



Published in final edited form as:

*Psychol Sci.* 2016 February ; 27(2): 138–149. doi:10.1177/0956797615612727.

## Large Cross-National Differences in Gene $\times$ Socioeconomic Status Interaction on Intelligence

Elliot M. Tucker-Drob<sup>#1</sup> and Timothy C. Bates<sup>#2</sup>

<sup>1</sup>University of Texas at Austin

<sup>2</sup>University of Edinburgh

# These authors contributed equally to this work.

### Abstract

A core hypothesis in developmental theory predicts that genetic influences on intelligence and academic achievement are suppressed under conditions of socioeconomic privation and more fully realized under conditions of socioeconomic advantage: a gene  $\times$  childhood-socioeconomic status ( $G \times SES$ ) interaction. Tests of this hypothesis have produced apparently inconsistent results. We meta-analyzed tests of  $G \times SES$  interaction on intelligence and academic achievement test scores, allowing for stratification by nation (US vs non-US), and conducted rigorous tests for publication bias and between-study heterogeneity. In US studies, we find clear support for moderately sized  $G \times SES$  effects. In studies from Western Europe and Australia, where social policies ensure more uniform access to high quality education and healthcare,  $G \times SES$  effects were zero or reversed.

---

One of the most important, but contentious, topics of the nature-nurture debate is the source of individual differences in intelligence (Galton, 1869; Jensen, 1969). Historically, the debate pit a hereditarian perspective, which views individual differences in intelligence as primarily genetic, against a sociological perspective, which views such differences as primarily rooted in environmental experience. Reports of recovery from IQ deficits among children rescued from severely adverse circumstances (e.g. Nelson et al., 2007) support a role for environmental experience on intellectual development. Yet these findings are seemingly undermined by the ubiquitous finding that heritability estimates from twin and pedigree studies of intelligence are large (Bouchard & McGue, 1981) and increase with age (Briley & Tucker-Drob, 2013), findings generally supported by heritability estimates from genome-wide data on unrelated individuals (Davies et al., 2011; Plomin et al., 2013). Understanding this apparent paradox is a central question for research on cognitive development. One potential rapprochement between hereditarian and sociological views is the hypothesis that natural potentials for adaptive functioning are more fully expressed in the

---

**Address Correspondences to:** Elliot Tucker-Drob, 108 E. Dean Keeton Stop A8000, Austin, Texas 78712, tuckerdrob@utexas.edu or Timothy Bates, 7 George Square, Edinburgh, Scotland, EH8 9JZ, tim.bates@ed.ac.uk.

#### Author Contributions

EMTD and TCB conceived the project, acquired and reanalyzed/standardized data, and wrote the paper; EMTD conducted the meta analyses and power analyses.

The data reported in the paper are archived at <https://osf.io/vkmep>.

The authors declare no competing financial interests.

context of more nourishing environmental experiences (Bronfenbrenner & Ceci, 1994). Known as the “*Scarr-Rowe*” hypothesis of gene  $\times$  socioeconomic status (SES) interaction (Turkheimer, Harden, D’Onofrio, & Gottesman, 2009), it was originally described by Scarr-Salapatek (1971, p. 1286) as follows:

“IQ scores within advantaged groups will show larger proportions of genetic variance and smaller proportions of environmental variance than IQ scores for disadvantaged groups. Environmental disadvantage is predicated to reduce the genotype-phenotype correlation in lower-class groups.” (p. 1286).

