Joint consideration of means and variances might change our understanding of etiology

Joint consideration of means and variances might change our understanding of etiology

S. Alexandra Burt[1] and Wendy Johnson[2]

[1]Department of Psychology, Michigan State University, USA, burts@msu.edu

[2]Department of Psychology, University of Edinburgh, UK, wendy.johnson@ed.ac.uk

Abstract

Twin and adoption studies compare the similarities of people with differing degrees of relatedness to estimate genetic and environmental contributions to trait population variance. The analytic workhorse of these kinds of variance-focused designs is the intraclass correlation, which estimates similarity between pairs of individuals. Group means, by contrast, play no overt role in estimating genetic and environmental influences. Although this focus on variance has made very important contributions to understanding psychological characteristics, we contend that the exclusion of mean effects from behavioral genetic designs may have obscured key environmental influences and impeded full appreciation of the ubiquity and nature of gene-environment interplay in human outcomes. We provide empirical examples already in the literature, as well as a theoretical framework for thinking through the incorporation of mean effects using largely forgotten, non-Mendelian theory regarding how genes influence human outcomes. We conclude that the field needs to develop models capable of fully incorporating mean effects into twin and adoption studies.

Key words: means; variances, environmental influences; genetic influences

Psychologists seek to understand the origins and development of human behavior, addressing the question "*Why do people behave/learn/emote as they do*?". They examine specific behaviors and mental activities ranging from emotional regulation to spatial reasoning, often in socially discernible lifespan periods such as infancy, adolescence, and 'late life'. In the process, psychologists working in different content areas tend to ask somewhat different kinds of questions and apply different study designs and statistical techniques, almost always using either observational/correlational approaches or experimental approaches but not both. This is a reasonable approach to science if researchers in different topic areas ultimately collaborate to piece together their various close-up views of the human mind as a whole, but all too often, they remain too busy examining further details of their own 'bits' to benefit from or inform the work of others.

Behavioral genetics is one such focal area. What distinguishes its approach from most other areas of psychology is explicit recognition that genetic influences need to be directly considered in approaching psychological questions. This has been easier said than done, however, both because for many years it was not possible to observe the genome directly and because genotype-environment interplay appears to permeate nearly every aspect of life, even those seemingly "environmental". In non-human animal research, genetic involvement has been addressed in many ways, including overt breeding, cross-fostering, cloning, and knockout experiments. Human experiments of this sort are obviously unethical, but it was recognized long ago (Burks, 1928; Galton, 1875) that so-called "experiments of nature" could provide information about the contributions of genetic and environmental influences to naturally-occurring human behavioral differences.

These classical natural experiments compare similarities of people with differing degrees of relatedness (genetic and/or environmental) to estimate genetic and environmental contributions to the variance in given characteristics. Twin pairs have been the relationship of choice for much of this work because they offer a particularly sharp genetic relatedness contrast coupled with common age and prenatal experiences, as well as similar pair-by-pair rearing situations in most cases. When monozygotic co-twins are more similar to each other on a focal characteristic than are dizygotic co-twins, we infer the presence of genetic influence on the characteristic's observed sample variance. Larger differences in similarity with zygosity imply larger levels of genetic involvement of some kind (direct or via genotype-environment interplay). Similar inferences about population-level genetic variance can be drawn when adopted offspring are shown to be more similar to biological relatives (e.g., parents, siblings raised by the biological parents) than to adoptive relatives.

Like all studies, twin and adoption studies rely on critical underlying assumptions, violations of which can distort study estimates (Falconer, 1960; Keller & Medland, 2008; Keller, Medland, & Duncan, 2010), and doubters of genetic influences on human behavior continue to criticize them on this basis (C. H. Burt & Simons, 2014). Fortunately, in the past 20 years or so, behavioral geneticists have also developed/ leveraged technology enabling the tabulation of unrelated individuals' molecular genetic variants (Visscher, Brown, McCarthy, & Yang, 2012; Yang, Lee, Goddard, & Visscher, 2011). These molecular approaches have been used to successfully quantify associations between variants among people's genomes and their psychological characteristics. In doing so, they have corroborated many conclusions from twin and adoption studies, and in particular the conclusion that polygenetic influences make substantial contributions to behavioral and psychological characteristics.

Twin and adoption designs thus allow researchers to simultaneously investigate genetic and environmental influences, and in doing so, offer key population-level information on how people behave/learn/emote they do. Conceptually, the analytic workhorse of these kinds of variance-focused twin-family studies is the intraclass correlation (ICC), which indexes similarity between paired entities such as twins (see Table 1). Like the interclass (Pearson) correlation, the ICC is a covariance standardized on its respective variances that quantifies the degree to which mean deviations of the members of pairs (groups) within data co-vary systematically. The core observation that makes it possible to distinguish genetic from environmental variance in behavioral genetic studies is thus the relative extent of *similarity in individual differences* around a variable's mean among pairs of individuals with specific amounts of genetic relatedness (monozygotic and dizygotic twins; full and half-siblings and cousins, etc.), with means themselves all but ignored. To estimate genetic variance, the basic formula is: A ('additive' genetic variance, assuming each genetic variant acts independently – itself a large and questionable assumption) = (*r*MZ-*r*DZ)*2, where *r* represents ICC (although in practice variance components are nearly always estimated using structural equation models of twin covariances). That is, we compute the similarity of MZ twins and that of DZ twins, respectively, and multiply the difference between their respective similarities by two (since DZ twins share an average of 50% of their segregating genes, whereas MZ twins share 100%). Analogous formulae reflecting their proportions of shared segregating genes are used with other pairs of relatives.

Group means thus play no overt role in traditional (or 'simple') behavioral genetic interpretations (see Table 1), which are instead focused on variances. Instead, means are usually presented as simple descriptive statistics to check the sample's ability to represent its underlying population, and/or included in structural equation models to ensure identification. This is

unfortunate, since means and variances index different statistical properties of variables' distributions, each with their own implications for interpretation. What's more, means and variances are statistically independent of one another when data are normally distributed (DeGroot, 2012).

The one exception to this lack of consideration of means in modern behavior genetics is seen in genotype-environment interaction (GxE) models (e.g., Purcell, 2002), which test whether one variable moderates another's variance components. Even here, however, researchers model means only to regress any covariance between the moderator and the outcome from the outcome. This is done to circumvent genotype-environment correlational confounds, in the same way that, for example, age or sex are removed from consideration in regressions by placing them in equations before the focal variables and examining the significance of change-in-F statistics. As a consequence, the moderator's influence in these GxE models is thus estimated exclusively on the variance unique to the outcome variance. The core interpretive focus is still centered on the variance, with means effectively ignored. This treatment "throws the baby out with the bathwater" in many respects. That is, what typically motivates people to apply GxE moderation models is awareness of association between the two variables, such that mean levels of the variable termed 'DV' vary systematically from one moderator interval level to the next. Whatever drives this is inevitably bound up in the two variables' covariance – with DV mean levels that vary systematically from one IV interval to the next – which the model explicitly regresses away to streamline interpretation of the interaction term.

