

March 30, 2018

In Press Manuscript at European Journal of Personality

Behavior Genetic Frameworks of Causal Reasoning for Personality Psychology

Daniel A. Briley¹, Jonathan Livengood², & Jaime Derringer¹

¹Department of Psychology, University of Illinois at Urbana-Champaign

²Department of Philosophy, University of Illinois at Urbana-Champaign

RUNNING HEAD: BEHAVIOR GENETICS AND CAUSAL REASONING

Address correspondence to: Daniel A. Briley, Department of Psychology, University of Illinois at Urbana-Champaign, 603 East Daniel Street, Champaign, IL 61822. E-mail:

dabriley@illinois.edu

Acknowledgements: The production of this manuscript was supported by a grant from the John Templeton Foundation (JTF58792).

Abstract

Identifying causal relations from correlational data is a fundamental challenge in personality psychology. In most cases, random assignment is not feasible, leaving observational studies as the primary methodological tool. Here, we document several techniques from behavior genetics that attempt to demonstrate causality. Although no one method is conclusive at ruling out all possible confounds, combining techniques can triangulate on causal relations. Behavior genetic tools leverage information gained by sampling pairs of individuals with assumed genetic and environmental relatedness or by measuring genetic variants in unrelated individuals. These designs can find evidence consistent with causality, while simultaneously providing strong controls against common confounds. We conclude by discussing several potential problems that may limit the utility of these techniques when applied to personality. Ultimately, genetically informative designs can aid in drawing causal conclusions from correlational studies.

Keywords: Behavior genetics; causal models; personality; Mendelian randomization

Behavior Genetic Frameworks of Causal Reasoning for Personality Psychology

Making causal inferences is a major scientific goal. Correct causal models may serve as guides for enacting social policies and for modifying individual behaviors. Experimental designs involving randomization and manipulation of a single potentially causal factor are thought to be the gold standard for making causal inferences, and some go so far as to suggest that randomization and manipulation are *necessary* for reliable causal inference (Bickman & Reich, 2009). Personality psychology relies primarily on correlational designs, since most hypotheses involve variables that are not manipulatable due to practical limitations or ethical concerns. Whereas random assignment is typically seen as washing away potential confounds (though see Deaton & Cartwright, 2017 for important caveats), correlational researchers must consider all the myriad strings of influence that may affect a variable (Cronbach, 1957). Several promising approaches to discern causal relations from correlational designs have emerged that use only patterns of observed correlations, although success is far from perfect (e.g., Chickering, 2002; Glymour, 2010; Mooij, Peters, Janzing, Zscheischler, & Schölkopf, 2016; Pearl, 2010; Shimizu, Hoyer, Hyvärinen, & Kerminen, 2006). When analyzing human behavior, causal inferences can be strengthened by basic facts of development, such as genetic variation taking temporal precedence over subsequent outcomes. Genetically informative data strengthens causal claims concerning psychological factors that are not easily subjected to random assignment.

Here, we describe three interrelated issues for personality psychologists interested in making causal claims. First, personality psychologists face an explanatory problem. Personality dimensions correlate with myriad important outcomes. Such correlations invite causal explanations. Philosophy provides us with options for thinking about how causation works and how it is related to the correlations that personality psychologists want to explain. Second, the

data collected in empirical personality studies are often underdetermined compared to the necessary information required for making a causal claim. Even in longitudinal correlational studies, the collected data are unable to distinguish the different plausible models (e.g., causal personality effects vs. a common cause confound), and are incapable of identifying when in development the causation occurred (e.g., a one-time scarring effect of personality vs. a small, persistently accumulating effect). Common empirical designs in behavior genetics provide additional sources of information to address the problem of underdetermination, by making use of family-based and other genetically informative samples to ground the orientation of potential causes in clearly defined genetic and environmental influences. Third, we offer specialized behavior genetic designs that can be applied to address particular issues of causal reasoning in personality research. These techniques will not solve all of the explanatory problems in personality psychology, but they provide additional important information about the causal processes that are plausible and which ones can be ruled out.¹

Mapping Preliminary Alternative Causal Models

Personality psychologists would like to make claims such as, “conscientiousness causes students to achieve more academically, avoid getting in trouble with the law, and live longer, healthier lives” (for references that document such correlations, see Poropat, 2009; Moffitt et al., 2011; Roberts et al., 2007). To the extent that each statement is true and conscientiousness is itself possible to manipulate, conscientiousness would represent an extremely important factor for public policy intervention. However, several causal models may generate data whereby

¹ Our goal is not to provide an overview of behavior genetic findings (see Johnson, Penke, & Spinath, 2011; Johnson, Vernon, & Feiler, 2008; Krueger & Johnson, 2008) or the causal logic of gene-personality associations (see Lee, 2012). Instead, we focus on the how behavior genetics can shed light on causally ambiguous correlations between personality and consequential life outcomes.

conscientiousness and life outcomes are correlated (see Rohrer, 2018). Perhaps conscientiousness does not cause student achievement, but rather students that achieve academically increase in conscientiousness. Or, it could be the case that growing up in a wealthy family independently increases conscientiousness and reduces the likelihood that someone engages in criminal behavior. In both cases, conscientiousness might be a byproduct or side effect, rather than a true causal factor. Were we to base public policy on raising conscientiousness, it might have no effect on important life outcomes because the causal relations do not flow along the expected pathways.

Personality psychologists invoke longitudinal designs as a way to demonstrate that personality precedes and predicts some later outcome, such as longevity, to demonstrate a causal effect (e.g., Roberts et al., 2007). However, conscientiousness (along with essentially every other characteristic of the individual; Polderman et al., 2015) is partly heritable, meaning genes play some role in the development of the trait. It could be the case that conscientiousness causes longevity, but it is also plausible that a set of common genetic factors influence conscientiousness and longevity independently. In this latter case, public policy designed around increasing conscientiousness would not be effective at increasing longevity because a common cause explains the correlation.

Each of these examples demonstrates the explanatory problem in personality psychology. We need additional information about causal structure to give a complete explanation for any observed correlation. Behavior genetics provides a wide-range of tools to test and eliminate these alternative causal models.

At its core, behavior genetics is a discipline that identifies sources of individual differences in an outcome, typically referred to as a phenotype (see Table 1 for a glossary of

common behavior genetic terms). Using pairs of individuals with known genetic and environmental similarity or by measuring individuals' genotypes, behavior genetic studies decompose (or split) the total variance in the phenotype observed among individuals in a sample into genetic and environmental sources. For example, the heritability of personality is approximately 40%, with the remaining 60% attributable to the nonshared environment, meaning unique life experiences that make individuals different from one another (regardless of their genetic similarity), including measurement error (Briley & Tucker-Drob, 2014; Johnson et al., 2008; Turkheimer, Pettersson, & Horn, 2014; Vukasović & Bratko, 2015). Little variance in personality is attributable to the shared environment, or influences that make reared-together individuals more similar to one another regardless of genetic similarity. Behavior genetic methods are also capable of decomposing the covariance (or correlation) among phenotypes. As another example, the covariance between measures of personality and academic achievement is primarily due to genetic sources of variance (Krapohl et al., 2014; Tucker-Drob, Briley, Engelhardt, Mann, & Harden, 2016).

Behavior genetic analyses rely on simple correlations between family members (i.e., quantitative behavior genetics) or regression models among unrelated individuals using genetic variants as an independent variable (i.e., molecular behavior genetics) that are no different from other correlations and regressions encountered in personality psychology. As with all associational statistical techniques, it is important to consider many possible causal models that could produce such results. For example, what causal model is suggested by the finding that individuals with a certain set of genetic variants tend to do well academically (Selzam et al., 2017)? Is the causal model different from the one suggested by the finding that students that score relatively high on conscientiousness tend to do well academically (Poropat, 2009)? Here,

we take the example of the link between personality and academic achievement primarily being mediated by genetic factors (Krapohl et al., 2014; Tucker-Drob et al., 2016). Figure 1 presents three alternative causal models that are each consistent with these empirical findings.

Personality and achievement may be correlated for genetic reasons due to pleiotropy (where a single genetic factor impacts multiple phenotypes through independent pathways), identical to a common cause confound (Figure 1a). Such a model implies that personality does not have a causal impact on achievement, only genes exert a causal influence. In this case, manipulating personality (while holding everything else fixed) would not produce a difference in achievement. Similarly, manipulating achievement would not produce any difference in personality.

Next, a causal chain could produce a genetic correlation between personality and achievement, where genes influence personality, which then influences achievement (Figure 1b). In this framework, personality does have a causal influence on achievement. Manipulating personality would produce changes in achievement. The association could also result from reverse causation, where genes influence achievement, which then influences personality (Figure 1c). Studies that include some aspect of temporal ordering could identify the direction of causation, given adequate care to the possibility of reciprocal causation and the timing of assessment waves to properly track the causal effects (e.g., Berry & Willoughby, 2017). In a cross-sectional study, however, the genetic link between personality and achievement is consistent with either a causal chain or reverse causation or a bidirectional model in which both directions occur simultaneously.

With these examples in mind, we turn now to exploring alternate conceptualizations of what we mean by the word “cause.” Personality psychologists implicitly rely on certain

definitions of causality over others. Explicit consideration of these alternatives allows us to formally plan research designs to address the particular causal problem(s) that we are most interested in disentangling when presented with a correlation between personality and some outcome of interest.

Philosophical Remarks on Causality

In the preceding discussion, we relied on an intuitive definition of causality – if one were to manipulate conscientiousness, would some other outcome be different. If conscientiousness is a cause, then the outcome will be different. If conscientiousness is spuriously correlated with the outcome, then there will be no difference. The most common intuition in psychology is that manipulability is strongly tied to causality, which is used to justify the elevated status of experimental designs. Personality psychologists tend to be more interested in causal structure, meaning building up a naturalistic model that includes all the necessary inputs and moderators to understand and predict the outcome. Interestingly, these are two separate causal reasoning problems that have divergent philosophical backing. Both tasks are important and establish different aspects of causality. Here, we briefly touch on this philosophical tradition before returning to methodological concerns.

With respect to the nature of causality, philosophers have been primarily concerned with determining the best metaphysical reduction basis for the causal relation, meaning identifying the fundamental building blocks of causality. Many philosophers maintain that the causal relation should be reduced to *something*, but there is essentially no agreement as to the correct reduction basis.² For example, some accounts argue that counterfactual dependence (similar to our intuitive

² Some candidates for reduction basis include regular succession (Hume, 1748; Psillos, 2009), laws of nature (Davidson, 1967), probability (Suppes, 1970; Glynn, 2011), counterfactual dependence (Lewis, 1973; Lewis, 2000),

definition above, where we may think about the cause of an outcome using a statement suggesting that “if not for” some causal factor, the outcome would have been different) is sufficient to define causality, but then this runs up against challenges, such as situations where we would want to say something is causal simply because it raises the probability of something else occurring (rather than necessarily requiring that the outcome occurs, in a one-to-one relationship).

Others argue that there is *no* good reduction basis for the causal relation. Taylor (1966, p. 40, quoted in Carroll 2009, p. 284) puts it elegantly when he writes, “To say of anything, then, that it was the cause of something else, means simply and solely that it was the cause of the thing in question, and there is absolutely no other conceptually clearer way of putting the matter except by the introduction of near synonyms for causation.” We largely agree, but also note that the elements of probability, counterfactual dependence, intervention, and mechanism give some guidance as to what causality looks like.

