Behavior Genetics (2019) 49:196–210
https://doi.org/10.1007/s10519-018-9939-6

ORIGINAL RESEARCH

Interpreting Behavior Genetic Models: Seven Developmental Processes to Understand

Daniel A. Briley[1] · Jonathan Livengood[2] · Jaime Derringer[1] · Elliot M. Tucker-Drob[3] · R. Chris Fraley[1] · Brent W. Roberts[1]

Received: 1 December 2017 / Accepted: 16 November 2018 / Published online: 22 November 2018

Abstract
Behavior genetic findings figure in debates ranging from urgent public policy matters to perennial questions about the nature of human agency. Despite a common set of methodological tools, behavior genetic studies approach scientific questions with potentially divergent goals. Some studies may be interested in identifying a complete model of how individual differences come to be (e.g., identifying causal pathways among genotypes, environments, and phenotypes across development). Other studies place primary importance on developing models with predictive utility, in which case understanding of underlying causal processes is not necessarily required. Although certainly not mutually exclusive, these two goals often represent tradeoffs in terms of costs and benefits associated with various methodological approaches. In particular, given that most empirical behavior genetic research assumes that variance can be neatly decomposed into independent genetic and environmental components, violations of model assumptions have different consequences for interpretation, depending on the particular goals. Developmental behavior genetic theories postulate complex transactions between genetic variation and environmental experiences over time, meaning assumptions are routinely violated. Here, we consider two primary questions: (1) How might the simultaneous operation of several mechanisms of gene–environment (GE)-interplay affect behavioral genetic model estimates? (2) At what level of GE-interplay does the 'gloomy prospect' of unsystematic and non-replicable genetic associations with a phenotype become an unavoidable certainty?

Keywords Gene–environment interplay · Human agency · Personality · Cognitive ability · Developmental genetics

Introduction

All behavior genetic models operate under the assumption that genetic and environmental processes combine to give rise to psychological phenotypes. The most detailed theoretical models posit multifaceted, dynamic developmental processes whereby genetic variation comes to be correlated with and statistically dependent on experience (Beam et al. 2015; Bronfenbrenner and; Ceci 1994; Dickens and Flynn 2001; Johnson 2007; Kandler and Zapko-Willmes 2017; Plomin et al. 1977; Scarr and McCartney 1983; Tucker-Drob 2017; Tucker-Drob et al. 2013). However, most empirical studies in behavior genetics use simple model specifications according to which genetic and environmental influences combine by addition. For some purposes, simple models are entirely appropriate. But for others, simple models may be misleading. Therefore, it is important to consider the research goals being pursued when interpreting behavior genetic results. We focus on two potential research goals: explanation and prediction.[1] Increases in predictive accuracy do not always

✉ Daniel A. Briley
dabriley@illinois.edu

[1] Department of Psychology, University of Illinois at Urbana-Champaign, 603 East Daniel Street, Champaign, IL 61820, USA

[2] Department of Philosophy, University of Illinois at Urbana-Champaign, Champaign, IL, USA

[3] Department of Psychology and Population Research Center, University of Texas at Austin, Austin, TX, USA

[1] Other aims that have been claimed as constitutive of scientific inquiry include having true answers to our questions (Kelly and Glymour 2004), obtaining knowledge (Nagel 1967), advancing empirically adequate theories (van Fraassen 1980, 1986), having understanding (de Regt 2015), and gaining the ability to control nature (Keller 1985). We invite the reader to think about how what we say in this paper matters with respect to these other aims as well, though we will not discuss them explicitly.

imply increases in explanatory power. For example, simple model specifications aggregate over the potential presence of complex underlying processes into relatively few parameters. In some circumstances, prediction could be maximized via simplifying assumptions, but explanation of the underlying processes may not be enhanced (or vice versa). Here, we outline seven theoretically and empirically pervasive developmental processes that should be considered when interpreting behavior genetic model results, regardless of whether such processes are formally modeled within any given investigation (see Table 1 for summary).

To make some progress on these issues, we begin by describing how research goals aimed at prediction compared to explanation differ. Next, we describe seven developmental processes necessary to keep in mind when interpreting behavior genetic results. The first six processes are relatively easy to understand, but the final process, simultaneous gene–environment interplay (GE-interplay), is quite difficult. We then consider how multiple developmental processes could generate the empirical data observed in behavior genetic studies, including the possibility of the "gloomy prospect," wherein the influences on human behavior are so incredibly idiosyncratic as to preclude the possibility of identifying a generalizable model (Plomin and Daniels 1987). Prospects of gloominess may be phenotype specific, implying that boundaries on gloominess could be established by a more complete accounting of the magnitude, timing, and interdependencies among developmental mechanisms guiding phenotype growth.

Identifying research goals

When specifying models or interpreting parameter estimates, researchers should first consider their goals. Although there are many potential goals, herein we focus on two broad categories which cover most current behavior genetic research: explanation and prediction. If one is primarily interested in understanding the causal processes that explain observed patterns of individual differences and their development, then the specified model would need to include the causal processes relevant for a phenotype. To the extent that GE-interplay is relevant to the development of a phenotype, it would be desirable to include these processes in a model. If, however, one is primarily interested in obtaining a useful model for predicting a phenotype, then the specified model may need not represent any causal processes. These processes would only need to be included to the extent that they would obscure prediction.

When interpretation of models conflates the goals of explanation and prediction—and especially when improvements in predictive accuracy are seen as improvements in explanatory power—misinterpretation may be likely. For example, accuracy of predicting neuroticism could be enhanced by including a biological variable reflecting sex (for example, presence or absence of a Y chromosome) in a regression equation; yet, the fact that sex is a statistically significant predictor does not explain individual differences with respect to neuroticism. Although sex differences in neuroticism are cross-culturally consistent (Schmitt et al. 2008), the fact that sex is associated with neuroticism does not tell us whether that association results from biological processes that stem directly from the measured genetic difference (e.g. sexual differentiation, and associated hormones, which stem from the presence or absence of the *SRY* gene), or from persistent cultural processes that are confounded with the same measured genetic difference (e.g. the historical power differential between the sexes, including assumed gender roles in society).

Similarly, it would be a mistake to infer from the fact that the R^2 for one polygenic risk score is greater than the R^2 for another that the first risk score is more informative about the data-generating mechanism. Bluntly, it would be a mistake to think that when the R^2 value goes up, we have better understanding of how the world works, when in fact, we only have a more accurate prediction. Our point here is not to disparage prediction. Predictive accuracy is a worthwhile goal, but we should not confuse it with explanatory power. As we will discuss, the possibility of GE-interplay may alter the number of parameters that plausibly should be included in a model or at least considered as to the impact on estimated parameters. Moreover, our ability to explain our data and to understand how the world works on the basis of models that do not include GE-interplay parameters will be limited to the extent that GE-interplay influences development, even if such models are very useful for prediction. For example, Lee et al. (2018, p. 1116) reminded readers that "it is inappropriate to interpret the polygenic score for educational attainment as a measure of genetic endowment" due to evidence of GE-interplay. The polygenic risk score is a useful aggregate of information, but it does not explain or identify "genetic endowment."

Likewise, the classical twin design specifies the A parameter. Falconer's formula (1960) approximates[2] the mathematical definition of A in a classical twin design: $2 \times (r\text{MZ} - r\text{DZ})$, or in other words, twice the difference in similarity between monozygotic and dizygotic twins. The A parameter is often described as representing additive genetic effects because additive genetic effects would lead monozygotic twins to be more phenotypically similar than dizygotic twins. The description of the A parameter is

[2] Modern behavior genetic models typically rely on structural equation modeling approaches, rather than simple multiplication and subtraction.