The Scarr-Rowe hypothesis thus predicts that the heritability of intelligence will be lower among those raised under conditions of greater childhood socioeconomic disadvantage. The first support for this hypothesis came from a small sample of school-aged Philadelphia twins, for whom data on sex but not zygosity was available, and for whom only census-level socioeconomic status, but not family-level socioeconomic status, was available (Scarr-Salapatek, 1971). These findings were sharply criticized for methodological flaws, with Eaves and Jinks (1972) concluding that “evidence previously analysed is insufficient to support the conclusions drawn” (p 84). After a 28-year period of virtually no new research on the topic, Rowe, Jacobson, and Van den Oord (1999) reported higher heritability of vocabulary IQ at higher levels of parental education among a population-based American sample of adolescent twin and sibling pairs. Some subsequent US research has replicated the interaction (Bates, Lewis, & Weiss, 2013; Harden, Turkheimer, & Loehlin, 2007; Tucker-Drob, Rhemtulla, Harden, Turkheimer, & Fask, 2011; Turkheimer, Haley, Waldron, D’Onofrio, & Gottesman, 2003). Other studies, however, have failed to replicate the finding (Bartels, van Beijsterveldt, & Boomsma, 2009; Kremen et al., 2005; Soden-Hensler, 2012; van der Sluis, Willemsen, de Geus, Boomsma, & Posthuma, 2008), including the largest sample to date (Hanscombe et al., 2012).

A number of factors might explain the inconsistencies among findings. First, reports of gene  $\times$  socioeconomic status ( $G \times$  SES) interaction on IQ might be false positives resulting from poorly powered studies conducted against a backdrop of bias towards publishing and citing positive findings. Alternatively,  $G \times$  SES interactions might be true effects, but researchers could fail to replicate them because of low power or poor methodology. Finally, multiple authors have suggested that Gene  $\times$  SES interactions may differ in strength across different populations or societies (Bates, Hansell, Martin, & Wright, under review; Bates et al., 2013; Hanscombe et al., 2012; Tucker-Drob, Briley, & Harden, 2013; Turkheimer & Horn, 2014). Those supporting this type of explanation have pointed to higher social stratification in access to education (Hauser, 1970) and relatively modest social health (Adler & Newman, 2002) and social welfare support (DeNavas-Walt, Proctor, & Bureau, 2014) in the United States compared to Australia and Western Europe.

In this meta-analysis, we sought robust and reliable answers to three questions: First, does the range of studies from the US support a positive estimate of  $G \times$  SES interaction on achieved IQ? Second, do studies on participants outside the US show a similar greater-than-zero  $G \times$  SES effect? Third, can a single estimate adequately account for all of the observed effect sizes, or are separate estimates necessary to represent effect sizes from the US vs. Western Europe and Australia? To answer these questions we collected the world’s

literature on  $G \times \text{SES}$  effects on IQ, undertaking or commissioning from the original authors a number of re-analyses of the data, and surfacing previously unpublished or in-press studies.

## Method

We meta-analyzed the world's literature on  $G \times \text{SES}$  interaction effects on IQ, seeking to (a) determine the true effect size, (b) examine bias in reporting effects, and (c) examine moderators of effect size, including age of subjects, nation (US versus non-US), test type (verbal, spatial, or non-IQ cognitive measures), SES assessment (education, income, wealth, and/or occupational status) and year of publication. We used sophisticated methods for estimating random effects meta-regressions in nested data (i.e., in a meta-analytic dataset that contained multiple effect sizes per study). Because this methodology required that all effect sizes and standard errors be calculated using a consistent modeling approach, we typically (a) obtained the raw data and re-analyzed it ourselves, (b) were provided with parameter estimates from re-analyses by the original study authors (all authors contacted complied with our request), or (c) re-analyzed the published sub-group covariance matrices (see Table 1).

## Study Identification

We identified published studies of  $G \times \text{SES}$  interaction on objective measures of intelligence and academic achievement by searching databases, reviewing citations to and by identified studies and citations from narrative reviews, and consulting with colleagues. We used Google Scholar to search for studies using combinations of the following search terms: twin, gene, socioeconomic status, education, income, achievement, intelligence, cognition, and *interaction*. Previous narrative reviews of this literature that we consulted were papers by Turkheimer and Horn (2014), Turkheimer et al. (2009), Nisbett et al. (2012), and Tucker-Drob et al. (2013). We also consulted a table of previous studies included in Hanscombe et al. (2012).