Going forward, we will use the term “cause” to refer to a (primitive) relation between two property universals (such as *Height* and *Weight*).³ We use random variables to represent such property universals, and we use arrows to represent causation. A directed graph over a set of random variables represents a *causal structure* with respect to that collection of property universals. Since we take causation to be a primitive relation, we do not think that it can be reduced to probability or counterfactual dependence or the like. But we can still say something informative about how causation works. Specifically, if one property is a *direct cause* of another

agency (Menzies and Price, 1993), transmission of a conserved quantity (Dowe, 1995; Salmon, 1997), and mechanism (Glennan, 1996). See Appendix A.1 and A.2 for additional discussion.

³ What we are calling “causation” is sometimes referred to as “structural causation” and distinguished from what is then called “actual causation.” See Appendix A.3 for additional discussion.

relative to some background, then there is some way of holding the background fixed such that if we could change the cause just so, the effect would change along with it.⁴

For example, suppose we randomly assigned individuals to several conditions – brushing their teeth twice daily vs. not, reading the newspaper in the morning vs. not, and attending psychotherapy vs. not. We then find those that attended psychotherapy tended to have decreased neuroticism (as in Roberts et al., 2017), conditional on the other experimental factors. In our example, psychotherapy is a direct cause of neuroticism, relative to the background factors of newspaper reading and teeth brushing.

The way we are thinking about “direct cause” here brings together insights from various attempts to give a reductive account: causes and effects tend to be statistically related (probability), if the cause had been different, then the effect would also have been different (counterfactual dependence), and if the cause were to be manipulated, then the effect would also be manipulated (intervention). Probability raising, counterfactual dependence, and control through intervention are *marks* or *signs* of causation. Moreover, causation involves mechanisms. Direct causal relations together with the properties they relate are the simple parts (relative to some background) that might be organized so as to constitute a mechanism, which potentially includes some hidden internal structure to be explored by measuring further properties.

In thinking about how to draw causal inferences from the marks or signs of causation, it will be helpful to distinguish two kinds of problem: causal structure learning and causal effect learning. A *causal structure learning problem* is similar to the traditional goals of personality psychology: determining which variables are connected (qualitative causal structure) and then

⁴ For us, claims about what causes what are always relative to the collection of properties under consideration. One property is a direct cause of another *relative to a background collection of properties* if no property in the collection mediates the influence of the first on the second. For a formally precise definition of *direct causation*, see Appendix A.2.

determining the direction and strength of those connections (quantitative causal structure). Both steps are naturally guided by careful measurement in randomized controlled trials. But in some cases, it is possible to solve a causal structure learning problem with significantly less informative data.⁵

In a causal effect learning problem, which is more similar to experimental psychology, we want to know what values a response variable would (or will) take on if some predictor variable were (to be set) one way as opposed to another. In solving a causal effect learning problem, we are trying to compare what would happen under treatment to what would happen under control. Or in cases where the response has already occurred, we are trying to compare what *actually happened* under treatment (control) to what *would have happened* under control (treatment). However, we cannot observe both potential outcomes. That is, we cannot observe both what happens under treatment and also what happens under control for the same individual at the same time. Holland (1986) calls this the *fundamental problem of causal inference* (see A.4 for additional nuances of the causal effect learning problem).

We are primarily interested in the causal structure learning problem, and we think that causal structure learning is both the most important problem faced by personality psychologists and the one that is (at least implicitly) the target of most research in personality psychology. However, it is easy to confuse causal structure learning with causal effect learning. After all, the two problems are related. Once we know the causal structure and the parameters of the functional relationships with respect to the variables of interest, we can read off the solutions to

⁵ See Pearl (2000), Spirtes, Glymour, and Scheines (2000), Chickering (2002), Glymour (2010), Pearl (2010), and Pearl, Glymour, and Jewell (2017) for detailed treatments of how reliable causal structure learning can be made to work with non-experimental data.

whatever causal effect learning problems confront us.⁶ But notice that inferences from causal effects to causal structure are not so obvious. If the effect of X on Y is non-zero, then we know that X structurally causes Y . But we do not thereby know whether X is a direct structural cause of Y . Nor do we know the functional form or parameterization on any of the specific causal paths from X to Y . Even worse, if there is path canceling, the estimated causal effect of X on Y could be zero despite the fact that X really is a structural cause of Y .⁷ In general, knowing the answer to a causal structure learning problem (i.e., the goal of personality psychology) is more informative than knowing the answer to a causal effect learning problem, since causes often bring about their effects by way of multiple paths.

These considerations highlight the underdetermination problem in personality psychology. Despite our best efforts, it is very difficult to include all possible potential confounds or relevant variables at all impactful developmental time points within one study. Our data underdetermine the hypotheses and causal structures we intend to test. In the remainder of the manuscript, we focus on behavior genetic estimation tools. Although we cannot offer the silver bullet to making foolproof causal claims from correlational data, the following methodological approaches add substantial causal information and significantly reduce the underdetermination problem when applied correctly.

Genetic and Environmental Sources of Variance and Covariance

⁶ Of course, if all we have is the solution to the qualitative step of a causal structure learning problem, then we do not know whether the effect of X on Y is large or small, positive or negative.

⁷ The classical example of path cancelling (for which, see Hesslow 1976) is the relationship between birth control and thrombosis (a blood clot obstructing circulation). Using birth control increases the risk of thrombosis directly. But since pregnancy increases the risk of thrombosis, using birth control decreases the risk of thrombosis indirectly by reducing the risk of pregnancy.

All behavior genetic methods aim to identify the extent to which individual differences in some phenotype are due to genetic and environmental influences.⁸ Quantitative behavior genetic methods rely on calculating the phenotypic similarity (i.e., correlation coefficients) for pairs of individuals with known genetic or environmental relatedness (Neale & Cardon, 1992; Posthuma et al., 2003). Common examples of quantitative genetic models include the classical twin design (i.e., comparing the similarity of identical twins to fraternal twins) and adoption studies (i.e., comparing adoptive siblings or adopted children and their biological parents). More commonly, these models may be identified as ACE models, indicating that variance is decomposed into additive genetic (A), shared or common environmental (C), and nonshared environmental (E) components. Thousands of empirical reports have been published based on these designs (Polderman et al., 2015).

As an example of this type of reasoning, identical twins correlate in their personality at approximately a level of 0.4, and fraternal twins correlate at approximately 0.2 (Vukasović & Bratko, 2015). Because identical twins share identical genotypes and fraternal twins share 50% of segregating genetic material on average, taking twice the difference between the correlations for these groups is an easy method to get a rough estimate of heritability. This variance component is referred to as additive genetic variance, or alternatively, narrow-sense heritability. The shared environmental component indexes how much of the observed individual differences occur due to between-family differences. To the extent that siblings growing up in the same family resemble one another in their psychological characteristics (more than would be predicted

⁸ Although this language sounds causal, it is only causal in a weak sense. Behavior genetic methods decompose all of the variance in phenotypes, so there must be some causal reason why people differ (even if it is purely random noise). When behavior geneticists use terms like genetic influences, what is really intended is that variance is statistically associated with genotypic variation in some manner, including all potential causal models.

on the basis of their genetic similarity alone), then this similarity implies that between-family variation in the population impacts the phenotype. Twin correlations for personality do not indicate substantial shared environmental effects, which would be implied by fraternal twin correlations more similar to identical twin correlations. Finally, the nonshared environment captures within-family variation, reflecting the extent to which individuals within a family are different from one another, after considering genetic differences. Identical twins are only correlated at 0.4 for personality, meaning the remaining variance must be due to the nonshared environment (the only thing that identical twins raised together do not share, by definition). This factor includes measurement error⁹ and other sorts of idiosyncratic effects, as well as individual-specific effects of objectively shared environments (e.g., parents having unique relationships with their children).

Multivariate techniques extend these basic univariate models to estimate genetic and environmental covariance. For example, Krapohl et al. (2014) and Tucker-Drob et al. (2016) present correlations between personality and academic achievement, broken down into genetic and environmental components. Specifically, Krapohl et al. (2014) estimated a phenotypic correlation of 0.28 between a composite of multiple personality measures and achievement, with 92% of this correlation due to genetic factors. Similarly, Tucker-Drob et al. (2016) estimated a correlation of 0.45 between latent variables representing personality and achievement, with essentially 100% of this correlation due to genetic factors. These techniques draw on differences in cross-twin, cross-trait correlations between genetically more related pairs of individuals (e.g., identical twins) compared to genetically less related pairs (e.g., fraternal twins). For example,

⁹ Here, we primarily focus on developmental, rather than psychometric, implications for behavior genetic models. Psychometrics also play an important role in behavior genetic estimates, such as broader traits potentially displaying larger estimates of heritability (see Johnson et al., 2011, p. 256-257).

Krapohl et al. (2014) found that one twin's personality correlated with the other twin's academic achievement at 0.25 for identical twins, but only 0.07 for fraternal twins. The logic is much the same as quantifying heritability, except using multivariate data (see Neale & Cardon, 1992 for methodological details).

More recently, methods have been developed that allow data originally collected as part of genome-wide association studies (GWAS), which test associations between specific single nucleotide polymorphisms (SNPs, pronounced "snips") across the genome and outcomes, to be used as an alternate method of estimating heritability (cf. Visscher, Brown, McCarthy, & Yang, 2012). Genome-wide complex trait analysis (GCTA; Yang, Lee, Goddard, & Visscher, 2011) extends the logic of a classical twin design to putatively unrelated individuals. Rather than estimating the phenotypic similarity of pre-specified pairs with assumed genetic relatedness (i.e. twins), genetic similarity between every possible pair of individuals in a sample is estimated based on observed genotypes. The extent to which minute variations in relatedness is associated with phenotypic similarity provides a heritability estimate. Molecular genetic designs do not typically distinguish between shared and nonshared environmental variance as family members are not sampled. Environmental variance is simply whatever variance is not associated with genetic variants.

Multivariate molecular genetic methods are also capable of estimating genetic and environmental correlations (see Table 1 for description of one common method, LD score regression)¹⁰. As examples, Okbay, Beauchamp et al. (2016) estimated a moderate genetic

¹⁰ For large, public repositories of such information as well as the ability to conduct analyses online, see <http://ldsc.broadinstitute.org> (Zheng et al., 2017) and <http://www.nealelab.is/blog/2017/7/19/rapid-gwas-of-thousands-of-phenotypes-for-337000-samples-in-the-uk-biobank>.

correlation between educational attainment and neuroticism (-0.41) and a strong genetic correlation between educational attainment and cognitive ability (0.75). Similarly, Nagel et al. (2017) estimated strong genetic correlations between neuroticism and depression, anxiety, and subjective well-being (all $|r| > 0.65$), meaning the genetic variants that influence each dimension are almost entirely overlapping.

Like all statistical methods, these approaches rely on certain underlying assumptions which may bias results if violated (see Appendix B.1). The assumptions differ across methods, suggesting that convergence of conclusions across different approaches is unlikely to be the result of violations of any one specific assumption. Across multiple methods and samples, behavior genetic research produces replicable findings (Plomin, DeFries, Knopik, & Neiderhiser, 2016). Having reviewed methods for estimating the relative influence of genes and environments on single and multiple phenotypes, we turn now to considering how these genetic and environmental sources of influence may not be truly independent of each other.

