

Special Issue Article

Gene–environment interaction using polygenic scores: Do polygenic scores for psychopathology moderate predictions from environmental risk to behavior problems?

Robert Plomin¹ , Agnieszka Gidziela², Margherita Malanchini² and Sophie von Stumm³

¹Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK, ²School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK and ³Department of Education, University of York, York, UK

Abstract

The DNA revolution has energized research on interactions between genes and environments (GxE) by creating indices of G (polygenic scores) that are powerful predictors of behavioral traits. Here, we test the extent to which polygenic scores for attention-deficit/hyperactivity disorder and neuroticism moderate associations between parent reports of their children's environmental risk (E) at ages 3 and 4 and teacher ratings of behavior problems (hyperactivity/inattention, conduct problems, emotional symptoms, and peer relationship problems) at ages 7, 9 and 12. The sampling frame included up to 6687 twins from the Twins Early Development Study. Our analyses focused on relative effect sizes of G, E and GxE in predicting behavior problems. G, E and GxE predicted up to 2%, 2% and 0.4%, respectively, of the variance in externalizing behavior problems (hyperactivity/inattention and conduct problems) across ages 7, 9 and 12, with no clear developmental trends. G and E predictions of emotional symptoms and peer relationship problems were weaker. A quarter (12 of 48) of our tests of GxE were nominally significant ($p = .05$). Increasing the predictive power of G and E would enhance the search for GxE.

Keywords: behavior problems; genotype–environment interaction; polygenic scores; twins

(Received 14 January 2022; revised 7 July 2022; accepted 7 July 2022; First Published online 23 September 2022)

Introduction

Gene–environment interaction (GxE) refers to environmental effects that depend on genetic effects, that is, genetic sensitivity to the environment (Kendler & Eaves, 1986). In relation to developmental psychopathology, GxE can be thought of as genetic moderation of the association between environmental factors, such as parenting, and children's behavior problems.

GxE is important because it recognizes that one size does not fit all and offers the possibility of personalized tailoring of children's environments based on their genetic propensities. Moreover, weak environmental effects in the population could have strong effects on children with particular genetic proclivities. GxE is the genetic extension of phenotypic research on differential reactivity to the environment (Garmezzy & Rutter, 1983; Slagt et al., 2016; Wachs & Gruen, 1982), such as research on the goodness-of-fit model of temperament (Thomas & Chess, 1977). GxE also has the virtue of moving beyond nature versus nurture to consider their interplay.

GxE is distinct conceptually from gene–environment correlation, which denotes experiences that are correlated with genetic propensities, that is, genetic exposure to environmental effects

and genetic mediation of associations between environmental factors and psychopathology (Plomin et al., 1977). In this paper, we define GxE as an interaction in the statistical sense that the effects of G and E are conditional on one another, independent of their main effects. There are other ways to construe the interplay between G and E (Rutter et al., 2006).

GxE is not *interactionism*, the view that environmental and genetic threads in the fabric of behavior are so tightly interwoven that they cannot be disentangled, as implied by the often-repeated phrases “the organism is a product of its genes and its past environment” (Anastasi, 1958, p. 197). Interactionism is a truism at the level of the individual, but it is not true for individual differences in a population. As discussed later, environmental effects on differences between children can exist without genetic effects, genetic effects can exist without environmental effects, and environmental and genetic effects can interact, that is, the effects of environments can depend on genes.

Although in this article we occasionally use words like *effect* and *explanation* instead of *correlation* and *prediction*, these words are only used in their statistical sense, not to imply causation. Our aim is to predict behavior problems from genes, environmental measures, and their interaction, without regard to causality (Plomin & von Stumm, 2021). That is, we use measures deemed to be measures of the environment, even though these measures show significant genetic influence, the issue of the “nature of nurture” (Plomin, 1994; Plomin & Bergeman, 1991). Similarly, we do not know the causal biological and environmental pathways through

Corresponding author. Robert Plomin, email: robert.plomin@kcl.ac.uk

Cite this article: Plomin, R., et al. (2022). Gene–environment interaction using polygenic scores: Do polygenic scores for psychopathology moderate predictions from environmental risk to behavior problems? *Development and Psychopathology* 34: 1816–1826, <https://doi.org/10.1017/S0954579422000931>

which inherited differences in DNA affect development (Pingault et al., 2018). As we shall see, our focus on the relatively simple issue of prediction raises complicated issues even though we do not attempt to address issues of causality.

Three stages of GxE research

For the first century of genetic research on human behavior, quantitative genetic designs such as identical and fraternal twins and adoptive and nonadoptive relatives were used to estimate genetic and environmental components of observed (phenotypic) variance. These are anonymous components of variance in the sense that specific genetic and environmental factors are not identified. Although quantitative genetic theory recognizes that GxE can contribute to phenotypic variance, quantitative genetic designs in themselves can only provide indirect glimpses of GxE (Jinks & Fulker, 1970; Plomin et al., 1977).

We differentiate three stages in GxE research.

Stage 1

The first stage of GxE research incorporated measures of the environment into twin and adoption designs. In the twin design, environmental measures made it possible to ask whether heritability differs as a function of environment, a limited type of GxE. For example, it has been reported that greater parental warmth and directiveness is associated with greater heritability of conduct problems (Burt et al., 2013). For this and similar findings, replication is needed because the power demands for detecting significant differences in heritability are daunting (Hanscombe et al., 2012).

Environmental measures can be integrated in more powerful ways in the adoption design (Plomin et al., 1977). For example, psychopathology of birth (biological) parents provides an estimate of their adopted-away children's genetic risk free of postnatal environmental influence of the parents. Parenting of the adoptive parents of these children can be used to estimate environmental risk free of parental genetic confounds. GxE can then be assessed as the statistical interaction between these G and E indices.

GxE is illustrated in Figure 1 in a 2x2 framework in which parenting (low vs. high risk) and genetic propensity (low vs. high risk) predict children's behavior problems. Main effects of parenting and genetics can occur without interaction (panel A). GxE can occur without main effects (panel D), which is called a disordinal (cross-over) interaction in that the environment has opposite effects depending on children's genetic propensities. For example, differential susceptibility theory predicts a type of disordinal interaction in which genetically sensitive children are more affected by their environments, both for better and for worse (Belsky & Pluess, 2009; Pluess & Belsky, 2010). However, a recent systematic analysis of differential susceptibility theory found little support for the theory (Cree et al., 2021).

The more likely type of GxE is an ordinal interaction, in which genetic effects magnify or diminish environmental effects. Panel B in Figure 1 shows an example in which high environmental risk disproportionately affects children who are at high genetic risk. In psychiatric genetics, this type of GxE is known as *diathesis-stress* (Gottesman, 1991; Monroe & Simons, 1991; Paris, 1999). That is, children at genetic risk for psychopathology, the diathesis, are especially sensitive to the effects of stressful environments. For behavioral problems, GxE also occurs in the opposite direction (panel C) in which high-risk environments overwhelm genetic propensities and low-risk environments allow genetic differences to be expressed.

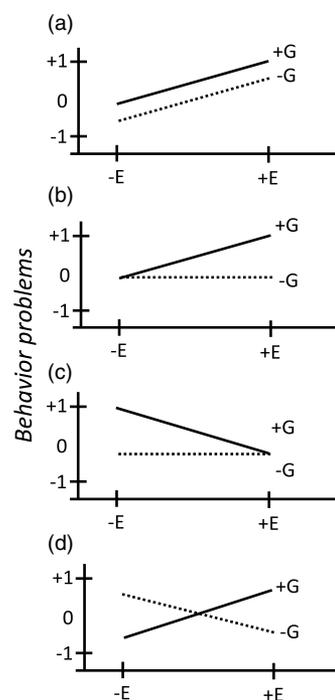


Figure 1. A 2x2 illustration of the prediction of children's behavior problems from environments (low vs. high risk), from a genetic estimate (low vs. high risk), and from their interaction (GxE). See text for explanation.

A better use of the data is to analyze the variables in a continuous rather than dichotomous manner with multiple regression models in which main effects of G and E are modeled before testing GxE. The GxE interaction term can be a "dummy" variable created by the product of the main variables of G and E. In the Colorado Adoption Project (CAP), some ordinal GxE effects were found in infancy (Plomin & DeFries, 1985) and early childhood (Plomin et al., 1988). As an example of an ordinal interaction shown in panel C, high parental control damped down genetic differences relevant to difficult temperament in infancy, whereas low parental control facilitated expression of genetic differences. Overall, however, only a chance number of significant GxE effects emerged from dozens of analyses of this type.

In contrast, a more recent adoption study, the Early Growth and Development Study (EGDS) (Leve et al., 2019), reported several GxE effects for behavior problems. Similar to CAP, the association between parental control and children's behavior problems was moderated by genetic risk (Leve et al., 2009). Significant GxE was also reported in EGDS for internalizing problems (Brooker et al., 2014) and externalizing problems (Lipscomb et al., 2014). Most GxE in EGDS was of the ordinal type shown in panel C in Figure 1 (Cree et al., 2021).