In sum, twin models (both simple and GxE) are focused almost exclusively on understanding the origins of variance around the mean, with next to no attention given to the mean itself. We argue below that this exclusion of mean effects from behavioral genetic

interpretations may have obscured key environmental influences and impeded full appreciation of the ubiquity and nature of gene-environment interplay in human outcomes. We begin by providing concrete examples in the literature. We then discuss prior work attempting to reconcile considerations of means and variance. We close by providing a broader framework for conceptualizing the different implications of means and variance for our understanding of genetic and environmental influences.

How our statistical focus constrains our causal inferences: Some concrete examples

To make abstract topics less esoteric, concrete examples are often helpful. There are many from which to choose. As discussed in Burt et al. (2019), biological males engage in significantly more aggression and antisocial behavior than do biological females across the lifespan, with typical Cohen's *d* effect sizes ranging from .4 to .8 (Archer, 2004). This male preponderance persists across numerous human societies and across most mammalian species, including humans' nearest phylogenetic cousins, the chimpanzee and much more peaceful bonobo (Archer, 2004). Despite this, variance decompositions in twin-family studies have nearly all indicated that 1) the same genetic and environmental influences underlie antisocial behavior in males and females, and 2) these genetic and environmental influences are equally influential in males and females (S. A. Burt, Slawinski, et al., 2019). Findings like these pose a conundrum: how do we make sense of pronounced mean differences if not via differences in the underlying genetic and environmental influences? One factor may well lie in the fact that behavioral genetic analyses have been restricted to decompositions of the variance around means without considering or interpreting the mean differences themselves. This matters since broad-scale underlying influences that do not differ in within-sample groups – for example, cultural norms for girls against aggression and greater male exposure to prenatal testosterone – are not

detectable in decompositions of variance, but may contribute to mean differences across sex. Focusing only on variances thus limits our interpretations in ways we may not realize. Focusing only on mean differences and ignoring the variance would be just as problematic, albeit in the other direction, since it is extraordinarily doubtful that all girls in the sample are socialized identically against aggressive behavior and that all male fetuses receive identical exposure to testosterone and/or are identically sensitive to it.

However, the effects in question may be more subtle as well. Consider poverty and environmental disadvantage, which predict many negative outcomes, including physical health problems, antisocial behavior, depression, anxiety, ADHD, poor academic performance, school delay/dropout, and low occupational attainment (Brooks-Gunn, Duncan, Klebanov, & Sealand, 1993; Holz, Laucht, & Meyer-Lindenberg, 2015; Leventhal & Brooks-Gunn, 2000; Lynam et al., 2000). Although correlations/regression coefficients/structural equation model parameters of this kind are often interpreted as if the two variables are moderately associated in everyone, it is in fact the case that the links between environmental disadvantage and deleterious outcomes are not observed in everyone to the same degree, and outcomes overlap considerably between them. Indeed, even the most cursory examination of the data scatterplot of a rather typical .35 correlation (see Figure 1; reprinted with permission from Turkheimer et al., 2017) makes clear that many points lie rather far from the correlation line, and quite a few of those points suggest no association at all between the two variables. The cases lying on or very close to the regression line thus compensate for these ‘wonky’ cases to generate the correlation (effectively the mean of the individual points’ indicated associations, assuming all else equal, which it rarely is).

Perhaps even more importantly, the amount of variance around the regression line can also vary with particular environmental features. Several studies (S. A. Burt, Pearson, Carroll,

Klump, & Neiderhiser, 2020; Hanscombe et al., 2012; Johnson et al., 2010; Eric Turkheimer, Beam, Sundet, & Tambs, 2017) have observed more outcome variability (and often considerably more variability) in disadvantaged contexts (or low levels of advantaged contexts, depending on how environmental quality is operationalized). This can also be seen clearly in Figure 1 – in fact, the figure's original point was that both the mean and the variance in offspring cognitive ability clearly co-varied with parental education. That is, parents with more education tended to have offspring with higher cognitive ability scores, but this association was also "*visibly heteroscedastic, with reduced variation around the regression line at higher levels of parental education…*" pg. 509, Turkheimer et al., 2017). Subsequent biometric analyses, which attempted to control censoring in the data to mitigate the effects of the heteroscedasticity, further suggested that the reduction in variance with greater parental education could be attributed to greater additive genetic variance, less shared/family-level environmental variance, or both. In other words, offspring cognitive ability appeared to be significantly more genetic and/or less shared environmental when parents were better educated. Other studies in the US have observed similar patterns when examining associations between parental SES and offspring cognitive ability (see meta-analysis by Tucker-Drob & Bates, 2015).

Samuelsson and colleagues (2008) also observed a disambiguation of means and variances, although there were in a different direction and were in response to at least partially distinct environmental experience (i.e., the onset of formal reading instruction). The authors longitudinally examined the heritability of reading ability at the ends of both Kindergarten and first grade in Australia, the United States, and Scandinavia (Samuelsson et al., 2008). Their reported means, standard deviations, and heritability estimates are presented in Table 2. As shown, mean reading skills increased dramatically during the one-year study period in all three

counties, although the magnitude of this increase varied by country (i.e., it was 4-fold in Scandinavia, 3-fold in the United States, and 2-fold in Australia). The variances in reading skills also increased during the one-year study period, but these increases were far less dramatic. Finally, although genetic influences were already proportionately important in Australia during Kindergarten, this was specific to that cultural context; genetic influences during kindergarten were less salient in Scandinavia (33%) and the United States (68%). A year later, however, heritabilities were high in all three countries (79-83%). Shared environmental influences displayed the opposite pattern. These findings were thought to have reflected the timings and intensities of early reading instruction in those countries: children in Australia receive compulsory reading instruction in kindergarten and first grade, whereas it does not begin until first grade for children in Scandinavia. This was interpreted to suggest that genetic influences on reading ability may be 'activated' by the onset of reading instruction. Unlike the Turkheimer, et al. (2017) example, however, in this case the greater genetic variances corresponded with greater mean reading skills and less so with greater total variances.