Gene-Environment Interplay

Previously, we described three preliminary causal models consistent with a genetic association between personality and academic achievement (Figure 1). Here, we elaborate on these models by considering gene-environment interplay. The umbrella term gene-environment interplay refers to any sort of dependence between genetic and environmental influences, such as gene-environment correlation and Gene \times Environment interaction (Bleidorn, Kandler, & Caspi, 2014; Briley & Tucker-Drob, 2017; Kandler, 2012; Krueger & Johnson, 2008; Nivard & Boomsma, 2016; Plomin, DeFries, & Loehlin, 1977; Scarr & McCartney, 1983). Several forms of gene-environment interplay would produce or amplify a genetic correlation between personality and academic achievement: evocative or active gene-environment correlation and

Gene \times Shared Environment interaction (see Appendix B.2 for technical details and examples). To the extent that these processes occur during development as part of the causal arrows in Figure 1, then we would expect a genetic association. For completeness, we will also describe passive gene-environment correlation, which would imply shared environmental associations, and Gene \times Nonshared Environmental interaction, which would imply nonshared environmental associations.

Evocative gene-environment correlation. Evocative gene-environment correlation occurs when others respond to observable, genetically influenced characteristics of the target individual. There may be genetically influenced individual differences in the tendency for children to be more active versus more calm. Parents may notice this and provide activities (sports versus books) in accordance with the child's heritable characteristic. Evocative gene-environment correlation could produce a genetic association (Figure 2a). This diagram assumes that teachers observe genetically influenced aspects of their students' personality, which shapes their teaching approach. Whereas effects of evocative gene-environment correlation are filtered through the mental state of individuals in the target's environment, the effect of active gene-environment correlation is filtered through the target's mental state.

Active gene-environment correlation. Active gene-environment correlation occurs when individuals actively create or select environmental experiences aligned with their genetically influenced preferences and desires. For example, some individuals may possess high levels of extraversion in part due to genetic influences, and these individuals may seek out environmental experiences that afford them greater opportunity for social interaction, which in turn leads them to be more interpersonally skilled and confident. Figure 2b displays a model for active gene-

environment correlation. The subtle difference here is that the student actively influences the teacher through their communication.

Both evocative and active gene-environment correlation build on the causal chain model (Figure 1b) by specifying an additional causal mechanism. The plausibility of these models demonstrates that heritability estimates are not useful in determining how much of any genetic association occurs “under the skin” compared to “outside the skin” (Kendler, 2001). Although it is common to label sources of variance as genetic, gene-environment interplay may guide development. Additional evidence is required to determine whether a genetic association might be due to purely “under the skin” mechanisms absent from environmental influence compared to socially-responsive genetic variance. For example, a strong case can be made that rapid increases in the heritability of cognitive ability are driven by environmental reinforcing of early differences (Briley & Tucker-Drob, 2013).

Passive gene-environment correlation. Passive gene-environment correlation occurs when parents pass on genes and correlated environmental experiences to their children. Parents may possess genetic influences that predispose them to keeping a messy house, which in turn establishes both the child’s environment and their genotype. Therefore, both genetic transmission and socialization to messy environments may influence the child. For this type of gene-environment interplay, it is more fitting to use cognitive ability as an example, as part of the association between cognitive ability and achievement is due to the shared environment (Krapohl et al., 2014). Passive gene-environment correlation could produce shared environmental variance linked with achievement (Figure 2c). For example, parents may pass on intellectually stimulating environments as well as genetic influences that predispose toward learning.

Gene × Environment interaction. Gene × Environment interaction refers to a situation where individuals respond differently to an environmental experience on the basis of genotype (or vice versa). For example, some genetic variants may predispose individuals to be especially sensitive to the environment or especially resilient, meaning the effect of the environment differs. The genetic disorder phenylketonuria (PKU) is classic example of Gene × Environment interaction (Ottman, 1996). Individuals with this disorder produce substantially less of an enzyme that converts phenylalanine to tyrosine. Here, we can think of diet as the environment. If individuals with PKU eat a special diet, then negative consequences are largely avoided. A failure to follow this diet results in severe cognitive impairment. PKU genotype would be a powerful predictor of cognitive ability if we had no knowledge of the disorder, but in the actual world, PKU genotype would not be predictive due to comprehensive medical testing and dietary intervention. As with all statistical interactions, Gene × Environment interactions can reflect various patterns (e.g., diathesis-stress, differential susceptibility, etc.; Roisman et al., 2012), with different implications for detection within statistical models.

Gene × Shared Environment interaction could produce a genetic link between personality and achievement (Figure 3a). If genetic influences on personality are dependent on aspects of the family-level environment, then this produces genetic variance in personality. It could be that this source of variance predicts achievement because it is relevant for the development of conscientiousness, such as whether or not parents provide structure interacting with the child's genetically influenced levels of achievement striving. Gene × Nonshared Environment interaction could produce nonshared environmental variance linked with achievement (Figure 3b). For example, genetic influences on cognitive ability may manifest differently dependent on one's peer group's interest in school, which could ultimately affect achievement.

As can be seen, the correlation between personality and consequential life outcomes could be due to many different causal models. Decomposing this association into genetic and environmental variance components moves somewhat closer to a causal explanation, but genetic and environmental correlations are subject to several potentially non-causal interpretations (e.g., common causes or reverse causation). Additional methodological information can strengthen the causal claim.

Behavior Genetic Models for Establishing Causality

Behavior genetic models estimating heritability versus environmentality and gene-environment interplay address some aspects of causality, particularly stemming from the assumption that, within a causal chain, genetic factors can typically be taken as having temporal precedence. More specialized approaches are also available that take advantage of genetically informative samples to specifically address questions of causality. Although no one method fully establishes the causal structure, each provides unique, incremental evidence for or against various models. Taken together, such research designs allow for powerful causal inferences, without requiring researcher manipulation of the causal variable(s) of interest.

Co-twin control design and within-family differences. The co-twin control design is an intuitive behavior genetic model based on comparing identical twins that grew up in the same home to one another on a pair of phenotypes (e.g., exposure to cigarette smoking and some health outcome). Since identical twins are virtually perfectly matched with respect to genetic and family-level environmental confounds, any differences between the twin who was exposed to some environment and the co-twin who was not exposed to that environment are due to the difference in environmental exposure. This logic is similar to counterfactual dependence. Figure 4a displays a hypothetical example comparing self-rated health for twins that smoke compared to

their identical twin that does not smoke. Here, it is clear that the hypothetical non-smoking twin reports better health in the boxplot, but formalized models are also available. Turning toward empirical examples, the identical twin who starts smoking cigarettes earlier in life tends to experience greater nicotine addiction (Kendler, Myers, Damaj, & Chen, 2013), and the twin that smokes more tends to have an elevated risk of schizophrenia (Kendler, Lönn, Sundquist, & Sundquist, 2015). There may be some unknown environmental common cause unique to the individual that influences both cigarette smoking and these outcomes, but this caveat is much smaller than all possible sources of confounding, raising confidence that smoking is a cause.

The classical twin design also provides an intuitive and powerful design for testing causal hypotheses by leveraging within-family effects. In these models, the nonshared environmental association between two phenotypes has a similar interpretation as the co-twin control design in that an association is tested holding constant genetic and shared environmental confounds. This model is displayed in Figure 4b, with the critical nonshared environmental pathway highlighted. An alternative model (Figure 4c) specifies that the effect of genetic and environmental influences on one variable (e.g., conscientiousness) flow through the predictor phenotypically to the outcome (e.g., academic achievement; Turkheimer & Harden, 2014, pg. 172). This model is in line with the phenotypic null hypothesis for personality, which states that causation emerges through the phenotype rather than hidden sources of genetic or environmental covariance (Turkheimer et al., 2014). Behaving conscientiously helps children in school; teachers do not care whether a child's behavior is due to genes or the environment. In this case, only the phenotype is observable and impactful.

If a within-family effect is found (that is, if the co-twins have different outcomes based on their different exposures), this result is fairly convincing evidence that a causal relation exists

in some form. However, a null finding is less informative, in contrast to common interpretations in the literature which view null results as evidence for a non-causal relation (see Appendix B.3). This common interpretation implies that genetic or family-wide environmental factors act solely as common causes (e.g., Figure 1a), omitting the possibility of causal chains flowing from these sources (e.g., Figure 1b). The association may very well be causal, but simply through an exclusively genetic or shared environmental pathway. If genetic sources of variance drive exploration of the environment, then it may be the case that much of the causal effect of the exposure is eliminated by co-twin control designs. Returning to the cigarette smoking example, all pathways resembling *Genes* → *Sensation Seeking* → *Smoking* → *Health* are removed from the estimated causal effect, despite the fact that there is reason to believe these pathways are substantial (Harden, Quinn, & Tucker-Drob, 2012). As a result, the co-twin control design may produce an even more biased estimate of the causal effect than an ordinary regression (Boardman & Fletcher, 2015). More generally, environmental exposures may causally reinforce phenotypes through genetic or shared environmental pathways.

If the phenotypic null hypothesis (that causation emerges through the phenotype rather than hidden sources of genetic or environmental covariance) is true, then the genetic and environmental covariance between a predictor and an outcome should be proportional to the magnitude of genetic and environmental influences on the predictor. In this case, a null within-pair finding would be convincing evidence against causality as all phenotypes are influenced by the nonshared environment, often more strongly than any other factor. Yet, a number of theoretical perspectives (e.g., Bouchard, 1997; Kandler & Zapko-Willmes, 2017; Scarr & McCartney, 1983) emphasize that genetic influences might lead people to certain environments, which in turn shape development. These sorts of systematic experiences may play an outsized

role in generating associations. In contrast, the nonshared environmental factors that within-family designs rely on may simply be too transient, idiosyncratic, chaotic, or otherwise random to accrue into meaningful associations (e.g., Dickens & Flynn, 2001). Additionally, co-twin control designs compound measurement error, reducing power to detect effects (McGue et al., 2010, pg. 551).

In addition to the above limitations, the causal direction is ambiguous in co-twin control models. For some phenotype pairings, the causal ordering may be noncontroversial (e.g., childhood maltreatment and adult psychopathology). Since personality is not a discrete event, matters are more complicated. There may be a directional causal relation from personality to the outcome, from the outcome to personality, or bidirectional causation.

Longitudinal twin and family designs. Temporal precedence can inform directionality of causal effects. If personality causes an outcome, then it should precede and predict the outcome. Common longitudinal designs, such as growth curves, can be adapted to behavior genetic designs (see Duncan et al., 2014; McArdle, 1986; McArdle & Hamagami, 2003). These models decompose variance in change into genetic and environmental components. McGue, Bacon, and Lykken (1993) found that stable variance in personality was predominantly due to genetic factors and change due to the environment. Subsequent meta-analyses have found that this differs across the lifespan, with the environment playing an increasing role in stability with age (Briley & Tucker-Drob, 2014; Kandler & Papendick, 2017). For individual differences in change, Bleidorn et al. (2009) estimated that approximately half of the individual-level variation around normative trajectories was due to genetic influences and the other half due to the nonshared environment.

These longitudinal models can integrate multiple phenotypes, in much the same way that cross-sectional behavior genetic models do. Kandler et al. (2012) used a variant of an

autoregressive model to test for the genetic and environmental stability of personality and life events. Put differently, genetic and environmental variance at an earlier time point was used to predict genetic and environmental variance at a later time point. Additionally, pathways between life events and personality were included. By including longitudinal data, the authors were able to distinguish selection effects (i.e., personality predicting life events) from reciprocal effects (i.e., personality predicting life events, which in turn predict later personality). Harden et al. (2012) applied growth curve modeling to data on sensation seeking and delinquency in childhood and adolescence. They found that changes in sensation seeking were primarily genetically influenced, and that these changes in sensation seeking predicted changes in delinquency. Here, the longitudinal information adds causal information to support models of genetically driven development of sensation seeking leading to delinquency. Luo et al. (2017) found that developmental changes in perceived stress were associated with changes in personality, with the entirety of genetically influenced change in stress shared with personality change. As a final example, Briley, Harden, and Tucker-Drob (2014) tested cross-temporal associations between parental educational expectations and child achievement. They found genetic and environmental correlations across time between expectations and child approaches toward learning in both directions. This result implies a bidirectional relation between parenting and child development.