Table 1 Seven developmental behavior genetic processes and implications for standard models

Process	Contributes to…	Evidence for process	Notes
(1) Phenotype growth	Means and variance	Mõttus et al. (2017)	Beyond controls for age, trajectories of means and variances are typically not considered in standard quantitative or molecular genetic approaches. For many phenotypes, however, age-trends are common. Means and variances may differ across cohorts or samples
(2) Independent genetic and environmental effects	ACE	Polderman et al. (2015) and Turkheimer and Waldron (2000)	This is a typical assumption in both classical quantitative genetic and molecular genetic approaches. In standard quantitative genetic designs, variation is decomposed into uncorrelated genetic, shared environmental, and nonshared environmental (typically including measurement error) variance components. In molecular genetic designs, genetic effects can be represented as SNP-effect sizes or as polygenic prediction, among other estimates, and the residual variance reflects a combination of unexplained genetic effects, estimation error (of the polygenic scores), measurement error, and environmental effects
(3) Non-random mating	A	D'Onofrio et al. (1999) and Yengo et al. (2018)	To the extent that partners positively assort, standard estimates of heritability in quantitative genetic studies may be downwardly biased. Positive assortment is more commonly reported than negative assortment. Negative assortment may produce upwardly biased estimates of heritability. Both types can limit generalizability of SNP associations. Extended family designs (e.g. Keller et al. 2010) may mitigate these biases, but themselves add on additional assumptions, such as continuity of genetic risk across age and cohorts
(4) Individual development	(In)stability of ACE	Harden et al. (2015) and Fraley et al. (2013)	Because developmental changes may occur as outcomes of individually-varying experiential or biological events that themselves vary to differing extents over time, genetic and environmental variance components may be stable or changing across time. When contexts are longstanding, or experiences recur, other processes (e.g. G×E) could accumulate across the lifespan (e.g., increasing estimates of heritability with age)

Table 1 (continued)

Process	Contributes to…	Evidence for process	Notes
(5) *r*GE	A or C	Kendler & Baker 2007 Krapohl & Plomin 2016	If genotypes are positively correlated with individually-varying environmental factors, initial genetically-influenced phenotypic differences may be magnified over time. Standard quantitative and molecular genetic models may attribute this source of variance to genetic factors. This pattern may result from active or evocative *r*GE (Plomin et al. 1977; Scarr and McCartney 1983). Passive *r*GE, whereby the environment magnifies phenotypic differences, may contribute to shared environmental variance in a quantitative design, but may contribute to the SNP effect size in a molecular genetic design. Environments that do *not* magnify initial phenotypic differences may lead to different implications. For example, compensatory processes (e.g., therapy) may reduce genetic effects in both standard quantitative and molecular genetic studies
(6) Gene × environment interaction	A or E	Tucker-Drob and Bates (2016)	In standard quantitative genetic designs, unmodeled interactions with the shared environment tend to produce genetic variance and interactions with the nonshared environment tend to produce nonshared environmental variance. In standard molecular genetic designs, interactions may alter SNP effect size estimates. The specific impact of unmodeled G × E on both quantitative and molecular standard models that do not estimate these effects is dependent on the form of the interaction and the distribution of the environmental variable in the sample
(7) Simultaneous GE-interplay	Dependent	Limited	Due to the complexity of identifying each of these processes simultaneously, there is limited understanding of how each process may function in the context of the others. The solution is likely dependent on the magnitudes of the other processes and their timing across the lifespan

A genetic effects, including SNP effect sizes, *C* shared environmental effects, *E* nonshared environmental effects

useful—it accurately describes a likely causal model that would lead monozygotic twins to be more similar than dizygotic twins. Yet, the description is not complete. As Purcell (2002) detailed, GE-interplay affects the A parameter. Therefore, the A parameter could be described as something much more general: any sort of causal process—not just an additive genetic one—that would lead monozygotic twins to be more phenotypically similar to each other than dizygotic twins are to each other.

Such distinctions may be more or less relevant for prediction compared to explanation. As an illustrative example, consider measures of personality, which are associated with academic achievement primarily through a genetic pathway (Krapohl et al. 2014; Tucker-Drob et al. 2016). How might GE-interplay need to be represented in this result? From a prediction perspective, the question may be irrelevant. Assuming a consistent environment in which the results of these studies hold and genetic markers of relevant

personality characteristics are available, early prediction of achievement would be possible.

However, if the goal is to explain our observations or to understand the way the world works, representing GE-interplay will often be important. For example, there may be small genetically influenced differences in personality which are detectable by parents, teachers, and peers, who in turn reinforce the personality characteristics in achievement-relevant ways. Teachers may be particularly attentive to children who sit still and pay attention, and therefore exacerbate potentially small early differences. Understanding the process requires recognizing that the environment may have played a larger role across development compared to genetic factors. It just so happens that genetic factors are correlated with environmental reinforcement.

In order to make predictions, however, we simply need to know the patterns of association. Presumably genetic markers possess beneficial qualities for prediction, such as being measurable early in the lifespan and being relatively simple (i.e., not requiring multiple measures of possible sources of environmental reinforcement unfolding across the lifespan). Yet, more can be gained by merging across these goals. For example, gaining control over the process would likely benefit from identifying a set of predictive markers, but also understanding the detailed causal mechanisms. Control does not need to focus on the genetic variant; a biological or pharmaceutical intervention may not be desired for numerous reasons. Instead, understanding the causal pathway points towards several intervention points, such as teacher–student interactions.

Disentangling multiple GE-interplay processes across development

One of the most complex empirical approaches to GE-interplay simultaneously estimates active or evocative gene–environment correlation (*r*GE) and quantitative Gene × Environment interaction (G × E) at a single point in time (van der Sluis et al. 2012).[3] An environment and a phenotype are measured once per family member. Heritability of the environment reflects *r*GE, and G × E is estimated by calculating the heritability of the phenotype at different levels of the environment. Although such studies represent substantial progress beyond conventional biometric variance decomposition approaches, they nevertheless represent static snapshots of developmental processes.

Developmental models are *underdetermined* relative to the data collected in behavior genetic studies (Earman 1993; Glymour 1970; Hausman et al. 2014; Norton 2008; Stanford 2001, 2006). In order to find useful models of phenotype development, the data must be able to identify each of the relevant parameters, some of which may dynamically change across the lifespan. Of course, practical limitations (e.g., economic resources, not to mention participant fatigue) impede our ability to conduct intensive studies to estimate each developmental parameter central to models of GE-interplay. Yet, the typical interpretation of empirical models of GE-interplay is premised on the idea that psychological outcomes emerge through slow, accumulating, developmental processes.

For example, Beam et al. (2015) hypothesized that small differences in early phenotypes between siblings will increasingly drift apart across development due to environmental reinforcement. If one sibling happens to have slightly higher cognitive abilities than the co-sibling, the difference between the siblings might be exacerbated by phenotype-matching behavior by teachers, peers, parents, and economics (i.e., job demands). The sibling differences typically observed in behavior genetic studies of adolescents or adults may not have always been present earlier in development, and the magnitude of the difference also may not be static across subsequent development. To make matters even more difficult, both *r*GE and G × E may take place simultaneously, and the effect of such interplay may wax and wane across the lifespan.

Theories of GE-interplay are highly complex and dynamic, making it nearly impossible, absent prohibitively intensive multivariate longitudinal data, to empirically determine what combinations of processes are responsible for observed phenotypic variation from among the universe of potential models. It may be the case that twin models are perfectly specified, and the only influences that matter for the development of phenotypes are additive and uncorrelated genetic, shared environmental, and nonshared environmental influences. If development works in this manner, then the interpretation of behavior genetic models would be much easier. To the extent that these sources of variance are correlated and non-additive, then the latent variance components and molecular genetic associations are potentially representative of these developmental processes (Purcell 2002). If development works in this manner, then the work of behavior geneticists requires thoughtful consideration of multiple plausible mechanisms that could lead genetically related individuals to resemble one another phenotypically. We describe seven processes that are relevant when interpreting behavior genetic results.

[3] In this manuscript, we focus on quantitative G × E, meaning the effect sizes of genes and environments are interdependent, rather than qualitative G × E, meaning different genes may operate across groups.

Phenotypic mean and variance shifts

Non-genetically informative studies indicate that phenotype means and variance shift across development, typically in tandem. Based on CDC growth charts (Kuczmarski et al. 2002), variance in human height increases by roughly 1200% from birth to adulthood which is also the time over which the mean increases. Mean levels of many psychological dimensions also change across development, such as cognitive ability (Tucker-Drob 2009), personality (Roberts et al. 2006), sensation seeking and delinquency (Harden et al. 2011), and internalizing disorders (Hankin et al. 2009). Concomitant with these mean-level changes, variance also increases for academic achievement (NWEA 2015), personality (Mõttus et al. in press, 2017), and psychopathology (Caspi et al. 2014).[4] Of course, mean-levels and variance do not increase continuously or uniformly across the lifespan (e.g., Mõttus et al. 2016; Tucker-Drob 2011). Individuals tend to grow and mature across childhood and adolescence, followed largely by stability of means and individual differences in adulthood and possible declines in old age. There may be some limit on development, at least relative to the prevailing environmental conditions.

