tindana, & Ramsay, 2021), advance methodology to increase the integration of diverse samples in GWASs (Mathur et al., 2022), improve discovery of within-family genetic effects (Howe et al., 2022), develop equitable partnerships among international institutions that promote resource sharing and shared-infrastructure development (Martin et al., 2022), and build more representative biobanks like the Trans-Omics for Precision Medicine Program that allow for genetic discovery within cultural subgroups (Popejoy & Fullerton, 2016). These initiatives represent important steps toward building more representative causal knowledge. Finally, we also echo **Tarchi et al.**'s warning that an absolute requirement for a genetic study to have a causal identification strategy or to provide mechanistic understanding might have undesirable consequences for representation in genetics: "to consider 'association studies' as secondary or even detrimental should be critically evaluated," lest we exacerbating problems of exclusion in genetic research.

R4. Is behavior genetics a qualitatively different endeavor?

Although most commentators offered critical refinements to our target article, suggesting ways to improve future studies or adding nuance to the interpretation of existing effects, **Burt** denounced the entire enterprise of behavior genetics as "impracticable." In particular, they argued, while genetics "has the potential to advance understanding of human health and disease," it was not "appropriate" to apply genetic methods to the study of "complex, social, non-disease achievements or behaviors." (**Kaplan & Bird** similarly distinguished between what they referred to as "disease GWAS" and "sociobehavioral GWAS.") In contrast, we find the distinction between "disease" and "complex, social (non-disease)" behavior artificial, and think it is conceptually incoherent to champion the virtues of medical genetics and criticize the utility of behavior genetics in the same breath.

Despite the distinction between "disease" and "complex, social non-disease" phenotypes being central to their argument, **Burt** gives only a circular, "I know it when I see it" non-definition of what should be, in their view, off-limits to genetic study: "complex social traits are defined by social context and thus irreducibly social." But, the distinction between "disease" and "social (non-disease)" phenotypes is as contentious, fluid, and historically and culturally contingent, as the distinction between art and obscenity. Consider again the example of substance use disorders (SUDs) (a common subject of behavioral genetic research). In the last two decades, it has become increasingly popular to view SUDs under a medical model (Leshner, 1997). The American Society of Addiction Medicine (ASAM) defines addiction as "a primary, chronic disease of brain reward, motivation, memory and related circuitry" (ASAM, 2017). Yet many scholars have pushed back on the medicalization of SUDs (not to mention other psychiatric conditions) (Borsboom, Cramer, & Kalis, 2019; Heilig et al., 2021), arguing that it is "reductively inattentive to individual values and social context" (Courtwright, 2010, p. 144). This debate is not without stakes: legal challenges to the incarceration of individuals with substance-related charges have hinged on the question of whether SUDs are "diseases" or "behaviors" (Commonwealth v. Eldred, 2018). The most commonly used diagnostic system in North American psychiatry, the *DSM-V*, offers no clarity, defining SUDs as a constellation of biological (e.g., physiological withdrawal), behavioral (e.g., disengaging from hobbies to protect use), and social (e.g., use interfering with relationships) symptoms. In the current debates about whether SUDs should be viewed as a disease or as an "irreducibly social" behavior, we hear echoes of previous and ongoing debates about how to best understand, for example, melancholy, Asperger's, psychosis, sexual orientation. Burt would have us believe that they can resolve these debates by fiat.

The distinction that **Burt** draws between "complex social (non-disease)" versus "disease" maps onto their distinction between "downward social causation," in which "sociocultural forces … sort and select individuals based on genetically influenced traits" versus "upward genetic causation," which operates "from genetic differences to trait differences through biological pathways." The former is said to produce "artificial" genetic associations; the latter "authentic" ones.

This pat story neglects the role of "downward social causation" in disease and disabilities. Monogenic retinitis pigmentosa, for instance, is a rare disease that causes progressive loss of sight; an article in *Genome Medicine* summarized that the "first symptoms are retinal pigment on fundus evaluation, …eventually leading to legal blindness in a few decades" (Ayuso & Millan, 2010, p. 1). What an interesting phrase, "legal blindness"! Blindness is defined by visual acuity. What makes blindness "legal" is whether legislators and policymakers deem the loss of visual acuity to be sufficiently severe enough that one, for example, qualifies for Social Security disability benefits, can no longer operate an automobile, gets special tax exemptions. A monogenic disorder with a well-defined biological pathway causes legal blindness; legal blindness is an artificially bounded category that is created when sociocultural forces – like the IRS – sort and select individuals based on their genetically influenced traits. Legal blindness is genetically caused *and* is "irreducibly social."