Through consulting with colleagues, as well as reviewing recent abstracts from the meetings of the *Behavior Genetics Association* and the *International Society for Intelligence Research*, we additionally identified a number of unpublished findings. K. Jacobson (an original author of Rowe et al. 1999) conducted a  $G \times \text{SES}$  analysis of a later wave of the Add Health Study, E. Tucker-Drob conducted a  $G \times \text{SES}$  analysis of a later wave of the ECLS-B data (analyses of previous waves were published in Tucker-Drob et al, 2011 and Rhemtulla & Tucker-Drob, 2012). Soden-Hensler reported  $G \times \text{SES}$  results in an unpublished dissertation that we identified in our electronic literature search. T. Bates provided results from a manuscript under review. M. Spengler shared an unpublished manuscript that she updated with additional data for the purposes of inclusion in the current meta-analysis.

## Inclusion and Exclusion Criteria

Studies included in the final meta analytic dataset are listed in Table 1. In order to qualify for inclusion in our meta-analysis, studies needed to meet the following criteria: (a)

intelligence or achievement was continuously measured using an objective performance-based test or tests, (b) inference of genetic influence was based on siblings (preferably twins) with varying degrees of genetic relatedness, (c) the degree of genetic relatedness was known to a high degree of certainty (for twins, this meant that zygosity was diagnosed either through physical similarity ratings or genotyping), (d) an ordered categorical or continuous measure of family SES during childhood (a single or composite measure of parental education, family income, and/or parental job prestige) was examined as a moderator of genetic variance in intelligence or achievement, and (e) participants were not specifically selected on the basis of their psychiatric or medical diagnoses, patient status, or (low or high) intelligence or achievement test scores. Thus, studies were excluded if they (a) used a categorical measure of intelligence or achievement, (b) the measure of intelligence or achievement was not an objective test, (c) only sex, but not genetic relatedness/zygosity information, was available for the sibling/twin pairs, (d) only adulthood SES was measured, (e) SES was measured at the school or neighborhood, but not family, level, or (f) participants were selected on the basis of their test scores, medical/psychiatric diagnoses, or patient status. We did not exclude studies on the basis of the age at which intelligence/achievement was measured.

When the original studies reported separate effect sizes for each of multiple cognitive measures, we included the effect sizes for each measure. When studies reported effect sizes for each of multiple measures of SES, we used the effect sizes associated with the composite SES measure, if available. If a composite measure was not available, we used effect sizes associated with parental education (mid-parent education preferred to maternal or paternal education) and/or family income. If both parental education and family income were available, but not a composite score, we used effect sizes from both. We made one exception. For the TEDS sample (Hanscombe et al., 2012), which used three different SES measures from three different waves of data collection (composite of parental education and occupation at 18 months and 7 years, and income at 9 years). Because each measure provided information about SES during a different period of childhood, we deemed it best to retain effect sizes from all three SES indices.

### Excluded Studies

Some notable published examinations of a  $G \times \text{SES}$  interaction did not meet our inclusion criteria. The following articles were excluded from our meta-analytic dataset for the reasons described below.

Scarr-Salapatek (1971) relied on a data from participants for whom only sex but not zygosity was available, and for whom only census-level socioeconomic status, but not family-level socioeconomic status, was available.

Using kinship data from the National Longitudinal Survey of Youth (NLSY), Van Den Oord and Rowe (1997) relied on full siblings, half siblings, and cousins to estimate genetic effects on academic achievement at different levels of a variety of measured environments, including parental education. A classification algorithm was used to determine sibling types from indirect parentage information. Genetic inference in this design rests on comparisons between sibling types that differ markedly in some of the very same variables that are

hypothesized to be moderators of heritability. For instance, A. K. Cheung, Harden, and Tucker-Drob (2014) report that, in the NLSY dataset, full siblings compared to half siblings were nearly twice as likely to be Caucasian and over twice as likely to have college-educated parents. We therefore decided that the NLSY kinship data were not appropriate for examining  $G \times \text{SES}$  interaction.

Nagoshi and Johnson (2005) examined SES differences in familial resemblance for cognitive abilities in the Hawaii Family Study of Cognition. However, as these authors did not capitalize on data from relatives of different degrees of genetic relatedness or from adoptees, the study cannot distinguish genetic from environmental sources of familial variance.