Behavior genetic models are rarely interpreted with these sorts of lifespan trends in mind. When evaluating the heritability of a phenotype in adulthood, it may be useful to consider the processes that could lead both to mean-level changes in the population and also to increases in variance of the phenotype. For instance, individuals could all grow in the same direction, but at different rates. Or, fan-shaped longitudinal patterns may emerge when some individuals decrease, while more individuals increase or increase to a greater extent. Just as phenotypic variance estimates for a phenotype reflect intermediate states that have resulted from an ongoing developmental process, behavior genetic variance components reflect a developmental process. The common practice to standardize phenotypes at each wave (e.g., as Z-scores) or treat age as a simple covariate masks this rich information.

Independent genetic and environmental effects

Genetic and environmental differences may influence phenotypes in an additive and independent manner. By "independent," we mean the effects are not tied to GE-interplay and instead have relatively direct influences on the phenotype. This definition is admittedly murky as all growth requires both genetic and environmental factors to be present. Even so, we can imagine some allele that increases height by some constant amount in all individuals with the allele, in every environment that we observe in the real world. Similarly, there may be environmental experiences (e.g., getting struck by lightning or bitten by a radioactive spider) that have a constant impact on development for all individuals that experience it, and individuals do not select or evoke the environment on the basis of their characteristics. How many genetic and environmental effects are of this variety, compared to effects that are partially dependent on GE-interplay? Unfortunately, even though this question is a critically important one in behavior genetics and goes back to the earliest debates in the field (e.g., R. A. Fisher and L. T. Hogben, see Tabery 2014), the answer remains unknown.

Despite this gap in knowledge, several things can be said concerning the structure of genetic and environmental influences. Twin and family studies provide estimates of heritability and environmentality for a wide variety of phenotypes (Polderman et al. 2015). Heritability is nontrivial for essentially all phenotypes, and estimates are also not particularly close to 0 or 100% for most common human individual differences. Turning to molecular genetic information, between 1 and 15% of common genetic variants may play a causal role in height, cognitive ability, and personality, indicating substantial pleiotropy (Zeng et al. 2018). The magnitude of these statistical associations is known to be incredibly small (Chabris et al. 2015). The results of genome-wide association studies (GWASs) remain correlational in nature. The results are unable to satisfy our definition of "independent" given above. Quite the contrary, it is well-established that polygenic scores derived from GWAS are at least partly correlated with environmental processes (e.g., Koellinger and Harden 2018). On the environmental side, a comprehensive scan of the relevant environments has not been conducted, although it is likely that many relevant environments exist with small effect sizes (Turkheimer and Waldron 2000).

Non-random mating

Assortative mating refers to the observation that individuals do not produce offspring with partners having random characteristics. Rather, individuals tend to mate with others that share similar characteristics (e.g., D'Onofrio et al. 1999; Eaves et al. 1999). Assortative mating may occur due to individuals actively selecting a particular trait in a partner (e.g., educational attainment), selecting partners on the basis of a phenotype correlated with some other trait (e.g., selecting on educational attainment also selects on cognitive ability), or selecting partners based on social categories related to access (e.g., being willing to partner with individuals who differ from one's educational attainment, but being exposed

[4] Caspi et al. (2014) report longitudinal data on the psychometric structure of psychopathology across ages 18–38 years with approximately five waves for 11 disorders. Supplemental Table 1 reports means and standard deviations for each wave. When mean levels of psychopathology increase from one wave to the next, variance in psychopathology also increases ($r = 0.78$).

to individuals with similar educational attainment, termed social homogamy, which may occur in the presence of other selection processes, e.g., McGue et al. 1989). Partner similarity may also be observed due to partners influencing each other's behavior (e.g., sharing a partner's enthusiasm for education may motivate an individual to pursue further schooling).

Molecular genetic data has made it easier to distinguish among these possibilities. For example, at the phenotypic level assortative mating for educational attainment is moderate in size ($r \sim 0.4$), and partly due to genetic assortment (Domingue et al. 2014; Hugh-Jones et al. 2016). Results for other phenotypes are similar (Conley et al. 2016). One limitation of these studies is that they typically rely on polygenic scores, which are themselves estimated with considerable error. This error limits the precision of the estimate of genetic assortment and tends to push the estimates toward zero. Yengo et al. (2018) developed a technique to infer assortative mating from patterns found in the genome. They found significant assortative mating based on genotype for educational attainment and height, but not for 30 other phenotypes (potentially due to limited sample size of the underlying GWAS). The estimates from this approach matched those found when comparing genomes of actual couples.

The implications of these types of assortment differ for behavior genetic parameters. If positive assortment (i.e., partners have similar characteristics) occurs on the basis of genetically influenced factors, then partners will be more genetically similar than two random members of the population. Therefore, the expectation that the genetic correlation between dizygotic twins will be on average 0.5 is incorrect; the average will be shifted upwards, resulting in an underestimate of heritability. If negative assortment (i.e., partners have dissimilar characteristics) occurs, then the expectation would be reversed. However, across a wide array of phenotypes, positive assortment is more common (D'Onofrio et al. 1999; Eaves et al. 1999). Extensions of the classical twin design, such as extended family designs (Keller et al. 2010), can be used to estimate assortative mating, but these designs carry their own assumptions concerning the consistency of genetic effects across cohorts and ages. For molecular genetic studies, assortative mating could hinder confirmation of SNP associations. Specifically, tests of within-family associations will be less powerful to the extent that assortative mating exists in the population. Evidence of assortative mating from molecular genetic data has been documented for educational attainment, where estimates of within-family effects are systematically smaller than between-family effects (Lee et al. 2018), and sibling polygenic risk scores for educational attainment are more similar than would be expected by chance (sibling PRS $r \sim 0.55$, Belsky et al. 2018). Results found in a population with assortative mating will also be less likely to transfer to a population were assortative mating does not take place.

Individual developmental trajectories

Over the lifespan, effects may persist from birth (e.g., cesarean sections and immune function; Cho and Norman 2013) or activate in response to some transition (e.g., puberty and genetic effects on rule-breaking; Harden et al. 2015). Effects may fade in importance (e.g., divorce and well-being; Lucas 2007) or sustain their importance (e.g., parental support and social competence; Fraley et al. 2013). Effects may be static (i.e., stable across development), innovative (i.e., come "online" later in development) or decay (i.e., decrease in importance as time passes). Further, it is unlikely that phenotypic growth can continue constantly in any direction (Waddington 1942). Put differently, there may be some sort of reaction range in which a phenotype may change within an individual, but the range is somewhat limited (Turkheimer and Gottesman 1991).

Longitudinal behavior genetic studies have identified lifespan trends in estimates that coincide with phenotype growth. For example, personality and cognitive ability increase in test–retest stability, heritability, and genetic and environmental stability with age (Briley and Tucker-Drob 2014; Kandler and Papendick 2017; Tucker-Drob and Briley 2014). The specifics of these trends differ substantially, however, potentially pointing toward contrasting developmental processes (Briley and Tucker-Drob 2017). Nearly the entire increase in test–retest stability of cognitive ability is driven by increasingly stable genetic influences. In contrast, nearly the entire increase in test–retest stability of personality is driven by increasingly stable environmental influences. Translating these aggregate trends into actual causal effects, such as those outlined at the beginning of this section, requires causal reasoning that is more detailed in terms of mechanism and developmental specificity (i.e., how, when, and where effects occur) than abstract proportions of latent variance (see Tucker-Drob and Briley in press).

Similarly, Haworth et al. (2010) demonstrated that the heritability of cognitive ability increases relatively linearly across childhood and adolescence, which is one of the most well-replicated results in behavior genetics (Plomin et al. 2016). However, there are many potential interpretations of this finding. New genetic influences may turn on and explain new variance, either in response to some sort of intrinsic maturational process or in response to a novel environment (e.g., novel genetically influenced characteristics guide development after entry into school compared to prior). Alternatively, perhaps environmental influences decay and no longer impact cognition. Examining unstandardized shifts in variance components might help in this context.

GE-interplay could also result in increasing heritability for cognitive ability. We consider two plausible possibilities next.