Direction-of-causation models. Inferences concerning the causal direction (or bidirectionality) can also be made from cross-sectional genetically informative data (Heath et al., 1993). By comparing the fit of a series of models that imply different causal structures, the most plausible causal direction can be inferred. Figure 5 displays the full version of this model, and reduced models are compared to determine the best fitting model. For example, Gillespie et al.

(2003) tested the causal direction between measures of parenting and psychological distress. They found that the best fitting model was one in which parenting led to psychological distress, not the other way around. Using a similar design, Olivares et al. (2016) found an absence of a causal relation between parental monitoring and substance use. Instead, the authors concluded that a common set of risk factors influence both the parenting environment and substance use (i.e., a common cause explanation).

The key assumption of these models is that causal effects would leave a trace on the other variable. If some portion of the variance in the outcome is causally explained by the phenotype, then the variance components of the phenotype should be represented in the outcome proportional to the observed association. If a phenotype that has a large shared environmental component is hypothesized to have a causal effect on an outcome, but the outcome did not reflect any shared environmental variance, then that is evidence that the phenotype with a large shared environmental component had not been a cause up to that point in development. By comparing multiple models that structure the direction of causation differently, the best model is selected.

However, the statistical power requirements for evaluating direction-of-causation models are substantial. To simply reject one causal direction, Heath et al. (1993) estimated that sample sizes of more than 10,000 twin pairs would be required under circumstances that roughly mirror common personality phenotypes. Additional power is needed to further differentiate a bidirectional model or a common cause model. As large-scale twin studies become more readily available, direction-of-causation models may be particularly useful for testing the association between personality and environments due to the large difference in terms of genetic and environmental structure. However, for other questions in personality psychology, such models

likely have limited utility due to the relative lack of systematic differences in genetic and environmental proportions of variance.

Mendelian randomization. Instrumental variable analysis is a common approach used in econometrics to infer causation (Angrist, Imbens, & Rubin, 1996). If one wanted to test whether cigarette smoking caused poor health, confounds could potentially produce an association, as well as poor health causing cigarette smoking. However, if an instrument can be found that influences health solely through cigarette smoking (i.e., some sort of shock that increases smoking), then the causal effect can be estimated, given a variety of assumptions are met. Instrumental variable analyses are subject to intense debate because a perfect instrument is seldom found. Genetic variants make promising instruments because fears of reverse causation are reduced, and if enough of the biological pathway from genotype to phenotype is known, then assumptions can realistically be met. This type of analysis is termed Mendelian randomization because which allele someone receives is essentially randomized at birth (Davey Smith & Ebrahim, 2003). Gage, Davey Smith, Ware, Flint, and Munafò (2016) highlight that this line of thinking creates a link between genotype and modifiable environmental exposures. Instead of GWAS uncovering biological mechanisms, more may be uncovered about the environment. They use the example of a SNP predicting lung cancer in the general population. However, when stratified by smoking status, this effect only emerges in smokers. In this example, the SNP can act as an instrument to demonstrate that smoking plays a causal role in the development of lung cancer. A fundamental problem of Mendelian randomization approaches is that individual genetic instruments will almost always be very weak explanatory variables, which can produce misleading results (e.g., Bound, Jaeger, & Baker, 1995).

In an era of massive, readily accessible genetic databases, the necessity to pre-select a specific SNP instrument and apply it to a narrow outcome is becoming less relevant (Hemani et al., 2016). Methodological improvements on Mendelian randomization occur frequently and draw on many of the techniques described above. Here, we describe two promising avenues for future development: direction-of-causation and mechanistic models.

Direction-of-causation variants of Mendelian randomization bear some similarity to other techniques for inferring causality from observational data. Some patterns of correlations provide evidence that the causal direction must flow one way rather than the reverse. Mooij et al. (2016) give the example of altitude causing temperature at weather stations, which is possible to infer from the scatterplot even if axes labels are omitted. Pickrell et al. (2016) lay out a similar approach using GWAS results. To determine whether *Genes* \rightarrow *Conscientiousness* \rightarrow *Achievement* is the proper causal chain, scatterplots of genetic variant effect sizes can be produced. For example, one could plot the genome-wide significant SNPs for conscientiousness against those same SNPs' effect sizes for achievement. If conscientiousness causes achievement, then SNPs that predict elevated levels of conscientiousness would predict elevated levels of achievement proportionally (i.e., the scatterplot would reflect a positive correlation). Now, consider the reverse. Presumably there are many genetic pathways that lead to achievement, and many of them likely have little to do with conscientiousness. Thus, the scatterplot of significant achievement SNPs' effect sizes would have no or minimal association with those same SNPs' effect sizes for conscientiousness. Figure 6 displays this scenario using hypothetical data. Pickrell et al. (2016) demonstrated the utility of this technique which identified BMI as a cause of triglyceride levels and type 2 diabetes, LDL cholesterol as a cause of coronary artery disease, and hypothyroidism as a cause of height. Methods development in the arena of Mendelian

randomization is advancing at a rapid pace, suggesting a powerful role for similar approaches to causal inferences in the future (see Appendix B.4 for some promising future avenues). For example, Zhu et al. (2018) formalized the Pickrell et al. (2016) approach with increased statistical power.

Richardson and colleagues (2017) provide an extensive demonstration of state-of-the-art techniques making use of mechanism elucidation. They tested the causal links between genome-wide SNPs, methylation (an epigenetic signal that impacts expression of the underlying genetic sequence), and 139 complex trait outcomes using multiple samples. By moving the exposure variable (in this case, methylation) closer to the genotype, it may be plausible to expect larger effect sizes. The causal mechanism may also be clearer at this level of analysis, compared to attempting to link abstract personality constructs with outcomes. Small incremental steps could aggregate into a coherent causal story.

It is unclear how likely this cumulative vision will manifest. For example, Linnér et al. (2017) published an epigenome-wide association study of educational attainment ($N = 10,767$) and Marzi et al. (2018) published an epigenome-wide association study of stress ($N = 2,232$). These studies do not offer strong support for the idea that epigenetic effects will be much larger than SNP effects or produce a revolution in genetic thinking. Linnér et al. (2017) did find that epigenetic effect sizes were larger than SNP effect sizes, but the difference for the biologically distal phenotype of educational attainment was not nearly as dramatic as other, more biologically proximal phenotypes, such as smoking. In fact, many of the effect sizes were substantially reduced and did not replicate out of sample after taking into account smoking. Similarly, Marzi et al. (2018) conducted numerous tests of developmentally contextualized stress exposure in a

well-characterized sample, but they found minimal evidence that even extreme forms of stress left lasting epigenetic traces.

The assumptions required for valid inferences from Mendelian randomization are complex. We would encourage researchers interested in pursuing this approach to follow Zhu et al. (2018) who provide a good example of the care necessary for interpreting the results. Further, more evidence is needed concerning the plausibility of the assumption that SNP effect sizes do not reflect confounds (Koellinger & Harden, 2018), as recent evidence suggests may be the more common than appreciated (Kong et al., 2018).

Mechanism elucidation. Mechanical accounts of causality proceed by linking small pieces in the chain of causation together. Rather than theorizing about the causal relation between conscientiousness and achievement, a mechanical approach would attempt to break down each bit of the association into as causally unambiguous pieces as possible. In psychology generally, mediation analysis is used with the inference that the mediator plays some sort of mechanistic role in linking the predictor with the outcome. Behavior genetic work has had mixed success laying out mechanisms. Early candidate gene work was largely premised on expert intuitions about mechanisms that influence some psychological variable, but these assumptions have largely proven inaccurate following large-scale replication attempts (e.g., Chabris et al., 2012). Expanding from the limited scope of traditional candidate gene approaches, specific genetic variants are located within specific genes, which are in turn located within pathways that perform certain functions, which tend to be expressed in certain organs in the body. Modern approaches take an agnostic perspective and instead query vast databases of genetic pathways (e.g., FUMA; Watanabe, Taskesen, van Bochoven, & Posthuma, 2017; MAGMA; de Leeuw,

Mooij, Heskes, & Posthuma, 2015; GTEx Consortium, 2015). This approach is based on the idea that aggregating levels of biological information can be synthesized into a coherent whole.

To demonstrate this approach, we highlight the largest ($N = 449,484$) GWAS of personality to date (Nagel et al., 2017). This report on neuroticism follows a standard implementation of all the molecular genetic tools developed to quantify genetic mechanisms. Pooling all genetic information together, the SNP-based heritability was 10%. Examining specific genetic markers, there were 136 independent SNPs that were genome-wide significant ($p < 5 \times 10^{-8}$). Most of these SNPs were introns, meaning they are not known to specifically code for a protein, but may still have regulatory function. Only four SNPs were in exon regions, meaning they have direct protein coding function. For example, the SNP with the largest effect results in the production of one of two amino acids, Tryptophan or Arginine, depending on which variant someone has, pointing toward the potential for downstream effects. The significant SNPs tended to be located in conserved regions (or areas of the genome that are very similar across many species). Approximately 500 genes were implicated as playing a role in neuroticism. These genes tended to be expressed in a variety of tissues, such as the aorta, esophagus, lung, thyroid, and adrenal gland. Gene-sets were associated with neurogenesis, neuron differentiation, and behavioral response to cocaine. Some cell types associated with neuroticism were dopaminergic neuroblasts, medium spiny neurons, and serotonergic neurons. The interested reader will note that these implicated cell types align nicely with existing theories of neurological processes underlying neuroticism and depression, suggesting the utility of such mechanistic approaches to identifying novel causal pathways underlying personality and related outcomes in the future.

One limitation of this approach, as is hopefully clear by now, is the incredible complexity of biological organisms. Our ability to infer mechanism from GWAS signals is limited by the state of basic biological knowledge. Current methods of pathway analysis rely on pre-specified pathways; although databases are available, definitions of pathways are ambiguous, vary across alternative sources, and are continually updated. Thousands of SNPs likely impact personality, each one with an incredibly tiny effect, limiting the ability to create a cohesive story (de Moor et al., 2012; Chabris et al., 2015). There are substantially fewer genes contained in the human genome compared to SNPs, but the story there is no easier, and pathways, expression, and all sorts of epigenetic phenomena make tracing these pathways a daunting task. At some level, accelerated progress on this front for personality psychology will proceed as the foundational biological annotations continue to advance.