And so is every other human phenotype. "Upward genetic causation" *and* "downward social causation" are always operating, on every human phenotype, because humans are social animals. Genes act on our bodies; society acts on our bodies. Society changes our biology; our biology changes how society responds to us. Every aspect of human life reflects these two streams of influence. Sometimes one stream is a raging torrent; sometimes the other is a trickle. But, regardless of their relative width and depth and ferocity, where the streams of influence mix and mingle

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

is, for us, the most scientifically interesting point of study. Confluences are sacred.

R5. Parting thoughts

In closing, let us revisit a quote from **Lynch et al.**: "The high causal stimulus heterogeneity is likely to produce non-uniform causal pathways from the very first steps, thus making it *difficult or impossible* to trace mechanisms from particular drug ingredients [i.e., from particular genetic loci] given only associations between treatments [i.e., PGSs] and outcomes" (emphasis added).

The "or" in their sentence captures a core disagreement surfaced in this set of commentaries: difficult or impossible? We think that it will, in practice, be *difficult* – very difficult – to trace the mechanisms by which polygenic causal signals make a difference for human behavior, and to leverage that knowledge to improve human lives. We do not think that it is, in principle, *impossible*.

Considered from one angle, our rejection of epistemological skepticism is not the least bit surprising: we are psychologists, a profession that, by definition, presupposes that we can know things scientifically about why humans think, feel, and behave the way they do, and strives to use that knowledge, even when it is incomplete and flawed, to change how humans think, feel, and behave. We are writing, after all, in a journal called *Behavioral and Brain Sciences*. What is striking about these commentaries is that some working scientists appear to embrace a highly selective epistemological skepticism: knowing, in their view, is impossible, but only knowing about genetic influences on behavior, not about genetic influences on "disease," or about environmental influences on behavior.

We think this selective skepticism is incoherent, because, as we've described in this reply, we see the conceptual and practical difficulties in understanding genetic influences on behavior as largely of a piece with other subfields. As Turkheimer (2012) so pithily summarized: "Genome-wide association studies of behavior *are* social science" (our emphasis added). If we are going to continue our audacious attempts to study free-range humans scientifically, then we will necessarily grapple with causal stimulus heterogeneity and trade-offs between internal and external validity and small effect sizes and opaque mechanisms and uncertain generalizability and "downward" social causation and causal complexity, whether our work seeks to understand genetic causes or environmental ones, whether our work focuses on diseases or behaviors.

If we treat complexity as a "dead end" (Plomin & Daniels, 1987), if we dismiss attempts to understand it as "far-fetched" (Kaplan & Turkheimer, 2021), if we arbitrarily declare some areas of inquiry "inappropriate," then we run the risk of missing out on meaningful progress. "Even if we never understand biology completely…we can understand enough to interfere" (Hayden, 2010, p. 667). The only way forward is to muddle through.

References

Almas, A. N., Degnan, K. A., Nelson, C. A., Zeanah, C. H., & Fox, N. A. (2016). IQ at age 12 following a history of institutional care: Findings from the Bucharest Early Intervention Project. *Developmental Psychology*, *52*, 1858.

American Society of Addiction Medicine (2017). Retrieved from https://www.asam.org/docs/default-source/advocacy/2-asam-general-information-flyer.pdf?sfvrsn=3c6f46c2_2

Angrist, J. D. (1990). Lifetime earnings and the Vietnam era draft lottery: Evidence from social security administrative records. *American Economic Review*, *80*, 313–336.

Ayuso, C., & Millan, J. M. (2010). Retinitis pigmentosa and allied conditions today: A paradigm of translational research. *Genome Medicine*, *2*, 1–11.

Bakermans-Kranenburg, M. J., & Van Ijzendoorn, M. H. (2015). The hidden efficacy of interventions: Gene×environment experiments from a differential susceptibility perspective. *Annual Review of Psychology*, *66*, 381–409.