Gene–Environment correlation

Following Plomin et al. (1977), we distinguish between passive *r*GE (parents pass on correlated genes and environments to their children), evocative *r*GE (individuals evoke a response from the environment based on genetically influenced characteristics), and active *r*GE (individuals select or create environments based on genetically influenced characteristics). Empirical evidence suggests that *r*GE is ubiquitous. Kendler and Baker (2007) estimated moderate heritabilities for a wide range of environmental factors, such as stressful life events, social support, peer relationships, and marital quality. Similarly, parenting behaviors are heritable, not only on the part of the parent, but also in response to genetically influenced characteristics of the child (Klahr and Burt 2014; Briley et al. 2014). These results from quantitative genetic studies are consistent with active or evocative *r*GE. Children may possess genetically influenced characteristics which their parents notice and respond to, or children may possess genetically influenced characteristics which lead to active influence on the parent. Passive *r*GE is somewhat easier to document with molecular genetic data. For example, Krapohl and Plomin (2016) found an association between family socioeconomic status and a polygenic score for educational attainment estimated for the child, both of which were also associated with actual academic achievement. Thus, parents provided an environment correlated with genetic material, and each of these factors are likely influential for child development (see also Krapohl et al. 2017). Passive *r*GE, in addition to assortative mating, may explain why within-family associations with molecular genetic data are weaker than between-family associations (Lee et al. 2018).

Strong theoretical models point to *r*GE as a central driver of development through the mechanism of selection into environments that match one's characteristics (e.g., Beam and Turkheimer 2013; Scarr and McCartney 1983). For example, *r*GE may be a likely candidate for the increases in heritability across age. Teachers might observe aspects of their students related to the ability to pick up material quickly and provide tailored instruction. Although this might be good pedagogical practice, a side effect is that initial differences may become magnified. Similarly, students may actively choose their educational experiences, whether that is paying attention in class, completing homework assignments, or studying for exams. Each of these behaviors likely has some sort of causal influence on learning. Again, the initial preferences for engaging in these behaviors might magnify differences across development, increasing estimates of heritability. In particular, the increase in heritability is driven by earlier genetic influences exerting a stronger impact on later cognitive ability (Briley and Tucker-Drob 2013). Importantly, these sorts of explanations for increasing heritability assume that the environment has a causal effect. If it did not, then a correlation between genes and environments would be found, but would not affect phenotype development.

In addition to *r*GE leading to the enhancement of one's characteristics, it is also possible for genetic variants and environments to act in opposite directions. For example, some individuals may possess genetic variants that increase levels of neuroticism, which is a risk factor for depression and anxiety disorders (Cuijpers et al. 2010). Individuals with relatively high levels of neuroticism may be more likely to seek psychotherapy, an environmental experience that appears effective at reducing levels of neuroticism (Roberts et al. 2017). This sort of *r*GE would *not* lead to increasing heritability estimates. Instead, the environmental experience may limit variance by moving extreme phenotypes toward the average. In some circumstances (i.e., if the intervention only acts on genetic sources of variance), heritability could be reduced. It is also plausible that the intervention is effective at the level of the phenotype, and therefore aggregate variance may be reduced without altering the genetic and environmental composition.

Gene × Environment interaction

Theoretical arguments in favor of the pervasive presence of G × E interaction are popular (Manuck and McCaffery 2014), although they are hindered by one limitation: there currently exists no single, confirmed, accepted-by-the-broader-scientific-community example of such an effect at the molecular level on any common human individual difference. The frequently used example, phenylketonuria, is relevant to only a tiny fraction of the population with a Mendelian genetic disorder. One major limitation in this area is that the expected magnitude of G × E effect sizes remains unknown, with nearly all studies likely having very low power to detect the effects. Identification of main effects via GWAS exploded after the first few successful studies allowed researchers to understand the likely magnitude of the effect sizes.

Evidence of G × E effects from twin and family studies (outside of developmental impacts on heritability) is limited, with few published attempts at direct replication (see Plomin et al. 2016, p. 4). It is challenging to know whether this is due to a preponderance of negative results, which are generally less likely to be published, or due to natural variation in areas of interest and data availability between research groups, leading to structural challenges in collaboration and replication. Where direct replication has been sought, results

appear inconsistently replicable across samples. This may be due to low statistical power, lack of true effects, or meaningful moderators. For example, the reported G × SES effects on cognitive ability emerge primarily in the US, but not other parts of the world, potentially due to differences in social services (Tucker-Drob and Bates 2016).

To the extent that such G × SES effects do impact development, heritability would be increased when the interaction effect is not modeled. This implication occurs due to the fact that monozygotic twins would respond similarly to a shared environment, but dizygotic twins would respond potentially differentially to the extent that genetic differences alter response to the environment. That process would magnify the difference between monozygotic and dizygotic twin similarity. Theoretically, genetically influenced characteristics may play a larger role in cognitive development in resource rich environments, and in contrast, such characteristics may not be able to play a role in development when there are considerable obstacles to development in low resource environments (Bronfenbrenner and Ceci 1994). It could be that this process early in life stratifies children's educational outcomes (e.g., achievement, but also motivation, values, and relationships with teachers). Then, the subsequent development of the child is guided by this early event. Or, it could be the case that the effect of such G × E slowly and incrementally accumulates across the lifespan.

Molecular genetic data, particularly candidate gene studies, have a poor replication record for G × E (Duncan and Keller 2011), although evaluation of the evidence is challenging in light of the quality and variety of available research (Duncan et al. 2014). Molecular G × E research has been hampered by incorrect methods (Keller 2014), but even large-sample investigations with appropriate statistical controls demonstrate conclusively that the specific variants selected for the focus of early candidate gene work are not consistently associated with outcomes of interest, either in terms of main or interaction effects (Chabris et al. 2012; Samek et al. 2016; Haberstick et al. 2014, 2016). It is unlikely that G × E effects should be expected to be much larger than main effects (and the distribution may, in fact, be considerably closer to zero). Very large samples, at least as large as main effects GWAS and likely much larger still due to lower power for interaction terms (e.g., $N >> 100{,}000$), will be necessary to detect interaction of specific genetic variant and environmental effects.

Simultaneous GE-interplay

As emphasized earlier, GE-interplay substantially complicates interpretation of behavior genetic parameters. It is sometimes stated that behavior genetic models assume that GE-interplay does not take place. We know this assumption is false. Therefore, the parameter estimates in standard models should be interpreted cautiously. To the extent that genetic influences are dependent on environmental context (G × E) or are systematically linked with environmental context (*r*GE), straightforward implications for development are difficult. For example, educational attainment is heritable. One interpretation is that genetic variants are associated with educational attainment in an invariant manner across all environments and all plausible environmental contexts. An equally plausible interpretation is that variants are associated only in certain environments (G × E), only when systematically exposed to certain environments (*r*GE), or are easily altered by some sort of intervention.

The impact of GE-interplay on behavior genetic parameter estimates can be reasoned out. As laid out by Purcell (2002) for twin studies, the typical interpretation states that correlation between genes and the shared environment results in shared environmental variance, and correlation between genes and the nonshared environment results in genetic variance. When genetic effects are dependent (i.e., G × E) on the shared environment (or vice versa), the result is genetic variance. When genetic effects are dependent on the nonshared environment (or vice versa), the result is nonshared environmental variance. However, these implications may not always be so clear. For example, if genes are correlated with nonshared environments that *decrease* the phenotype (e.g., neuroticism and psychotherapy), then the result would not be an increase in genetic variance. Similarly, there are likely many potential causal processes that impact phenotype development that may not fit into the standard interpretation. Reasoning out the implications of a certain process in isolation requires understanding the phenotype.

For molecular genetic studies, the implications are somewhat different. To the extent that genes and environments are correlated in a way that the environment magnifies the initial genetic difference, the genetic variant effect size is increased. For example, if a SNP predisposes individuals to start smoking, then that SNP will likely be associated with lung cancer due to the environmental impact of smoking (e.g., Thorgeirsson et al. 2008). Similarly, if a SNP predisposes parents to behave in a certain way toward their children and then this SNP is passed on to their children, then the SNP will index both the environmental pathway and any other genetic pathway that may occur. A similar pattern has been demonstrated in recent studies of non-transmitted alleles displaying a statistical association with child variables (Bates et al. 2018; Kong et al. 2018). Again, these implications assume that environments match and amplify genetic effects. To the extent that correlated environments mask genetic effects, SNP effect sizes may be decreased.