Limits on Utility for Personality Psychology

If there were few, large SNP effects on personality, then it would be relatively easy to identify them, implement Mendelian randomization, and determine their causal chains, but such effects do not exist (Chabris et al., 2015). Incredibly small effect sizes and high degrees of polygenicity are the rule in behavior genetics, but especially so for personality. For some reason, personality is the primary domain in which GWAS progress remains slow despite substantial sample sizes. As a comparison, the largest GWAS of cognitive ability (Davies et al., 2017, see also Savage et al., 2017) had a sample size a little over half that of the sample for the current largest neuroticism GWAS (Nagel et al., 2017), but estimated a much larger SNP-based heritability and roughly similar numbers of SNP and gene hits. In most domains, increases in sample size have led to exponential gains in discovery. Similarly, Cheesman et al. (2018) tested the gap between twin-based heritability and SNP-based heritability in the same sample. They

found the largest gap for phenotypes most similar to personality (across both self- and parent-report), with a considerably smaller gap for cognitive and anthropometric traits. There are three nonexclusive possibilities that may explain the slow progress for personality: psychometric structure, departure from additivity, and developmental complexity.¹¹

Psychometric structure. It may be the case that early factor analytic work in personality identified convenient phenotypic factors, but these factors do not adequately capture the underlying genetic architecture of personality. There is no absolute reason why phenotypic structure should necessarily match genetic structure, and models for testing the specific genetic structure exist (Franić et al., 2013). For example, some analyses indicate that genetic influences on the Big Five are not structured coherently around unifying factors (Briley & Tucker-Drob, 2012; Johnson & Krueger, 2004; Kandler, Riemann, Spinath, & Angleitner, 2010). Others have found that the HEXACO dimensions do reflect coherent genetic structures (Lewis & Bates, 2014). It is unclear whether the proper level of analysis for identifying genetic associations is at more narrow levels (e.g., facets) or more broad levels (e.g., superfactors), or alternatively, whether the identification of factor structures derived from genetically informative data would yield more useful dimensions for genetic association (Mõttus, Kandler, et al., 2017).

Departure from additivity. Molecular genetic methods overwhelmingly assume additivity, meaning the effects of SNPs add up across environments, across the genome, and across alleles within a locus. Although there is relatively little information concerning a lack of additivity with respect to environments (cf. Krueger, South, Johnson, & Iacono, 2008), there is

¹¹ It is noteworthy that the limitations we identified were also the focus of Baumert et al. (2017). Specifically, they laid out the empirical evidence (or lack thereof) concerning the psychometric structure and development of personality, particularly for making more fine-grained distinctions than the Big Five. It is encouraging that these issues are seen as critical in general for personality psychology, even beyond the relevance for genetic approaches.

fairly substantial evidence that other sorts of additivity may not hold for personality in contrast to other phenotypes. In large scale twin studies (Möttus et al., *in press*; Rimfeld et al., 2016; van den Berg et al., 2014), identical twins routinely correlate more than twice as strongly as fraternal twins for the Big Five and other personality dimensions. Under a purely additive model, it is impossible for identical twins to be correlated more than twice as strong as fraternal twins because that is the expectation of the linear (additive) impact of the additional genetic similarity. When identical twins are correlated more than twice as strongly as fraternal twins, it is an indication of dominant genetic influences (i.e., within-locus interactions), epistatic genetic influences (i.e., between-locus interactions), or non-additivity due to some other reason. As a simplified example, eye color is known to display dominant genetic influences, whereby two blue-eyed alleles are required for blue vs. non-blue categorization (Sturm et al., 2008). However, eye color varies beyond this simple distinction. Interactions among genes, meaning epistatic effects, further determine the specific shade of eye color, whether blue, brown, green, or hazel (Pośpiech et al., 2011). Beyond the common example of eye color, non-additivity is rare across a wide range of traits (Polderman et al., 2015); its pervasiveness in personality may explain why GWAS successes for personality lag behind other phenotypes.

Developmental complexity. Non-additivity may also result from complex developmental processes, rather than narrowly defined mechanisms like dominance or epistasis. It may be the case that lower order personality facets all mutually interact with one another to produce higher-order dimensions like the Big Five. If this is the case, then there may be complex interdependencies among genetically influenced facets or nuances that aggregate in a way that could obscure main effects of SNPs (e.g., Lykken, McGue, Tellegen, & Bouchard, 1992; Möttus & Allerhand, *in press*). Additionally, multiple processes of gene-environment interplay may all

operate simultaneously, building on top of one another (Kandler & Zapko-Willmes, 2017; Tucker-Drob & Briley, *in press*). To the extent that multiple sources of gene-environment interplay operate, either concurrently or sequentially, this could potentially explain why measured genetic and environmental predictors of personality development have been very difficult to find. Adding to the struggle, development is inherently random and chaotic, even under conditions where genes and the environment are essentially constant (Molenaar, Boomsma, & Dolan, 1993)

Potential solutions and recommendations. We encourage personality psychologists to adopt the methods outlined in this manuscript. Twin and family models are used somewhat, but molecular designs are used relatively infrequently among personality psychologists who do not explicitly identify as behavior geneticists. This situation is unfortunate given the wide-ranging applicability of the methods for answering causal questions, rather than merely estimating genetic and environmental variance components. As the examples presented in this brief review illustrate, behavior geneticists have an expansive history of interest in personality-relevant outcomes, as well as access to existing datasets and expertise in developing new data collection efforts. Collaboration has long been a hallmark of the behavior genetics field, and we encourage personality psychologists with an interest in causality to seek out such collaborative relationships.

As an example, Möttus, Realo et al. (2017) investigated the interconnection between genetic influences on educational attainment and personality using molecular, multi-informant, facet-level data. This approach allowed for innovative analyses concerning whether associations were present at the domain-level or uniquely at the facet-level and whether associations were consistent across self- and informant-report. Yet, the facet structure and developmental

mechanisms driving differentiation of personality is far from established (Baumert et al., 2017). Future work stemming from these promising methodological synergies can provide strong tests of personality structure at multiple levels, particularly when joined with statistically powerful family-based designs (see Franić et al., 2013). We anticipate that providing a satisfying explanation to structural issues may also help issues of non-additivity.

Massively increasing sample size is another plausible approach. Rather than careful measurement of a phenotype, GWAS have advanced simply by adding more participants. It is conceivable that public interest in genetics and willingness to complete online surveys will produce large datasets capable of overcoming psychometric difficulties simply through size alone. If researchers are willing to be a bit creative with the specific items used and willing to pore through gigantic datasets which ask all sorts of medical and social questions, then extremely large datasets are already available (see Footnote 10). Of course, it would be ideal if these two tracks, careful psychometric assessment and large-scale data collection, were mutually informative.

Currently available twin and family studies have sufficient power for many of the highlighted designs. These designs are not intended to estimate heritability, but rather to test any sort of hypothesis for which a personality psychologist can pull together the data. Such data can be used to make stronger causal claims than non-genetically informative data. We encourage researchers to carefully consider possible confounds or assumptions in these models, not least of which is the possibility that gene-environment interplay may play a role in the estimates. Nonetheless, the wide range of approaches presented here illustrates the potential power for a multi-method approach to causal reasoning in personality. Convergence of implied causal

structures across multiple methods could reveal strong evidence for (or against) certain causal paths of interest to personality psychologists and society as a whole.

Conclusions

Each of the tools available for analyzing genetically informative data provide different information about the underlying causal structures that produced the data. No one method can identify causality, but consilience of multiple techniques all pointing in the same direction can provide a solid foundation for causal inferences (Whewell, 1840). We recommend that researchers carefully consider what sorts of confounds certain methods can rule out and at the same time what sorts of potentially causal pathways may also be omitted. As more information concerning the functioning of biological systems and the developmental structuring of personality is uncovered, behavior genetic techniques should allow personality psychology to move from correlation to causation.

References

- Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*, 444-455.
- Baumert, A., Schmitt, M., Perugini, M., Johnson, W., Blum, G., ... & Wrzus, C. (2017). Integrating personality structure, personality process, and personality development. *European Journal of Personality*, *31*, 503-528.
- Berry, D., & Willoughby, M. T. (2017). On the practical interpretability of cross-lagged panel models: Rethinking a developmental workhorse. *Child Development*, *88*, 1186-1206.
- Bleidorn, W., Kandler, C., & Caspi, A. (2014). The behavioural genetics of personality development in adulthood – Classic, contemporary, and future trends. *European Journal of Personality*, *28*, 244-255.
- Bleidorn, W., Kandler, C., Riemann, R., Angleitner, A., & Spinath, F. M. (2009). Patterns and sources of adult personality development: Growth curve analyses of the NEO PI-R scales in a longitudinal twin study. *Journal of Personality and Social Psychology*, *97*, 142-155.
- Bickman, L., & Reich, S. M. (2009). Randomized controlled trials: Gold standard with feet of clay? In S. I. Donaldson, C. A. Christie, and M. M. Mark (Eds.), *What counts as credible evidence in applied research and evaluation practice?* (pp. 51-77). Thousand Oaks, CA: Sage.
- Boardman, J. D., & Fletcher, J. M. (2015). To cause or not to cause? That is the question, but identical twins might not have all the answers. *Social Science & Medicine*, *127*, 198-200.

- Bouchard, T. J. (1997). Experience producing drive theory: How genes drive experience and shape personality. *Acta Paediatrica*, *86*, 60-64.
- Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, *90*, 443-450.
- Briley, D. A., & Tucker-Drob, E. M. (2012). Broad bandwidth or high fidelity? Evidence from the structure of genetic and environmental effects on the facets of the five factor model. *Behavior Genetics*, *42*, 743-763.
- Briley, D. A., & Tucker-Drob, E. M. (2013). Explaining the increasing heritability of cognitive ability across development: A meta-analysis of longitudinal twin and adoption studies. *Psychological Science*, *24*, 1704-1713.
- Briley, D. A., & Tucker-Drob, E. M. (2014). Genetic and environmental continuity in personality development: A meta-analysis. *Psychological Bulletin*, *140*, 1303-1331.
- Briley, D. A., & Tucker-Drob, E. M. (2017). Comparing the developmental genetics of cognition and personality over the lifespan. *Journal of Personality*, *85*, 51-64.
- Briley, D. A., Harden, K. P., & Tucker-Drob, E. M. (2014). Child characteristics and parental educational expectations: Evidence for transmission with transaction. *Developmental Psychology*, *50*, 2614-2632.
- Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., ... & Neale, B. M. (2015). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, *47*, 1236-1241.

- Carroll, J. (2009). Anti-reductionism. In H. Beebe, C. Hitchcock, & P. Menzies (Eds.), *The Oxford handbook of causation* (pp. 279-298). Oxford, Oxford University Press.
- Cartwright, N. (1979). Causal laws and effective strategies. *Nous*, 419-437.
- Chabris, C. F., Hebert, B. M., Benjamin, D. J., Beauchamp, J., Cesarini, D., Van der Loos, M., ... & Freese, J. (2012). Most reported genetic associations with general intelligence are probably false positives. *Psychological Science*, 23, 1314-1323.
- 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.
- Cheesman, R., Selzam, S., Ronald, A., Dale, P. S., McAdams, T. A., Eley, T. C., & Plomin, R. (2017). Childhood behavior problems show the greatest gap between DNA-based and twin heritability. *Translational Psychiatry*, 7, 1284.
- Chickering, D. M. (2002). Optimal structure identification with greedy search. *Journal of Machine Learning Research*, 3, 507-554.
- Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12, 671-684.
- 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.
- Davidson, D. (1967). Causal relations. *The Journal of Philosophy*, 64, 691-703.
- Davies, G., Lam, M., Harris, S. E., Tampush, J., Luciano, M., Hill, W. D., ... & Deary, I. J. (2017). Ninety-nine independent genetic loci influencing general cognitive function

- include genes associated with brain health and structure (N = 280,360). *Preprint on bioRxiv*, doi: <https://doi.org/10.1101/176511>
- Deaton, A., & Cartwright, N. (2017, December 25). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*. Advanced online publication. doi: 10.1016/j.socscimed.2017.12.005
- de Leeuw, C. A., Mooij, J. M., Heskes, T., & Posthuma, D. (2015). MAGMA: Generalized gene-set analysis of GWAS data. *PLoS Computational Biology*, *11*, 1-19.
- de Moor, M. H. M., Costa, P. T., Terracciano, A., Krueger, R. F., de Geus, E. J. C., ... Boomsma, D. I. (2012). Meta-analysis of genome-wide association studies for personality. *Molecular Psychiatry*, *17*, 337-349.
- Dickens, W. T., & Flynn, J. R. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. *Psychological Review*, *108*, 346-369.
- Dowe, P. (1995). Causality and conserved quantities: A reply to Salmon. *Philosophy of Science*, *62*, 321-333.
- Duncan, G. E., Mills, B., Strachan, E., Hurvitz, P., Huang, R., Moudon, A. V., & Turkheimer, E. (2014). Stepping towards causation in studies of neighborhood and environmental effects: How twin research can overcome problems of selection and reverse causation. *Health & Place*, *27*, 106-111.
- Franić, S., Dolan, C. V., Borsboom, D., Hudziak, J. J., van Beijsterveldt, C. E. M., & Boomsma, D. I. (2013). Can genetics help psychometrics? Improving dimensionality assessment through genetic factor modeling. *Psychological Methods*, *18*, 406-433.