Borsboom, D., Cramer, A. O., & Kalis, A. (2019). Brain disorders? Not really: Why network structures block reductionism in psychopathology research. *Behavioral and Brain Sciences*, *42*, E2. doi:10.1017/S0140525X17002266

Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour*, *5*, 980–989.

Chetty, R., Hendren, N., & Katz, L. F. (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, *106*, 855–902.

Commonwealth v. Eldred (2018). 480 Mass. 90. MA: Supreme Judicial Court. Retrieved from https://law.justia.com/cases/massachusetts/supreme-court/2018/sjc-12279.html

Courtwright, D. T. (2010). The NIDA brain disease paradigm: History, resistance and spinoffs. *BioSocieties*, *5*, 137–147.

Emerson, R. W. (1993). *Self-reliance, and other essays*. Dover (original work published in 1841).

Fisher, R. (1952). Statistical methods in genetics. *Heredity*, *6*, 1–12 (reprinted in 2010, *International Journal of Epidemiology*, **39**, 329–335). https://doi.org/10.1093/ije/dyp379

Gulsuner, S., Stein, D. J., Susser, E. S., Sibeko, G., Pretorius, A., Walsh, T., … McClellan, D. J. (2020). Genetics of schizophrenia in the South African Xhosa. *Science (New York, N.Y.)*, *367*, 569–573.

Hall, W., Carter, A., & Forlini, C. (2015). The brain disease model of addiction: Is it supported by the evidence and has it delivered on its promises? *The Lancet Psychiatry*, *2*, 105–110.

Hassan, A. N., Ragheb, H., Malick, A., Abdullah, Z., Ahmad, Y., Sunderji, N., & Islam, F. (2021). Inspiring Muslim minds: Evaluating a spiritually adapted psycho-educational program on addiction to overcome stigma in Canadian Muslim communities. *Community Mental Health Journal*, *57*, 644–654.

Hayden, E. C. (2010). Human genome at ten: Life is complicated. *Nature*, *464*, 664–667.

Heiler, P., & Knaus, M. C. (2021). Effect or treatment heterogeneity? Policy evaluation with aggregated and disaggregated treatments. *arXiv preprint arXiv:2110.01427*.

Heilig, M., MacKillop, J., Martinez, D., Rehm, J., Leggio, L., & Vanderschuren, L. J. (2021). Addiction as a brain disease revised: Why it still matters, and the need for consilience. *Neuropsychopharmacology*, *46*, 1715–1723.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*, 945–960. doi:10.1080/01621459.1986.10478354

Howe, L. J., Nivard, M. G., Morris, T. T., Hansen, A. F., Rasheed, H., Cho, Y., … Davies, N. M. (2022). Within-sibship genome-wide association analyses decrease bias in estimates of direct genetic effects. *Nature Genetics*, *54*, 581–592.

Kaplan, J. M., & Turkheimer, E. (2021). Galton's Quincunx: Probabilistic causation in developmental behavior genetics. *Studies in History and Philosophy of Science*, *88*, 60–69.

Kendler, K. S., Turkheimer, E., Ohlsson, H., Sundquist, J., & Sundquist, K. (2015). Family environment and the malleability of cognitive ability: A Swedish national home-reared and adopted-away cosibling control study. *Proceedings of the National Academy of Sciences*, *112*, 4612–4617.

Kinney, D. (2019). On the explanatory depth and pragmatic value of coarse-grained, probabilistic, causal explanations. *Philosophy of Science*, *86*, 145–167.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*, 241–253.

Lehrman, D. S. (1970). Semantic and conceptual issues in the nature–nurture problem. In L. R. Aronson (Ed.), *Development and evolution of behavior. Essays in memory of TC Schneirla* (pp. 17–52). WH Freeman.

Leshner, A. I. (1997). Addiction is a brain disease, and it matters. *Science (New York, N.Y.)*, *278*, 45–47.

Levins, R. (1966). The strategy of model building in population biology. *American Scientist*, *54*, 421–431.

Lewontin, R. C. (1974). Annotation: The analysis of variance and the analysis of causes. *American Journal of Human Genetics*, *26*, 400–411.

Martin, A. R., Stroud, R. E., Abebe, T., Akena, D., Alemayehu, M., Atwoli, L., … Chibnik, L. B. (2022). Increasing diversity in genomics requires investment in equitable partnerships and capacity building. *Nature Genetics*, *54*, 740–745.