The influence of G × E on molecular genetic associations is dependent on the form that the interaction takes, and only in cases of pure cross-over interactions will main effects be completely obscured. In all other cases of statistical

interaction and with a large enough sample size, the main effect should be detectable even in the absence of knowledge of the moderating environment.

Multiple developmental processes could generate the empirical data observed in behavior genetic studies because GE-interplay processes may not be independent. Mechanisms like *r*GE and G×E may synergistically guide development (Tucker-Drob et al. 2013). Multiple forms of GE-interplay may lead to the increasing heritability of cognitive ability with age simultaneously. Stratified educational opportunity may limit opportunities for active *r*GE to occur, resulting in G×E. Although it is statistically straightforward to identify the implications of typical examples of *r*GE and G×E in isolation, we have little knowledge about the relative magnitude of each process, when in development these mechanisms exert a causal influence, or how the true cocktail of developmental inputs interact with one another. The developmental picture may be very complex. For example, Brant et al. (2013) found that the heritability of cognition increased more quickly for individuals with lower ability compared to high ability, with high ability adolescents similar to low ability children in terms of the magnitude of genetic influences. What combination of developmental inputs would produce such a potentially counter-intuitive finding?

Despite having some knowledge of the traces that *r*GE and G×E leave on behavior genetic estimates, it is not known how much each process contributes to any given estimate of heritability. In part, this gap occurs because each process is layered on top of the others, resulting in a complex developmental history of phenotype growth.

Complexity, compression, and the gloomy prospect

As a field, behavior genetics has produced substantial knowledge concerning replicable patterns of genetic and environmental influences across the lifespan (Plomin et al. 2016). Heritability is substantial (Turkheimer 2000), but each SNP explains a tiny portion of variance (Chabris et al. 2015). There is some evidence of GE-interplay, even if the empirical data to this point have not identified many replicable examples for G×E. Genetic and environmental effects shift across the lifespan as phenotypes become more stable. Although the statistical and interpretational implications of GE-interplay processes are well-known, the magnitude of each process is not well-known. Worse still, the factors that affect behavior genetic estimates all occur potentially simultaneously and continuously across development, and they may even interact with one another in a nonlinear and highly complex fashion. Researchers can increase the reasonableness of their inferences from behavior genetic models by gaining clarity on what is known and unknown concerning processes that influence parameter estimates. Ruling out potential processes can substantially shrink the number of possible interpretations.

Some basic questions remain difficult to address: what processes led to an estimate of 40% heritability? Was it additive and independent genetic effects, *r*GE reinforcing initial differences associated with genotype, or some form of G×E? Would heritability have been 40% if the sample was 10 years younger? Would heritability actually be 50% if assortative mating was correctly handled? Numerous papers have been written on the interpretive problems of heritability (e.g., Johnson et al. 2011; Keller et al. 2010; Turkheimer 1998). Our point here is not to retread this ground, but instead to point out the number of considerations required. Each of these considerations can be deconstructed in isolation to infer what the impact would be on behavior genetic models. The real world combines them all simultaneously in different quantities for each phenotype.

In the face of such taxing complexity, a framework with which to visualize the impact of different combinations of structural inputs would be useful. A successful model could generate phenotype levels from the ground up, starting with partners producing offspring with synthetic genomes and environments. One goal could be to identify what sets of model parameters can fill in the gaps identified in this review. As noted, there are likely several plausible sets of developmental parameters that could lead to the empirical results found in the literature. It might be the case that several potential models could produce similar observed trends, such as increasing heritability with age. We view this as a useful demonstration of the potential for equifinality in behavior genetic models, a limitation of the models that could be overlooked due to implicit assumptions about the data-generating mechanisms. A simulation approach would force these assumptions to be explicit and would allow them to be contrasted with other plausible assumptions.

In this context, we may think of phenotype development or the task of individual-level prediction as falling along a continuum of complexity. At one end is perfect simplicity: a change in an input leads to a change in the output every time, and researchers are able to make accurate predictions with easily obtainable and cognizable information. At the other end, it may be the case that there is such complexity that a description of development requires the full history of all variables at all points in time; the data stream is incapable of any compression. Under this scenario, the best anyone can do is record what happens. There is no more efficient way to express the observations, and the observations do not support any interesting predictions. Although behavior geneticists widely acknowledge that the phenotypes under study are *complex* (i.e., not having a single cause or simple set of causes), less consideration has been given to the potential

compressibility of the phenotypes across individuals relative to the set of available variables (e.g., Li and Vitányi 1997; Wallace and Freeman 1987). By "compression," we mean the ability to represent some large set of information in a more compact manner (Braddon-Mitchell 2001; Sayood 2005; Wheeler 2016). To what extent can behavior genetics move from thousands of genetic associations toward a cognizable and useful model of development (see Kendler 2008)? This type of question has emerged most clearly in the literature surrounding the "gloomy prospect."

The need to empirically evaluate the gloomy prospect

Under the limitations of empirical data collection, little behavior genetic research exists that explicitly considers the possibility of the gloomy prospect. Plomin and Daniels (1987, p. 8) described the gloomy prospect as a situation in which "the salient environment might be unsystematic, idiosyncratic, or serendipitous events," ultimately minimizing the possibility that much scientific progress can be made. Turkheimer and Gottesman (1996) used a simulation approach to illustrate the gloomy prospect; small shifts in environmental context completely removed all specific phenotype–environment associations. Turkheimer (2000, p. 163) applied the same gloomy outlook to molecular genetic associations in the real world due to the inherent complexity of development and noted that "the underlying complex causal processes would cause the apparent results [of molecular genetic studies] to be small, and to change unpredictably from one experiment to the next."

The gloomy prospect is discouraging from an empirical standpoint as it implies that the upper limit for scientific progress in predicting and explaining future behavior at the individual-level may already have been reached or be reached without substantially more meaningful progress. If phenotype development is driven by genetic effects that manifest differently across environments that are peculiar to a given individual, then identifying the effect that a genetic variant has on development will necessarily also be idiosyncratic. If true, the clinical utility of genetic or environmental information about individuals will be largely worthless, since a plethora of interdependent factors (many of which are inaccessible due to a failure of measurement over development) must be known before reasonable predictions can be made.

Gloominess falls on a continuum, and how gloomy the prospect of giving an informative behavior genetic account depends on the phenotype. For example, it may be that things are a bit gloomier for personality compared to cognitive ability or anthropometric traits (e.g., Cheesman et al. 2017). If there is no GE-interplay and no other potentially biasing factors, then molecular genetic associations will replicate and the prospects for giving an informative account is not gloomy at all. But if, on the other hand, GE-interplay is extremely large and the effects of any genetic variant are entirely dependent on the (potentially random) environmental context, then it is unlikely that any genetic effect will replicate. This situation would be maximally gloomy. However, most phenotypes likely fall somewhere between these extremes.

We suggest that a plausible starting point for identifying the "gloominess" of a phenotype is to investigate the seven developmental processes highlighted in this manuscript. Put differently, a greater understanding of phenotype processes (i.e., how the phenotype influences engagement with the environment), structure (i.e., how phenotypes covary), and development (i.e., how phenotypes respond to engagement with the environment in the context of other relevant phenotypes across the lifespan; see Baumert et al. 2017). Each of these questions can be addressed with behavior genetic methodology. For example, the field has established the genetic and environmental structure of many related phenotypes. We suggest that gains can be made in overcoming the gloomy prospect by better understanding our phenotypes, that is to say, gaining knowledge not only of genetic and environmental structure, but also of the processes that led to such a structure across developmental time. This work toward explanation is directly relevant to researchers interested primarily in prediction as the gloomy prospect may imply some upper limit on prediction. Evaluating simultaneous GE-interplay will be challenging, but such work could provide important insight into the mechanisms of phenotype growth.