Friend, DeFries, and Olson (2008) and Friend et al. (2009) tested for differences in the heritability of categorical outcomes: reading disability and high reading ability, respectively. These samples were also specifically selected on the basis of the outcome under study, i.e., at least one twin meeting criteria for proband status.

Finally, in a poster presented at the 2014 *Behavior Genetics Association* meeting, Prescott, McArdle, Achorn, Kaiser, and Lapham (2014) estimated biometric models of  $G \times \text{SES}$  interaction using archival twin and sibling data from Project Talent of 1960. As in Scarr-Salapatek (1971), only sex, but not zygosity, was available from participants, and genetic inference required reliance on differences in intraclass correlations across same-sex twins, opposite-sex twins, and non-twin siblings.

### Obtaining Effect Size Estimates

Because our analysis plan required that the results from each study be estimated from the same model, we often either reanalyzed data to which we had access, or we obtained reanalyzed data from the original study authors.

For each study, we obtained a full set of parameter estimates and associated standard errors for an untrimmed version of Purcell's (2002) biometric model of latent Gene  $\times$  Measured Shared Environment interaction, where the measured shared environment is the SES index and the phenotype is the index of intelligence or achievement. This model, displayed in Figure S1, is an extension of the standard ACE model, which decomposes phenotypic variation into additive genetic effects (A), shared environmental effects (C), and non-shared environmental effects (E). The parameters produced are the main effects of A, C, E, and SES, along with  $A \times \text{SES}$ ,  $C \times \text{SES}$ , and  $E \times \text{SES}$  interactions. These parameters are termed  $a$ ,  $c$ ,  $e$ ,  $s$ , and  $a'$ ,  $c'$ , and  $e'$ , respectively. Of central interest is the  $A \times \text{SES}$  interaction, labeled  $a'$ . In order to ensure that effect sizes were placed on a metric that was comparable across studies, we ensured that phenotypes and continuous SES measures were Z-transformed (based on the  $M$  and  $SD$  from data pooled across all twins and siblings of all relatedness types) prior to fitting the Purcell (2002) model.

For studies that provided separate mean and covariance matrices for each level of an ordered categorical SES variable, we used parametric cross-group constraints to obtain estimates of  $a$ ,  $c$ ,  $e$ ,  $s$ ,  $a'$ ,  $c'$ , and  $e'$  that are directly comparable to those obtained from the Purcell (2002)

applied to continuous SES data. To do this, we determined the SES Z-score for each discrete group by using the proportions of participants in each category to calculate normal distribution thresholds for group membership, and then computing the mean Z-score between the calculated thresholds based on the assumption of a continuous normal underlying SES distribution. The assumption that a continuous normal distribution underlies ordered categorical SES distributions was considered reasonable, because the studies that did employ continuous measures of SES typically transformed their SES measure to normality if it was not already approximately normally distributed.

### Meta-Analyzing Effect Sizes

After compiling a meta-analytic database of study-specific effect sizes, we fit random effects meta-regression models in *Mplus* (Muthén & Muthén, 1998-2012), using the general procedure described by M. W. L. Cheung (2008), separately for each parameter of interest ( $a$ ,  $c$ ,  $e$ ,  $s$ ,  $a'$ ,  $c'$ , and  $e'$ ). Random effects models do not require the traditional assumption that all effect sizes were derived from the same population with a single true effect size. Rather, they allow for population variation in the true effect size above and beyond variation attributable to sampling error. In practice, this means that standard errors and significance levels for meta-analytic parameters for random effects models are more conservative than they would be for fixed effects models. We included multiple effect sizes per sample (e.g. for different intelligence measures, different waves, or different measures of SES), weighted meta-regression models by the reciprocal number of effect sizes included for the corresponding sample and by the inverse of each parameter's sampling variance, and implemented a sandwich estimator that corrects standard errors for non-independence of data points derived from the same sample.