Adoption studies approaching similar etiological questions by evaluating mean differences between siblings raised in separate households have had an entirely different analytic and interpretative focus (Capron & Duyme, 1989; Schiff, Duyme, Dumaret, & Tomkiewicz, 1982; van IJzendoorn & Juffer, 2005). Schiff et al. (1982), for example, obtained indices of cognitive ability (IQ) in 32 children of unskilled workers who had been adopted away at an average age of 4 months into wealthy families, and compared their IQ scores and academic progress to those of children who had remained in their birth families. Among these were biological half-siblings of some of the adopted children who had remained with at least one biological parent. The adopted children scored on average 14 points higher than did their

biological half-siblings and were only 25% as likely to be required to repeat a grade. Consistent with these results, a meta-analysis of the few available studies on this topic (Van Ijzendoorn, Juffer, & Poelhuis, 2005) suggested that mean IQ differences between adopted children and their non-adopted biological siblings or same-aged peers (who stayed in the institutions in which all were initially placed) were very large, on the order of 1.17 standard deviations.

Although fascinating, none of these experimental studies examined sibling similarities in individual differences around the means, restricting their focus to the means themselves. Put another way, their research focus was on 'main effects', or those involved in mean differences between groups, which were interpreted as if they operate uniformly causally on everyone – all else assumed equal – despite omnipresent variance around the category means and virtually always substantively overlapping group distributions. In other words, these kinds of experimentally-based studies typically interpreted the unmeasured variance as if it were simply noise (rather than evidence of individual genetic influences).

In short, both means and variances clearly vary with particular exposures but do so in different ways. Despite this, twin-family studies restrict their analytic focus to variance differences alone, evaluating the extent to which variance around a variable's mean in pairs of individuals grouped by extent of familial relationship fall short of completely randomly occurring expectations, *without regard for its mean or potential systematic differences in the means among either pair members or pair groups*. This focus differs entirely from that in traditional experimental studies, which center on the extents of difference among the means of condition-manipulated groups, *with little or no regard for the variances around those means*.

Learning from our scientific history

Current behavioral genetic theory is silent on the possibility that we may be able to obtain additional information from means, but this has not always been the case. Decades ago, during the height of the nature-nurture debate, the different inferences obtained from evaluations of mean differences and those of correlations formed the empirical core of the Two Realms Hypothesis (E. Turkheimer, 1991). Under this proposal, mean developmental functions and individual differences around the mean were conceptualized as causally independent phenomena, such that the environment accounts for group differences (as captured by mean differences across reared-apart relatives) and genotype accounts for individual differences (as captured by correlations between reared-apart and reared-together relatives). In this view, group differences were viewed as highly malleable whereas individual differences were less so. By making space for both the 'nature' and the 'nurture' sides of the debate, the Two Realms hypothesis effectively provided an "*escape from the dilemma presented by individual- and group-difference results of adoption studies*" (Turkheimer,1991; pg. 393).

Although emotionally appealing, the Two Realms theory was quickly shown to be implausible on both statistical and conceptual grounds. Turkheimer (1991) astutely noted that mean 'outcome' differences can be represented just as easily by dichotomous group membership (or environmental feature) variable that is regressed on the outcome. What's more, Turkheimer (1991) argued that it was not feasible for group means to be inherently more malleable than individual differences among the members of those groups, since both types of studies measure a single etiologic process – namely, the influences of genotypes and environments on individual outcomes. So convincing were Turkheimer's arguments that the Two Realms Hypothesis has since faded into the annals of history.

But, as also noted by Turkheimer (1991) and as demonstrated in the examples above, the discrepancy between the two types of studies very much remains – mean differences between reared-apart relatives are routinely larger than would be expected based on the very small within-group environmental relations observed in correlational data from adoption studies. Such observations, point to very different kinds and extents of environmental impacts. We thus agree with Turkheimer (1991) that a more complete understanding of causality requires analyses that incorporate and analyze both mean differences and correlations/variance decompositions. However, we further argue that although both variance decompositions and studies of means are measuring the influences of genotypes and environments on individual outcomes – the 'single etiologic process' outlined by Turkheimer (1991) – the two statistical moments are in fact capturing different elements and levels of causal impacts. Namely, 'cause' is inferred via sources of *individual differences* in variance-based studies, but by experimentally-manipulated environmental sources of relative *group similarities* in means-based studies. The focus on variance among individuals in the former ignores intervention and macro-level environmental effects that could be important at a population level (e.g., effects of governmental policies, widespread pollutant exposures, climate change, epidemics, broad cultural forces, species-wide but sex-specific organizational effects of prenatal hormones, etc.). In sharp contrast, the means-only focus relies on the assumption that all genetic influences have been randomized during sampling, and thus eliminated from consideration (which is extraordinarily doubtful since genetic and environmental influences are rarely independent in naturalistic settings and gene-environment correlation is very common and socially powerful). In short, by restricting their examinations only to particular statistical moments and ignoring other statistical moments, we

contend that both variance-focused and means-focused studies can and probably often are interpreted in overly simplistic and sometimes misleading ways.

Unravelling the etiologic implications of links between means and variances

The differences in the respective analytic foci of behavioral genetic and experimental studies thus have potentially enormous consequences for their causal inferences regarding etiology. Despite this, different patterns in the means and variances of a given set of data are rarely even noted in studies of genetic and environmental influences, much less considered important in understanding the development of associations in question. We suspect that this blind spot is made possible by the prevailing Mendelian/Fisherian view (Fisher, 1919), which assumes that (among other things) *1)* genetic variants have direct effects independent of environments, and *2)* heritability (the proportion of population variance attributable to genetic variance) is a reasonable indicator of the aggregate extent to which genes "matter" for a given trait. Neither of these assumptions has stood the test of time. It is now well known that gene-environment interplay is a key contributing factor to psychopathology, psychological characteristics, and even physical features such as eye and hair color (Johnson, 2007; Ridley & Pierpoint, 2003; West-Eberhard, 2003). What's more, there is no evidence that highly heritable traits are more likely to be *directly and additively influenced* by genes than are less heritable traits. Heritability estimates for breast cancer, for example, are only around 27% (Lichtenstein et al., 2000; Möller et al., 2016), yet particular genetic variants have been identified that appear strongly to predict breast cancer for the individuals who have them (i.e., BRCA1 and BRCA2). Even here, however, these variants are not 'deterministic' – not all carriers get breast cancer – and these variants are quite rare, so overall they account for very little population variance in breast cancer occurrence. A typical rationale for discrepancy between heritability 'magnitudes'

and ability to identify influential genetic variants is that some variants are more 'penetrant' (i.e., are more directly causal) than others.