Stage 2

The DNA revolution fueled the second and third stages of GxE research by making it possible to incorporate direct DNA measures of genetic propensities of children (Plomin, 2018). The huge benefit of these direct DNA measures is that they can be used in research on any sample of unrelated or related children, circumventing the need for special samples of twins and adoptees.

Until a decade ago, it was necessary to genotype one DNA variant at a time for each individual. The time and expense of this process meant that researchers could only genotype a few DNA

variants. This led to the second stage of GxE research, which genotyped a few “candidate” genes, usually genes coding for neurotransmitters presumed to be related to psychopathology. Two of the most cited papers in behavioral genetics reported GxE of the diathesis-stress type (Panel B in Figure 1). The first study showed that childhood maltreatment was only associated with adult anti-social behavior for individuals who carried a specific version (allele) of the monoamine oxidase A gene that causes lower levels of the monoamine oxidase A, which is involved in metabolizing a broad range of neurotransmitters (Caspi *et al.*, 2002). Similarly in the second study, the effect of stressful life events on depression was strongest for individuals who carried an allele that increased serotonin transport (Caspi *et al.*, 2003).

These early candidate-gene GxE findings triggered an explosion of research on GxE (Keers & Pluess, 2017), but many studies were underpowered to detect reasonable effect sizes, and failures to replicate accumulated (Dick *et al.*, 2015; Duncan & Keller, 2011). Some journals imposed a ban on GxE reports unless they included a replication study (Hewitt, 2012).

Stepping back from GxE, the fundamental problem with a candidate-gene strategy is that we now know that single DNA variants hardly ever account for as much as one percent of the variance in the population. Most reported single-variant associations with behavioral traits are false (Chabris *et al.*, 2012).

Stage 3

About 15 years ago, a technological advance sparked the DNA revolution by making it possible to genotype hundreds of thousands of DNA variants quickly and inexpensively. The technology was a small DNA array that genotyped the most common type of DNA variant, a single-nucleotide polymorphism (SNP). This tool, called a SNP chip, facilitated a strategy that is the opposite of the candidate-gene approach. Genome-wide association (GWA) is a hypothesis-free method that looks for associations using millions of variants across the entire genome genotyped for each individual on a SNP chip. GWA research has shown that the heritability of complex traits and common disorders is caused by thousands of inherited DNA differences, each with miniscule effects. Success in GWA research came only after sample sizes in the tens and hundreds of thousands were achieved to reach the power needed to detect these tiny effects (Plomin, 2018).

Thousands of GWA with a particular trait can be aggregated in a genetic index of that trait for each individual, called a polygenic score. Polygenic scores have many labels, but we prefer genome-wide polygenic scores (GPS) to distinguish polygenic scores that are derived from all SNPs, usually tens of thousands, throughout the genome that are associated with the trait from polygenic scores that are based on only a few selected SNPs (Plomin & von Stumm, 2018). Although GWA studies require huge samples, associations from these GWA studies can be used to create polygenic scores for any sample (Choi *et al.*, 2020). In contrast to candidate genes, polygenic scores can predict up to six percent of the variance of behavior problems in childhood (Gidziela *et al.*, 2021).

Polygenic scores mark the third stage of GxE research. Lessons can be learned from candidate-gene GxE research to avoid the pitfalls of underpowered studies and questionable research practices (Domingue *et al.*, 2020). It has been estimated that large GxE effects, whether indexed by a single gene or a polygenic score, account for about 1% of the variance, which requires sample sizes of about 600 to reach 80 percent power to detect them as calculated by Duncan and Keller (2011). For moderate GxE effects accounting

for 0.1 percent of the variance, sample sizes in the tens of thousands are needed to reach 80 percent power.

Most behavioral GxE research using polygenic scores has involved adult psychiatric disorders, although no solidly replicated GxE findings have emerged as yet (e.g., Arnau-Soler *et al.*, 2019; Bogdan *et al.*, 2018; Colodro-Conde *et al.*, 2018; Kandaswamy *et al.*, 2021; Robinson & Bergen, 2021). Reports of GxE in developmental psychopathology are beginning to emerge. Two early studies using a polygenic score for major depressive disorder reported GxE for childhood adversity predicting adult depression (Mullins *et al.*, 2016; Peyrot *et al.*, 2014), but a later study could not replicate the finding (Peyrot *et al.*, 2018). In another study, no GxE was found for childhood adversity predicting adult psychotic disorders (Trotta *et al.*, 2016).

Recent reports of significant GxE in developmental psychopathology in which G was indexed by polygenic scores include critical parenting and depression (Nelemans *et al.*, 2021), family and neighborhood stress on conduct problems (Bares *et al.*, 2020), and childhood adversity and emotion dysregulation and psychosis proneness (Pries *et al.*, 2020). However, at least as many studies reported no significant GxE, including research on peer victimization and resilience (Armitage *et al.*, 2021), maltreatment and attention-deficit/hyperactivity disorder (ADHD) symptoms (He & Li, 2021), and parenting and adolescent externalizing problems (Ksinan *et al.*, 2021).

The present study

In this paper, we test GxE for behavior problems in childhood using data from up to 6687 twins from the Twins Early Development Study (Rimfeld *et al.*, 2019). To avoid associations driven by rater bias or contemporaneous events, we focused on parent reports of their children’s environments at 3 and 4 years as they predict teacher ratings of behavior problems at 7, 9 and 12 years. Based on previous analyses (Gidziela *et al.*, 2021), we selected polygenic scores for ADHD and neuroticism that were found to be most predictive of teacher ratings of childhood behavior problems. We also selected environmental measures at ages 3 and 4 known to predict teacher ratings of behavior problems (Gidziela *et al.*, 2022). Our primary aim was to compare effect sizes for G, E and GxE in predicting behavior problems as rated by teachers. A secondary aim was to explore G, E and GxE effects on behavior problems developmentally across childhood.

Method

Participants

Our sample consists of twins born in England and Wales between 1994 and 1996 who have participated in the Twins Early Development Study (TEDS; Rimfeld *et al.*, 2019). TEDS is a longitudinal study designed to investigate the development of behavior, cognition and communication, as well as developmental problems. The original TEDS sample involved over 13,000 twin pairs, from whom data collection took place when the twins were aged 2, 3, 4, 7, 8, 9, 10, 12, 14, 16, 18 and 21. The sample of TEDS twins is representative of the UK population in terms of ethnicity and socioeconomic status (SES) (Rimfeld *et al.*, 2019).

In addition to phenotypic data, a subsample of 10,346 TEDS twins (i.e., one twin per pair for 3706 twin pairs and 3320 pairs of DZ twins) were genotyped using one of two genotyping platforms (AffymetrixGeneChip 6.0 and Illumina

HumanOmniExpressExome-8v1.2) in two waves, 5 years apart. A detailed genotyping protocol is available (Selzam et al., 2018).

The present analyses included a subsample of TEDS twins with complete genotype data along with parent reports of environmental risk and discipline in preschool (ages 3 and 4) and teacher ratings of behavior problems in childhood and early adolescence (ages 7, 9 and 12), resulting in up to 6687 individuals included in our analyses. The sample included only one member of a twin pair for all monozygotic twins and for half of the dizygotic twin pairs (4561 unrelated twins). We also included 2126 dizygotic co-twins to increase the power of our analyses; the inclusion of co-twins slightly affects statistical significance (which were not adjusted for lack of independence) but does not affect estimates of effect size, which is the focus of our analyses.

Measures

In this section, we outline our measures of environmental risk, behavior problems, and polygenic scores. Details about the measures are available in the TEDS data dictionary (<https://www.teds.ac.uk/datadictionary/home.htm>). We selected environmental measures in early childhood and polygenic scores that we knew from previous analyses predicted behavior problems in the school years (Gidziela et al., 2021; Gidziela et al., 2022) in order to increase the likelihood of finding ordinal GxE (Duncan & Keller, 2011).

Environmental measures

Parent-reported environmental measures in early childhood were selected based on their prediction of teacher-rated behavior problems in childhood, following a procedure described in detail elsewhere (Gidziela et al., 2022). Using two criteria, this procedure reduced the number of environmental items from several hundred to eight items using two criteria. The first criterion was a phenotypic correlation greater than 0.20 between the environmental measure and at least one of the behavior problem measures. The second criterion excluded highly correlated environmental measures using a penalized elastic net regularization with training and test iterations.

We began with two environmental risk measures at age 3 and at age 4: a family-general measure based on variables that are the same for co-twins and a twin-specific measure based on variables that differentiated between co-twins. The family-general environmental risk composite was computed as a standardized mean of five standardized scores: family SES based on both parents' education and occupation and mothers' age at first contact, prenatal and perinatal medical risk, household chaos (the Confusion, Hubbub and Order Scale; Matheny et al., 1995), maternal postnatal depression (the Edinburgh Postnatal Depression Scale (Cox et al., 1987), and life events such as changes to marital status, new siblings, mother's pregnancy, job changes and serious illness/accident).