- Gage, S. H., Davey Smith, G., Ware, J. J., Flint, J., Munafò, M. (2016). G = E: What GWAS can tell us about the environment. *PLOS Genetics*, *12*, e1005765.
- Gillespie, N. A., Zhu, G., Neale, M. C., Heath, A. C., & Martin, N. G. (2003). Direction of causation modeling between cross-sectional measures of parenting and psychological distress in female twins. *Behavior Genetics*, *33*, 383-396.
- Glennan, S. S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, *44*, 49-71.
- Glymour, C. (2004). Review of Woodward's *Making Things Happen*. *The British Journal for the Philosophy of Science*, *55*, 779-790.
- Glymour, C. (2010). What is right with 'Bayes net methods' and what is wrong with 'hunting causes and using them'? *The British Journal for the Philosophy of Science*, *61*, 161-211.
- Glymour, C., & Glymour, M. R. (2014). Commentary: race and sex are causes. *Epidemiology*, *25*, 488-490.
- Glynn, L. (2011). A probabilistic analysis of causation. *The British Journal for the Philosophy of Science*, *62*, 343-392.
- GTEx Consortium (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*, *348*, 648-660.
- Harden, K. P., Quinn, P. D., & Tucker-Drob, E. M. (2012). Genetically influenced change in sensation seeking drives the rise of delinquent behavior during adolescence. *Developmental Science*, *15*, 150-163.
- Heath, A. C., Kessler, R. C., Neale, M. C., Hewitt, J. K., Eaves, L. J., & Kendler, K. S. (1993). Testing hypotheses about direction of causation using cross-sectional family data. *Behavior Genetics*, *23*, 29-50.

- Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, *103*, 2052-2086.
- Hemani, G., Zheng, J., Wade, K. H., Laurin, C., Elsworth, B., Burgess, S., ... & The MR-Base Collaboration (2016). MR-Base: A platform for systematic causal inference across the phenome using billions of genetic associations. *Preprint on bioRxiv*. doi: <https://doi.org/10.1101/078972>
- Holland, P. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*, 945-960.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*, 945-960.
- Hume, D. (1748). *An enquiry concerning human understanding*.
- Johnson, A. M., Vernon, P. A., & Feiler, A. R. (2008). Behavioral genetic studies of personality: An introduction and review of the results of 50+ years of research. In G. Boyle, G. Matthews, & D. Saklofske (Eds.), *Handbook of personality theory and assessment* (pp. 145-173). London, England: Sage.
- Johnson, W., & Krueger, R. F. (2004). Genetic and environmental structure of adjectives describing the domains of the Big Five Model of personality: A nationwide US twin study. *Journal of Research in Personality*, *38*, 448-472.
- Johnson, W., Penke, L., & Spinath, F. M. (2011). Heritability in the era of molecular genetics: Some thoughts for understanding genetic influences on behavioural traits. *European Journal of Personality*, *25*, 254-266.

- Kandler, C. (2012). Nature and nurture in personality development: The case of neuroticism and extraversion. *Current Directions in Psychological Science, 21*, 290-296.
- Kandler, C., & Papendick, M. (2017) *Behavior genetics and personality development: A methodological and meta-analytic review*. In J. Specht (Ed.), *Personality development across the life-span* (pp. 473–95). San Diego, CA: Elsevier.
- Kandler, C., & Zapko-Willmes, A. (2017). Theoretical perspectives on the interplay of nature and nurture in personality development. In J. Specht (Ed.), *Personality Development across the Lifespan* (pp. 101-115). Elsevier Academic Press.
- Kandler, C., Bleidorn, W., Riemann, R., Angleitner, A., & Spinath, F. M. (2012). Life events as environmental states and genetic traits and the role of personality: A longitudinal twin study. *Behavior Genetics, 42*, 57-72.
- Kandler, C., Riemann, R., Spinath, F. M., & Angleitner, A. (2010). Sources of variance in personality facets: A multiple-rater twin study of self-peer, peer-peer, and self-self (dis-) agreement. *Journal of Personality, 78*, 1565-1594.
- Kendler, K. S. (2001). Twin studies of psychiatric illness: An update. *Archives of Genetic Psychiatry, 58*, 1005-1014.
- Kendler, K. S., Lönn, S. L., Sundquist, J., & Sundquist, K. (2015). Smoking and schizophrenia in population cohorts of Swedish women and men: A prospective co-relative control study. *The American Journal of Psychiatry, 172*, 1092-1100.
- Kendler, K. S., Myers, J., Damaj, M. I., & Chen, Z. (2013). Early smoking onset and risk for subsequent nicotine dependence: A monozygotic co-twin control study. *American Journal of Psychiatry, 170*, 408-413.

- Koellinger, P. D., & Harden, K. P. (2018). Using nature to understand nurture. *Science*, *359*, 386-387.
- Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsdottir, B. J., Young, A. I., Thorgeirsson, T. E., ... & Stefansson, K. (2018). The nature of nurture: Effects of parental genotypes. *Science*, *359*, 424-428.
- Krapohl, E., & Plomin, R. (2016). Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. *Molecular Psychiatry*, *21*, 437-443.
- Krapohl, E., Rimfeld, K., Shakeshaft, N. G., Trzaskowski, M., McMillan, A., Pingault, J.-B., Asbury, K., Harlaar, N., Kovas, Y., Dale, P. S., & Plomin, R. (2014). The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 15273-15278.
- Krueger, R. F., & Johnson, W. (2008). Behavioral genetics and personality: A new look at the integration of nature and nurture. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed.). Guilford.
- Krueger, R. F., South, S., Johnson, W., & Iacono, W. (2008). The heritability of personality is not always 50%: Gene-environment interactions and correlations between personality and parenting. *Journal of Personality*, *76*, 1485-1522.
- Lee, J. J. (2012). Correlation and causation in the study of personality. *European Journal of Personality*, *26*, 372-390.

- Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M., & Wray, N. R. (2012). Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics*, 28(19), 2540-2542.
- Lewis, D. (1973). Causation. *The Journal of Philosophy*, 70, 556-567.
- Lewis, D. (2000). Causation as influence. *The Journal of Philosophy*, 97, 182-197.
- Lewis, G. J., & Bates, T. C. (2014). How genes influence personality: Evidence from multi-facet twin analyses of the HEXACO dimensions. *Journal of Research in Personality*, 51, 9-17.
- Linnér, R. K., Marioni, R. E., Rietveld, C. A., Simpkin, A. J., Davies, N. M., Watanabe, K., ... & Benjamin, D. J. (2017). An epigenome-wide association study meta-analysis of educational attainment. *Molecular Psychiatry*, 22, 1680-1690.
- Luo, J., Derringer, J., Briley, D. A., & Roberts, B. W. (2017). Genetic and environmental pathways underlying personality traits and perceived stress: Concurrent and longitudinal twin studies. *European Journal of Personality*, 31, 614-629.
- Lykken, D. T., McGue, M., Tellegen, A., & Bouchard, T. J. (1992). Emergenesis: Genetic traits that may not run in families. *American Psychologist*, 47, 1565-1577.
- Marzi, S. J., Sugden, K., Arseneault, L., Belsky, D. W., Burrage, J., Corcoran, D. L., ... & Caspi, A. (2018). Analysis of DNA methylation in young people: Limited evidence for an association between victimization stress and epigenetic variation in blood. *American Journal of Psychiatry*. Advanced online publication. doi: 10.1176/appi.ajp.2017.17060693

- McArdle, J. J. (1986). Latent variable growth within behavior genetic models. *Behavior Genetics, 16*, 163-200.
- McArdle, J. J., & Hamagami, F. (2003). Structural equation models for evaluating dynamic concepts within longitudinal twin analyses. *Behavior Genetics, 33*, 137-159.
- McGue, M., Bacon, S., & Lykken, D. T. (1993). Personality stability and change in early adulthood: A behavioral genetic analysis. *Developmental Psychology, 29*, 94-109.
- Menzies, P., & Price, H. (1993). Causation as a secondary quality. *The British Journal for the Philosophy of Science, 44*, 187-203.
- Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., ... & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 2693-2698.
- Molenaar, D., Middeldorp, C., van Beijsterveldt, T., & Boomsma, D. I. (2015). Analysis of behavioral and emotional problems in children highlights the role of Genotype \times Environment interaction. *Child Development, 86*, 1999-2016.
- Molenaar, P. C. M., Boomsma, D. I., & Dolan, C. V. (1993). A third source of developmental differences. *Behavior Genetics, 23*, 519-524.
- Mooij, J. M., Peters, J., Janzing, D., Zscheischler, J., & Schölkopf, B. (2016). Distinguishing cause from effect using observational data: Methods and benchmarks. *Journal of Machine Learning Research, 17*, 1-102.
- Möttus, R., & Allerhand, M. (*in press*). Why do traits come together? The underlying trait and network approaches. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *SAGE handbook of*

personality and individual differences: Volume 1. The science of personality and individual differences. London: SAGE.

Mõttus, R., Briley, D. A., Zheng, A., Mann, F. D., Engelhardt, L. E., Tackett, J. L., Harden, K.

P., Tucker-Drob, E. M. (*in press*) Kids becoming less alike: A behavioral genetic analysis of developmental increases in personality variance from ages 8 to 18. *Journal of Personality and Social Psychology*.

Mõttus, R., Kandler, C., Bleidorn, W., Riemann, R., & McCrae, R. R. (2017). Personality traits below facets: The consensual validity, longitudinal stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology*, *112*, 474-490.

Mõttus, R., Realo, A., Vainik, U., Allik, J., & Esko, T. (2017). Educational attainment and personality are genetically intertwined. *Psychological Science*, *28*, 1631-1639.

Nagel, M., Jansen, P. R., Stringer, S., Watanabe, K., de Leeuw, C. A., Bryois, J., ... & Posthuma, D. (2017). GWAS meta-analysis of neuroticism (N=449,484) identifies novel genetic loci and pathways. *Preprint on bioRxiv*, doi: <https://doi.org/10.1101/184820>

Neale, M. C., & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Netherlands: Springer.

Nivard, M. G., & Boomsma, D. I. (2016). Genetics: From molecule to society. *Current Biology*, *26*, R1194-R1196.

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.