However, there are other possible explanations that rely on a developmental, non-Mendelian/Fisherian theory of genetic influences (Pigliucci, 2003, 2005; Schmalhausen, 1949). Ivan Schmalhausen was an ecological evolutionary geneticist who recognized that environments continuously change, both permanently (e.g., climactically, via earthquakes and volcanoes, via human actions such as deforestation for construction and/or agriculture) and cyclically (e.g., diurnally, seasonally, predator-prey imbalances), relative to organismic lifespans. These changes demand adaptations that all individual organisms must make or perish. Organisms' adaptations to permanent change stabilize in relatively few generations (as observed in *Drosophila*; Waddington, 1942) but do so to different degrees depending on relevant population genetic/environmental structure (e.g., gene-environment stratification), involved trait distributions (variances, skews, etc.), extents of change and potential for multiple ways of adapting, and relative frequencies of relevant genetic variants. For their part, many cyclical changes (e.g., seasonal, diurnal) are so common as to be pervasive facts of life. They tend to stabilize ecologies dynamically, in transactional equilibria between environmental cycles and population behavioral variations, separating the genetic variants underpinning them from these regular fluctuations and, in recompense, increasing regulatory complexity underlying those variants' expression and its variability.

This 'regulatory complexity' consists, Schmalhausen (1946) suggested, of genetic 'redundancies' – multiple genetic variants that can, together or alone – bring about emergence of 'typical' developmental trajectories. From this perspective, our genes would be better conceptualized as toolboxes of rather non-descript, multiply combinable components that, like

houses built using Lego bricks, can be cobbled together in all kinds of different ways to bring about any one ‘outcome’ such as pubertal development, effective management or marketing skills, or ways to answer matrix reasoning problems. Evolutionarily, preserving many genetic variants that primarily regulate others’ expression is much more efficient than requiring environments to select out variants that disrupt critical development. Under Schmalhausen’s theory, the extreme polygenicity (Boyle, Li, & Pritchard, 2017) we now routinely observe in psychological and many physical characteristics (Chabris, Lee, Cesarini, Benjamin, & Laibson, 2015) would thus be inevitable.

Evolutionary and developmental genetic experiments with model organisms have long offered evidence that Schmalhausen (1946) was ‘on the right track’. Schmalhausen posited that apparent genetic variance in developmental characteristics important in survival and reproductive capacity can vary with environmental conditions such that harsh conditions demand larger “teams” of expressing variants to maintain normal development, and benign conditions allow more either to remain silent or to direct expression towards taking greater advantage of pre-existing “outcome” opportunities via other characteristics. What’s more, Schmalhausen (1946) suggested that when populations face severe stresses such as war, economic crises, or epidemics, they become more vulnerable to small perturbations in other aspects of their environments. This occurs because putatively separate environmental contexts are often interrelated, and the developmental disturbances created by the major stress ripple through them all. This undermines their stability and recruits usually silent genes into expression to meet immediate coping needs and imminent developmental milestones, without regard for longer-term consequences of that expression. The new genetic ‘kludge’ destabilizes developmental mechanisms, which, in turn, creates its own vulnerabilities, launching a vulnerability cascade. Furthermore, individuals

within the population inevitably vary in vulnerability to the primary stress, and in extent and nature of their experiences of and responses to that stress. The end result is greater population variance in outcomes, a process later termed "Schmalhausen's Law" (Lewontin & Levins, 2000). As indicated in the example below, however, this "Law" may not hold when sources of stress are so strong as to preclude genes' ability to kludge together effective coping responses.

Johnson (2012) further integrated Schmalhausen's (1946) theory with modern understandings of genotype-environment interplay, proposing that, when the main effects of environmental conditions are strong enough across individuals, they act to both alter mean trait levels (increasing or decreasing, as the case may be) and to suppress genetic sources of variance in the population. The remaining variance would thus appear to be environmental in origin due to underlying gene-environment correlation (and resulting population genetic stratification). When main environmental effects are less extreme or less consistent among organisms, however, they have weaker effects on the mean and release otherwise unexpressed population genetic variance.

Viewed from Schmalhausen's perspective, interpretations of Turkheimer et al.'s (2017) and Samuelsson et al.'s (2008) results are as follows: advantageous contexts such as high parental education and reading instruction released otherwise unexpressed genetic influences that fostered development of cognitive potential. Disadvantaged environments, in contrast, suppressed expression of fostering genetic influences. That is, the notion that genetic variance is greater when environments are uniform only goes so far for these outcomes: Relatively uniform disadvantage had strongly destructive social "force" on gene expression, but similarly uniform advantage did no more than offer expressive opportunity. Even here, however, although both these studies indicated similar increases in genetic influences, the pattern of change in variance

differed across the environment in question (high parental education suppressed total variance whereas exposure to reading instruction released more variance).

In sum, both similarities and differences in the social forces and structures affording and constraining environmental exposures and opportunities were indicated. Moreover, there clearly was covariance – otherwise, moderators and outcomes would not have been correlated – but this covariance was not modeled, and the authors did not offer any empirically-based measure of total outcome variance along the moderating dimensions. Thus, although Schmalhausen's (1946) perspective offers a much richer interpretation than the standard Fisher-Mendel genetic framework, we do not yet have the statistical tools needed to fully model such processes.

Statistical techniques that can accommodate this duality to some degree

Current approaches to evaluating etiology focus almost exclusively on only means or only variances – but not both – to make their causal inferences, potentially hamstringing our etiologic conclusions. There is thus a clear need for novel design and analytic strategies that leverage information from *both* means and variances simultaneously. One design already available (but still underutilized) is the yoked adoption study. Such studies assess adoptive children and their adoptive and biological family members, and are effectively natural experimental interventions (van IJzendoorn & Juffer, 2005). One child is reared by the biological mother, while a biological sibling is adopted and raised by different parents, in a different neighborhood, with different siblings and different schools. These differences are augmented by the fact that adoptive and biological families tend to differ quite a bit from one another in many ways. Adoptive parents are often highly selected (via both niche-picking and adoption agencies) relative to the general population of parents, and tend to be older, married, better-educated, and earning higher income (McGue et al., 2007). This point has been frequently noted to highlight

potential problems with range restriction (McGue et al., 2007; Stoolmiller, 1999), which can blunt statistical associations between adoptive family members (although they did not do so in McGue et al., 2007) or can exaggerate them in some cases; (Johnson, Deary, & Bouchard Jr, 2018). What is less frequently observed, however, is that adopted children's birth parents are also selected relative to the general parental population. They tend to be younger, unmarried, less educated, and/or to come disproportionately from disadvantaged backgrounds (Leve et al., 2019). In short, the yoked adoption design is well-suited both for incorporating considerations of advantage/disadvantage into behavioral genetic studies, and for studies of mean effects more generally (see Figure 2).