The twin-specific environmental risk measure was computed as a standardized mean of the same variables that were included in the family-general environmental risk measure, with the addition of standardized scores of twin medical risk factors (4 items), parental discipline scale (6 items) and a parental feelings scale (7 items) (Deater-Deckard et al., 1998). At each age we also included a separate item about smacking and shouting because of its predictive-ness (Gidziela et al., 2022).

Because the environmental measures were correlated with one another across age 3 and 4 (see Supplementary Figure 1), we used exploratory factor analysis (EFA) as a data reduction technique (see Supplementary Note 1). EFA analysis yielded two clear factors

(see Supplementary Figure 2). The first factor was a general environmental risk factor that included both twin-specific and family-general environmental risk composites at ages 3 and 4, as well as SES (see Supplementary Figure 3). The second factor was a discipline factor including parental discipline composites and smacking/shouting items at ages 3 and 4.

The environmental risk and discipline factors correlated moderately (0.40). Using the two-factor structure suggested by the EFA, we created factor scores derived from a confirmatory factor analysis (CFA) because CFA can account for data missingness using Full Information Maximum Likelihood. Although we were simply using CFA as a data reduction technique to handle data missingness, we achieved semi-independence of the EFA and CFA by conducting EFA on one randomly selected member of each twin pair and conducting CFA on the other twin (see Supplementary Note 2). Results of the CFA are illustrated in Supplementary Figure 4 and model fit indices presented in Supplementary Table 1. The two factors, which we refer to as environmental risk and discipline, were used in our subsequent analyses.

Behavior problems

We assessed teacher ratings of hyperactivity/inattention, conduct problems, emotional symptoms, and peer relationship problems at ages 7, 9 and 12 using the Strengths and Difficulties Questionnaire (Goodman, 1997). Teacher ratings of behavior problems were obtained via mail and included a total of 20 items at each age, that is, five items for each of the four scales. The items were rated on a three-point Likert scale (certainly true; sometimes true; not true), with items scored in the direction of greater problems or reversed where necessary so that higher scale scores indicated more behavior problems.

Polygenic scores

GPS were obtained from well-powered GWA studies of ADHD (Demontis et al., 2019; $N = 55,374$) and neuroticism (Luciano et al., 2018) ($N = 329,821$). These GPS were previously found to predict childhood behavior problems (Gidziela et al., 2021). Quality control procedures and construction of polygenic scores have been described elsewhere (Selzam et al., 2018).

Analysis

Environmental and behavior problems variables were residualized for the effects of age and sex. Both polygenic scores were corrected for the effects of genotyping chip and 10 principal components of ancestry prior to downstream analyses. All regression analyses were conducted using stats for R (R Core Team, 2021).

Main effects of polygenic scores and environment on behavior problems

The effects of genotype and environment on teacher-rated behavior problems at ages 7, 9 and 12 were examined using regression analysis, with the GPS (ADHD or neuroticism) and environment (environmental risk or discipline) as predictors. Effects of each of the four predictors were examined separately to estimate the variance they predicted in teacher-rated hyperactivity/inattention, conduct problems, emotional symptoms, and peer relationship problems at ages 7, 9 and 12. Although these single-variable regressions are the same as simple correlations, we present the results in the regression model to facilitate comparisons with the multiple regressions we used to test for GxE.

Interaction between polygenic scores and environment

To estimate the proportion of variance explained by the interaction between genotype and environment (GxE), we used multiple regression analysis, testing for GxE after controlling for the joint effects of GPS and environmental factors (i.e., G+E). We first estimated G+E prediction of each of the four teacher-rated behavior problems at ages 7, 9 and 12 from the GPS (ADHD or neuroticism) and environmental factors (environmental risk or discipline). Subsequently we ran another regression model with the GPS, environmental factor, and the interaction between them as predictors. To estimate the proportion of variance in behavior problems predicted by GxE independent of G+E, we subtracted the variance explained (R^2) by the G+E model from the R^2 of the GxE model, which is a hierarchical multiple regression analysis.

Results

Descriptive statistics, including means, standard deviations and sample sizes for the environmental measures and behavior problems are presented in Supplementary Table 2. The full correlation matrix between the environmental measures, GPS, and behavior problems across the three ages is shown in Supplementary Figure 1.

Main effects of polygenic scores and environment on behavior problems

Figure 2 summarizes results for the prediction of behavior problems from the environmental factors and GPS, with details of the regression analyses in Supplementary Table 3. The strongest GPS prediction was for the ADHD GPS predicting hyperactivity/inattention at age 7, accounting for 2.0% of the variance. Results were similar for teacher ratings of behavior problems at ages 7, 9 and 12. The strongest E prediction was for the environmental risk factor predicting hyperactivity/inattention at age 9, which accounted for 2.7% of the variance. The environmental risk factor accounted on average for 2.3% of the variance in hyperactivity/inattention across ages and 2.1% of the variance in conduct problems, but less than 1% for emotional symptoms and peer relationship problems. The discipline factor yielded a similar degree of prediction, accounting for 2.1% variance in hyperactivity/inattention across ages, 1.5% in conduct problems, 0.2% for emotional symptoms and 0.4% for peer relationship problems.

The environmental and GPS predictions were higher for externalizing problems (hyperactivity/inattention and conduct problems) than for internalizing problems (emotional symptoms and peer relationship problems). For GPS, prediction from the ADHD GPS was greater than from the neuroticism GPS: 1.8% versus 0.1% for hyperactivity/inattention, 1.2% versus 0.1% for conduct problems, and 0.2% versus 0.1% for peer relationship problems, averaged across the three ages. The one exception was that the neuroticism GPS explained slightly more variance in emotional symptoms than the ADHD GPS (0.4% vs. 0.2% across ages). No clear developmental trends emerged across childhood.

Interactions between polygenic scores and environment

Figure 3 presents the proportion of variance in behavior problems explained jointly by the GPS and environmental factors (G+E) and the variance explained by the interaction between them (GxE) after controlling for G+E. Details of the multiple regression results are presented in Supplementary Table 5.

The joint G+E prediction in Figure 3 reflects the individual G and E predictions shown in Figure 2. G and E together predict less variance than the sum of variance explained by G and E separately (Figure 2) because G and E are correlated (see Supplementary Figure 1). The R^2 estimate of the joint effect of G and E discounts the covariance between G and E.

The strongest G+E predictions are for hyperactivity/inattention. For the ADHD GPS and environmental risk factor, G and E together predict 3.7% of the variance in hyperactivity/inattention on average across the three ages. Similar results were obtained for discipline (3.6%). For the other three Strengths and Difficulties Questionnaire scales, the ADHD GPS and the environmental risk factor jointly predicted 3% of the variance in conduct problems and 0.8% for both emotional symptoms and peer relationship problems. Again, the neuroticism GPS was less predictive. Neuroticism GPS and environmental risk jointly predicted 2.3% of the variance for hyperactivity/inattention, 2.1% for conduct problems, 0.9% for emotional symptoms, and 0.7% for peer relationship problems. Results for the discipline factor were similar: 2.1% for hyperactivity/inattention, 1.6% for conduct problems, 0.5% for emotional symptoms and 0.4% for peer relationship problems. No developmental trends were apparent.

GxE results are also shown in Figure 3. The most striking aspect of Figure 3 is how little variance is predicted by GxE beyond the joint effects of G and E. The proportions of variance explained by G+E as well as GxE are listed in Supplementary Table 4. Tests of significance from analysis of variance are presented in Supplementary Table 5. Overall, 12 of the 48 tests of GxE were nominally significant ($p < .05$). Eight of these significant GxE tests involved the ADHD GPS. However, the significant GxE that predicted the most variance, the ADHD GPS and the discipline factor predicting conduct problems at age 12, only accounted for 0.4% of the variance. Because the variance explained by GxE is less than our power to detect GxE (0.5%), caution is warranted until these interactions are replicated. No developmental trends emerged.

Figure 4 plots these 12 significant GxE interactions in 2x2 analyses in which children were selected in the +/- 1 SD quadrants of G and E. Analysis of variance results for these 2x2 analyses are included in Supplementary Table 6.

The 2x2 plots generally show main effects of G and E but only hint at GxE. Eight of the 12 main effects for G were significant and, in each case, the high G group had more behavioral problems than the low G group. Ten of the 12 main effects for E were significant and there were more behavioral problems in the high E group than the low E group. However, for GxE, only three of the 12 analyses yielded significant GxE, even though these 12 comparisons were selected because they yielded significant GxE in the continuous analyses (Supplementary Table 5). In these three cases, as well as most of the other GxE, the GxE were ordinal and of the diathesis-stress type. That is, stressful environments exacerbated genetic risk for behavior problems. For example, in the significant GxE in the upper left-hand corner of Figure 4, high parental discipline at ages 3 and 4 predicted more hyperactivity/inattention problems at age 9 primarily for children with high GPS scores for hyperactivity.