Additionally, progress toward identifying the boundaries of the gloomy prospect could be made by drawing more heavily on animal models. Although the strength of animal models is typically seen as exerting control over environmental experiences, an increasing number of studies use designs in which GE-interplay is possible (Bell and Saltz 2017; Freund et al. 2013). For example, social niche construction refers to the tendency of certain organisms to form social groups partially based on genetic differences (i.e., *r*GE; Saltz and Foley 2011; Saltz and Nuzhdin 2014). This behavioral tendency has also been found to be context dependent (Saltz 2011) and influence development (Saltz 2013, 2014). More generally, animals exhibit repeatable behavioral syndromes (Bell et al. 2009; Sih et al. 2004), similar to human personality, and a host of tools are available to better explain and predict these patterns (Bengston et al. 2018). This work may be better situated to address major unanswered questions in human behavior genetics, such as potential sources of Gene × Environment interaction. Lee et al. (2018) found relatively few leads on why genetic associations with educational attainment might vary across contexts (although, see Tropf et al. 2017 for an analysis with

individual-level data), but the animal literature may offer further clues (see Saltz et al. 2018). Of course, evidence from animal models may be difficult to extrapolate to a phenotype like educational attainment, but the ability to track the effect of GE-interplay on development dynamically and consistently across the lifespan is a major advantage of animal models.

Communicating complexity

Given the complexity of development, disseminating results to researchers outside of behavior genetics or to the lay public is difficult due to perceptions about genetic influences that may not be warranted, such as strict genetic determinism (Dar-Nimrod and Heine 2011). In order to head off misinterpretations, some researchers have written publicly-accessible responses to frequently asked questions as accompaniments to major publications (e.g., Rietveld et al. 2013). In addition to answering frequently asked questions, researchers should describe competing mechanisms by which the observed data could have been generated, which would help other researchers and the public to better interpret the results. For example, we might find a non-trivial heritability of educational attainment (Lee et al. 2018). One possibility is that this heritability relates to fixed genetic effects. However, if the heritability of educational attainment were entirely filtered through environmentally mediated processes that were quite distant from the genome (e.g., via reinforcement due to parental cognitive stimulation; Tucker-Drob and Harden 2012) and effectively modifiable by intervention (e.g., universal preschool; Tucker-Drob 2012), then different implications would be drawn. This descriptive example could be an effective tool for communicating with the general public that high (or non-zero) heritability estimates do not imply an absence of environmental processes. A better understanding of phenotype inputs across development could aid in distinguishing these potential mechanisms.

Conclusion

Behavior genetics has a public communication problem, partly due to the disconnect between our simplistically presented models and the long list of required caveats and assumptions. Common intuitions concerning genetics (e.g., Dar-Nimrod and Heine 2011) likely lead to shortcuts about the causal relation between some genetic variant and an outcome. For example, a person may intuitively believe that an association between a SNP and educational attainment is deterministic and not sensitive to any sort of environmental input. Vague statements about GE-interplay and other caveats of the models (e.g., behavior genetic models reflect what is, not what could be) may not be maximally effective. Identifying and interrogating the most likely and robust models that are plausibly involved in human development may allow for a more nuanced discussion of phenotypes, both among researchers and between researchers and the general public. Further, we suggest that discussion among researchers could benefit from clarifying the motivation of scientific inquiry, whether aiming at understanding nature, predicting individual-level outcomes, or gaining control over development. Some of the complexity we highlight may be particularly relevant to researchers aiming at understanding, rather than prediction or control. Unifying across these goals may improve behavior genetic theory and utility.

In this report we have laid out a set of known empirical behavior genetic results, and at the same time, the interpretive ambiguity that accompanies these results. Although the implications of GE-interplay for various analytic models are relatively straightforward (Purcell 2002), a difficulty emerges when applied to development, where multiple interdependent inputs exert pressure on phenotypes. As genetically-informative models of development move toward specifying small-scale, mechanistic inputs (Briley et al. 2018; Nivard and Boomsma 2016; Tucker-Drob and Briley in press) in addition to broad-scale inputs (e.g., Plomin et al. 1977; Scarr and McCartney 1983), we encourage nuanced thinking concerning the causal chain that leads to estimates of heritability and molecular genetic associations. Such considerations may lead to different solutions to causal reasoning problems or judgments of human agency (see Lynch 2017).

All phenotypes of interest to behavior geneticists are *complex* by one measure or another. The question becomes *how* complex, whether doomed to a true gloomy prospect, whereby the intractably complex developmental processes that lead to the outcome of interest are so unique as to be essentially ungeneralizable beyond a single individual, or rather may be placed along a spectrum of relative gloominess. That is, relative complexity is bound to vary, with some phenotypes being simpler to disentangle than others. Existing empirical trends provide a tool to narrow down the likely candidates, from a universe of nearly infinite possibilities.

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

Compliance with ethical standards

Conflict of interest Daniel A. Briley, Jonathan Livengood, Jaime Derringer, Elliot M. Tucker-Drob, R. Chris Fraley, and Brent W. Roberts-declare that they have no conflict of interest.

Human and animal rights and informed consent This article does not contain any studies with human participants or animals performed by any of the authors.

References

Bates TC, Maher BS, Medland SE, McAloney K, Wright M, Hansell NK et al (2018) The nature of nurture: using a virtual-parent design to test parenting effects on children's educational attainment in genotyped families. Twin Res 21:73–83

Baumert A, Schmitt M, Perugini M, Johnson W, Blum G et al (2017) Integrating personality structure, personality process, and personality development. Eur J Personal 31:503–528

Beam CR, Turkheimer E (2013) Phenotype–environment correlations in longitudinal twin models. Dev Psychopathol 25:7–16

Beam CR, Turkheimer E, Dickens WT, Davis DW (2015) Twin differentiation of cognitive ability through phenotype to environment transmission: the Louisville Twin Study. Behav Genet 45:622–634

Bell AM, Saltz JB (2017) Profitable interplay between the study of human and nonhuman personality. Eur J Personal 31:532–533

Bell AM, Hankison SJ, Laskowski KL (2009) The repeatability of behaviour: a meta-analysis. Anim Behav 31:532–533

Belsky DW, Domingue BW, Wedow R, Arseneault L, Boardman JD, Caspi A et al (2018) Genetic analysis of social-class mobility in five longitudinal studies. Proc Natl Acad Sci USA. https://doi.org/10.1073/pnas.1801238115

Bengston SE, Dahan RA, Donaldson Z, Phelps SM, van Oers K, Sih A, Bell AM (2018) Genomic tools for behavioural ecologists to understand repeatable individual differences in behaviour. Nat Ecol Evol. https://doi.org/10.1038/s41559-017-0411-4

Braddon-Mitchell D (2001) Lossy laws. Noûs 35:260–277

Brant AM, Munakata Y, Boomsma DI, DeFries JC, Haworth CMA, Keller MC et al (2013) The nature and nurture of high IQ: an extended sensitive period for intellectual development. Psychol Sci 24:1487–1495

Briley DA, Tucker-Drob EM (2013) Explaining the increasing heritability of cognitive ability across development: a meta-analysis of longitudinal twin and adoption studies. Psychol Sci 24:1704–1713

Briley DA, Tucker-Drob EM (2014) Genetic and environmental continuity in personality development: a meta-analysis. Psychol Bull 140:1303–1331

Briley DA, Tucker-Drob EM (2017) Comparing the developmental genetics of cognition and personality over the lifespan. J Personal 85:51–64

Briley DA, Harden KP, Tucker-Drob EM (2014) Child characteristics and parental educational expectations: evidence for transmission with transaction. Dev Psychol 50:2614–2632

Briley DA, Livengood J, Derringer J (2018) Behavior genetic frameworks of causal reasoning for personality psychology. Eur J Personal 32:202–220

Bronfenbenner U, Ceci SJ (1994) Nature-nurture reconceptualized in developmental perspective: a bioecological model. Psychol Rev 101:568–586

Caspi A, Houts RM, Belsky DW, Goldman-Mellor SJ, Harrington H, Israel S et al (2014) The p factor: one general psychopathology factor in the structure of psychiatric disorders? Clin Psychol Sci 2:119–137

Chabris CF, Hebert BM, Benjamin DJ, Beauchamp J, Cesarini D, Van der Loos M et al (2012) Most reported genetic associations with general intelligence are probably false positives. Psychol Sci 23:1314–1323

Chabris CF, Lee JJ, Cesarini D, Benjamin DJ, Laibson DI (2015) The fourth law of behavior genetics. Curr Dir Psychol Sci 24:304–312

Cheesman R, Selzam S, Ronald A, Dale PS, McAdams TA, Eley TC, Plomin R (2017) Childhood behavior problems show the greatest gap between DNA-based and twin heritability. Transl Psychiatry 7:1284

