

## PERSPECTIVE OPEN



# Nurture might be nature: cautionary tales and proposed solutions

Sara A. Hart<sup>1,2</sup> , Callie Little<sup>2,3</sup> and Elsje van Bergen<sup>4</sup>

Across a wide range of studies, researchers often conclude that the home environment and children's outcomes are causally linked. In contrast, behavioral genetic studies show that parents influence their children by providing them with both environment and genes, meaning the environment that parents provide should not be considered in the absence of genetic influences, because that can lead to erroneous conclusions on causation. This article seeks to provide behavioral scientists with a synopsis of numerous methods to estimate the direct effect of the environment, controlling for the potential of genetic confounding. Ideally, using genetically sensitive designs can fully disentangle this genetic confound, but these require specialized samples. In the near future, researchers will likely have access to measured DNA variants (summarized in a polygenic scores), which could serve as a partial genetic control, but that is currently not an option that is ideal or widely available. We also propose a work around for when genetically sensitive data are not readily available: the Familial Control Method. In this method, one measures the same trait in the parents as the child, and the parents' trait is then used as a covariate (e.g., a genetic proxy). When these options are all not possible, we plead with our colleagues to clearly mention genetic confound as a limitation, and to be cautious with any environmental causal statements which could lead to unnecessary parent blaming.

*npj Science of Learning* (2021)6:2; <https://doi.org/10.1038/s41539-020-00079-z>

Most parents spend hours fretting over decisions about the environment they provide to their children. The scientific literature mirrors this idea. Across a wide range of studies from many psychological domains, researchers often conclude that the environment parents provide and children's outcomes are causally linked, through environmental transmission (see Box 1). For example, a study examining the association of having a home library as an adolescent and later adult literacy, numeracy and technology skills drew our attention because of in-depth coverage in the Guardian (<https://www.theguardian.com/books/2018/oct/10/growing-up-in-a-house-full-of-books-is-major-boost-to-literacy-and-numeracy-study-finds>). This study used a very rich and well-powered dataset, and found a correlation between the number of books in adolescents' homes and literacy performance in adulthood. They conclude that "growing up with home libraries boosts adult skills", inferring a causal connection<sup>1</sup>. This is depicted in Fig. 1. Here we discuss how the correlation between the environments parents provide, the "rearing environment", and their children's outcomes can indeed be fully due to a causal association, or importantly, can also be partly or fully due to a genetic confounding, illustrated in Fig. 2 (see Footnote 1 in the Supplementary Notes). After highlighting the problem, we suggest ways that psychological scientists can examine research questions related to the rearing environment and children's outcomes in ways that account for, or at least acknowledge, genetic confounding.

## GENETIC CONTROL OF EXPOSURE TO THE ENVIRONMENT

Decades of work from behavioral genetics show that children's traits are influenced by both genetic and environmental effects<sup>2,3</sup>. Likely more surprising to hear for most is that genetic influences

are often seen on measures of the "environment", suggesting that the contexts surrounding children are partly under genetic control<sup>4</sup>. For example, a meta-analysis found cumulative support for genetic influences on the parenting children received<sup>5</sup>. This idea, that there is genetic influence on exposure to environments, is called a gene–environment correlation. A gene–environment correlation describes the process by which a person's genotype influences their exposure to the environment<sup>6</sup>. It is certainly not the case that genes are doing this directly, but instead genotypes matter for aspects of our personality, behaviors and cognitions, which then influence how we interact with our environment and how others interact with us<sup>7</sup>. This concept of an individual purposely and dynamically interacting with their surrounding environment is not limited to behavioral genetics; similar processes have been described in other literatures, for example person-centered interactions<sup>8</sup> and the Selective Optimization with Compensation<sup>9</sup>.

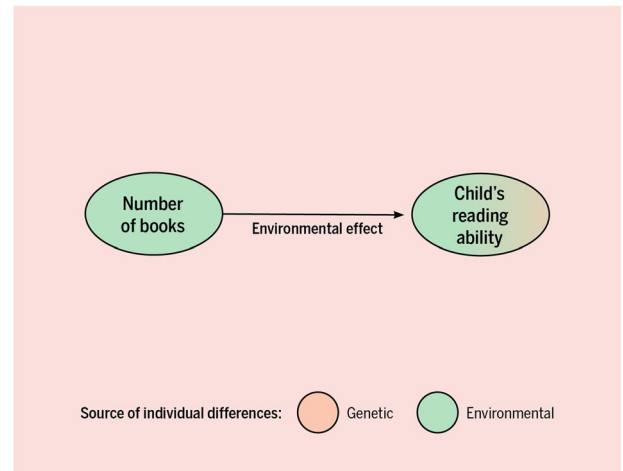
Specifically, there are three types of gene–environment correlations that can result in genetic confounds<sup>6</sup>. First, a "passive gene–environment correlation" describes the association between the genotype a child inherits from their parents and the environment the child is raised in. Another way to think of it is that genes are a third variable which influence both the rearing environment a child receives as well as the child's own traits, via genetic transmission from parents to child. This means it is not possible to draw causal conclusions between the rearing environment and children's traits. For example, home environments have been found to be less chaotic for children with high effortful control, with results indicating that the same genes in parents which contribute to the levels of structure in their home (i.e., factors such as absence of noise and crowding, as well as presence of structure and routine) are also transmitted to their

<sup>1</sup>Department of Psychology, Florida State University, Tallahassee, FL, USA. <sup>2</sup>Florida Center for Reading Research, Florida State University, Tallahassee, FL, USA. <sup>3</sup>Department of Psychology, University of New England, Armidale, NSW, Australia. <sup>4</sup>Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. ✉email: [hart@psy.fsu.edu](mailto:hart@psy.fsu.edu)

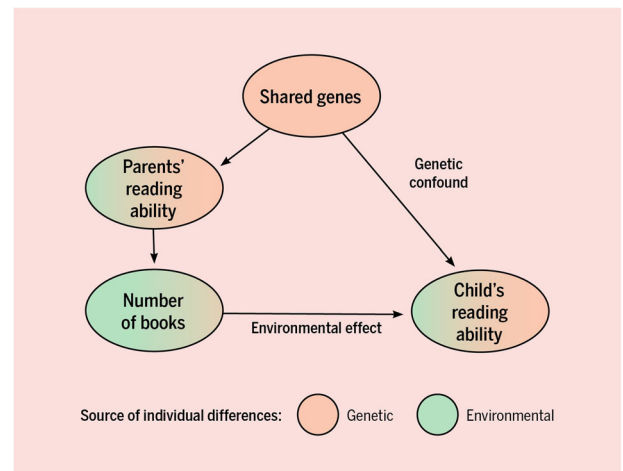
**Box 1** A glossary of some key terms

Environmental transmission (also called: cultural or phenotypic transmission):	Transmission of traits from parents to their children by non-genetic means. It is used to describe when parents' traits impact their child's traits through the environment they create.
Familial Control Method:	Using a measure of the same trait in the parent as the child as a covariate in models estimating the effect of the rearing environment. That covariate then serves as a proxy control for the genetic transmission effect.
Familial transmission:	Transmission of traits from parents to their children, both by genetic and non-genetic means. Familial transmission gives rise to parent-child resemblance.
Gene-environment correlation:	Genetic influence on the exposure to the environment. There are three types: passive, evocative, and active (see text).
Genetic confounding:	Confounding due to gene-environment correlation. Here we focus on confounding due to a <i>passive</i> gene-environment correlation, describing a situation where the influence of parental traits on children's traits is not (solely) due to environmental transmission.
Genetic transmission:	Transmission of traits from parents to their children by genetic means (i.e., children inherit genes from their parents for a given trait).
Genotype:	An individual's complete heritable information. A combination of alleles for a specific gene or across the whole genome.
GWAS:	Genome-wide association study. Identifies genetic variants (i.e., SNPs) across the genome that are linked to a trait.
Phenotype:	An individual's observable traits, like eye color, reading ability, or parenting style.
SNP (pronounced "snip"):	Single nucleotide polymorphisms (SNPs). A single position in a DNA sequence that varies among individuals. For example, if a particular SNP can be nucleotide G or nucleotide C, then individuals can have GG, GC, or CC (one nucleotide from each parent). SNPs are the basis for genome-wide association studies.

children and contribute to effortful control<sup>10</sup>. Second, an "evocative gene-environment correlation" is when a person's genetically influenced trait elicits, or evokes, a specific response from others in the environment. For example, it has been found that a person's genes are associated with being rated as "more likeable" by others, meaning how others perceive you as a social partner, and then likely interact with you, is influenced by your genes<sup>11</sup>. Third, an "active gene-environment correlation" describes the association of a person's genetically influenced traits and the environments they select. For example, the genetically influenced personality trait of socialization, measured in childhood, was associated with exposure to risky environments related to substance abuse in adolescence, in that children with



**Fig. 1** An example of a direct environmental transmission effect. Number of books in the home is thought to be an environmental causal effect on children's reading ability. Figure by ref. <sup>66</sup> available at <https://bit.ly/3gl8MVk> under a CC BY 4.0 license.