Discussion

To illustrate issues involved in GxE analysis, we explored GxE in the prediction of teacher ratings of behavior problems at ages 7, 9 and 12 from parent reports of children's environments at ages 3 and 4. We found significant but modest GxE with no developmental trends across childhood. Although one quarter (12 of 48) of our

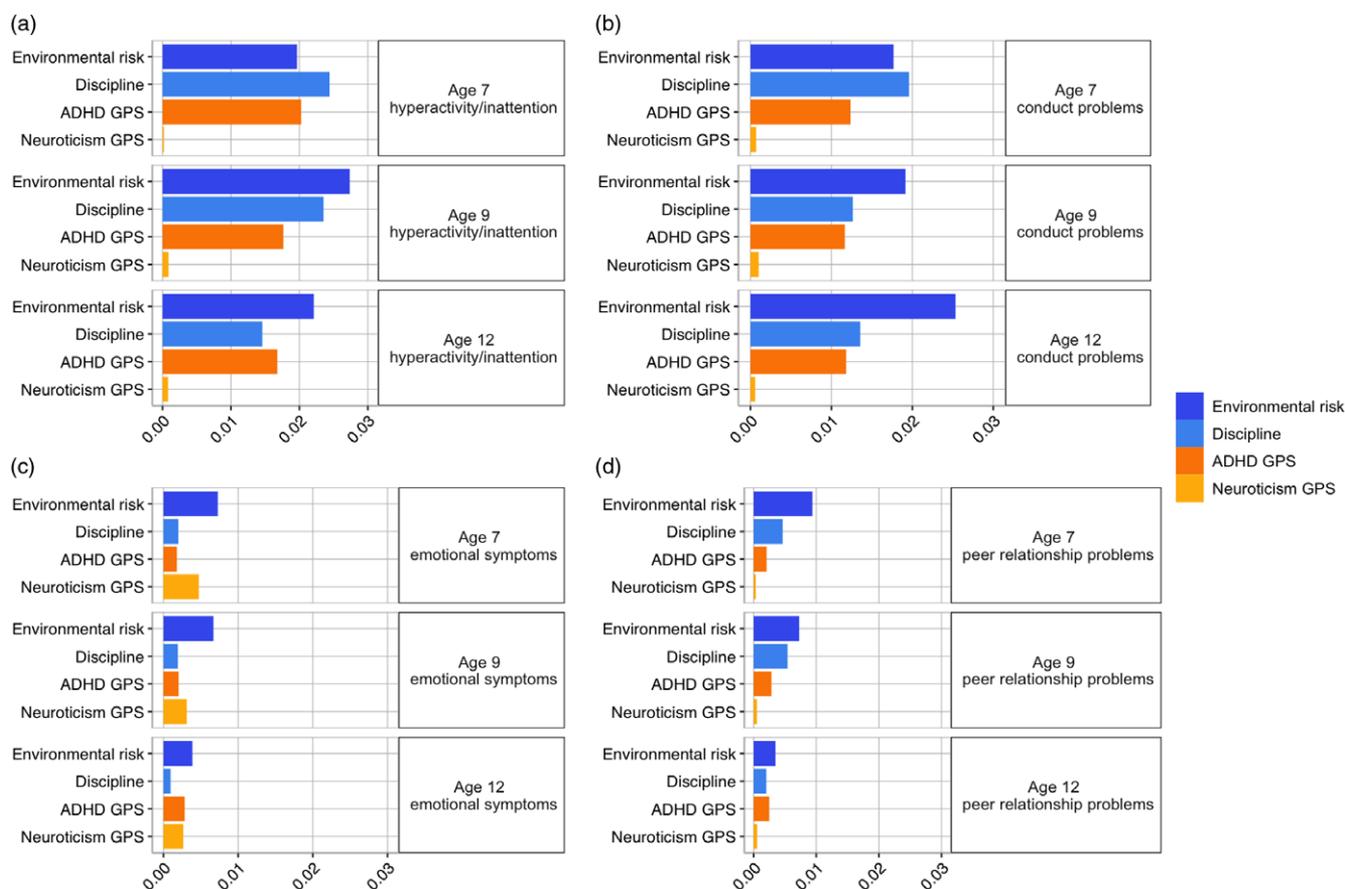


Figure 2. Variance explained by G (ADHD and neuroticism GPS) and E (environmental risk and discipline factors) predicting teacher-rated hyperactivity/inattention (a), conduct problems (b), emotional symptoms (c) and peer relationship problems (d) at ages 7, 9 and 12.

GxE tests were significant, the largest GxE effect accounted for only 0.4% of the variance. This GxE was an example of ordinal diathesis-stress interaction in that the difference in conduct problems between children with high and low ADHD GPS scores at age 12 was greater for children whose parents were high in discipline, as shown in Figure 4. That is, children with high ADHD GPS scores showed disproportionately more conduct problems when their parents were disciplinarians.

The average effect size of all 12 significant GxE is 0.2%, which represents 50% power for a nominal p value of 0.05. Because our sample size provides substantial power to detect “large” GxE effects that account for 1% of the variance (Duncan & Keller, 2011), we can safely exclude the possibility of GxE accounting for more than 1% of the variance for these measures of E, G and behavior problems at these ages in our sample. Smaller GxE effects could be useful in understanding causal pathways between G and E (Götz et al., 2021), but they are of limited utility from the perspective of prediction. The goal of prediction is to account for as much variance as possible without regard for explanation; the goal of explanation is to deduce causality, usually without regard for prediction (Shmueli, 2010; Yarkoni & Westfall, 2017). Although causal explanation will remain the long-term goal for psychology, prediction has immediate practical utility for identifying individuals at risk and is a necessary first step towards explanation (Plomin & von Stumm, 2021). One example is the success of genomic research after candidate-gene analyses were supplanted by hypothesis-free GWA analyses. Polygenic scores flourished because they are

constructed to maximize the prediction of a target trait. Another example for the value of prediction without explanation is artificial intelligence in which machine learning explicitly eschews explanation.

When G and E are combined to predict behavior problems (Allegrini et al., 2020), they predict less variance than the sum of their individual predictions because G and E indices correlate. For example, the ADHD GPS and discipline factor correlate 0.14 and 0.16, respectively, with teacher ratings of hyperactivity/inattention at age 7 (see Supplementary Figure 1). The sum of the square of these correlations (0.05) exceeds the R^2 from the multiple regression (0.04) because R^2 adjusts for the correlated effect of G and E on the trait. GE correlation is an important topic beyond the scope of this article (Krapohl et al., 2017); its relevance to GxE is that it is necessary to control for the correlation between G and E in GxE analyses (Keller, 2014), which is what the multiple regression does.

The potential for identifying ordinal GxE depends on the effect sizes of G and E (Duncan & Keller, 2011). For this reason, a larger issue that emerges from our results is that the effect sizes of our G and E predictors of behavior problems are modest. The strongest E predictor was the environmental risk factor predicting hyperactivity/inattention at age 9, although this prediction accounted for only 2.7% of the variance. The strongest G predictor, the ADHD polygenic score predicting hyperactivity/inattention at age 7, accounted for 2.0% of the variance.

Much recent research has been directed towards increasing the effect size of polygenic score predictions. Different methods for

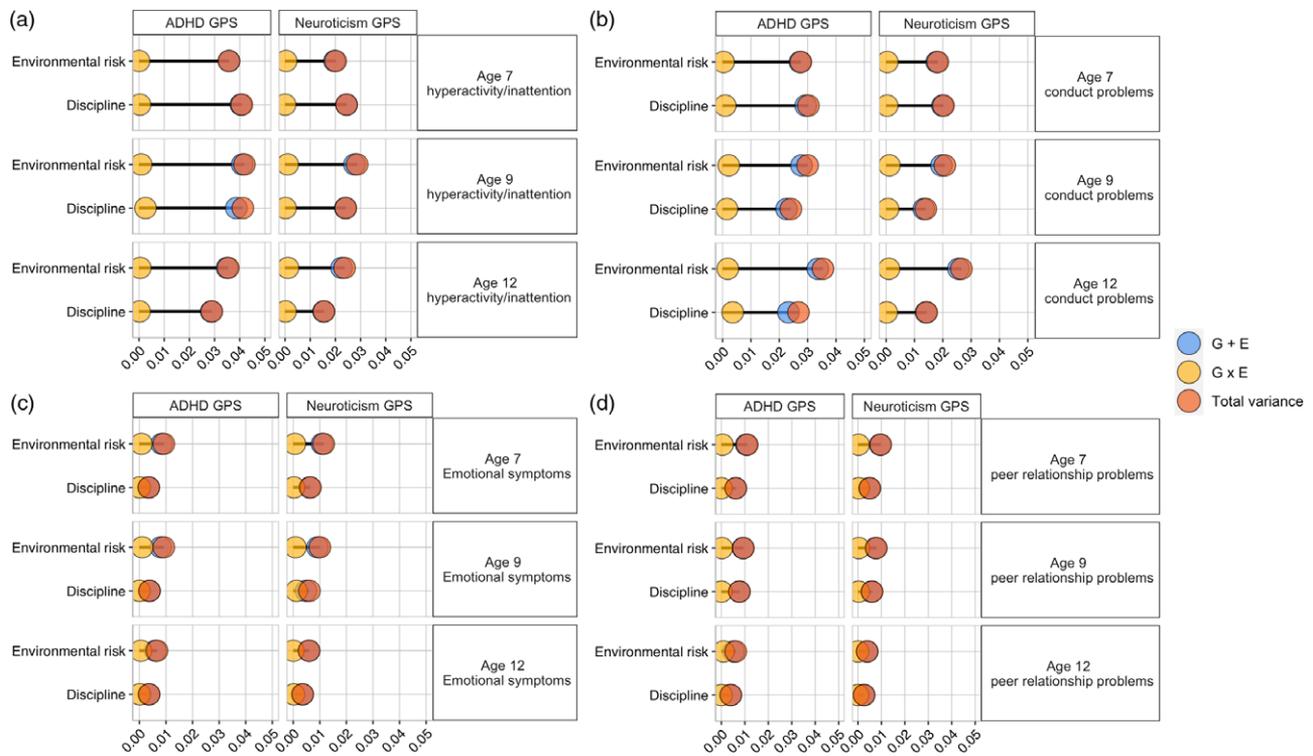


Figure 3. G+E and GxE prediction of teacher-rated hyperactivity/inattention (a), conduct problems (b), emotional symptoms (c) and peer relationship problems (d) at ages 7, 9 and 12.