Cho CE, Norman M (2013) Cesarean section and development of the immune system in the offspring. Am J Obstet Gynecol 208:249–254

Conley D, Laidley T, Belsky DW, Fletcher JM, Boardman JD, Domingue BW (2016) Assortative mating and differential fertility by phenotype and genotype across the 20th century. Proc Natl Acad Sci USA 113:6647–6652

Cuijpers P, Smit F, Penninx BWJH, de Graaf R, ten Have M, Beekman ATF (2010) Economic costs of neuroticism: a population-based study. Arch Gen Psychiatry 67:1086–1093

D'Onofrio BM, Eaves LJ, Murelle L, Maes HH, Spilka B (1999) Understanding biological and social influences on religious affiliation, attitudes, and behaviors: a behavior genetic perspective. J Personal 67:953–984

Dar-Nimrod I, Heine SJ (2011) Genetic essentialism: on the deceptive determinism of DNA. Psychol Bull 137:800–818

De Regt HW (2015) Scientific understanding: truth or dare? Synthese 192:3781–3797

Dickens WT, Flynn JR (2001) Heritability estimates versus large environmental effects: the IQ paradox resolved. Psychol Rev 108:349–369

Domingue BW, Fletcher J, Conley D, Boardman JD (2014) Genetic and educational assortative mating among US adults. Proc Natl Acad Sci USA 111:7996–8000

Duncan LE, Keller MC (2011) A critical review of the first 10 years of candidate Gene-by-Environment Interaction research in psychiatry. Am J Psychiatry 168:1041–1049

Duncan LE, Pollastri AR, Smoller JW (2014) Mind the gap: why many geneticists and psychological scientists have discrepant views about gene–environment interaction (G × E) research. Am Psychol 69:249–268

Earman J (1993) Underdetermination, realism, and reason. Midwest Studies Philos 18:19–38

Eaves L, Heath A, Martin N, Maes H, Neale M et al (1999) Comparing the biological and cultural inheritance of personality and social attitudes in the Virginia 30,000 study of twins and their relatives. Twin Res 2:62–80

Falconer DS (1960) An introduction to quantitative genetics. Ronald, New York

Forster M (2002) Predictive accuracy as an achievable goal of science. Philos Sci 69:S124–S134

Forster M, Sober E (1994) How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions. Br J Philos Sci 45:1–35

Fraley RC, Roisman GI, Haltigan JD (2013) The legacy of early experiences in development: formalizing alternative models of how early experiences are carried forward over time. Dev Psychol 49:109–126

Freund J, Brandmaier AM, Lewejohann L, Kirste I, Kritzler M et al (2013) Emergence of individuality in genetically identical mice. Science 340:756–759

Glymour C (1970) Theoretical equivalence and the semantic view of theories. Philos Sci 80:286–297

Haberstick BC, Lessem JM, Hewitt JK, Smolen A, Hopfer CJ, Halpern CT et al (2014) MAOA genotype, childhood maltreatment, and their interaction in the etiology of adult antisocial behaviors. Biol Psychiatry 75:25–30

Haberstick BC, Boardman JD, Wagner B, Smolen A, Hewitt JK, Killeya-Jones LA et al (2016) Depression, stressful life events, and the impact of variation in the serotonin transporter: findings from the National Longitudinal Study of Adolescent to Adult Health (Add Health). PLoS ONE 11:e0148373

Hankin BL, Oppenheimer C, Jenness J, Barrocas A, Shapero BG, Goldband J (2009) Developmental origins of cognitive vulnerabilities to depression: review of processes contributing to stability and change across time. J Clin Psychol 65:1327–1338

Harden KP, Quinn PD, Tucker-Drob EM (2011) Genetically influenced change in sensation seeking drives the rise of delinquent behavior during adolescence. Dev Sci 15:150–163
Harden KP, Patterson MW, Briley DA, Engelhardt LE, Kretsch N et al (2015) Developmental changes in genetic and environmental influences on rule-breaking and aggression: age and pubertal development. J Child Psychol Psychiatry 56:1370–1379
Hausman DM, Stern R, Weinberger N (2014) Systems without a graphical causal representation. Synthese 19:1925–1930
Haworth CMA, Wright MJ, Luciano M, Martin NG, de Geus EJC, van Beijsterveldt CEM et al (2010) The heritability of general cognitive ability increases linearly from childhood to young adulthood. Mol Psychiatry 15:1112–1120
Hugh-Jones D, Verweij KJH, Pourcain B, Abdellaoui A (2016) Assortative mating on educational attainment leads to genetic spousal resemblance for polygenic scores. Intelligence 59:103–108
Johnson W (2007) Genetic and environmental influences on behavior: capturing all the interplay. Psychol Rev 114:423–440
Johnson W, Penke L, Spinath FM (2011) Heritability in the era of molecular genetics: some thoughts for understanding genetic influences on behavioural traits. Eur J Personal 25:254–266
Kandler C, Papendick M (2017) Behavior genetics and personality development: a methodological and meta-analytic review. In: Specht J (ed) Personality development across the lifespan. Academic, Cambridge, p 473–495
Kandler C, Zapko-Willmes A (2017) Theoretical perspectives on the interplay of nature and nurture in personality development. In: Specht J (ed) Personality development across the lifespan. Academic, Cambridge, p 101–115
Keller EF (1985) Reflections on gender and science. Yale University Press, New Haven
Keller MC (2014) Gene × environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. Biol Psychiatry 75:18–24
Keller MC, Medland SE, Duncan LE (2010) Are extended twin family designs worth the trouble? A comparison of the bias, precision, and accuracy of parameters estimated in four twin family models. Behav Genet 40:377–393
Kelly K, Glymour C (2004) Why probability does not capture the logic of scientific justification. In: Hitchcock C (ed) Contemporary debates in the philosophy of science. Blackwell, Oxford
Kendler KS (2008) Explanatory models for psychiatric illness. Am J Psychiatry 165:695–702
Kendler KS, Baker JH (2007) Genetic influences on measures of the environment: a systematic review. Psychol Med 37:615–626
Klahr AM, Burt SA (2014) Elucidating the etiology of individual differences in parenting: a meta-analysis of behavioral genetic research. Psychol Bull 140:544–586
Koellinger PD, Harden KP (2018) Using nature to understand nurture. Science 359:386–387
Kong A, Thorleifsson G, Frigge ML, Vilhjalmsson BJ, Young AI, Thorgeirsson TE, et al (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. Mol Psychiatry 21:437–443
Krapohl E, Rimfeld K, Shakeshaft NG, Trzaskowski M, McMillan A, Pingault J-B et al (2014) The high heritability of educational achievement reflects many genetically influences traits, not just intelligence. Proc Natl Acad Sci USA 42:15273–15278
Krapohl E, Hannigan LJ, Pingault J-B, Patel H, Kadeva N, Curtis C et al (2017) Widespread covariation of early environmental exposures and trait-associated polygenic variation. Proc Natl Acad Sci USA 114:11727–11732
Kuczmarski RJ, Ogden CL, Guo SS, Grummer-Strawn LM, Flegal KM, Mei Z et al (2002) 2000 CDC growth charts for the United States: methods and development. DHHS Publication No. (PHS) 2002-1696
Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M et al (2018) Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet 50:1112–1121
Li M, Vitányi P (1997) An introduction to Kolmogorov complexity and its applications, 2nd edn. Springer, New York
Lucas RE (2007) Adaptation and the set-point model of subjective well-being: does happiness change after major life events? Curr Dir Psychol Sci 16:75–79
Lynch KE (2017) Heritability and causal reasoning. Biol Philos 32:25–49
Manuck SB, McCaffery JM (2014) Gene–environment interaction. Annu Rev Psychol 65:41–70
McGue M, Wette R, Rao DC (1989) Path analysis under generalized marital resemblance: evaluation of the assumptions underlying the mixed homogamy model by the Monte Carlo method. Genet Epidemiol 6:373–388
Mõttus R, Allik J, Hřebíčková M, Kööts-Ausmees L, Realo A (2016) Age differences in the variance of personality characteristics. Eur J Personal 30:4–11
Mõttus R, Soto CJ, Slobodskaya HR (2017) Are all kids alike? The magnitude of individual differences in personality characteristics tends to increase from early childhood to early adolescence. Eur J Personal 31:313–328
Mõttus R, Briley DA, Zheng A, Mann FD, Engelhardt LE, Tackett JL et al. Kids becoming less alike: a behavioral genetic analysis of developmental increases in personality variance from childhood to adolescence. J Personal Soc Psychol. https://www.researchgate.net/publication/323112699_Kids_becoming_less_alike_A_behavioral_genetic_analysis_of_developmental_increases_in_personality_variance_from_childhood_to_adolescence?enrichId=rgreq-6517f0bb889b3e63a2cf940f156fd788-XXX&enrichSource=Y292ZXJQYWdlOzMyMzExMjY5OTtBUzo2MDQyMjA4Njg4MTY4OTZAMTUyMTA2ODg4OTAwNA%3D%3D&el=1_x_2&_esc=publicationCoverPdf
Nagel E (1967) The nature and aim of science. In: S Morgenbesser (ed) Philosophy of science today. Basic Books, New York
Nivard MG, Boomsma DI (2016) Genetics: from molecule to society. Curr Biol 26:R1194–R1196
Norton J (2008) Must evidence underdetermine theory? In: Carrier M, Howard D, Kourany J (eds) The challenge of the social and the pressure of practice: science and values revisited. University of Pittsburgh Press, Pittsburgh, p 17–44
NWEA (2015) 2015 NWEA MAP growth normative data. https://www.nwea.org/content/uploads/2015/11/Normative-Data-2015.pdf
Plomin R, Daniels D (1987) Why are children from the same family so different from one another? Behav Brain Sci 10:1–60
Plomin R, DeFries JC, Loehlin JC (1977) Genotype–environment interaction and correlation in the analysis of human behavior. Psychol Bull 84:309–322
Plomin R, DeFries JC, Knopik VS, Neiderhiser JM (2016) Top 10 replicated findings from behavioral genetics. Perspect Psychol Sci 11:3–23
Polderman TJC, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, Posthuma D (2015) Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nat Genet 47:702–709
Popper K (1957) The aim of science. Ratio 1:24–35
Purcell S (2002) Variance components models for gene–environment interaction in twin analysis. Twin Res Hum Genet 5:554–571
Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW et al (2013) GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. Science 340:1467–1471