Kendler et al. (2015) offered a peek as to what might be accomplished by incorporating mean differences into standard adoption analyses. They linked multiple Swedish nationwide registries to identify 436 full-sibling pairs in which only one of the two siblings was adopted. They observed that the adopted siblings had on average 4.4 additional IQ points relative to their reared-apart full siblings. They then replicated these observations in 2,341 half-sibling pairs. In both cases, however, these IQ differences varied with the extents of educational differences between the adoptive and biological parents, such that larger differences in parental education were associated with larger differences in offspring IQ. Moreover, adoptive offspring IQ continued to be correlated with both the non-rearing biological parents and the adoptive parents (correlations ranged from .18 to .20), and with those of their reared-apart biological siblings (.30 for full siblings and .27 for half-siblings). As Kendler, et al. noted (p. 4616), such observations provided evidence that, "*despite being demonstrably related to genetic endowment, cognitive ability is environmentally malleable, and the malleability shows plausible dose-response relations with the magnitude of the environmental differences*".

The kind of yoked adoptive- and biological-family sample Kendler et al. (2015) used can be even more valuable when household and individual psychological characteristic information is available. Burt and colleagues did just this, adding specific measures of the adoptive and biological home literacy environments to estimate environmental effects with more precision (S. A. Burt et al., submitted). They found that birth mothers' academic achievement scores were as similar to their adopted children as they were to the adopted child's biological siblings, despite the fact that birthmothers were raising the latter and not the former. By contrast, there was little-to-no evidence for rank-order similarity between adopted children's academic achievement and that of their adoptive mothers. They also observed small-to-modest zero-order correlations between measures of the adoptive home literacy environment and adopted children's achievement, tentatively suggesting that environmental influences on standardized achievement tests may be marginally linked to homes' literacy-promotive features. Such interpretations are consistent with the last several decades of behavior genetic studies in this area, indicating robust genetic influences and smaller, less durable environmental influences (Tucker-Drob, Briley, & Harden, 2013).

Once Burt and colleagues also considered mean differences among reared-apart relatives, however, their interpretation changed rather considerably. They observed moderate-to-large mean differences between adopted children's academic achievement and their reared-apart family members, as high as 15+ points (just over one SD; represented in Figure 2). Furthermore, multilevel modeling analyses indicated that both birth mothers' cognitive ability and the availability of reading materials in rearing homes accounted for a significant proportion of these mean differences in standardized achievement test scores. They thus concluded that, while genetic influences were present regardless of how the analyses were done, environmental

influences on achievement were more clearly revealed when analyzing mean differences across reared-apart relatives alongside intraclass correlations.

In short, it is certainly possible (and in our view, likely) that we can extract additional – and potentially quite different – information about underlying etiologic processes by incorporating the mean into traditional correlation- and variance-focused behavioral genetic studies of etiology. And although the yoked adoption design described above moves closer to this ideal for adoption studies, we are not yet able to analytically incorporate the mean into variance-focused twin studies in any meaningful way. Indeed, the most that twin researchers can do as of this publication is *1)* explicitly identify heteroscedasticity in associations between moderators and 'outcomes' (although as noted, very few studies do even this; for a rare exception, see Johnson, Kyvik, Skytthe, Deary, & Sørensen, 2011), and *2)* discuss how this heteroscedasticity affects interpretations of specific sets of GxE results. These most valuably center around interplay between population-level social forces/structures and individual characteristics that constrain and afford opportunities – the very patterns of environmental movement Schmalhausen (1946) outlined. For example, a standard GxE analysis might indicate that the magnitude of genetic variance in X outcome was greater at the low end of environment Y than at its high end. Mean levels of X could be lower or higher at the low end of Y, which could have different implications for policy involving X.

While these kinds of discussions would enrich current research considerably, what is really needed is a formal analytic model and/or additional sorts of quasi-experimental designs that can fully leverage changes in both means and variances to illuminate individual etiologies and population forces/structures better. These advances would have several positive downstream implications. First and foremost, as argued throughout this article, it seems likely to improve

(perhaps substantially) our understanding of gene-environment interplay. This would be especially important since, as argued here, gene-environment interplay is both so complex and yet so common and widespread that the current (simplistic) models of gene-environment interplay are likely only scratching the surface.

Second, traditional behavior genetics' correlational approach to science also constrains the field's focus to examining *currently existing* population situations. The field thus has virtually nothing to say about etiology in environments that *could exist* (perhaps following successful interventions, cataclysmic events, or implementation of new governmental policies), but do not (see Burt et al., 2019). As such, traditional behavioral genetic approaches provide no actionable information about how society might intervene to change a given outcome, nor how unforeseen natural events might alter population characteristics (Lewontin, 1974a, 1974b). Adding a focus on *what could be* (rather than *what is*) would allow behavioral genetics to better inform efforts to change behavior or social inequalities, and to illuminate how unforeseen natural events might alter population characteristics (Lewontin, 1974a, 1974b).

Genetically-informed studies that explicitly incorporate interventions (e.g., Burgoyne et al., 2020; S. A. Burt, Plaisance, & Hambrick, 2019) or leverage naturally-occurring cultural innovations over time (e.g., the recent introduction of the internet and social media; S. A. Burt, in press) would likely get us far closer to this lofty ideal. For example, we might experimentally manipulate technological features of the online environment (e.g., anonymity) and evaluate both differences in means and decompositions of variance, which could greatly enrich inferences for further testing. Alternately, we might ask whether and how the origins of a given outcome vary before and after naturally occurring treatments. We could also leverage natural experiments and providential occurrences to assess whether and how a given outcome and its etiology have

changed over the last 20-30 years. Several twin studies have been conducted in ongoing fashions across the last few decades, collecting both cross-sectional and longitudinal data on many thousands of twins, and are thus ripe for this kind of analysis. The core challenge to this work would center on disambiguating age, cohort, and period effects. Such analytic techniques are certainly available for phenotypic data, although they have never been applied to behavioral genetics analyses to our knowledge. Finally, researchers could link twin data to broad-scale cultural shifts (e.g., introduction of social media into popular culture), which may well have altered how adolescents manifest many outcomes (e.g., shifting away from in-person aggression and towards online aggression). In sum, joint considerations of mean changes (e.g., before and after interventions or introduction of novel cultural forces) alongside individual differences is poised to make incredibly provocative and important advances in understanding.