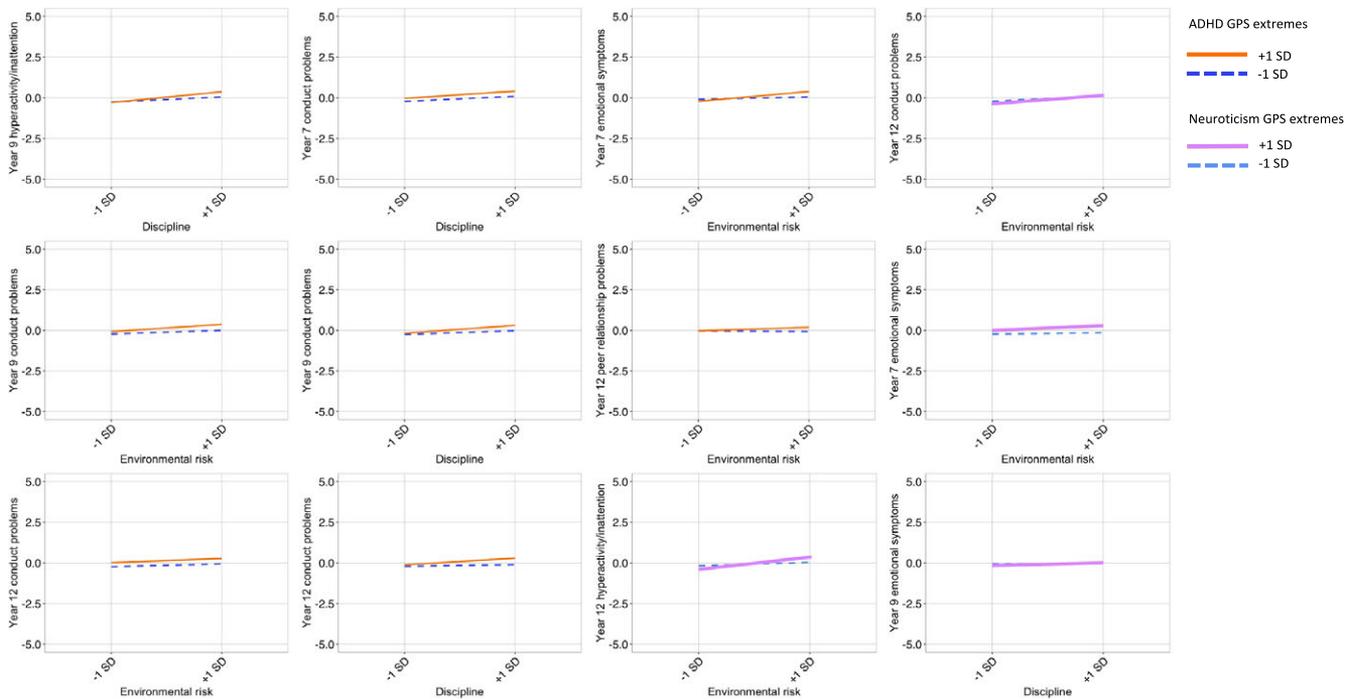


Figure 4. Significant GxE interactions in 2x2 analyses of extreme (+/- 1 SD) ADHD GPS, neuroticism GPS and environmental factor scores. Solid line = +1 SD for GPS; dashed line = -1 SD for GPS.

constructing polygenic scores have not had much effect (Allegrini *et al.*, 2019; Pain *et al.*, 2020), but using multiple polygenic scores improves prediction of cognitive traits (Allegrini *et al.*, 2020; Grotzinger *et al.*, 2019; Krapohl *et al.*, 2018) and behavior problems

(Gidziela *et al.*, 2021). By far the most effective boost to predictive power has come from increasing the sample sizes of GWA studies so that they are powered to detect small effect sizes of individual DNA variants. For example, the most predictive polygenic score

in the behavioral sciences is based on GWA analyses of educational attainment (years of education). The predictive power of this polygenic score in independent samples not included in the GWA study increased from 2% with a GWA sample of 125,000 (Rietveld et al., 2013) to 3% with a sample size of 294,000 (Okbay et al., 2016) to more than 10% with a sample size of 1.1 million (Lee et al., 2018) to about 14% with a sample size of 3 million (Okbay et al., 2022).

GWA samples for behavior problems are a long way from these sizes. At present, the most powerful polygenic score predictor of behavior problems is derived from a GWA analysis of 20,000 diagnosed ADHD cases and 35,000 controls (Demontis et al., 2019). This ADHD polygenic score, used in the current study, predicted 5.5% of the liability variance (Nagelkerke's R^2 for analysis of cases versus controls) in an independent sample (Demontis et al., 2019). On the other hand, the polygenic score for neuroticism that was also used in our study was derived from a GWA analysis with a sample size of nearly 330,000 adults but predicted only 2.8% of the variance of self-reported neuroticism in an independent sample (Luciano et al., 2018). These earlier findings on out-of-sample prediction of polygenic scores align with our results that the ADHD polygenic score predicted hyperactivity/inattention (1.8%) better than the neuroticism polygenic score predicted emotional problems (0.4%).

For environmental measures, less effort has been directed towards increasing – or even noticing – predictive power. Much environmental research, and psychological research in general, still focuses on statistical significance rather than effect size (Funder & Ozer, 2019). Despite a century of research on environmental predictors of psychopathology, surprisingly little can be concluded about prediction effect sizes. For example, even for the most well studied disorders of schizophrenia and especially bipolar disorder, it is not possible to say how much variance in liability can be predicted by environmental factors (Robinson & Bergen, 2021). Studies that have examined the combined prediction of multiple environmental risk factors on the development of behavioral problems point to their weak effect size. For example, a study across four large European cohorts showed that perinatal factors (family SES, maternal drinking and smoking during pregnancy, maternal stress, breast-feeding, and gestational age) accounted for less than 1.5% of the variance in aggressive behavior in childhood and adolescence (Malanchini et al., 2020).

The fundamental problem for prediction is that environmental research has nothing remotely comparable to the genetic code in DNA sequence or the technological advance of SNP chips that assess millions of DNA variants quickly, accurately, and inexpensively. A comparison between the genome and what could be called the *environome* is instructive (von Stumm & d'Apice, 2022). In the genome, millions of inherited differences in DNA sequence have been identified and their tiny individual associations with a trait can be summed to create polygenic scores. Another huge advantage of polygenic scores as predictors is that they do not change after conception because SNPs are inherited DNA variants. This is why a polygenic score derived from a GWA of adult psychopathology can be usefully applied to investigate psychopathology in childhood, as in our study. In contrast, for the *environome*, there is no fundamental unit of transmission and the *environome* changes throughout development.

Pushing the analogy with DNA further, much environmental research is still anchored in the “candidate environment” stage, with researchers looking for a few reasonable factors, like parenting, that are assumed to have major effects on children's outcomes.

But what if, like genetics, environmental influences involve thousands of miniscule effects? It has been argued that it is unrealistic to expect large effects for psychological phenomena (Götz et al., 2021). One way to increase effect size is to aggregate small effects (Funder & Ozer, 2019), which was key to the success of polygenic scores. For example, environmental risks across events and across time have long been used as cumulative risk indices in developmental research (Rutter, 1981; Widaman, 2021). But what environmental variables should be included such composites? As a parallel to the SNP chip, it has been suggested that digital technologies could be used to collect naturalistic observations in real time, “capturing the *environome* across levels, dimensions, and time in unprecedented depth and detail” (von Stumm & d'Apice, 2022, p. 6).