## Results

### Summary of Meta-Analytic Dataset

As displayed in Table 1, our meta-analytic dataset consisted of  $k = 43$  effect sizes from a total of  $N = 24,926$  pairs of twins and siblings (approximately 50,000 individuals) participating in  $K = 14$  independent studies. Data were relatively evenly split across US ( $k = 18$ ,  $K = 8$ ,  $N = 10,831$ ) and Non-US (Western Europe and Australia) samples ( $k = 25$ ,  $K = 6$ ,  $N = 14,095$ ). Because all variables were placed on a Z-scale, the magnitude of the  $G \times SES$  interaction parameter ( $a'$ ) represents the expected difference in the regression effect of the additive genetic factor on intelligence for each SD difference in SES.

### G $\times$ SES Effects in US and Non-US Samples

In an unconditional meta-analytic model, the meta-analytic mean for the  $G \times SES$  interaction was non-significant ( $a' = .029$ ,  $SE = .019$ ,  $p = .136$ ), but there was substantial heterogeneity in the effect, with significant variation in the random effect (the standard deviation of the random effect was .048;  $SE = .019$ ,  $p = .011$ ). A model that added a dummy coded nation variable (0=non-US, 1=US) to the meta-regression model indicated that there was a significant difference between  $G \times SES$  effect sizes from the US compared to those from Western Europe and Australia ( $b = .101$ ,  $SE = .032$ ,  $p = .001$ ). There was a significant  $G \times SES$  effect for US studies ( $a' = .074$ ,  $SE = .020$ ,  $p < .0005$ ), but a trivial and non-

significant  $G \times \text{SES}$  effect for non-US studies ( $a' = -.027$ ,  $SE = .022$ ,  $p = .223$ ). The standard deviation of the random effect was .029 ( $SE = .012$ ,  $p = .022$ ), indicating residual heterogeneity in the  $G \times \text{SES}$  effect. Estimates from this model, as applied to all 7 parameters, are presented in Table 2.

The source of residual heterogeneity in  $a'$  primarily stemmed from the non-US studies ( $\tau = .033$ ,  $SE = .017$ ,  $p = .047$ ). Among US studies, heterogeneity was not significant ( $\tau = .014$ ,  $SE = .040$ ,  $p = .735$ ). Post-hoc analysis revealed that, exclusive of effect sizes from the Netherlands, residual heterogeneity was nil in the non-US group ( $\tau = .005$ ,  $SE = .517$ ,  $p = .993$ ), and there continued to be a significant difference between  $G \times \text{SES}$  effect sizes from the US compared to those from Western Europe and Australia ( $b = .069$ ,  $SE = .012$ ,  $p < .0005$ ). The model excluding studies from the Netherlands still indicated that there was a significant  $G \times \text{SES}$  effect for US studies ( $a' = .068$ ,  $SE = .021$ ,  $p = -.001$ ), but no  $G \times \text{SES}$  effect for non-US studies ( $a' = -.001$ ,  $SE = .033$ ,  $p = .979$ ).

Figure 1 presents genetic and environmental variance components as a function of SES, as implied by the meta-analytic parameter estimates for the US. Genetic variance in intelligence increases from .24 at 2 SD below the mean SES to .61 at 2 SD above the mean SES. As instantaneous proportions of variance, these correspond to heritability estimates of 26% at 2 SD's below the mean SES, and 61% at 2 SD's above the mean SES. There is also some indication that the shared and non-shared environmental variance components decrease with SES, although these interactions were not statistically significant.

### Additional Moderator Analyses

We next tested additional moderators of the  $a'$  effect size. In our first moderation model, we compared effect size estimates for studies that measured cognition in adulthood (ages > 20 years;  $k = 7$ ) to those that measured cognition in childhood (ages < 20 years;  $k = 36$ ). Because there were a small number of studies of adult cognition, we restricted the remaining moderator analyses to the childhood subsample. We examined each of the following study characteristics as moderators of the  $a'$  effect size in turn, each time controlling for whether or not effect sizes came from US samples: childhood age, single vs. composite measure of SES, Achievement/Knowledge test score versus intelligence test score, single cognitive ability vs. composite cognitive measure. None of these moderators achieved statistical significance. Results are presented in Table 3.