**Fig. 2** An example of how genetic confounding works (note, only one parent drawn, for simplicity). Parents share genes related to reading ability with their children, and also control the number of books in their home. This creates gene-environment interplay. It is important to note that the environmental effect may still have a causal role, even with gene-environment interplay. If genes play a role but are not modeled (as in Fig. 1), the correlation between the environmental measure and the child's trait is genetically confounded. Here, the role of genes is modeled, allowing for an estimation of the genetic effect and the environmental effect. Figure by ref. <sup>66</sup> available at <https://bit.ly/31c52z9> under a CC BY 4.0 license.

low socialization were exposed to more risky environments<sup>12</sup>. All three have the potential to cloud the true combination of genetic and environmental influences transmitted between parents and children (i.e., genetic confounding), but it is theorized that passive gene-environment correlations have a greater effect in childhood<sup>13</sup>, and as such passive gene-environment correlations are the focus of our review.

To give an example of how (passive) gene-environment correlations can result in genetic confounding in studies focused on the rearing environment, a high impact finding reported that parents with higher math anxiety have children with higher math anxiety, solely due to the home environment<sup>14</sup>. The authors attribute helping with math homework as the causal environmental

factor, concluding that parents with high math anxiety should not help with their children's math homework. This causal connection could exist, but equally parents with math anxiety also pass on genetic (and environmental) risks related to both lower math cognition and higher math anxiety<sup>15</sup>. Because this genetic transmission was not controlled for, causal claims and associated parenting advice are not justified.

Another example, this time from the medical literature, examined the intergenerational transmission of smoking behavior from parents to adolescents, concluding that "the attitudes, beliefs, and behaviors toward [adolescent] cigarette use are learned through [parent] modeling"<sup>16</sup>. Again, this study focused on the environmental transmission from parents to offspring without accounting for the transmission of genes related to smoking behaviors and risk-taking behaviors<sup>17</sup>. Furthermore, the authors conclude that smoking cessation interventions in adults can reduce smoking in subsequent generations. Parent-centered interventions might help to reduce adolescent smoking, but what is overlooked is that children carry their own genetic risks for smoking, and direct intervention with the adolescents<sup>18</sup> could more strongly influence their smoking behaviors.

We are certainly not the first to point out this familial transmission confound within the ecological literature. Indeed, nearly 40 years ago, Scarr and McCartney proposed that "the human experience and its effects on development depend primarily on the evolved nature of the human genome"<sup>13</sup>, and nearly 30 years ago Plomin and Bergeman<sup>19</sup> addressed the prevalence of genetic confounding by illustrating that genetic influences are found on most if not all environmental measures. Since then, several reviews have pointed to multiple examples from parental warmth to alcohol use to depression where causal pathways from parent behavior to child outcomes are reported, without accounting for genetic confounding<sup>20–22</sup>. These reviews have called for researchers to use caution with causal statements,

and to address genetic confounding in their limitations. Further, they have asked for journal editors and reviewers to be better watch-dogs in this endeavor; to insist that manuscripts adhere to these standards. However, based on our experience listening to conference presentations and reading press releases and newspaper articles, we believe these guidelines are not yet being met.

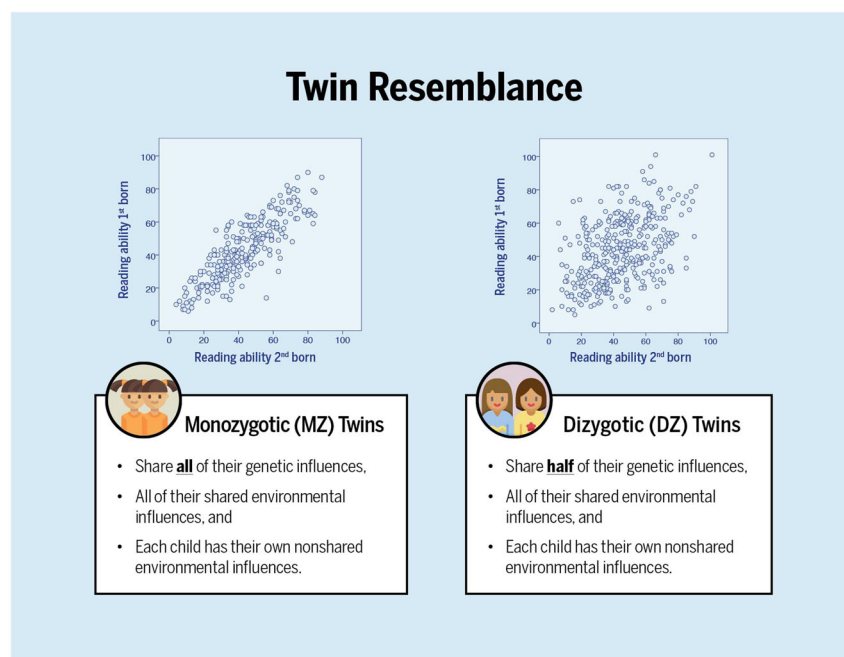
We believe a reason why these previous reviews have not successfully changed minds and methods is because they have not given actionable correlational design solutions to researchers outside of behavior genetics. Therefore, when faced with not doing the work or publishing work with only a potential genetic confound, researchers have chosen the latter. Therefore, in the following section we will give many possible solutions, from genetically sensitive designs to design solutions that work in lieu of genetically sensitive data, and finally, a renewed call for changes in reporting standards.

## WHAT RESEARCHERS CAN DO

The designs that we discuss below present a not all-encompassing but global overview of genetically sensitive designs and polygenic-scores (PGS) designs, and include a genetic-proxy control design (the "Familial Control Method"), which we recommend when genetically sensitive data are not available, as well as several other proxy control designs. These designs vary in how well they disentangle the genetic confound and in how challenging they are in terms of obtaining and analyzing the data.

### Genetically sensitive designs

Genetically sensitive designs are ideal for studying genetic and environmental influences and their interplay. These designs take advantage of samples of related individuals that differ in genetic relatedness (e.g., monozygotic and dizygotic twins; Fig. 3) or differ in environmental exposure (e.g., monozygotic twins reared apart).



**Fig. 3 The logic behind twin research.** The scatter plots depict how much the two types of (reared-together) twins resemble their co-twin on reading ability. Each dot represents the reading scores of both children within a pair. It can be seen that monozygotic twins are much more alike. From this, it can be concluded that differences between children are largely due to genetic differences. The data come from van Bergen et al.<sup>34</sup> and represent word-reading fluency test scores in Grade 2 of twin pairs with complete data. The score is the number of words read correctly within 1 min. In this sample, the monozygotic and dizygotic twin correlations were 0.84 and 0.46, respectively, which yield estimates using the Falconer formulas<sup>67</sup> of  $A = 0.76$ ,  $C = 0.08$ , and  $E = 0.16$  (see Fig. 4). Figure by ref.<sup>66</sup> available at <https://bit.ly/3k4w2Ji> under a CC BY 4.0 license.