More fundamental issues about assessing the environment go beyond the scope of this paper. For instance, what is the environment and what is its relationship to experience? Much discussion about environmental assessment assumes that the environment is “out there,” passively imposed on children. This view suggests that what might be most important is children's active construction of their experience (Plomin, 1994), but this is not usually assessed by current environmental measures. Supporting this perspective is a recent study showing that subjective rather than objective maltreatment in childhood predicted psychopathology later in life (Danese & Widom, 2020).

Another issue is that genetic research has shown that environmental influences on behavior problems are not shared by children growing up in the same family, called nonshared environment (Plomin, 2011). This important clue about the salient environmental influences has not yet been used in the construction of current measures of the environment. In the present study, we took a step in this direction by considering environmental measures that were to some extent specific to each twin separately from measures that are the same for all children in the family, such as SES. However, using available environmental measures that can differ between children in a family is a long way from environmental measures designed to maximize experiential differences between siblings (Daniels & Plomin, 1985).

Children's active construction of their experience might also explain why environmental influences on behavior problems are experienced differently by children growing up in the same family as events are filtered through their unique constellation of perceptions, cognitions, and emotions. Active construction of experience could result in nonshared environments that are largely idiosyncratic and stochastic (Plomin, 2018).

Ultimately what will be needed are large-scale collaborations, along the lines of GWA consortia, that can reach the statistical power needed to detect small effects of G, E and their interplay. GWA consortia could be a practical way to begin by incorporating environmental measures in large biobanks that already have GWA genotype data, especially new longitudinal cohorts that could track developmental changes in the *environome* (von Stumm & d'Apice, 2022). The difficult trade-off will be between quality and depth of environmental assessments versus the need for very large samples.

Increasing the power of G and E to predict behavior problems is valuable in itself for the field of developmental psychopathology. It might also eventually lead to the identification of robust GxE.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0954579422000931>

Funding statement. We gratefully acknowledge the ongoing contribution of the participants in the Twins Early Development Study (TEDS) and their families. TEDS has been supported by a program grant to R.P. from the UK Medical

Research Council (MR/M021475/1 and previously G0901245), with additional support from the US National Institutes of Health (AG046938) and the European Commission (602768; 295366). This work was supported by a Nuffield Foundation award (EDO/44110) to S.v.S, who is a recipient of a British Academy Mid-Career Fellowship (MF21/210017).

Conflicts of interest. None.

References

- Allegrini, A. G., Karhunen, V., Coleman, J. R. I., Selzam, S., Rimfeld, K., von Stumm, S., Pingault, J.-B., & Plomin, R. (2020). Multivariable G-E interplay in the prediction of educational achievement. *PLOS Genetics*, *16*, e1009153. <https://doi.org/10.1371/journal.pgen.1009153>
- Allegrini, A. G., Selzam, S., Rimfeld, K., von Stumm, S., Pingault, J. B., & Plomin, R. (2019). Genomic prediction of cognitive traits in childhood and adolescence. *Molecular Psychiatry*, *24*, 819–827. <https://doi.org/10.1038/s41380-019-0394-4>
- Anastasi, A. (1958). Heredity, environment, and the question 'how?' *Psychological Review*, *65*, 197–208. <https://doi.org/10.1037/h0044895>
- Armitage, J. M., Wang, R. A. H., Davis, O. S. P., & Haworth, C. M. A. (2021). A polygenic approach to understanding resilience to peer victimisation. *Behavior Genetics*, *52*, 1–12. <https://doi.org/10.1007/s10519-021-10085-5>
- Arnau-Soler, A., Macdonald-Dunlop, E., Adams, M. J., Clarke, T.-K., MacIntyre, D. J., Milburn, K., Navrady, L., Hayward, C., McIntosh, A. M., Thomson, P. A., Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., Agerbo, E., Air, T. M., Andlauer, T. F. M., & Sullivan, P. F. (2019). Genome-wide by environment interaction studies of depressive symptoms and psychosocial stress in UK Biobank and generation Scotland. *Translational Psychiatry*, *9*, 14. <https://doi.org/10.1038/s41398-018-0360-y>
- Bares, C. B., Chartier, K. G., Karriker-Jaffe, K. J., Aliev, F., Mustanski, B., & Dick, D. (2020). Exploring how family and neighborhood stressors influence genetic risk for adolescent conduct problems and alcohol use. *Journal of Youth and Adolescence*, *49*, 1365–1378. <https://doi.org/10.1007/s10964-019-01098-9>
- Belsky, J., & Pluess, M. (2009). Beyond diathesis stress: Differential susceptibility to environmental influences. *Psychological Bulletin*, *135*, 885–908. <https://doi.org/10.1037/a0017376>
- Bogdan, R., Baranger, D. A. A., & Agrawal, A. (2018). Polygenic risk scores in clinical psychology: bridging genomic risk to individual differences. In T. Widiger & T. D. Cannon (Eds.), *Annual review of clinical psychology* (Vol. 14, pp. 119–157). Annual Reviews. <https://doi.org/10.1146/annurev-clinpsy-050817084847>
- Brooker, R. J., Neiderhiser, J. M., Ganiban, J. M., Leve, L. D., Shaw, D. S., & Reiss, D. (2014). Birth and adoptive parent anxiety symptoms moderate the link between infant attention control and internalizing problems in toddlerhood. *Development and Psychopathology*, *26*, 347–359. <https://doi.org/10.1017/S095457941300103X>
- Burt, S. A., Klahr, A. M., Neale, M. C., & Klump, K. L. (2013). Maternal warmth and directiveness jointly moderate the etiology of childhood conduct problems. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *54*, 1030–1037. <https://doi.org/10.1111/jcpp.12095>
- Caspi, A., McClay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., Taylor, A., & Poulton, R. (2002). Role of genotype in the cycle of violence in maltreated children. *Science (New York, N.Y.)*, *297*, 851–854. <https://doi.org/10.1126/science.1072290>
- Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., McClay, J., Mill, J., Martin, J., Braithwaite, A., & Poulton, R. (2003). Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene. *Science (New York, N.Y.)*, *301*, 386–389. <https://doi.org/10.1126/science.1083968>
- Chabris, C. F., Hebert, B. M., Benjamin, D. J., Beauchamp, J., Cesarini, D., van der Loos, M., Johannesson, M., Magnusson, P. K. E., Lichtenstein, P., Atwood, C. S., Freese, J., Hauser, T. S., Hauser, R. M., Christakis, N., & Laibson, D. (2012). Most reported genetic associations with general intelligence are probably false positives. *Psychological Science*, *23*(11), 1314–1323. <https://doi.org/10.1177/0956797611435528>
- Choi, S. W., Mak, T. S.-H., & O'Reilly, P. F. (2020). Tutorial: A guide to performing polygenic risk score analyses. *Nature Protocols*, *15*, 2759–2772. <https://doi.org/10.1038/s41596-020-0353-1>
- Colodro-Conde, L., Couvy-Duchesne, B., Zhu, G., Coventry, W. L., Byrne, E. M., Gordon, S., Wright, M. J., Montgomery, G. W., Madden, P. A. F., Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Ripke, S., Eaves, L. J., Heath, A. C., Wray, N. R., Medland, S. E., & Martin, N. G. (2018). A direct test of the diathesis-stress model for depression. *Molecular Psychiatry*, *23*, 1590–1596. <https://doi.org/10.1038/mp.2017.130>
- Cox, J. L., Holden, J. M., & Sagovsky, R. (1987). Detection of postnatal depression: Development of the 10-item Edinburgh postnatal depression scale. *The British Journal of Psychiatry*, *150*, 782–786. <https://doi.org/10.1192/bjp.150.6.782>
- Cree, R. A., Liu, C., Gueorguieva, R., Neiderhiser, J. M., Leve, L. D., Connell, C. M., Shaw, D. S., Natsuaki, M. N., Ganiban, J. M., Beekman, C., Smith, M. V., & Reiss, D. (2021). Using an adoption design to test genetically based differences in risk for child behavior problems in response to home environmental influences. *Development and Psychopathology*, *33*, 1229–1247. <https://doi.org/10.1017/S0954579420000450>
- Danese, A., & Widom, C. S. (2020). Objective and subjective experiences of child maltreatment and their relationships with psychopathology. *Nature Human Behaviour*, *4*, 811–818. <https://doi.org/10.1038/s41562-020-0880-3>
- Daniels, D., & Plomin, R. (1985). Differential experience of siblings in the same family. *Developmental Psychology*, *21*, 747–760. <https://doi.org/10.1037/0012-1649.21.5.747>
- Deater-Deckard, K., Dodge, K. A., Bates, J. E., & Pettit, G. S. (1998). Multiple risk factors in the development of externalizing behavior problems: Group and individual differences. *Development and Psychopathology*, *10*, 469–493. <https://doi.org/10.1017/s0954579498001709>
- Demontis, D., Walters, R. K., Martin, J., Mattheisen, M., Als, T. D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Churchhouse, C., Dumont, A., Eriksson, N., Gandal, M., Goldstein, J. I., Grasby, K. L., Grove, J., ... Neale, B. M. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature Genetics*, *51*, 63–75. <https://doi.org/10.1038/s41588-018-0269-7>
- Dick, D. M., Agrawal, A., Keller, M. C., Adkins, A., Aliev, F., Monroe, S., Hewitt, J. K., Kendler, K. S., & Sher, K. J. (2015). Candidate gene–environment interaction research: Reflections and recommendations. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *10*, 37–59. <https://doi.org/10.1177/1745691614556682>
- Domingue, B., Trejo, S., Armstrong-Carter, E., & Tucker-Drob, E. (2020). Interactions between polygenic scores and environments: Methodological and conceptual challenges. *Sociological Science*, *7*, 465–486. <https://doi.org/10.15195/v7.a19>
- Duncan, L. E., & Keller, M. C. (2011). A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *American Journal of Psychiatry*, *168*, 1041–1049. <https://doi.org/10.1176/appi.ajp.2011.11020191>
- Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*, 156–168. <https://doi.org/10.1177/2515245919847202>
- Garnezy, N., & Rutter, M. (1983). *Stress, coping, and development in children*. McGraw-Hill.
- Gidziela, A., Malanchini, M., Rimfeld, K., McMillan, A., Ronald, A., Viding, E., Asbury, K., Eley, T. C., von Stumm, S., & Plomin, R. (2022). How much variance in behaviour problems can be predicted by measured nonshared environment from preschool to adulthood? *Submitted for Publication*.
- Gidziela, A., Rimfeld, K., Malanchini, M., Allegrini, A. G., McMillan, A., Selzam, S., Ronald, A., Viding, E., Stumm, S., Eley, T. C., & Plomin, R. (2021). Using DNA to predict behaviour problems from preschool to adulthood. *Journal of Child Psychology and Psychiatry*, *51*, 781–792. <https://doi.org/10.1111/jcpp.13519>
- Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, *38*, 581–586.