Finally, we examined whether there were childhood age differences in the  $a$  parameter, representing the main effect of genes. Consistent with previous meta analyses of  $G \times \text{Age}$  interactions (e.g. Briley & Tucker-Drob, 2013; Haworth et al., 2009), there was significant support for age differences in the  $a$  and  $c$  parameters, indicating that genetic variance increases and shared environmental variance decreases with age (Figure S2). This finding, which replicates previously documented age trends in genetic and environmental variance components, indicates that studies in the current meta-analysis are representative of the wider literature.

## Probing for Publication Bias

**Funnel Asymmetry**—Figure 2 presents diagnostic funnel plots of  $G \times$  SES interaction effect sizes, separately for US studies (black) and European/Australian studies (red).

This figure plots effect size on the  $x$ -axis against the SE of effect size on the  $y$ -axis. This type of figure is named a “funnel plot” because, when there is no (within-group) heterogeneity in population effect sizes, 95% of the meta-analytic effect sizes should fall within the inverted triangle-shaped region, with greater horizontal spread around estimates with higher standard errors. Biased reporting is indicated when asymmetrical scatter around the point estimate is observed; commonly, effects with large SEs (i.e., studies with low power) will be conspicuously absent from the region of the funnel nearer to the null. Evidence for bias was lacking. The only conspicuously empty area of the US funnel is toward the apex, indicating an absence of effect sizes from US studies with small standard errors. An alternative version of this funnel plot, with datapoints replaced by study acronyms, can be found in the online supplement (Figure S3).

Formal tests of funnel asymmetry also revealed no evidence of publication bias. In the US studies, neither the SE nor the square SE was related to  $a'$  effect sizes, both when estimated with ordinary least squares ( $p = .368$  and  $.578$ , respectively) or weighted least squares ( $p = .123$  and  $.211$ , respectively, with the weighting based on the reciprocal sampling variance). Similarly, in the non-US studies, neither the SE nor the square SE was related to  $a'$  effect sizes, both when estimated with ordinary least squares ( $p = .291$  and  $.505$ , respectively) or weighted least squares ( $p = .232$  and  $.560$ , respectively with the weighting based on the reciprocal sampling variance). In the entire meta-analytic dataset, controlling for a dummy-coded variable representing US vs. non-US studies, neither the SE nor the square SE was related to  $a'$  effect sizes, both when estimated with ordinary least squares ( $p = .595$  and  $.651$ , respectively) or weighted least squares ( $p = .951$  and  $.998$ , respectively) with the weighting based on the reciprocal sampling variance.

**p-Curve Analysis**—We submitted the  $a'$  effect sizes to a  $p$ -curve analysis (Simonsohn, Nelson, & Simmons, 2014). using the online  $p$ -checker app (<http://shinyapps.org/apps/p-checker/>).  $P$ -curve analysis tests whether the distribution of statistically significant  $p$  values is right skewed, as would be expected if the data reflects a true non-null effect. In the absence of a right-skewed distribution of significant  $p$ -values (i.e. when the distribution of significant  $p$ -values is either uniformly distributed or left skewed), one cannot rule out the null hypothesis of no true effect.

When only US studies were entered into  $p$ -curve analysis,  $p$ -checker indicated that the  $p$ -curve is, in fact, right skewed, indicating that these studies contain evidential value ( $z = -3.427$ ,  $p < .001$ ). It also indicated that the  $p$ -curve is not flatter than one would expect if studies were powered at 33% ( $z = 1.253$ ,  $p = .895$ ). If significant, this would have indicated that results have no evidential value. Finally, there was no evidence of left skew ( $z = 3.427$ ,  $p = 1.000$ ), which would have indicated  $p$ -hacking or selective reporting.

In contrast, when only non-US studies were entered into  $p$ -curve analysis,  $p$ -checker did not indicate significant right skew ( $z = -1.264$ ,  $p = .103$ ), indicating that there was no evidence



for a non-null  $a'$  effect outside of the US. It also indicated that the  $p$ -curve is not flatter than one would expect if studies were powered at 33% ( $z = .122, p = .548$ ), and that there was no evidence of left skew ( $z = 1.264, p = .897$ ).