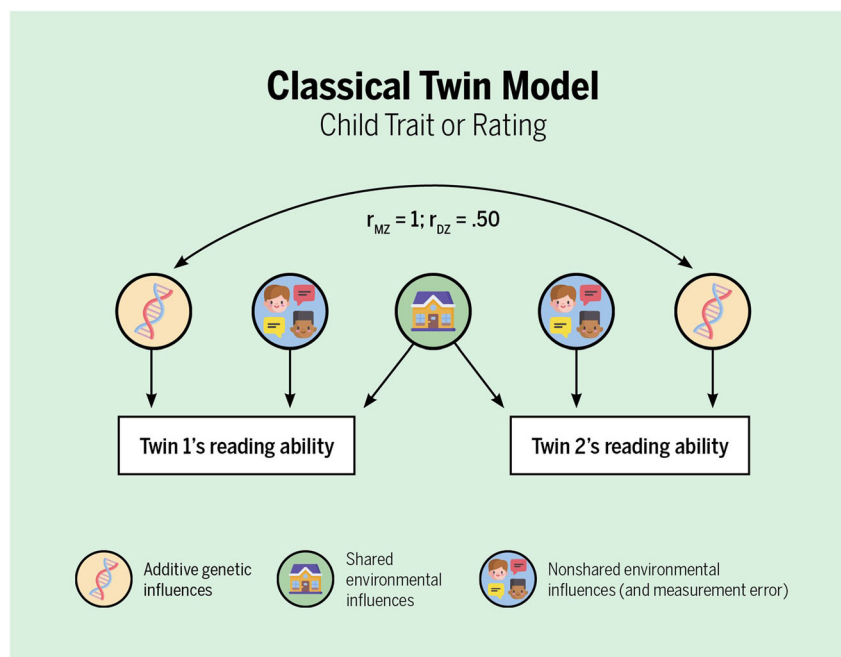
By far the most commonly used genetically sensitive design is the classical twin design. This design works because twins share either all (identical or monozygotic twins) or half (non-identical or dizygotic twins) of their genes<sup>23</sup>. Both types of twins share some parts of their environment such as their home, school, and neighborhood (referred to as common or shared environmental influences), and experience some aspects of their environments separately from each other such as peer groups, hobbies, or illness (referred to as unique or non-shared environmental influences). By comparing the average correlation between the two twins in a twin pair on a trait for monozygotic versus dizygotic twins, variance can be partitioned into additive genetic influences or heritability, shared environmental influences, and non-shared environmental influences (see Footnote 2 in the Supplementary Notes) (Fig. 4). Heritability of a trait is indicated if the correlation between monozygotic twins is higher than that of dizygotic twins. Shared environmental influences are estimated by subtracting the heritability estimate from the monozygotic twin correlation, and the estimated shared environmental influences is larger when the correlation coefficient between monozygotic versus dizygotic twins are close in magnitude. Finally, non-shared environmental influences are estimated by subtracting the monozygotic twin correlation from one.

With regard to disentangling possible genetic confounds and instead studying the direct effect of specific aspects of the rearing environment, classical twin studies are limited because both type of twins commonly share their rearing environments. For example, twin children growing up together are exposed to the same home library or household income, so monozygotic and dizygotic twin resemblance cannot be compared for these types of environmental measures. However, a classical twin study can begin to separate the direct effect of the home environment in two cases. First, child twins can be asked to individually rate their own rearing environment (Fig. 4). Since individual experiences are correlated with genetic predisposition, monozygotic twins often

rate their experiences of the home environment more similarly with each other than dizygotic twins do. Therefore, when child twins can report their own ratings of their rearing environment, these estimates can serve to differentiate monozygotic and dizygotic twins. Twins who are children might not differ in how many books they have in the home, but they will likely differ in how much their parents read to them, or how much their parents monitor their reading. In these cases, the extent to which aspects of children's rating of their rearing environment do not show entirely environmental influences, in other words, some heritability is measured on the "rearing environment", this infers that there is a genetic confound, via a passive gene–environment correlation<sup>4,19</sup>. Using child twin ratings of their rearing environment, Hanscombe et al.<sup>24</sup> found that 22% of the variance of chaos in the home was attributable to genetic factors, and moreover, 37% of association between chaos in the home and school achievement was due to shared genes. This suggests that this "environmental" variable of chaos in the home, measuring noise and lack of structure in the home, is partially genetically confounded. This means that chaos in the home does not have a completely direct, or causal role, on children's school achievement.

The second way that the classical twin model can be used to identify the direct effect of the home environment, free of genetic confounding, is by focusing on the environment that adult twins create (see Fig. 4, but replace "reading ability" with "books in their home" or the like). When twins are adults they can differ in how many books they own and the income of their household, so genetic and environmental influences on their home environments can be studied. These studies quantify genetic and environmental influences on the home environment that twins create<sup>4</sup>, but they do not quantify the influence of these home characteristics on outcomes in their offspring.

Other genetically sensitive designs that can address the direct effect of the rearing environment, after accounting for genetic



**Fig. 4** Simplified representation of the classical twin model. In behavioral-genetic models, the three sources of influences on individual differences are commonly labeled by the letters A, C, and E, respectively, stemming from **A**dditive genetic influences (also known as heritability, and sometimes represented by an  $h^2$  instead of an A), **C**ommon environmental influences (also known as shared environmental influences), and **E**nonshared environmental influences (and measurement **E**rror). Note that the latter are by definition uncorrelated between twins. See for a detailed representation of the classical twin model, for example, Figure A.9 in ref. <sup>23</sup>;  $r_{MZ}$  = monozygotic twin correlation;  $r_{DZ}$  = dizygotic twin correlation. Figure by ref. <sup>66</sup> available at <https://bit.ly/2Xkr29P> under a CC BY 4.0 license.



confounding, are adoption studies, within-family sibling studies, and twin-family studies<sup>25</sup>. In an adoption design, resemblance between adopted children and their biological parents is due to heritability (plus the prenatal environment). In contrast, resemblance between adopted children and their adoptive parents is fully due to the environment that the parents have provided. Another way to examine the rearing environment while partially controlling for genetic confounding is to use non-twin biological siblings within a family<sup>26,27</sup>. Because biological siblings, like dizygotic twins, share half of their genes, sibling resemblance on a trait suggests the influence of genetic factors along with some shared environmental influences. However, non-twin siblings can differ on several aspects of the rearing environment such as family size or parental health and age at birth, therefore, any dissimilarity between siblings can help to determine the influences of these non-shared aspects of the rearing environments on a given trait. Sibling designs have also recently incorporated the use of genome-wide PGS which strengthen their ability to control for confounding by disentangling direct genetic influences from gene–environment correlations<sup>28,29</sup>. PGS designs are discussed in more detail, below.

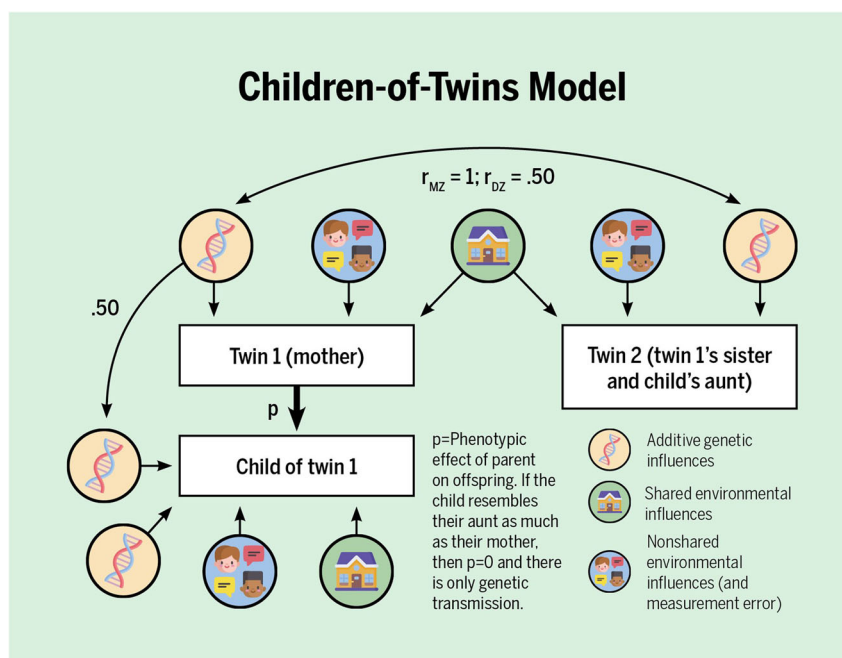
Twin-family studies include twins and their family members, like young twins and their parents, or adult twins and their children. The latter, referred to as children-of-twins design (Fig. 5), is particularly suitable to study the effect of children's rearing environment, free of genetic confounding<sup>21</sup>. Put simply, consider a mother who has an identical twin sister. The mother's son shares half of his genetic variants with his mother, but also with his aunt. If for a given trait he resembles his mother as much as his aunt, this suggests that the resemblance is fully due to shared genes. Conversely, if he is more like his mother than aunt, this demonstrates that the resemblance between mother and son is at least partially due to the environment provided by his mother (see Footnote 3 in the Supplementary Notes). There are even more complex extended twin family designs, described well in Keller et al.<sup>30</sup> and McAdams et al.<sup>31</sup>.