Mathur, R., Fang, F., Gaddis, N., Hancock, D. B., Cho, M. H., Hokanson, J. E., … Johnson, E. O. (2022). GAWMerge expands GWAS sample size and diversity by combining array-based genotyping and whole-genome sequencing. *Communications Biology*, *5*, 1–9.

Meghir, C., & Palme, M. (2005). Educational reform, ability, and family background. *American Economic Review*, *95*, 414–424.

Mill, J. S. (2002). *A system of logic*. University Press of the Pacific (original work published in 1843).

Pereira, L., Mutesa, L., Tindana, P., & Ramsay, M. (2021). African genetic diversity and adaptation inform a precision medicine agenda. *Nature Reviews Genetics*, *22*, 284–306.

Pingault, J. B., Richmond, R., & Davey Smith, G. (2022). Causal inference with genetic data: Past, present, and future. *Cold Spring Harbor Perspectives in Medicine*, *12*, a041271.

Plomin, R., & Daniels, D. (1987). Why are children in the same family so different from one another? *Behavioral and Brain Sciences*, *10*, 1–16.

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press

Popejoy, A. B., & Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature*, *538*, 161–164.

Ray, L. A. (2012). Clinical neuroscience of addiction: Applications to psychological science and practice. *Clinical Psychology: Science and Practice*, *19*, 154–166. https://doi.org/10.1111/j.1468-2850.2012.01280.x

Ross, L. N. (2021). Causal concepts in biology: How pathways differ from mechanisms and why it matters. *The British Journal for the Philosophy of Science*, *72*, 131–158. doi:10.1093/bjps/axy078

Ross, L. N. (2022). *Cascade versus mechanism: The diversity of causal structure in science. PhilSci Archive*. Retrieved from http://philsci-archive.pitt.edu/20215/1/Ross_Cascade.pdf

Schneider, B., & Bradford, L. (2020). What we are learning about fade-out of intervention effects: A commentary. *Psychological Science in the Public Interest*, *21*, 50–54. doi:10.1177/1529100620935793

Smith, J. A. (2022). Treatment effect heterogeneity. *Evaluation Review*, *46*, 652–677. doi:10.1177/0193841X221090731

Stein, Z., Susser, M., Saenger, G., & Marolla, F. (1972). Nutrition and mental performance: Prenatal exposure to the Dutch famine of 1944–1945 seems not related to mental performance at age 19. *Science (New York, N.Y.)*, *178*, 708–713.

Strauss, G. M., Herndon, J. E., Maddaus, M. A., Johnstone, D. W., Johnson, E. A., Harpole, D. H., … Green, M. R. (2008). Adjuvant paclitaxel plus carboplatin compared with observation in stage IB non-small-cell lung cancer: CALGB 9633 with the Cancer and Leukemia Group B, Radiation Therapy Oncology Group, and North Central Cancer Treatment Group Study Groups. *Journal of Clinical Oncology*, *26*, 5043–5051.

Turkheimer, E. (2012). Genome wide association studies of behavior are social science. In K. S. Plaisance & T. A. C. Reydon (Eds.), *Philosophy of behavioral biology* (pp. 43–64). Springer.

VanderWeele, T. J. (2022). Constructed measures and causal inference: Towards a new model of measurement for psychosocial constructs. *Epidemiology (Cambridge, Mass.)*, *33*, 141–151.

VanderWeele, T. J., & Hernan, M. A. (2013). Causal inference under multiple versions of treatment. *Journal of Causal Inference*, *1*, 1–20.

Volkow, N. D., & Koob, G. (2015). Brain disease model of addiction: Why is it so controversial? *The Lancet Psychiatry*, *2*, 677–679.

Watkins-Hayes, C. (2009). *The new welfare bureaucrats: Entanglements of race, class, and policy reform*. University of Chicago Press.

Woodward, J. (2005). *Making things happen: A theory of causal explanation* (paperback). Oxford University Press.

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, *45*, E1. doi:10.1017/S0140525X20001685

Yeager, D. S., & Dweck, C. S. (2020). What can be learned from growth mindset controversies? *American Psychologist*, *75*, 1269–1284.

https://doi.org/10.1017/S0140525X22000681 Published online by Cambridge University Press