- Olivares, E. L., Kendler, K. S., Neale, M. C., & Gillespie, N. A. (2016). The genetic and environmental association between parental monitoring and risk of cannabis, stimulants, and cocaine initiation in a sample of male twins: Does parenting matter? *Twin Research and Human Genetics, 19*, 297-305.
- Ottman, R. (1996). Gene-environment interaction: Definitions and study designs. *Preventative Medicine, 25*, 764-770.
- Pearl, J. (2000). *Causality*. Cambridge, UK: Cambridge University Press.
- Pearl, J. (2001). Bayesianism and causality, or, why I am only a half-Bayesian. In D. Corfield & J. Williamson (Eds.), *Foundations of bayesianism* (pp. 19-36). Dordrecht, Netherlands: Springer.
- Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). New York: Cambridge University Press.
- Pearl, J. (2010). An introduction to causal inference. *The International Journal of Biostatistics, 6*, Article 7.
- Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: a primer*. New York, NY: John Wiley & Sons.
- Pickrell, J. K., Berisa, T., Liu, J. Z., Ségurel, L., Tung, J. Y., & Hinds, D. A. (2016). Detection and interpretation of shared genetic influences on 42 human traits. *Nature Genetics, 48*, 709-717.
- 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.

- Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2016). Top 10 replicated findings from behavioral genetics. *Perspectives on Psychological Science, 11*, 3-23.
- Polderman, T. J. C., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, V., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics, 47*, 702-709.
- Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135*, 322-338.
- Pośpiech, E., Draus-Barini, J., Kupiec, T., Wojas-Pelc, A., & Branicki, W. (2011). Gene-gene interactions contribute to eye colour variation in humans. *Journal of Human Genetics, 56*, 447-455.
- Posthuma, D., Beem, A. L., de Geus, E. J., van Baal, G. C., von Hjelmberg, J. B., Iachine, I., & Boomsma, D. I. (2003). Theory and practice in quantitative genetics. *Twin Research and Human Genetics, 6*, 361-376.
- Psillos, S. (2009). Regularity theories. In H. Beebe, C. Hitchcock, & P. Menzies (Eds.), *The Oxford handbook of causation* (pp. 131-157). Oxford, Oxford University Press.
- Purcell, S. (2002). Variance components models for gene-environment interaction in twin analysis. *Twin Research and Human Genetics, 5*, 554-571.
- Richardson, T. G., Haycock, P. C., Zheng, J., Timpson, N. J., Gaunt, T. R., Davey Smith, G., Relton, C. L., & Hermani, G. (2017). Systematic Mendelian randomization framework elucidates hundreds of genetic loci which may influence disease through changes in DNA methylation levels. *Preprint on bioRxiv*. doi: <http://dx.doi.org/10.1101/189076>

- Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, 2, 313-345.
- Roberts, B. W., Luo, J., Briley, D. A., Chow, P. I., Su, R., & Hill, P. L. (2017). A systematic review of personality trait change through intervention. *Psychological Bulletin*, 143, 117-141.
- Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, 1, 27-42.
- Roisman, G. I., Newman, D. A., Fraley, R. C., Haltigan, J. D., Groh, A. M., & Haydon, K. C. (2012). Distinguishing differential susceptibility from diathesis-stress: Recommendations for evaluating interaction effects. *Development and Psychopathology*, 24, 389-409.
- Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39, 33-38.
- Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39, 33-38.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*, 66, 688-701.

- Salmon, W. C. (1997). Causality and explanation: A reply to two critiques. *Philosophy of Science*, 64, 461-477.
- Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C. A., ... & Posthuma, D. (2017). GWAS meta-analysis (N=279,930) identifies new genes and functional links to intelligences. *Preprint on bioRxiv*, doi: <https://doi.org/10.1101/184853>
- Scarr, S., & McCartney, K. (1983). How people make their own environments: A theory of genotype → environment effects. *Child Development*, 54, 424-435.
- Schaffer, J. (2005). Contrastive causation. *The Philosophical Review*, 114, 327-358.
- Selzam, S., Krapohl, E., von Stumm, S., O'Reilly, P. F., Rimfeld, K., Kovas, Y., ... & Plomin, R. (2017). Predicting educational achievement from DNA. *Molecular Psychiatry*, 22, 267-272.
- Shimizu, S., Hoyer, P. O., Hyvärinen, A., & Kerminen, A. (2006). A linear non-Gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7, 2003-2030.
- Spirtes, P., Glymour, C. N., & Scheines, R. (2000). *Causation, prediction, and search*. Cambridge, MA: MIT press.
- Sturm, R. A., Duffy, D. L., Zhao, Z. Z., Leite, F. P. N., Stark, M. S., Hayward, N. K., ... & Montgomery, G. W. (2008). A single SNP in an evolutionary conserved region within intron 86 of the *HERC2* gene determines human blue-brown eye color. *The American Journal of Human Genetics*, 82, 424-431.
- Suppes, P. (1970). *A probabilistic theory of causality*. Amsterdam: North-Holland Publishing Company.
- Taylor, R. (1966). *Action and purpose*. Englewood Cliffs, NJ: Prentice-Hall.

- Trzaskowski, M., Davis, O. S., DeFries, J. C., Yang, J., Visscher, P. M., & Plomin, R. (2013). DNA evidence for strong genome-wide pleiotropy of cognitive and learning abilities. *Behavior Genetics, 43*, 267-273.
- Tucker-Drob, E. M., & Briley, D. A. (*in press*). Theoretical concepts in the genetics of personality development. In D. P. McAdams, R. L. Shiner, and J. L. Tackett (Eds.). *The handbook of personality development* (Eds., McAdams, Shiner, & Tackett).
- Tucker-Drob, E. M., Briley, D. A., Engelhardt, L. E., Mann, F. D., & Harden, K. P. (2016). Genetically-mediated associations between measures of childhood character and academic achievement. *Journal of Personality and Social Psychology, 111*, 790-815.
- Turkheimer, E., & Harden, K. P. (2014). Behavior genetic research methods: Testing quasi-causal hypotheses using multivariate twin data. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in personality and social psychology* (2nd ed., pp. 159-187). New York, NY: Cambridge University Press.
- Turkheimer, E., & Waldron, M. (2000). Nonshared environment: A theoretical, methodological, and quantitative review. *Psychological Bulletin, 126*, 78-108.
- Turkheimer, E., D'Onofrio, B. M., Maes, H. H., & Eaves, L. J. (2005). Analysis and interpretation of twin studies including measures of the shared environment. *Child Development, 76*, 1217-1233.
- Turkheimer, E., Pettersson, E., & Horn, E. E. (2014). A phenotypic null hypothesis for the genetics of personality. *Annual Review of Psychology, 65*, 515-540.

- van den Berg, S. M., de Moor, M. H. M., McGue, M., Pettersson, E., Terracciano, A., Verweij, K. J. H., ... & Boomsma, D. I. (2014). Harmonization of neuroticism and extraversion phenotypes across inventories and cohorts in the Genetics of Personality Consortium: An application of Item Response Theory. *Behavior Genetics*, *44*, 295-313.
- Visscher, P. M., Brown, M. A., McCarthy, M. I., Yang, J. (2012). Five years of GWAS discovery. *American Journal of Human Genetics*, *90*, 7-24.
- Vukasović, T., & Bratko, D. (2015). Heritability of personality: A meta-analysis of behavior genetic studies. *Psychological Bulletin*, *141*, 769-785.
- Watanabe, K., Taskesen, E., van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations. *Preprint on bioRxiv*, doi: <https://doi.org/10.1101/110023>
- Whewell, W. (1840). *The philosophy of the inductive sciences, founded upon their history* (Vol. 2). London, England: John W. Parker.
- Woodward, J. (2003). *Making things happen: A causal theory of explanation*. Oxford, UK: Oxford University Press.
- Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, *88*, 76-82.
- Zheng, J., Erzurumluoglu, A. M., Elsworth, B. L., Kemp, J. P., Howe, L., Haycock, P. C., ... & Warrington, N. M. (2017). LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*, *33*, 272-279.

Zhu, Z., Zheng, Z., Zhang, F., Wu, Y., Trzaskowski, M., ... & Yang, J. (2018). Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nature Communications*, 9, 224.

Table 1. Overview of behavior genetic terminology relevant for causal reasoning

Term	Definition	Implication for estimates	Relevant Reference
Twin and Family Study	Studies in which the psychological similarity of genetically (e.g., identical and fraternal twins) or environmentally (e.g., adoptive siblings) related individuals is assessed. This information is used to make inferences about the proportion of variance due to genetic and environmental influences.	Capable of estimating population-level variance components, not individual prediction. Tend to estimate larger estimates of heritability as all sorts of non-additive effects are included.	Polderman et al., (2015)
Molecular Genetic Study	Studies in which the specific genetic differences are measured in a large sample of unrelated individuals. These genetic differences are then tested for association with a psychological variable.	Capable of estimating both population-level and individual-level information, but effect sizes are extremely small, limiting utility currently.	Nagel et al. (2017)
Phenotype	The observed or displayed characteristic of the individual. The manifestation of the underlying genetic and environmental influences.	As with all assessment, the psychometrics of the phenotype are critical. Measurement error and other sorts of validity problems may confound estimates (but may be overcome with large sample sizes).	Okbay et al., (2016)
Heritability	The proportion of variance in a phenotype due to genetic influences. How much variance is associated with differences in genetic relatedness?	Heritability is not an assessment of malleability or innateness. It is a variance component reflective of how much genotype (and all things that genotype is a proxy for) could possibly predict.	Vukasovic & Bratko (2015)
Shared Environmentality	The proportion of variance in a phenotype due to family-level environmental influences. How much variance is associated with between-family differences?	Captures effective shared environments, meaning effects that make siblings more similar (e.g., parents, peers, and teachers causally treating children the same).	Turkheimer et al., (2005)

Nonshared Environmentality	The proportion of variance in a phenotype due to individual-level environmental influences. How much variance is left over after genetic and between-family differences are taken into account?	Captures effective nonshared environments, meaning effects that lead siblings to differ (e.g., unique relationships between a child and their parents, peers, and teachers).	Turkheimer & Waldron (2000)
SNP	Single Nucleotide Polymorphism. A location in the genetic code at which individuals differ in terms of which nucleotide (a single rung on the ladder of DNA) is present.	As more knowledge of the genome has accumulated, common SNP chips that measure 2.5 million markers can be used to reliably impute over 17 million markers.	Nagel et al., (2017)
GWAS	Genome Wide Association Study. A study that correlates SNPs with a phenotype. Summary statistics are sometimes made publicly available which lists the association between each SNP and the phenotype.	Due to the sheer size of the genome, the GWAS significance level is set to $p < 5 \times 10^{-8}$.	Chabris et al., (2015)
GCTA	Genome Wide Complex Trait Analysis. A technique that uses SNP data to estimate the genetic relatedness of traditionally unrelated individuals (i.e., 5th cousins or more distant). Then, this genetic relatedness matrix is used to estimate SNP-based heritability.	Only captures measured (i.e. genotyped or imputed) additive genetic influences, in contrast to twin and family studies which capture all sorts of broadly genetic influences.	Yang et al., (2011)
Linkage Disequilibrium (LD)	SNPs are not independent of one another. With knowledge of a set of marker SNPs, one can predict other SNPs with extremely high confidence (>99.9%).	LD is the reason the strict significance cutoff in GWAS studies is not even more strict; the cutoff is based on the number of possible independent tests across the genome. This information allows genetic correlations to be estimated simply with GWAS summary statistics and LD information.	Bulik-Sullivan et al., (2015)
Epigenetics	Things that occur above the static genome; typically refers to environmental factors that alter the manner with which the static genome is read.	Despite the name, epigenetics would ordinarily appear as an environmental effect in behavior genetic studies. Epigenetic effects are not likely to be substantially larger than GWAS effects.	Linnér et al., (2017)