In sum, genetically sensitive designs can assess whether the rearing environment is influencing children's outcomes, outside of genetic confounds. Although they are observational and hence cannot establish causality, or the absence thereof, they can strongly infer causality above and beyond the majority of typical observational studies. An important point to make is if a genetically sensitive study suggests no direct causality of the rearing environment on children's outcomes, it does not imply that intervening is pointless. Successful parenting interventions are able to experimentally induce changes in parents' skills or behaviors, which then causally improve child outcomes. Thus, observational studies, such as all the genetically sensitive designs described here, and experimental studies (preferably randomized controlled trials<sup>32</sup>) answer related but different questions: the first on "what is", so causality in the natural situation, and the second on "what could be", so causality due to intervening<sup>33,34</sup>.

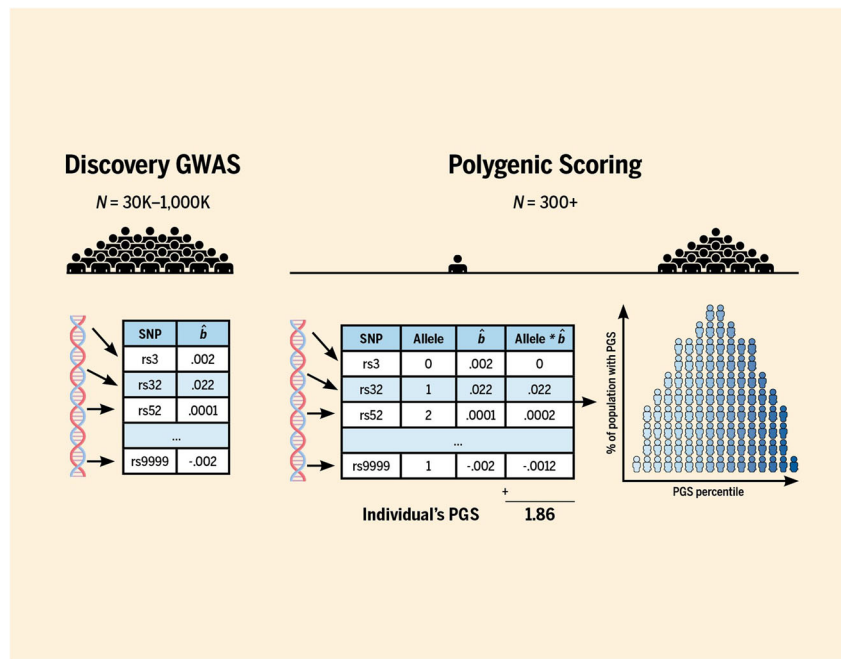
Returning to genetically sensitive designs, the disadvantage is that they require access to such data, which are challenging to collect and analyze. We note for a reader interested in using twin data to better answer their questions about the direct role of the rearing environment, twin datasets are increasingly becoming publically available. For example, TwinLife (<https://www.twin-life.de/en>), TEDS (<http://www.teds.ac.uk/researchers>), NLSY kinship links (<http://nlsy-links.github.io/NlsyLinks/>), Netherlands Twin Register (<http://tweelingenregister.vu.nl/research>), and others are available online or via application. In addition, there is a data sharing culture in the behavioral genetics community, and most will likely share when asked. We suggest that researchers consider using these resources to better test their research questions.

#### PGS designs

A new avenue to study intertwined genetic and environmental effects employs genome-wide PGS. This method relies on genome-wide association studies (GWASs) which pinpoint genetic variants (i.e., single nucleotide polymorphisms (SNPs)) that are linked to a trait (Fig. 6). The most powerful GWAS to date ( $N > 1$



**Fig. 5** Simplified representation of the children-of-twins model. In the given example, the (adult) twins are sisters. The genetic transmission (left hand side) is fixed at 0.50 because parents and children share 50% of their genome. The other set of genes that influence the child trait (bottom left) are genetic influences that explain variance in the child trait but not the parent trait. The crucial test for presence of environmental transmission is whether the  $p$ -path is significant. Note that 'child' can refer to child or adult offspring. See, for the full and detailed model, ref. <sup>31</sup>. Figure by ref. <sup>66</sup> available at <https://bit.ly/2D0aNYJ> under a CC BY 4.0 license.



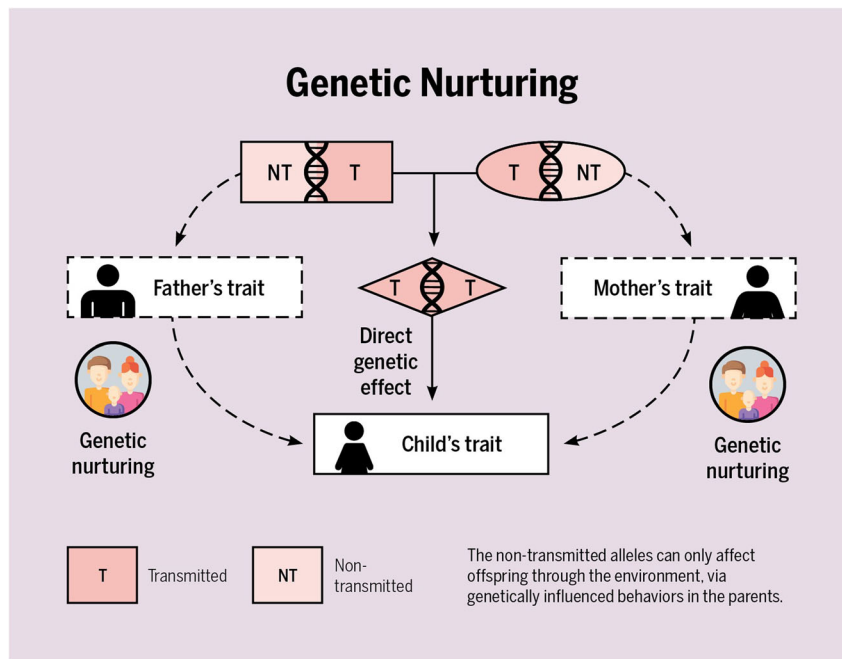
**Fig. 6** A polygenic score (PGS) indexes an individual's genetic predisposition for a certain trait or disease (see also<sup>68</sup>). Left panel: A published genome-wide association study (GWAS) serves as an external database. In an extremely large sample, a GWAS estimates tiny associations ( $\hat{b}$ ) between the trait of interest and millions of genetic variants. Specifically, the genetic variants studied are single-nucleotide polymorphisms (SNPs), located across the genome. Middle panel: Polygenic scoring can be done in a sample that was not part of the GWAS. For each individual in this sample, the SNP effects ( $\hat{b}$ ) are multiplied by the number of trait-associated alleles (0, 1, or 2) the person carries. These values are summed across all SNPs to arrive at the individual's PGS. Right panel: The resulting PGSs across individuals in that sample are normally distributed. If the trait of interest is a disorder, like ADHD, the individuals in the right tail have the highest genetic risk for developing ADHD. PGSs are not yet strong enough for predictions at the individual level, but see the main text for examples of how PGSs advance science at the group level. Figure adapted from ref. <sup>69</sup>. Figure by ref. <sup>66</sup> available at <https://bit.ly/2BPcCXP> under a CC BY 4.0 license.

million) has identified 1271 genetic variants associated with educational attainment<sup>35</sup>. Each of them has a tiny effect, but these tiny effects can be summed in a PGS. The PGS, calculated for all (unrelated) individuals in an independent sample, explains 12% of the variance in educational attainment. Note that twin studies estimate the heritability of educational attainment at 40%<sup>36</sup>, so the PGS currently captures less than one-third of this; the remainder is the “missing heritability”<sup>37</sup>.