Finally, incorporation of means into novel or traditional behavioral genetic designs could inform our conceptual understanding of genetic influences more broadly – starting with, are they Fisherian or Schmalhausian? Statistical methods that can leverage both the first and second statistical moments in the data may be able to test competing theories of how and when social forces constrain and liberate gene-environment interplay of various kinds and when and to what degree genes act more like Schmalhausen or Mendel proposed, with downstream implications for (or against) Fisherian and Schmalhausian models of gene-environment interplay. Ultimately, such work could offer foundationally important contributions to the literature.

Table 1. Core statistics and their use in traditional behavioral genetic studies of etiology

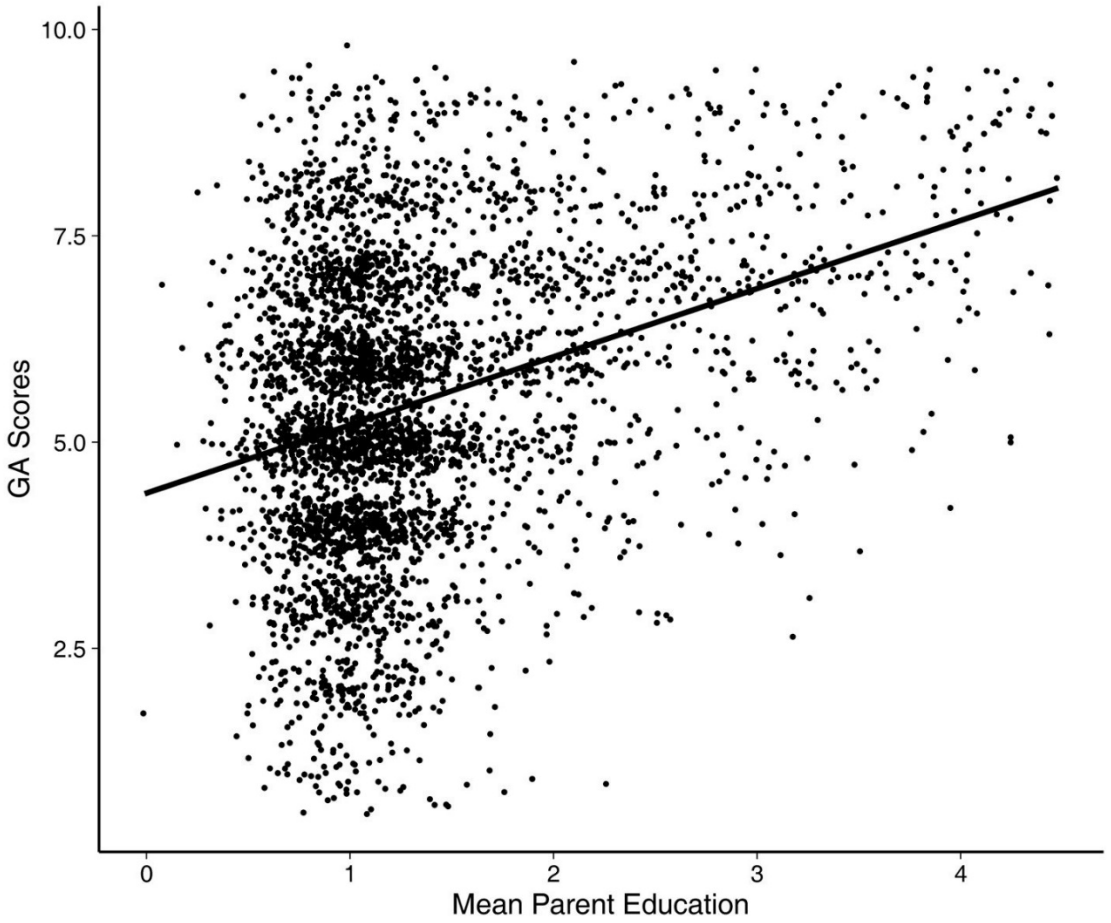
Statistic	**Formula**	**What it measures**	**Common usage in family-based studies**
Mean	$X = \sum x/N$	The average value in a given dataset; indexes the central tendency of the data	None; usually presented to describe the observed raw data and address its normativity
Variance	$s^2 = \sum(x- X)^2/N-1$	The average amount of variability around the mean; indexes the data's range and extent of clustering	The overall observed variance is decomposed into its genetic and environmental components based on patterns of similarity (as indexed via twin covariances in structural equation models) in pairs with differing degrees of genetic and environmental relatedness
Covariance/ Correlation	$Cov_{XY} = \sum(x- X)(y-Y)/N-1$ $rXY = Cov_{XY}/\sqrt{(s^2_X * s^2_Y)}$	The extent to which individual differences in variable X covary with those in variable Y. A correlation is simply a standardized version of the covariance. ICCs index the standardized covariance on pairs of individuals on the same variable	Studies leverage interclass covariances/correlations and intraclass ICCs/twin covariances as the central measures in their analytic approach

Table 2. Reading skill results from Samuelsson et al., 2008.

Country	Grade	Observed		Modelled proportionate variance components		
		Mean	SD	% A	% C	% E
Australia	End of Kindergarten	14.2	10.7	84*	9	8*
	End of first grade	31.4	14.1	80*	2	18*
United States	End of Kindergarten	8.6	9.1	68*	25*	07*
	End of first grade	27.7	13.3	83*	7	11*
Scandinavia	End of Kindergarten	5.3	9.2	33*	52*	15*
	End of first grade	21.1	13.1	79*	7	14*

Note. Reprinted with permission from Samuelsson et al. (2008). %A, %C, and %E indicated the proportion of additive genetic variance, shared environmental variance and nonshared environmental variance, respectively. The phenotypic mean and standard deviation (SD) of reading skill measures are presented, separately by country and grade level, in the middle columns. The univariate heritabilities are presented on the right side of the table.