- Gottesman, I. I. (1991). *Schizophrenia genesis: The origins of madness*. W. H. Freeman.
- Götz, F. M., Gosling, S. D., & Rentfrow, P. J. (2021). Small effects: The indispensable foundation for a cumulative psychological science. *Perspectives on Psychological Science*, 17, 205–215. <https://doi.org/10.1177/1745691620984483>
- Grotzinger, A. D., Rhemtulla, M., de Vlaming, R., Ritchie, S. J., Mallard, T. T., Hill, W. D., Ip, H. F., Marioni, R. E., McIntosh, A. M., Deary, I. J., Koellinger, P. D., Harden, K. P., Nivard, M. G., & Tucker-Drob, E. M. (2019). Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nature Human Behaviour*, 3, 513–525. <https://doi.org/10.1038/s41562-019-0566-x>
- Hanscombe, K. B., Trzaskowski, M., Haworth, C. M. A., Davis, O. S. P., 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, e30320. <https://doi.org/10.1371/journal.pone.0030320>
- He, Q., & Li, J. J. (2021). A gene–environment interaction study of polygenic scores and maltreatment on childhood ADHD. *Research on Child and Adolescent Psychopathology*, 50, 309–319. <https://doi.org/10.1007/s10802-021-00873-2>
- Hewitt, J. K. (2012). Editorial policy on candidate gene association and candidate gene-by-environment interaction studies of complex traits. *Behavior Genetics*, 42, 1–2. <https://doi.org/10.1007/s10519-011-9504-z>
- Jinks, J. L., & Fulker, D. W. (1970). Comparison of the biometrical genetical, MAVA, and classical approaches to the analysis of the human behavior. *Psychological Bulletin*, 73, 311–349. <https://doi.org/10.1037/h0029135>
- Kandaswamy, R., Allegrini, A., Nancarrow, A. F., Cave, S. N., Plomin, R., & von Stumm, S. (2021). Predicting alcohol use from genome-wide polygenic scores, environmental factors, and their interactions in young adulthood. *Psychosomatic Medicine*, 84, 244–250. <https://doi.org/10.1097/PSY.0000000000001005>
- Keers, R., & Pluess, M. (2017). Childhood quality influences genetic sensitivity to environmental influences across adulthood: A life-course Gene x Environment interaction study. *Development and Psychopathology*, 29, 1921–1933. <https://doi.org/10.1017/S0954579417001493>
- Keller, M. C. (2014). Gene x Environment interaction studies have not properly controlled for potential confounders: The problem and the (simple) solution. *Biological Psychiatry*, 75, 18–24. <https://doi.org/10.1016/j.biopsych.2013.09.006>
- Kendler, K. S., & Eaves, L. J. (1986). Models for the joint effect of genotype and environment on liability to psychiatric illness. *American Journal of Psychiatry*, 143, 279–289.
- Krapohl, E., Hannigan, L. J., Pingault, J.-B., Patel, H., Kadeva, N., Curtis, C., Breen, G., Newhouse, S. J., Eley, T. C., O'Reilly, P. F., & Plomin, R. (2017). Widespread covariation of early environmental exposures and trait-associated polygenic variation. *Proceedings of the National Academy of Sciences*, 114, 11727–11732. <https://doi.org/10.1073/pnas.1707178114>
- Krapohl, E., Patel, H., Newhouse, S., Curtis, C. J., von Stumm, S., Dale, P. S., Zabaneh, D., Breen, G., O'Reilly, P. F., & Plomin, R. (2018). Multi-polygenic score approach to trait prediction. *Molecular Psychiatry*, 23, 1368–1374. <https://doi.org/10.1038/mp.2017.163>
- Ksinan, A. J., Smith, R. L., Barr, P. B., & Vazsonyi, A. T. (2021). The associations of polygenic scores for risky behaviors and parenting behaviors with adolescent externalizing problems. *Behavior Genetics*, 52, 26–37. <https://doi.org/10.1007/s10519-021-10079-3>
- Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghziyan, O., Zacher, M., Nguyen-Viet, T. A., Bowers, P., Sidorenko, J., Karlsson Linnér, R., Fontana, M. A., Kundu, T., Lee, C., Li, H., Li, R., Royer, R., Timshel, P. N., Walters, R. K., Willoughby, E. A., ... Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, 50, 1112–1121. <https://doi.org/10.1038/s41588-018-0147-3>
- Leve, L. D., Harold, G. T., Ge, X., Neiderhiser, J. M., Shaw, D., Scaramella, L. V., & Reiss, D. (2009). Structured parenting of toddlers at high versus low genetic risk: Two pathways to child problems. *Journal of the American Academy of Child and Adolescent Psychiatry*, 48, 1102–1109. <https://doi.org/10.1097/CHI.0b013e3181b8bfc0>
- 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, 716–727. <https://doi.org/10.1017/thg.2019.66>
- Lipscomb, S. T., Laurent, H., Neiderhiser, J. M., Shaw, D. S., Natsuaki, M. N., Reiss, D., & Leve, L. D. (2014). Genetic vulnerability interacts with parenting and early care education to predict increasing externalizing behavior. *International Journal of Behavioral Development*, 38, 70–80. <https://doi.org/10.1177/0165025413508708>
- Luciano, M., Hagenaars, S. P., Davies, G., Hill, W. D., Clarke, T.-K., Shirali, M., Harris, S. E., Marioni, R. E., Liewald, D. C., Fawns-Ritchie, C., Adams, M. J., Howard, D. M., Lewis, C. M., Gale, C. R., McIntosh, A. M., & Deary, I. J. (2018). Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nature Genetics*, 50, 6–11. <https://doi.org/10.1038/s41588-017-0013-8>
- Malanchini, M., Rimfeld, K., Allegrini, A. G., Ritchie, S. J., & Plomin, R. (2020). Cognitive ability and education: How behavioural genetic research has advanced our knowledge and understanding of their association. *Neuroscience and Biobehavioral Reviews*, 111, 229–245. <https://doi.org/10.1016/j.neubiorev.2020.01.016>
- Matheny, A. P., Wachs, T. D., Ludwig, J. L., & Phillips, K. (1995). Bringing order out of chaos: Psychometric characteristics of the confusion, hubbub, and order scale. *Journal of Applied Developmental Psychology*, 16, 429–444. [https://doi.org/10.1016/0193-3973\(95\)90028-4](https://doi.org/10.1016/0193-3973(95)90028-4)
- Monroe, S. M., & Simons, A. D. (1991). Diathesis-stress theories in the context of life stress research: Implications for the depressive disorders. *Psychological Bulletin*, 110, 406–425. <https://doi.org/10.1037/0033-2909.110.3.406>
- Mullins, N., Power, R. A., Fisher, H. L., Hanscombe, K. B., Euesden, J., Iniesta, R., Levinson, D. F., Weissman, M. M., Potash, J. B., Shi, J., Uher, R., Cohen-Woods, S., Rivera, M., Jones, L., Jones, I., Craddock, N., Owen, M. J., Korszun, A., Craig, I. W., ... Lewis, C. M. (2016). Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. *Psychological Medicine*, 46, 759–770. <https://doi.org/10.1017/S0033291715002172>
- Nelemans, S. A., Boks, M., Lin, B., Oldehinkel, T., van Lier, P., Branje, S., & Meeus, W. (2021). Polygenic risk for major depression interacts with parental criticism in predicting adolescent depressive symptom development. *Journal of Youth and Adolescence*, 50, 159–176. <https://doi.org/10.1007/s10964-020-01353-4>
- Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., Turley, P., Chen, G.-B., Emilsson, V., Meddens, S. F. W., Oskarsson, S., Pickrell, J. K., Thom, K., Timshel, P., de Vlaming, R., Abdellaoui, A., Ahluwalia, T. S., Bacelis, J., Baumbach, C., ... Benjamin, D. J. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, 533, 539–542.
- Okbay, A., Wu, Y., Wang, N., Jayashankar, H., Bennett, M., Nehzati, S. M., Sidorenko, J., Kweon, H., Goldman, G., Gjorgjieva, T., Jiang, Y., Hicks, B., Tian, C., Hinds, D. A., Ahlsgog, R., Magnusson, P. K. E., Oskarsson, S., Hayward, C., Campbell, A., ... Young, A. I. (2022). Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nature Genetics*, 54, 437–449. <https://doi.org/10.1038/s41588-022-01016-z>
- Pain, O., Glanville, K. P., Hagenaars, S. P., Selzam, S., Fürtjes, A. E., Gaspar, H. A., Coleman, J. R. I., Rimfeld, K., Breen, G., Plomin, R., Folkers, L., & Lewis, C. M. (2020). Evaluation of polygenic prediction methodology within a reference-standardized framework. *PLOS Genomics*, 17, e100921. <https://doi.org/10.1101/2020.07.28.224782>
- Paris, J. (1999). *Genetics and psychopathology: Predisposition-stress interactions*. American Psychiatric Press.
- Peyrot, W. J., Milaneschi, Y., Abdellaoui, A., Sullivan, P. F., Hottenga, J. J., Boomsma, D. I., & Penninx, B. W. J. H. (2014). Effect of polygenic risk scores on depression in childhood trauma. *The British Journal of Psychiatry: The Journal of Mental Science*, 205, 113–119. <https://doi.org/10.1192/bjp.bp.113.143081>
- Peyrot, W. J., Van der Auwera, S., Milaneschi, Y., Dolan, C. V., Madden, P. A. F., Sullivan, P. F., Strohmaier, J., Ripke, S., Rietschel, M., Nivard, M. G., Mullins, N., Montgomery, G. W., Henders, A. K., Heat, A. C., Fisher, H.