As we speak, novel methods are being designed to disentangle nature and nurture that draw on PGS. Below, we list some examples of recent developments. First, Dolan et al.<sup>38</sup> bring PGS into the classical twin design. By doing so, one can estimate the gene–environment correlation, rather than assume it is absent. Second, Lee et al.<sup>35</sup> and Selzam et al.<sup>29</sup> found that for cognitive traits, the predictive power of a PGS within a family was about 50% lower than across unrelated individuals. The attenuation of the PGS' predictive power within families suggests that passive gene–environment correlations (as captured by the PGS) contribute to children's cognitive development. As a third example, both Kong et al.<sup>39</sup> and Bates et al.<sup>40</sup> separately proposed the same design incorporating parental and offspring PGS to disentangle environmental transmission from genetic transmission (i.e., account for the genetic confound between the home environment and child outcomes). In both cases, the researchers split the genetic variants of the parents in half—those that the parent had and had not transmitted to the offspring—and calculated for each half the PGS for educational attainment (Fig. 7). The researchers do this because of the biological fact that a parent only transmits a random half of their genes to their child. And for this design to work, both parents and their child must be included (but see ref. <sup>41</sup> for a work around). Amazingly, what the researchers found was both sets of parent PGS predicted adult offspring's educational attainment. The predictive value of the transmitted PGS was

unsurprising, as this captures directly transmitted genetic effects. But the predictive value of the non-transmitted PGS was not certain. If non-transmitted PGS influence children's traits, this effect must be environmental, likely acting through rearing behaviors that affect the child's development. Kong et al.<sup>39</sup> aptly coined this genetic effect through the rearing environment “genetic nurturing”. Belsky et al.<sup>42</sup> did a similar analysis but with an updated PGS score. Interestingly, when this design is expanded to include grandparents, there is little evidence for genetic nurturing from the grandparent generation<sup>43</sup>. Fourth, Wertz et al.<sup>44</sup> incorporated both PGS of mothers and children, as well as direct measures of parenting. They showed that mothers' cognitive stimulation explained the relation of the maternal non-transmitted PGS to child educational attainment. This indicated that there is a direct environmental transmission of parenting on children's outcomes, unconfounded by correlated genetic transmission. Finally, de Zeeuw et al.<sup>45</sup> and Willoughby et al.<sup>46</sup> both used the full genetic-nurturing design (employing DNA of children and both parents) and found (thereby replicated) genetic-nurturing effects on adults' educational attainment. Crucially, for outcomes in childhood, academic achievement and ADHD-symptoms, Zeeuw et al.<sup>45</sup> only found direct genetic effects; no genetic nurturing. They concluded that a large contributor to why the rearing environment predicts child outcomes may well be intergenerational transmission of genetic effects.

At the moment, measured genetic variants only explain small proportions of variance and the papers mentioned above may be seen by behavioral researchers as only a proof-of-principle. Nevertheless, these exciting developments will gain in strength when increasingly larger GWASs of all sorts of traits yield more refined PGS. By this we mean that PGSs will begin to explain more and more portions of the variance in outcomes we are interested in, with the hope that eventually they will reach the theoretical



**Fig. 7 Simplified representation of the genetic nurturing design.** In this design, one needs genotypes of parents and offspring, and a measured trait in the offspring generation only. The trait in the parents, for example educational attainment, is unobserved and indexed by a polygenic score of, in this example, educational attainment. The child receives half of the genotypes of father (top left) and mother (top right) and these transmitted alleles influence the child trait directly. The parental alleles that the child does not receive can still influence the child trait indirectly, via genetically influenced behaviors in the parents (denoted by the dotted genetic-nurturing paths). Genetic nurturing is present if the polygenic score of the *untransmitted* alleles explains a significant proportion of the variance in the child trait. The proportion of variance explained by the polygenic score of the *transmitted* alleles include both genetic nurturing and direct effects. Note that 'child' can refer to child or adult offspring. T = transmitted, NT = non-transmitted. Figure adapted from ref. <sup>39</sup>. Figure by ref. <sup>66</sup> available at <https://bit.ly/2PjpkRu> under a CC BY 4.0 license.

upper limit of SNP heritability. Even then, using them as a genetic control (i.e., as a covariate) will continue to underestimate the total genetic effects we are looking to control. PGSs account for only one type of genetic effect, namely common variants. There is increasing evidence that traits such as educational attainment are influenced by not only common variants, but also rare variants<sup>47,48</sup>. Another concern is that new work is indicating that a PGS is not a measure of only genetic variance. Instead, it likely represents not only causal genetic effects, but genetic ancestry, assortative mating, gene x environment interactions, direct environmental influences (i.e., genetic nurture), and environmental confounds from, for example, SES<sup>49,50</sup>. Therefore, a measure of genes (i.e., a PGS) can predict trait variance via environmental routes. This parallels our earlier notion that a measure of the environment can predict trait variance via genetic routes.

In summary, at the moment the PGS is not a perfect "genetic control", as it does not account for all of the genetic effects and also accounts for other effects, including the very environment we are interested in. But, we believe that next to no control, using a PGS as a statistical control is still better. Costs of genotyping are falling and the number of cohorts with genotype data is growing<sup>51</sup>. We predict that in the not-so-far future, using simply and cheaply collected genotypic information will become a regular part of the behavioral researchers' data collection protocol, especially as the predictive validity of the PGSs increases. This means that PGSs will allow researchers to partly control for genetic confounds in their models. We foresee that it will be easier to use PGSs than rely on genetically sensitive designs.

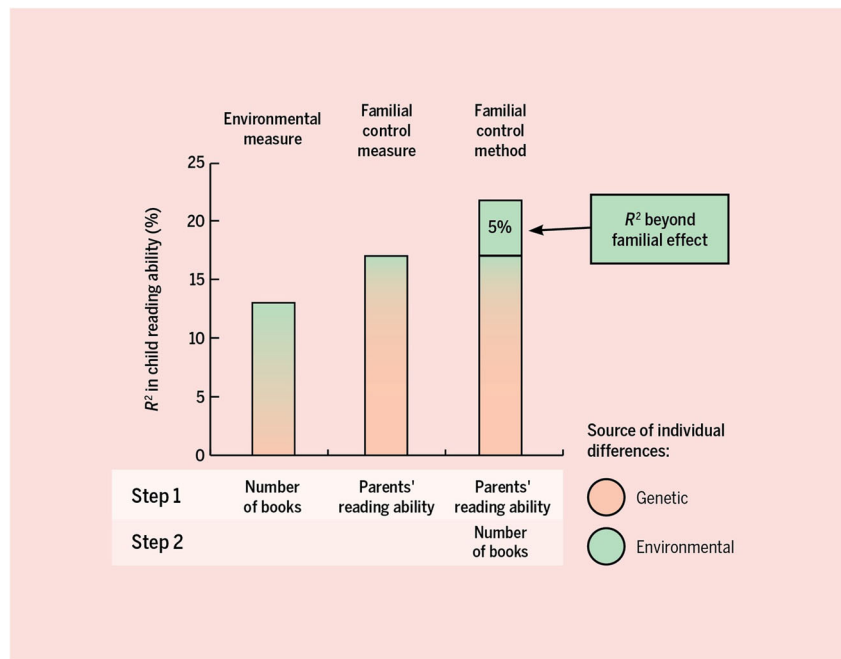
#### Genetic-proxy control designs: the Familial Control Method

The designs discussed above are the current gold-standards. However, as these types of samples discussed above are not

currently easy to collect and analyze, we propose here a useful work around<sup>52,53</sup>. Our colleagues can measure the same trait in both the parents as the child, and use the paternal and maternal traits as covariates. We advise to assess the traits in both parents (but acknowledge the challenge that brings), because the child shares only 50% of their genes with one parent, but all of their genes with both parents. Hence, both parents are needed to best tag the child's genetic liability. The parental traits, included as two covariates, then serve as a proxy for the familial transmission, including genetic transmission. In doing so, you have a proxy control for the familial effect. Hence, we term this method the Familial Control Method.

The Familial Control Method is designed for traits that are mostly transmitted from parent to child through genes rather than the environment, like reading ability<sup>54,55</sup> (see Footnote 4 in the Supplementary Notes). Van Bergen et al.<sup>53</sup> capitalized on this in studying whether children's reading ability is influenced by the home literacy environment, like reading habits of the parents and the number of books in the home. Analyses consisted of straight-forward step-wise regression analyses, illustrated in Fig. 8. The home literacy environment correlated with children's reading ability, but for most home-literacy indicators the effect was no longer significant after accounting for the reading ability of the parents. This suggests genetic confound rather than a genuine environmental effect. The one exception was the number of books children grow up with, which did explain variance over and beyond parents' reading skills (Fig. 8). This suggests a genuine environmental effect on children's reading by the number of books itself, or something related, like the value that the family places on reading<sup>45</sup>.

A similar approach was taken by Hart et al.<sup>52</sup> in studying the effect of the home numeracy environment on children's math ability. When a parent's math ability was included in the model,



**Fig. 8 Visualization of the Familial Control Method, in which a child outcome is predicted in a step-wise regression, with in the first step the familial control measure (i.e., the trait in both parents) and in the second step the measure of the environment.** The findings that are depicted here come from van Bergen et al.<sup>53</sup>. The key question is whether the environmental measure explains variance beyond the familial effect, as this indicates a genuine environmental effect. In the example given, this was 5% and significant. This was negligible and non-significant for the other environmental measures reported in ref.<sup>53</sup>. Figure by ref.<sup>66</sup> available at <https://bit.ly/2Pfjelh> under a CC BY 4.0 license.

some effects of aspects of the home numeracy environment on children's math ability were attenuated, but most held up. A note of caution is that the skills of only one parent could be obtained and controlled for, so the study lacked a proxy for the genetic liability passed on by the other parent. The authors concluded that doing more math-related activities with your children does seem to directly boost their math.