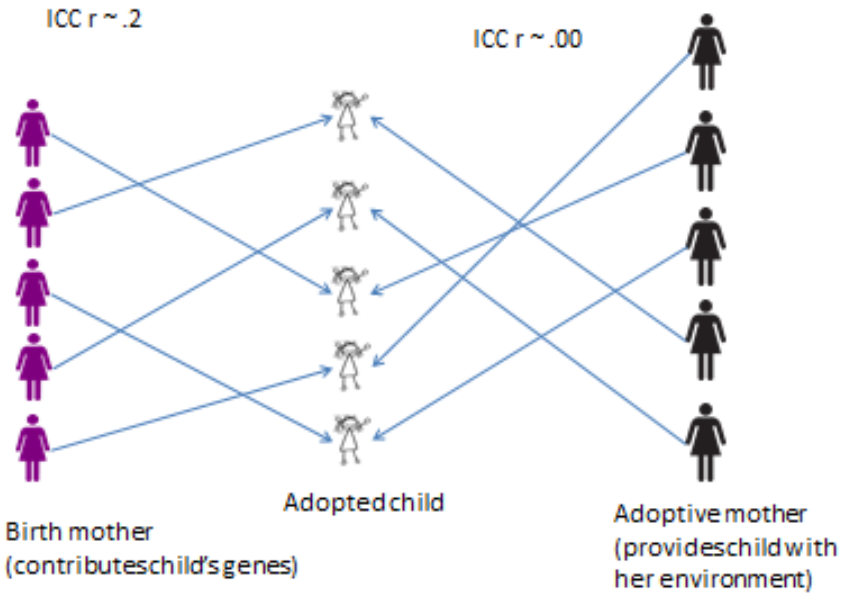
Figure 1.



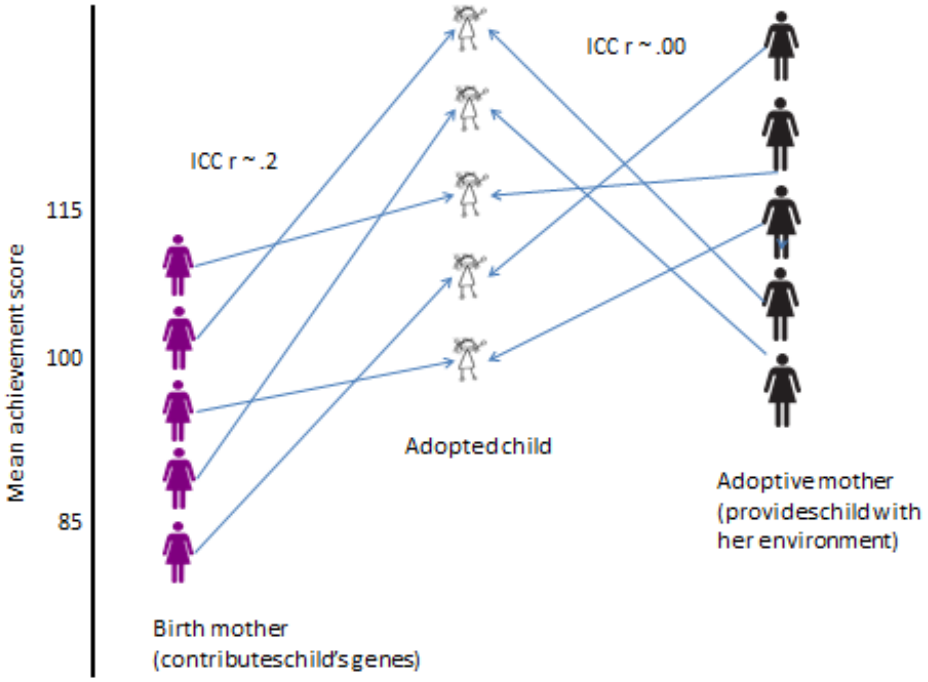
Note. Reprinted with permission from Turkheimer et al. (2017). Scatterplot of jittered conscript general ability (GA) scores plotted against jittered mid-parent education level with fitted ordinary least squares regression line.

Figure 2. Schematics of current and fully realized adoption designs (reprinted with permission from Burt et al., submitted)

1a) Traditional adoption design, focused on patterns of correlations



1b) Fully realized adoption design, focused on patterns of means and correlations



References

Archer, J. (2004). Sex differences in aggression in real-world settings: a meta-analytic review. *Review of general Psychology, 8*(4), 291.

Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An expanded view of complex traits: from polygenic to omnigenic. *Cell, 169*(7), 1177-1186.

Brooks-Gunn, J., Duncan, G. J., Klebanov, P. K., & Sealand, N. (1993). Do neighborhoods influence child and adolescent development? *American Journal of Sociology, 99*, 353-395.

Burgoyne, A. P., Carroll, S., Clark, D. A., Hambrick, D. Z., Plaisance, K. S., Klump, K. L., & Burt, S. A. (2020). Can a brief intervention alter genetic and environmental influences on psychological traits? An experimental behavioral genetics approach. *Learning and Motivation, 72*, 101683.

Burks, B. (1928). Chapter X: The relative influence of nature and nurture upon mental development: A comparative study of foster-parent offspring child resemblance and true parent-true child resemblance. *Yearbook of the National Society for the Sudy of Education, 27*, 219-316.

Burt, C. H., & Simons, R. L. (2014). Pulling back the curtain on heritability studies: Biosocial criminology in the postgenomic era. *Criminology, 52*(2), 223-262.

Burt, S. A. (in press). The Genetic, Environmental, and Cultural Forces influencing Youth Antisocial Behavior are Tightly Intertwined. *Annual Review of Clinical Psychology*.

Burt, S. A., Johnson, W., Leve, L. D., Natsuaki, M. N., Reiss, D., Shaw, D. S., . . . Neiderhiser, J. M. (submitted). The common understanding of environmental influences on academic achievement may not be correct.

Burt, S. A., Pearson, A. L., Carroll, S., Klump, K. L., & Neiderhiser, J. M. (2020). Child Antisocial Behavior Is more Environmental in Origin in Disadvantaged Neighborhoods: Evidence Across Residents' Perceptions and Geographic Scales in Two Samples. *Journal of Abnormal Child Psychology, 48*(2), 265-276.

Burt, S. A., Plaisance, K. S., & Hambrick, D. Z. (2019). Understanding "What Could Be": a call for 'experimental behavioral genetics'. *Behavior Genetics, 49*(2), 235-243.

Burt, S. A., Slawinski, B. L., Carsten, E. E., Harden, K. P., Hyde, L. W., & Klump, K. L. (2019). How should we understand the absence of sex differences in the genetic and environmental origins of antisocial behavior? *Psychological Medicine, 49*(10), 1600-1607.

Capron, C., & Duyme, M. (1989). Assessment of effects of socioeconomic status on IQ in a full cross-fostering study. *Nature, 340*, 552-554.

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*(4), 304-312.

DeGroot, M. H. (2012). *Probability and statistics*: Pearson.

Falconer, D. (1960). Introduction to quantitative genetics. Oliver and Boyd. *Edinburgh, 365p*.

Fisher, R. A. (1919). XV.—The correlation between relatives on the supposition of Mendelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh, 52*(2), 399-433.