Genetic Correlation/Covariance	The extent to which genetic influences on one phenotype are shared with another phenotype (e.g., do the genetic influences that make one high on extraversion make one higher or lower on agreeableness?).	Can be estimated using twin and family designs, GCTA with individual-level data, and LD Score regression using only GWAS summary statistics.	Tucker-Drob et al., (2016)
Environmental Correlation/Covariance	Similar to the genetic correlation, but for environmental factors (e.g., do the environmental factors that make one high on conscientiousness make one higher or lower on neuroticism?).	Twin and family studies typically distinguish shared and nonshared environmental correlations. Molecular genetic designs typically estimate the environmental correlation as a residual (i.e., whatever the additive genetic correlation does not account for).	Kandler et al., (2012)
Gene-Environment Interplay	Umbrella term for any way that genes and environments are dependent and non-additive causal factors.	Strict interpretation of the specification of behavior genetic models would be that genes and environments are uncorrelated and additive influences on the phenotype. Luckily, implications of deviations from this assumption are predictable.	Plomin et al., (1977)
Gene-Environment Correlation	Refers to the non-random experience of environments as a function of genotype. Can be passive, evocative, or active.	In family studies, passive gene-environment correlation contributes shared environmental variance, and evocative or active gene-environment correlation contributes genetic variance. In molecular studies, gene-environment correlation contributes genetic variance.	Scarr & McCartney (1983)
Gene \times Environment Interaction	Refers to the dependence of genetic and environmental influences on one another. Genetic influences may manifest differently depending on environmental context, or equivalently, some environmental risk factors may only affect individuals with a certain genetic polymorphism.	In family studies, Gene \times Shared Environment interaction contributes genetic variance, and interaction with the nonshared environment contributes nonshared environmental variance. In molecular genetic studies, the implications are dependent on the specific technique and form of interaction.	Purcell (2002)

Pleiotropy	Genes may act as a common cause of multiple phenotypes, meaning the phenotypes are not likely to be causally related. Environments may act in a similar way, but pleiotropy specifically refers to genes.	Pleiotropy can be inferred from genetic correlations.	Keller et al., (2013)
Causal Chain	Genes or environments may cause some intermediary phenotype, which in turn, causally influences a separate phenotype.	Causal chains can also be inferred from genetic correlations, highlighting the importance of clear causal reasoning.	Gage et al., (2016)
Co-twin control design	Comparing whether (identical) twins that are discordant for some risk factor (e.g., drug use) also differ on some phenotype of interest (e.g., cognitive development). Because siblings grew up in the same home environment and share genetic material, confounds are minimized.	If co-twin control designs find a significant association, then one still needs to worry about some sort of nonshared environmental common cause. If co-twin control designs do not find a significant association, then it is possible that two variables are causally connected through a genetic or shared environmental causal chain.	McGue et al., (2010)
Mendelian Randomization	Estimates causal effects of one phenotype on another, under the assumption that which SNP an individual was assigned at birth is independent of other confounding factors. This technique requires a variety of assumptions that may or may not be well-supported.	The best practices for this approach are constantly evolving. The utility of the method is dependent on properly estimating and interpreting GWAS effect sizes.	Davey Smith & Ebrahim (2003)

Figure 1. Causal models that could produce a genetic association between personality and achievement.

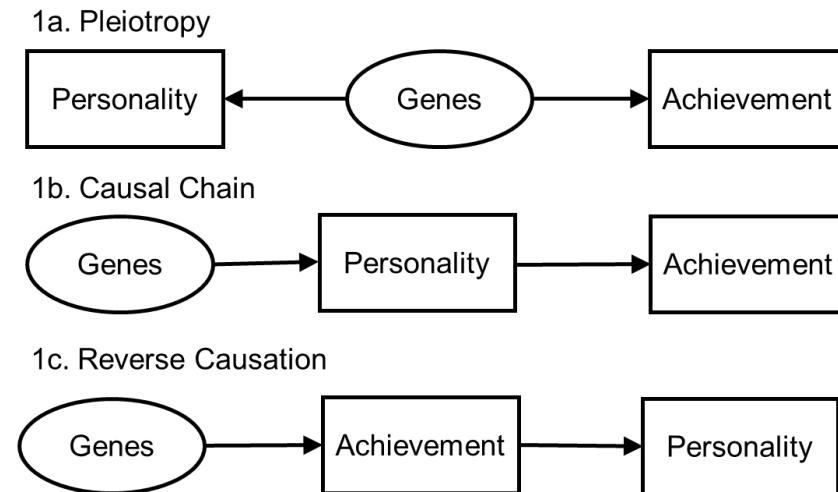
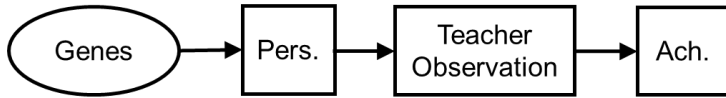
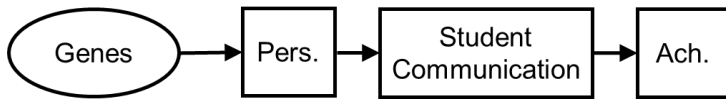


Figure 2. Causal models due to gene-environment correlation. Pers. = Personality. Ach. = Achievement. Env. = Environment.

2a. Evocative Gene-Environment Correlation



2b. Active Gene-Environment Correlation



2c. Passive Gene-Environment Correlation

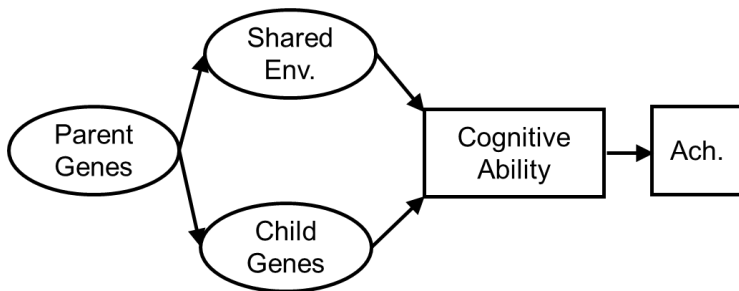
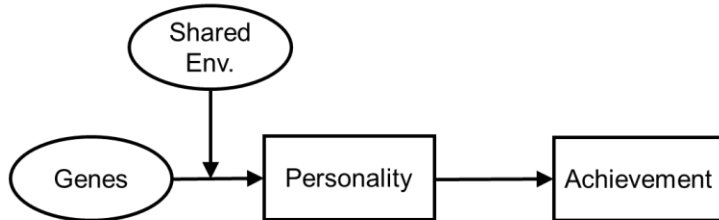


Figure 3. Causal models due to Gene \times Environment interaction. Env. = Environment. NonS. = Nonshared.

3a. Gene \times Shared Environment Interaction



3b. Gene \times Nonshared Environment Interaction

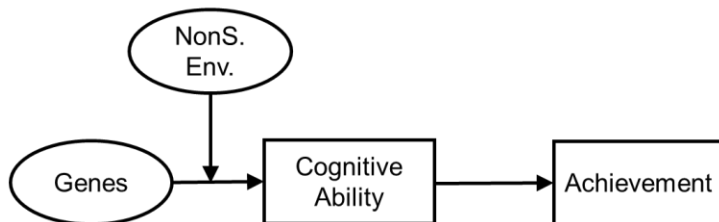
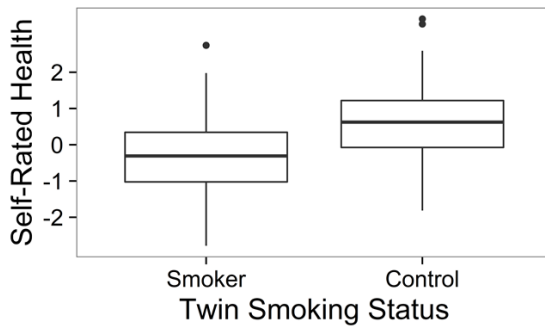
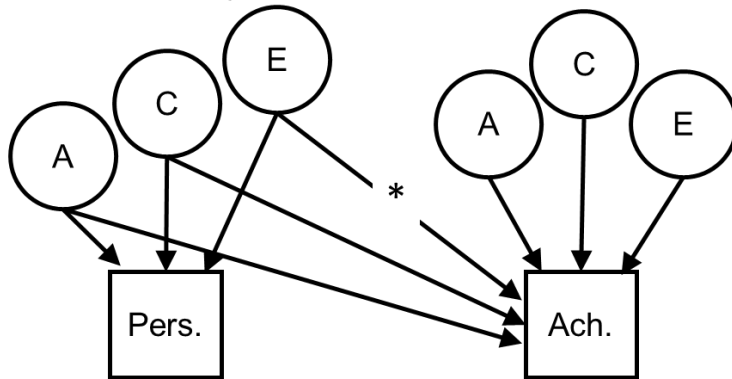


Figure 4. Twin and family designs intended to demonstrate causality. The boxplot demonstrating co-twin control is based on simulated data. A = additive genetic influences. C = shared environmental influences. E = nonshared environmental influences. Pers. = Personality. Ach. = Achievement.

4a. Co-twin control



4b. Within-family difference



4c. Phenotypic filtering

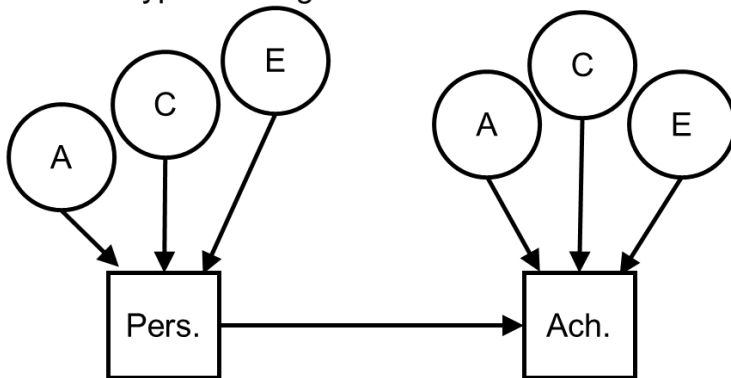


Figure 5. Direction of causation twin model. Note that this model is not identified as currently specified. Model comparisons are conducted based on dropping parameters (see Heath et al., 1993). A = additive genetic influences. C = shared environmental influences. E = nonshared environmental influences. Pers. = Personality. Ach. = Achievement.

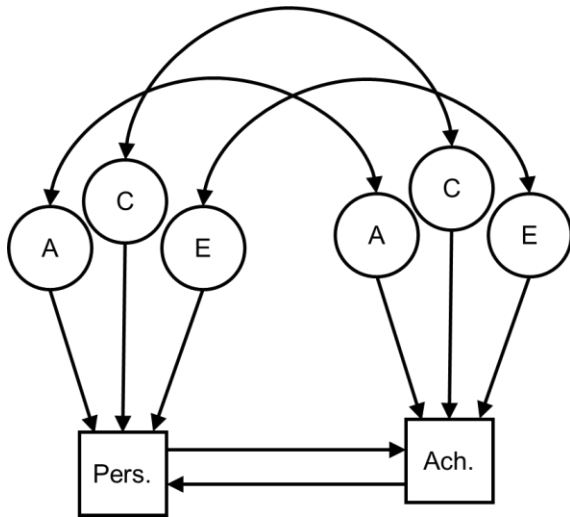


Figure 6. Simulated scatterplots of SNP effect sizes for personality and achievement. The scatterplot on the left displays a strong relation between the SNP effect sizes on personality and achievement for the top 100 personality associated SNPs. The scatterplot on the right displays a null relation between the SNP effect sizes on personality and achievement for the top 100 achievement associated SNPs. In this circumstance, one would infer that personality is a cause of achievement, not vice versa.