We advise researchers who are interested in applying the Familial Control Method to search first in the literature for adoption and twin-family studies. Such studies with the outcome trait of interest (e.g., reading ability) assessed in both the parent and the offspring generation, test whether parent to offspring transmission is mainly genetic or environmental in nature<sup>54,55</sup>. However, such studies are scarce. If such studies for the trait of interest do not exist, a good starting point are classical twin studies. Traits with no or a small influence of the shared environment (referred to as C), like neurological traits, are more likely to be transmitted just genetically compared to traits with large shared-environment influences, like social values. Results of meta-analyses of twin studies on a very large number of traits can be found in Polderman et al.<sup>2</sup> and the accompanying webtool (<http://match.ctglab.nl/>).

The Familial Control Method, using a parental trait as genetic proxy, is not watertight, and certain assumptions must be made for it to be effective for your research question (see Footnote 5 in the Supplementary Notes). First, if, for a certain trait, parent-child resemblance is not only due to genetic transmission but also environmental transmission, the Familial Control Method can be too conservative, as it also takes away some of the variance due to true environmental effects. However, one could argue that for many situations, being slightly too cautious in causal claims about environmental influences is less harmful than being too lax. However, this might not be the case for all researchers, and we encourage behavioral researchers to consider if being too conservative is actual harmful (e.g., for the effect of an unsafe home environment on child psychopathology). Second, as mentioned earlier, the trait measured in the parents should be

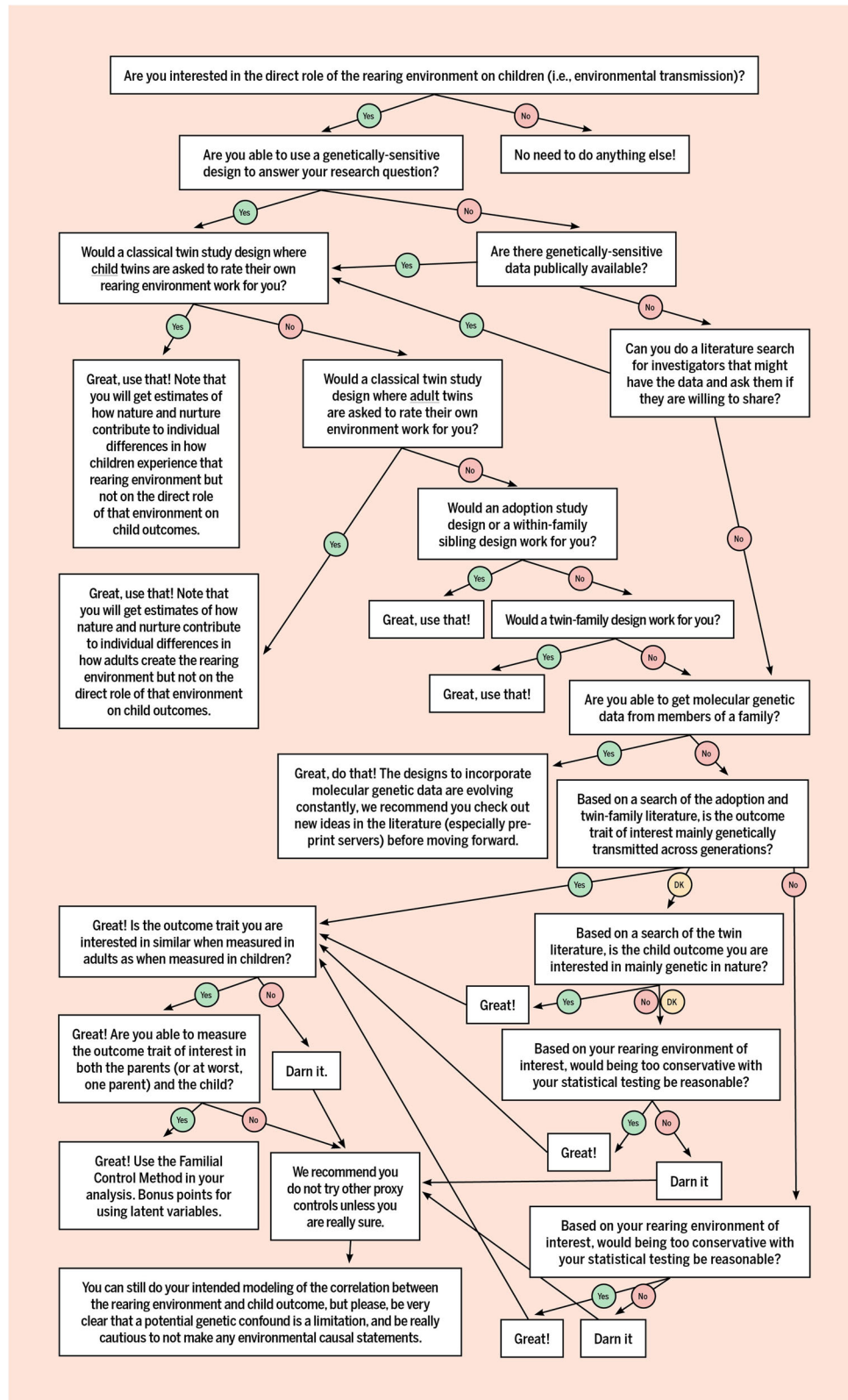
the same or highly similar as the trait measured in the child. This means that traits which are not at least reasonably the same in childhood as adulthood (i.e., across birth cohorts and across the lifespan) would not work in this design. So the trait should be at least reasonably measurement invariant and relatedly, show reasonable genetic stability. Fortunately, for many phenotypes, children's phenotypes are simply developmental precursors to the adult phenotypes (e.g., for reading ability<sup>56</sup>, and for ADHD<sup>57</sup>). A researcher must decide if the mentioned assumptions are appropriate for their trait of interest, but fortunately we do believe that these assumptions are reasonable for most to make. Third, correlations among the parent and child trait and the environmental measure of interest will be attenuated by measurement error. To reduce measurement error, one can do regressions according to the Familial Control Method in a structural equation modeling framework, with multiple indicators per construct. One can fit a model with as the outcome the latent child trait of interest, and as (correlated) predictors, the latent traits of both parents and the environmental measure of interest. If dropping the regression path 'environmental measure → child outcome' leads to a significantly worse model fit, this implies that the environmental measure is associated with the outcome above and beyond the familial effect. If the trait of interest is genetically transmitted, this equates to above and beyond genetic confounding, so suggests a direct environmental influence. The effect size here is given by the difference in explained variance in the child outcome of the models with and without the 'environmental measure → child outcome' path. Adopting a structural equation modeling framework with latent variables is especially advisable for constructs that are notoriously hard to measure reliably. Another advantage of this framework, compared to stepwise regressions, is that families with missing data can be retained.

It is likely the case that for most behavioral researchers interested in the direct role of the rearing environment, the Familial Control Method is currently the most feasible proxy genetic control. It does not require data of twin or adoption families, nor collecting DNA samples. In terms of prediction,



parental traits capture more of the variance in children's outcome than polygenic scores, so likely also capture more of the genetic confound. For the example of reading ability, it has been found

that the abilities of both of the parents explain 21% of the ability of children<sup>58</sup>. In comparison, polygenic scores (based on the educational-attainment GWAS) have been found to explain only



**Fig. 9** Decision flowchart for determining how to control for genetic confounding when examining the rearing environment. DK = don't know. Figure by ref. <sup>66</sup> available at <https://bit.ly/3gkM6Et> under a CC BY 4.0 license.

2–5% of reading ability in children<sup>59</sup>, and more recently 5–14% of “educational achievement” (including reading, writing, speaking, listening, and mathematics) in ages 7 to 16 years<sup>60</sup>. Certainly this proportion of variance explained from simple polygenic scores is not trivial, and the predictive ability of polygenic scores is anticipated to increase in the coming years. However, for most behavioral scientists the trait in the parents is not only easier to measure, but currently also a better predictor. On a related note, the value of parental traits as predictors of child outcomes has been used for decades in studying precursors of developmental disorders. In such family-risk studies, children with a family history of say dyslexia, attention-deficit/hyperactivity disorder or autism are followed from an early age, before the disorder manifests itself. These children have an increased risk to develop the disorder<sup>61,62</sup>.