Galton, F. (1875). The history of twins, as a criterion of the relative powers of nature and nuture. *Fraser's Magazine, 12*, 566-576.

Hanscombe, K. B., Trzaskowski, M., Haworth, C. M., Davis, O. S., Dale, P. S., & Plomin, R. (2012). Socioeconomic status (SES) and children's intelligence (IQ): In a UK-

representative sample SES moderates the environmental, not genetic, effect on IQ. *PLoS ONE, 7*(2), e30320.

Holz, N. E., Laucht, M., & Meyer-Lindenberg, A. (2015). Recent advances in understanding the neurobiology of childhood socioeconomic disadvantage. *Current opinion in psychiatry, 28*(5), 365-370.

Johnson, W. (2007). Genetic and environmental influences on behavior: capturing all the interplay. *Psychological Review, 114*(2), 423.

Johnson, W., Deary, I. J., & Bouchard Jr, T. J. (2018). Have standard formulas correcting correlations for range restriction been adequately tested?: Minor sampling distribution quirks distort them. *Educational and Psychological Measurement, 78*(6), 1021-1055.

Johnson, W., Kyvik, K. O., Mortensen, E. L., Skytthe, A., Batty, G. D., & Deary, I. J. (2010). Education reduces the effects of genetic susceptibilities to poor physical health. *International journal of epidemiology, 39*(2), 406-414.

Johnson, W., Kyvik, K. O., Skytthe, A., Deary, I. J., & Sørensen, T. I. (2011). Education modifies genetic and environmental influences on BMI. *PLoS ONE, 6*(1), e16290.

Keller, M. C., & Medland, S. E. (2008). *Evaluation of the cascade model - a new extended twin family model.* Paper presented at the Behavioral Genetics Association, Louisville, KY.

Keller, M. C., Medland, S. E., & Duncan, L. E. (2010). Are extended twin family designs worth the trouble? A comparison of the bias, precision, and accurracy of parameters estimated in four twin family models. *Behavior Genetics, 40*, 377-393.

Leve, L. D., Neiderhiser, J. M., Ganiban, J. M., Natsuaki, M. N., Shaw, D. S., & Reiss, D. (2019). The early growth and development study: A dual-family adoption study from birth through adolescence. *Twin Research and Human Genetics, 22*(6), 716-727.

Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. *Psychological Bulletin, 126*(2), 309-337.

Lewontin, R. C. (1974a). The analysis of variance and the analysis of causes. *American Journal of Human Genetics, 26*, 400-411.

Lewontin, R. C. (1974b). Annotation: the analysis of variance and the analysis of causes. *American Journal of Human Genetics, 26*(3), 400.

Lewontin, R. C., & Levins, R. (2000). Schmalhausen's law.

Lichtenstein, P., Holm, N. V., Verkasalo, P. K., Iliadou, A., Kaprio, J., Koskenvuo, M., . . . Hemminki, K. (2000). Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *New England Journal of Medicine, 343*(2), 78-85.

Lynam, D. R., Caspi, A., Moffitt, T. E., Wikstrom, P. H., Loeber, R., & Novak, S. (2000). The interaction between impulsivity and neighborhood context on offending: The effects of impulsivity are stronger in poorer neighborhoods. *Journal of Abnormal Psychology, 109*, 563-574.

McGue, M., Keyes, M., Sharma, A., Elkins, I., Legrand, L., Johnson, W., & Iacono, W. G. (2007). The environment of adopted and non-adopted youth: Evidence of range restriction from the Sibling Interaction and Behavior Study (SIBS). *Behavioral Genetics, 37*, 449–462.

Möller, S., Mucci, L. A., Harris, J. R., Scheike, T., Holst, K., Halekoh, U., . . . Holm, N. V. (2016). The heritability of breast cancer among women in the Nordic Twin Study of Cancer. *Cancer Epidemiology and Prevention Biomarkers, 25*(1), 145-150.

Pigliucci, M. (2003). Phenotypic integration: studying the ecology and evolution of complex phenotypes. *Ecology letters, 6*(3), 265-272.

Pigliucci, M. (2005). Evolution of phenotypic plasticity: where are we going now? *Trends in ecology & evolution, 20*(9), 481-486.

Ridley, M., & Pierpoint, G. (2003). *Nature via nurture: Genes, experience, and what makes us human* (Vol. 19): HarperCollins New York.

Samuelsson, S., Byrne, B., Olson, R. K., Hulslander, J., Wadsworth, S., Corley, R., . . . DeFries, J. C. (2008). Response to early literacy instruction in the United States, Australia, and Scandinavia: A behavioral-genetic analysis. *Learning and individual differences, 18*(3), 289-295.

Schiff, M., Duyme, M., Dumaret, A., & Tomkiewicz, S. (1982). How much could we boost scholastic achievement and IQ? A direct answer from a French adoption study. *Cognition, 25*, 165-196.

Schmalhausen, I. I. (1949). Factors of evolution: the theory of stabilizing selection.

Stoolmiller, M. (1999). Implications of the restricted range of family environments for estimates of heritability and nonshared environment in behavior–genetic adoption studies. *Psychological Bulletin, 125*(4), 392.

Tucker-Drob, E. M., & Bates, T. C. (2015). Large cross-national differences in gene× socioeconomic status interaction on intelligence. *Psychological Science*, 0956797615612727.

Tucker-Drob, E. M., Briley, D. A., & Harden, K. P. (2013). Genetic and environmental influences on cognition across development and context. *Current Directions in Psychological Science, 22*(5), 349-355.

Turkheimer, E. (1991). Individual and group differences in adoption studies of IQ. *Psychological Bulletin, 110*(3), 392.

Turkheimer, E., Beam, C. R., Sundet, J. M., & Tambs, K. (2017). Interaction between parental education and twin correlations for cognitive ability in a Norwegian conscript sample. *Behavior Genetics, 47*(5), 507-515.

van IJzendoorn, M. H., & Juffer, F. (2005). Adoption is a successful natural intervention enhancing adopted children's IQ and school performance. *Current Directions in Psychological Science, 14*, 326-330.

Van Ijzendoorn, M. H., Juffer, F., & Poelhuis, C. W. K. (2005). Adoption and cognitive development: a meta-analytic comparison of adopted and nonadopted children's IQ and school performance. *Psychological Bulletin, 131*(2), 301.

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *The American Journal of Human Genetics, 90*(1), 7-24.

Waddington, C. H. (1942). Canalization of development and the inheritance of acquired characters. *Nature, 150*(3811), 563-565.

West-Eberhard, M. J. (2003). *Developmental plasticity and evolution*: Oxford University Press.

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics, 88*(1), 76-82.