Roberts BW, Walton KE, Viechtbauer W (2006) Patterns of mean-level change in personality traits across the life course: a meta-analysis of longitudinal studies. Psychol Bull 132:1–25

Roberts BW, Luo J, Briley DA, Chow PI, Su R, Hill PL (2017) A systematic review of personality trait change through intervention. Psychol Bull 143:117–141

Saltz JB (2011) Natural genetic variation in social environment choice: context-dependent gene–environment correlation in *Drosophila melanogaster*. Evolution 8:2325–2334

Saltz JB (2013) Genetic composition of social groups influence male aggressive behaviour and fitness in natural genotypes of *Drosophila melanogaster*. Proc R Soc Lond B 280:20131926

Saltz JB (2014) Genetic variation in social environment construction influences the development of aggressive behavior in *Drosophila melanogaster*. Heredity 118:340–347

Saltz JB, Foley BR (2011) Natural genetic variation in social niche construction: social effects of aggression drive disruptive sexual selection in *Drosophila melanogaster*. Am Nat 177:645–654

Saltz JB, Nuzhdin S (2014) Genetic variation in niche construction: implications for development and evolutionary genetics. Trends Ecol Evol 29:8–14

Saltz JB, Bell AM, Flint J, Gomulkiewicz R, Hughes KA, Keagy J (2018) Why does the magnitude of genotype-by-environment interaction vary? Ecol Evol 8:6342–6353

Samek DR, Bailey J, Hill KG, Wilson S, Lee S, Keyes MA et al (2016) A test-replicate approach to candidate gene research on addiction and externalizing disorders: a collaboration across five longitudinal studies. Behav Genet 46:608–626

Sayood K (2005) Introduction to data compression. Elsevier, San Francisco

Scarr S, McCartney K (1983) How people make their own environments: a theory of genotype → environment effects. Child Dev 54:424–435

Schmitt D, Realo A, Voracek M, Allik J (2008) Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures. J Personal Soc Psychol 94:168–182

Sih A, Bell A, Johnson JC (2004) Behavioral syndromes: an ecological and evolutionary overview. Trends Ecol Evol 19:372–378

Stanford PK (2001) Refusing the devil's bargain: what kind of underdetermination should we take seriously? Philos Sci 68:S1–S12

Stanford PK (2006) Exceeding our grasp: science, history, and the problem of unconceived alternatives. Oxford University Press, New York

Tabery J (2014) Beyond versus: the struggle to understand the interaction of nature and nurture. MIT Press, Boston

Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP et al (2008) A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature 452:638–642

Tropf FC, Lee SH, Verwij RM, Stulp G, van der Most PJ et al (2017) Hidden heritability due to heterogeneity across seven populations. Nat Hum Behav 1:757–765

Tucker-Drob EM (2009) Differentiation of cognitive abilities across the life span. Dev Psychol 45:1097–1118

Tucker-Drob EM (2011) Global and domain-specific changes in cognition throughout adulthood. Dev Psychol 47:331–343

Tucker-Drob EM (2012) Preschools reduce early academic achievement gaps: a longitudinal twin approach. Psychol Sci 23:310–319

Tucker-Drob EM (2017) Motivational factors as mechanisms of gene–environment transactions in cognitive development and academic achievement. In: Elliot A, Dweck C, Yeager D (eds) Handbook of competence and motivation, 2nd edn. Guilford Press, New York

Tucker-Drob EM, Bates TC (2016) Large cross-national differences in Gene × Socioeconomic Status interaction on intelligence. Psychol Sci 27:138–149

Tucker-Drob EM, Briley DA (2014) Continuity of genetic and environmental influences on cognition across the life span: a meta-analysis of longitudinal twin and adoption studies. Psychol Bull 140:949–979

Tucker-Drob EM, Briley DA (in press) Theoretical concepts in the genetics of personality development. In McAdams D, Shiner R, Tackett JL (eds), The Handbook of Personality Development

Tucker-Drob EM, Harden KP (2012) Early childhood cognitive development and parental cognitive stimulation: evidence for reciprocal gene–environment transactions. Dev Sci 15:250–259

Tucker-Drob EM, Briley DA, Harden KP (2013) Genetic and environmental influences on cognition across development and context. Curr Dir Psychol Sci 22:349–355

Tucker-Drob EM, Briley DA, Engelhardt LE, Mann FD, Harden KP (2016) Genetically-mediated associations between measures of childhood character and academic achievement. J Personal Soc Psychol 111:790–815

Turkheimer E (1998) Heritability and biological explanation. Psychol Rev 105:782–791

Turkheimer E (2000) Three laws of behavior genetics and what they mean. Curr Dir Psychol Sci 9:160–164

Turkheimer E, Gottesman II (1991) Individual differences and the canalization of human behavior. Dev Psychol 27:18–22

Turkheimer E, Gottesman II (1996) Simulating the dynamics of genes and environment in development. Dev Psychopathol 8:667–677

Turkheimer E, Waldron M (2000) Nonshared environment: a theoretical, methodological, and quantitative review. Psychol Bull 126:78–108

van Fraassen BC (1980) The scientific image. Oxford University Press, Oxford

van Fraassen BC (1986) Aim and structure of scientific theories. Stud Logic Found Math 144:307–318

van der Sluis S, Posthuma D, Dolan CVD (2012) A note on false positives and power in G × E modelling of twin data. Behav Genet 42:170–186

Waddington CH (1942) Canalization of development and the inheritance of acquired characters. Nature 150:563–565

Wallace C, Freeman P (1987) Estimation and inference by compact coding. J R Stat Soc B 49:240–251

Wheeler B (2016) Simplicity, language-dependency and the best system account of laws. Theoria 31:189–206

Yengo L, Robinson MR, Keller MC, Kemper KE, Yang Y, Trzaskowski M et al (2018) Imprint of assortative mating on the human genome. Preprint posted on bioRxivorg. https://doi.org/10.1101/300020

Zeng J, de Vlaming R, Wu Y, Robinson M, Lloyd-Jones L, Yengo L et al (2018) Signatures of negative selection in the genetic architecture of human complex traits. Nat Genet 50:746–753