#### Other proxy control designs

Other proxy controls such as sociodemographic factors (e.g., SES) have been used ubiquitously, but these statistical adjustments are not capable of accounting for genetic confounding as adequately as the Familial Control Method, for several reasons. First, although sociodemographic factors such as educational attainment have been significantly associated with genetic factors through twin studies (40%<sup>36</sup>) and GWAS (~14%<sup>35</sup>), the estimates are less than unity which indicates that genetic influences are not entirely responsible for individual differences in SES. Indeed, work examining the intergenerational transmission of SES has suggested that both genetic and environmental transmission occurs<sup>63</sup>. In this scenario where SES is transmitted through genetic and environmental pathways, when controlling for SES in data analyses, a proportion of variance attributable to other background or environmental factors is also being controlled for in the model, unintentionally leading to reduced associations between potentially important family-level predictors and child outcomes. In other words, you’d be throwing the baby out with the bathwater.

On the other hand, controlling for SES does not actually control for all of the genetic confounding. Say a researcher is interested in controlling for genetic confounding when examining the direct influence of books in the home on children’s reading. Parental SES is a proxy for parental reading skill, but not a perfect correlate (average correlation is 0.26<sup>64</sup>). Controlling for parental SES would not control for all of the potential genetic confounds on the association between books in the home and children’s reading ability. In conclusion, when controlling for SES, other potential sources of environmental variance are also being removed from the prospective models, and at the same time would not capture the extent of genetic confounding. We believe this would happen with other proxy control measures as well, outside of the Familial Control Method described above.

#### CONCLUSION

Here we have laid out numerous ways that genetic confounding can be controlled for when examining the rearing environment, summarized in a decision flowchart (Fig. 9). We can certainly foresee times that none of these options are possible. Therefore, we conclude that in those instances, our colleagues need to clearly mention the possible genetic confounding as a limitation, and to be cautious with any environmental causal statements which could lead to unnecessary parent blaming or to interventions that are a waste of time and resources. To return to our first example, expecting all homes to have plenty of books is an idealistic goal, as it would surround all children with the opportunity to read if they wished. But unfortunately, having the opportunity to read as one wishes does not unlock the code of reading for all children. Reading is a skill that requires direct

instruction and practice, and children with a family history of dyslexia themselves have a 45% chance of dyslexia despite adequate instruction and practice<sup>61</sup>. Simply having books around the home is not enough<sup>65</sup>, yet the message that parents are getting is that it is. The take home messages from that are that either parents who do not have the resources for a home library are hurting their children, or parents with children struggling to read are at blame because they did not have quite enough books in the home. This is unfair and inaccurate. In the end, we believe that it is important to discover true environmental effects as well as how genes and environments interplay, especially when malleable, because then we can focus as a field on creating and testing interventions that have a greater chance of directly improving children’s outcomes.

#### Reporting summary

Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

#### DATA AVAILABILITY

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

#### CODE AVAILABILITY

Code sharing not applicable to this article as no codes were generated during the current study.

Received: 25 November 2019; Accepted: 12 November 2020;

Published online: 08 January 2021

#### REFERENCES

1. Sikora, J., Evans, M. & Kelley, J. Scholarly culture: how books in adolescence enhance adult literacy, numeracy and technology skills in 31 societies. *Soc. Sci. Res.* **77**, 1–15 (2019).
2. Polderman, T. J. et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **47**, 702 (2015).
3. Turkheimer, E. Three laws of behavior genetics and what they mean. *Curr. Dir. Psychol. Sci.* **9**, 160–164 (2000).
4. Kendler, K. S. & Baker, J. H. Genetic influences on measures of the environment: a systematic review. *Psychol. Med.* **37**, 615 (2007).
5. Klahr, A. M. & Burt, S. A. Elucidating the etiology of individual differences in parenting: a meta-analysis of behavioral genetic research. *Psychol. Bull.* **140**, 544 (2014).
6. Plomin, R., DeFries, J. C. & Loehlin, J. C. Genotype-environment interaction and correlation in the analysis of human behavior. *Psychol. Bull.* **84**, 309–322 (1977).
7. Prinzie, P., Stams, G. J. J., Deković, M., Reijntjes, A. H. & Belsky, J. The relations between parents’ Big Five personality factors and parenting: a meta-analytic review. *J. Pers. Soc. Psychol.* **97**, 351 (2009).
8. Magnusson, D., & Stattin, H. In *Handbook of Child Psychology: Theoretical Models of Human Development* (eds Damon, W. & Lerner, R. M.) 685–759 (John Wiley & Sons Inc., 1998).
9. Baltes, P. B. & Baltes, M. M. In *Successful Aging: Perspectives From the Behavioral Sciences* (eds Baltes, P. B. & Baltes, M. M.) (Cambridge University Press, 1990).
10. Lemery-Chalfant, K., Kao, K., Swann, G. & Goldsmith, H. H. Childhood temperament: passive gene-environment correlation, gene-environment interaction, and the hidden importance of the family environment. *Dev. Psychopathol.* **25**, 51 (2013).
11. Burt, S. A. Genes and popularity: evidence of an evocative gene-environment correlation. *Psychol. Sci.* **19**, 112–113 (2008).
12. Hicks, B. M. et al. Gene-environment correlation in the development of adolescent substance abuse: selection effects of child personality and mediation via contextual risk factors. *Dev. Psychopathol.* **25**, 119 (2013).
13. Scarr, S. & McCartney, K. How people make their own environments: a theory of genotype → environment effects. *Child Dev.* **54**, 424–435 (1983).
14. Maloney, E. A., Ramirez, G., Gunderson, E. A., Levine, S. C. & Beilock, S. L. Intergenerational effects of parents’ math anxiety on children’s math achievement and anxiety. *Psychol. Sci.* **26**, 1480–1488 (2015).