- L., Dunn, E. C., Byrne, E. M., Air, T. A., Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, . . . Penninx, B. W. J. H. (2018). Does childhood trauma moderate polygenic risk for depression? A meta-analysis of 5765 subjects from the psychiatric genomics consortium. *Biological Psychiatry*, *84*, 138–147. <https://doi.org/10.1016/j.biopsych.2017.09.009>
- Pingault, J.-B., O'Reilly, P. F., Schoeler, T., Plouhidis, G. B., Rijdsdijk, F., & Dudbridge, F. (2018). Using genetic data to strengthen causal inference in observational research. *Nature Reviews Genetics*, *19*, 566–580. <https://doi.org/10.1038/s41576-018-0020-3>
- Plomin, R. (1994). *Genetics and experience: The interplay between nature and nurture*. Sage Publications.
- Plomin, R. (2011). Commentary: Why are children in the same family so different? Non-shared environment three decades later. *International Journal of Epidemiology*, *40*, 582–592. <https://doi.org/10.1093/ije/dyq144>
- Plomin, R. (2018). *Blueprint: How DNA makes us who we are*. Allen Lane/Penguin Books. <https://www.penguin.co.uk/books/298/298391/blueprint/9780141984261.html>
- Plomin, R., & Bergeman, C. S. (1991). The nature of nurture: Genetic influence on “environmental” measures. *Behavioral and Brain Sciences*, *14*, 373–386. <https://doi.org/10.1017/S0140525X00070278>
- Plomin, R., & DeFries, J. C. (1985). *Origins of individual differences in infancy: The Colorado adoption project*. Academic Press.
- Plomin, R., DeFries, J. C., & Fulker, D. W. (1988). *Nature and nurture during infancy and early childhood*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511527654>
- Plomin, R., DeFries, J. C., & Loehlin, J. C. (1977). Genotype–environment interaction and correlation in the analysis of human behavior. *Psychological Bulletin*, *84*, 309–322. <https://doi.org/10.1037/0033-2909.84.2.309>
- Plomin, R., & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics*, *19*, 148–159. <https://doi.org/10.1038/nrg.2017.104>
- Plomin, R., & von Stumm, S. (2021). Polygenic scores: Prediction versus explanation. *Molecular Psychiatry*, *27*, 49–52. <https://doi.org/10.1038/s41380-021-01348-y>
- Pluess, M., & Belsky, J. (2010). Differential susceptibility to parenting and quality child care. *Developmental Psychology*, *46*, 379–390. <https://doi.org/10.1037/a0015203>
- Pries, L.-K., Klingenberg, B., Menne-Lothmann, C., Decoster, J., van Winkel, R., Collip, D., Delespaul, P., De Hert, M., Derom, C., Thiery, E., Jacobs, N., Wichers, M., Cinar, O., Lin, B. D., Luyckx, J. J., Rutten, B. P. F., van Os, J., & Guloksuz, S. (2020). Polygenic liability for schizophrenia and childhood adversity influences daily-life emotion dysregulation and psychosis proneness. *Acta Psychiatrica Scandinavica*, *141*, 465–475. <https://doi.org/10.1111/acps.13158>
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., Westra, H.-J., Shakhbazov, K., Abdellaoui, A., Agrawal, A., Albrecht, E., Alizadeh, B. Z., Amin, N., Barnard, J., Baumeister, S. E., Benke, K. S., Bielak, L. F., Boatman, J. A., Boyle, P. A., . . . Koellinger, P. D. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science (New York, N.Y.)*, *340*, 1467–1471. <https://doi.org/10.1126/science.1235488>
- Rimfeld, K., Malanchini, M., Spargo, T., Spickernell, G., Selzam, S., McMillan, A., Dale, P. S., Eley, T. C., & Plomin, R. (2019). Twins early development study (TEDS): A genetically sensitive investigation into behavioral and cognitive development from infancy to emerging adulthood. *Twin Research and Human Genetics*, *22*, 508–513. <https://doi.org/10.1017/thg.2019.56>
- Robinson, N., & Bergen, S. E. (2021). Environmental risk factors for schizophrenia and bipolar disorder and their relationship to genetic risk: Current knowledge and future directions. *Frontiers in Genetics*, *12*, 999. <https://doi.org/10.3389/fgene.2021.686666>
- Rutter, M. (1981). Stress, coping and development: Some issues and some questions. *Journal of Child Psychology and Psychiatry*, *22*, 323–356. <https://doi.org/10.1111/j.1469-7610.1981.tb00560.x>
- Rutter, M., Moffitt, T. E., & Caspi, A. (2006). Gene–environment interplay and psychopathology: Multiple varieties but real effects. *Journal of Child Psychology and Psychiatry*, *47*, 226–261. <https://doi.org/10.1111/j.1469-7610.2005.01557.x>
- Selzam, S., McAdams, T. A., Coleman, J. R. I., Carnell, S., O'Reilly, P. F., Plomin, R., & Llewellyn, C. H. (2018). Evidence for gene–environment correlation in child feeding: Links between common genetic variation for BMI in children and parental feeding practices. *PLOS Genetics*, *14*, e1007757. <https://doi.org/10.1371/journal.pgen.1007757>
- Shmueli, G. (2010). To explain or to predict? *Statistical Science*, *25*, 289–310. <https://doi.org/10.1214/10-STS330>
- Slagt, M., Dubas, J. S., Dekovic, M., & van Aken, M. A. G. (2016). Differences in sensitivity to parenting depending on child temperament: A meta-analysis. *Psychological Bulletin*, *142*, 1068–1110. <https://doi.org/10.1037/bul0000061>
- Thomas, A., & Chess, S. (1977). *Temperament and development*. Brunner-Routledge.
- Trotta, A., Iyegbe, C., Forti, M. D., Sham, P. C., Campbell, D. D., Cherny, S. S., Mondelli, V., Aitchison, K. J., Murray, R. M., Vassos, E., & Fisher, H. L. (2016). Interplay between schizophrenia polygenic risk score and childhood adversity in first-presentation psychotic disorder: A pilot study. *PLOS ONE*, *11*, e0163319. <https://doi.org/10.1371/journal.pone.0163319>
- von Stumm, S., & d'Apice, K. (2022). From genome-wide to environment-wide: Capturing the environome. *Perspectives on Psychological Science*, *17*, 30–40. <https://doi.org/10.1177/1745691620979803>
- Wachs, T. D., & Gruen, G. (1982). *Early experience and human development*. Plenum.
- Widaman, K. F. (2021). Interrogating the validity of cumulative indices of environmental and genetic risk for negative developmental outcomes. *Development and Psychopathology*. Online ahead of print. <https://doi.org/10.1017/S0954579421001097>
- Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*, 1100–1122. <https://doi.org/10.1177/1745691617693393>