15. Wang, Z. et al. Who is afraid of math? Two sources of genetic variance for mathematical anxiety. *J. Child Psychol. Psychiatry* **55**, 1056–1064 (2014).
16. Gilman, S. E. et al. Parental smoking and adolescent smoking initiation: an intergenerational perspective on tobacco control. *Pediatrics* **123**, e274–e281 (2009).
17. Vink, J. M., Willemsen, G. & Boomsma, D. I. Heritability of smoking initiation and nicotine dependence. *Behav. Genet.* **35**, 397–406 (2005).
18. Thomas, M. S., Kovas, Y., Meaburn, E. & Tolmie, A. What can the study of genetics offer to educators? *Mind Brain Educ.* **9**, 1–9 (2015).
19. Plomin, R. & Bergeman, C. S. The nature of nurture: genetic influence on “environmental” measures. *Behav. Brain Sci.* **14**, 373–386 (1991).
20. D’Onofrio, B. M., Turkheimer, E. N., Eaves, L. J. & Corey, L. A. The role of Children of Twins design in elucidating causal relationships between parent characteristics and child outcomes. *J. Child Psychol. Psychiatry* **44**, 1130–1144 (2003).
21. McAdams, T. A. et al. Accounting for genetic and environmental confounds in associations between parent and child characteristics: a systematic review of children-of-twins studies. *Psychol. Bull.* **140**, 1138 (2014).
22. Sherlock, J. M. & Zietsch, B. P. Longitudinal relationships between parents’ and children’s behavior need not implicate the influence of parental behavior and may reflect genetics: comment on Waldinger and Schulz (2016). *Psychol. Sci.* **29**, 154–157 (2018).
23. Knopik, V. S., Neiderhiser, J. M., DeFries, J. C. & Plomin, R. *Behavioral Genetics* (Macmillan Higher Education, 2017).
24. Hanscombe, K. B., Haworth, C., Davis, O. S. P., Jaffee, S. R. & Plomin, R. Chaotic homes and school achievement: a twin study. *J. Child Psychol. Psychiatry* **52**, 1212–1220 (2011).
25. D’Onofrio, B. M., Lahey, B. B., Turkheimer, E. & Lichtenstein, P. Critical need for family-based, quasi-experimental designs in integrating genetic and social science research. *Am. J. Public Health* **103**, S46–S55 (2013).
26. Petersen, A. H. & Lange, T. What is the causal interpretation of sibling comparison designs? *Epidemiology* **31**, 75–81 (2020).
27. Rowe, D. C. & Herstand, S. E. Familial influences on television viewing and aggression: a sibling study. *Aggress. Behav.* **12**, 111–120 (1986).
28. Brumpton, B. et al. Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. *Nat. Commun.* **11**, 1–13 (2020).
29. Selzam, S. et al. Comparing within-and between-family polygenic score prediction. *Am. J. Hum. Genet.* **105**, 351–363 (2019).
30. Keller, M. C., Medland, S. E. & Duncan, L. E. 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 (2010).
31. McAdams, T. A. et al. Revisiting the children-of-twins design: improving existing models for the exploration of intergenerational associations. *Behav. Genet.* **48**, 397–412 (2018).
32. Shadish, W. R., Cook, T. D. & Campbell, D. T. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (Houghton Mifflin Company, 2002).
33. Plomin, R., Shakeshaft, N. G., McMillan, A. & Trzaskowski, M. Nature, nurture, and expertise. *Intelligence* **45**, 46–59 (2014).
34. van Bergen, E. et al. Why do children read more? The influence of reading ability on voluntary reading practices. *J. Child Psychol. Psychiatry* **59**, 1205–1214 (2018).
35. Lee, J. J. et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
36. Branigan, A. R., McCallum, K. J. & Freese, J. Variation in the heritability of educational attainment: an international meta-analysis. *Soc. Forces* **92**, 109–140 (2013).
37. Maher, B. The case of the missing heritability. *Nature* **465**, 18–21 (2008).
38. Dolan, C. V., Huijskens, R. C., Minica, C. C., Neale, M. C. & Boomsma, D. I. Incorporating polygenic scores in the twin model to estimate genotype-environment covariance: exploration of statistical power. Preprint at <https://doi.org/10.1101/702738> (2019).
39. Kong, A. et al. The nature of nurture: effects of parental genotypes. *Science* **359**, 424–428 (2018).
40. Bates, T. C. et al. The nature of nurture: using a virtual-parent design to test parenting effects on children’s educational attainment in genotyped families. *Twin Res. Hum. Genet.* **21**, 73–83 (2018).
41. Young, A. I. et al. Mendelian imputation of parental genotypes for genome-wide estimation of direct and indirect genetic effects. Preprint at <https://doi.org/10.1101/2020.07.02.185199> (2020).
42. Belsky, D. W. et al. Genetic analysis of social-class mobility in five longitudinal studies. *Proc. Natl Acad. Sci. USA* **115**, E7275–E7284 (2018).
43. Liu, H. Social and genetic pathways in multigenerational transmission of educational attainment. *Am. Sociol. Rev.* **83**, 278–304 (2018).
44. Wertz, J. et al. Using DNA from mothers and children to study parental investment in children’s educational attainment. *Child Dev.* **91**, 1745–1761 (2020).
45. de Zeeuw, E. L. et al. Intergenerational transmission of education and ADHD: effects of parental genotypes. *Behav. Genet.* **50**, 221–232 (2020).
46. Willoughby, E. A., McGue, M., Iacono, W. G., Rustichini, A. & Lee, J. J. The role of parental genotype in predicting offspring years of education: evidence for genetic nurture. *Mol. Psychiatry* 1–9, <https://www.nature.com/articles/s41380-019-0494-1> (2019).
47. Cesarini, D. & Visscher, P. M. Genetics and educational attainment. *npj Sci. Learn.* **2**, 1–7 (2017).
48. Kendall, K. M. et al. Cognitive performance and functional outcomes of carriers of pathogenic copy number variants: analysis of the UK Biobank. *Br. J. Psychiatry* **214**, 297–304 (2019).
49. Trejo, S. & Domingue, B. W. Genetic nature or genetic nurture? Introducing social genetic parameters to quantify bias in polygenic score analyses. *Biodemography Soc. Biol.* **64**, 187–215 (2018).
50. Mostafavi, H. et al. Variable prediction accuracy of polygenic scores within an ancestry group. *Elife* **9**, e48376 (2020).
51. Evans, D. M., Moen, G.-H., Hwang, L.-D., Lawlor, D. A. & Warrington, N. M. Elucidating the role of maternal environmental exposures on offspring health and disease using two-sample Mendelian randomization. *Int. J. Epidemiol.* **48**, 861–875 (2019).
52. Hart, S. A., Ganley, C. M. & Purpura, D. J. Understanding the home math environment and its role in predicting parent report of children’s math skills. *PLoS ONE* **11**, e0168227 (2016).
53. van Bergen, E., van Zuijlen, T., Bishop, D. & de Jong, P. F. Why are home literacy environment and children’s reading skills associated? What parental skills reveal. *Read. Res. Q.* **52**, 147–160 (2017).
54. Swagerman, S. C. et al. Genetic transmission of reading ability. *Brain Lang.* **172**, 3–8 (2017).
55. Wadsworth, S. J., Corley, R. P., Hewitt, J. K., Plomin, R. & DeFries, J. C. Parent-offspring resemblance for reading performance at 7, 12 and 16 years of age in the Colorado Adoption Project. *J. Child Psychol. Psychiatry* **43**, 769–774 (2002).
56. Wadsworth, S. J., Corley, R. P., Hewitt, J. K. & DeFries, J. C. Stability of genetic and environmental influences on reading performance at 7, 12, and 16 years of age in the Colorado Adoption Project. *Behav. Genet.* **31**, 353–359 (2001).
57. Kan, K.-J. et al. Genetic and environmental stability in attention problems across the lifespan: evidence from the Netherlands twin register. *J. Am. Acad. Child Adolesc. Psychiatry* **52**, 12–25 (2013).
58. van Bergen, E., Bishop, D., van Zuijlen, T. & de Jong, P. F. How does parental reading influence children’s reading? A study of cognitive mediation. *Sci. Stud. Read.* **19**, 325–339 (2015).
59. Selzam, S. et al. Genome-wide polygenic scores predict reading performance throughout the school years. *Sci. Stud. Read.* **21**, 334–349 (2017).
60. von Stumm, S. et al. Predicting educational achievement from genomic measures and socioeconomic status. *Dev. Sci.* **23**, e12925 (2020).
61. Snowling, M. J. & Melby-Lervåg, M. Oral language deficits in familial dyslexia: a meta-analysis and review. *Psychol. Bull.* **142**, 498 (2016).
62. Musser, E. D. et al. Shared familial transmission of autism spectrum and attention-deficit/hyperactivity disorders. *J. Child Psychol. Psychiatry* **55**, 819–827 (2014).
63. Björklund, A., Jäntti, M. & Solon, G. Nature and nurture in the intergenerational transmission of socioeconomic status: evidence from Swedish children and their biological and rearing parents. *B.E. J. Econ. Anal. Policy* **7**, <https://www.nber.org/papers/w12985> (2007).
64. Sirin, S. R. Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Rev. Educ. Res.* **75**, 417–453 (2016).
65. Kim, J. S. The effects of a voluntary summer reading intervention on reading activities and reading achievement. *J. Educ. Psychol.* **99**, 505 (2007).
66. Hart, S. A., Little, C. & van Bergen, E. Genetic confounding figures. *Figshare*. <https://figshare.com/account/home#/collections/5080187> (2020).
67. Falconer, D. S. *Introduction to Quantitative Genetics* (Ronald Press, NY, New York, 1960).
68. Rayner, C. P. is for polygenic risk scores. *EDITLab Blog, KCL*. <https://blogs.kcl.ac.uk/editlab/2018/07/31/p-is-for-polygenic-risk-scores/> (2018).
69. Belsky, D. W. & Harden, K. P. Phenotypic annotation: using polygenic scores to translate discoveries from genome-wide association studies from the top down. *Curr. Dir. Psychol. Sci.* **28**, 82–90 (2019).

## ACKNOWLEDGEMENTS

S.A.H. is supported by Eunice Kennedy Shriver National Institute of Child Health & Human Development Grants HD052120 and HD095193. Views expressed herein are those of the authors and have neither been reviewed nor approved by the granting agencies. E.v.B. is supported by NWO VENI fellowship 451-15-017 (“Decoding the gene-environment interplay of reading ability”) and ZonMw grant 531003014 (“Genetics as a research tool: A natural experiment to elucidate the causal effects of social mobility on health”). She is a member of the Consortium on

Individual Development (CID; NWO Gravitation grant 024.001.003) and of Research Institute LEARN!.

## AUTHOR CONTRIBUTIONS

The first and last author contributed equally to this work.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41539-020-00079-z>.

**Correspondence** and requests for materials should be addressed to S.A.H.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021