### Robustness Checks

We conducted a series of further analyses to verify the robustness of our results. First, we ensured that the results persisted after removing effects sizes from the TEDS study and the MTFS study from analyses. The TEDS study, which is based on a UK sample, is the largest sample in the meta-analysis and reports a null  $G \times SES$  effect. The MTFS study, which reports a significant  $G \times SES$  result, is the second largest study from the US studies included in the meta-analysis, and reports the smallest standard error for the  $G \times SES$  effect of all the US studies. Thus, it was possible that the difference between results for Western Europe/Australia and the US was simply driven by a difference between TEDS results and the US results, or a difference between TEDS and MTFS results more specifically. We ruled these possibilities out. A model that excluded all effect sizes from TEDS still yielded a null  $G \times SES$  effect in non-US samples ( $a' = -.052, SE = .035, p = .133$ ), a significant  $G \times SES$  effect in the US ( $a' = .076, SE = .019, p < .0005$ ), and a significant difference between US and non-US studies for the  $G \times SES$  effect (difference = .128,  $SE = .040, p = .002$ ). A model that excluded all effect sizes from both TEDS and MTFS yielded a null  $G \times SES$  effect in non-US samples ( $a' = -.052, SE = .034, p = .127$ ), a significant  $G \times SES$  effect in the US ( $a' = .083, SE = .023, p < .0005$ ), and a significant difference between US and non-US studies (difference = .135,  $SE = .042, p = .001$ ).

Second, we ensured that results persisted in a model that excluded effect sizes from our own primary investigations of  $G \times SES$  (i.e., those from the ECLS-B, MIDUS, in the US which reported a positive  $G \times SES$  effect, and those from the BATS sample in Australia which reported a null  $G \times SES$  effect). This model yielded a null  $G \times SES$  effect in non-US samples ( $a' = -.033, SE = .028, p = .240$ ), a significant  $G \times SES$  effect in the US ( $a' = .074, SE = .025, p = .003$ ), and a significant difference between US and non-US samples (difference = .107,  $SE = .039, p = .006$ ). A very similar pattern occurred when (rather than excluding our own primary investigations) we controlled for a dummy-coded indicator of whether results came from our own primary investigations.

Third, we ensured that results persisted in a model that excluded all effect sizes from primary investigations of  $G \times SES$  authored by Eric Turkheimer, whose 2003 paper is the most highly cited and reports the largest  $G \times SES$  effect of all articles included in the meta-analysis. A model that excluded these effect sizes (reported in Harden et al., 2007; Tucker-Drob et al., 2011; Turkheimer et al., 2003), yielded a null  $G \times SES$  effect in non-US samples ( $a' = -.027, SE = .022, p = .220$ ), a significant  $G \times SES$  effect in the US ( $a' = .058, SE = .020, p = .003$ ), and a significant difference between US and non-US samples (difference = .085,  $SE = .030, p = .005$ ). A very similar pattern was obtained when (rather than excluding primary investigations by Turkheimer) we used a dummy-coded indicator to control for whether results came from primary investigations by Turkheimer, and also when we controlled for a dummy-coded indicator of whether results came from primary investigations by Turkheimer, Bates, or Tucker-Drob.

In conclusion, all robustness checks indicated the presence of a significantly positive  $G \times$  SES interaction in the US but not in Western Europe/Australia, with the difference between US and non-US studies being statistically significant. These results cannot be attributed to a few select studies with either disproportionately large samples or disproportionately large (or small) effect sizes.

### Cross-National Differences in Racial Diversity

One possibility that we were unable to fully test in our meta-analytic dataset is the extent to which the  $G \times$  SES interactions detected in the US were driven by the greater racial and ethnic diversity of the US samples. Evidence from individual studies included in the meta-analysis, however, suggests that this is not the case. For instance, Tucker-Drob et al. (2011) reported that  $G \times$  SES effects persisted in a nationally representative sample of American children even when race, and its interaction with the biometric components, was controlled. Additionally, Harden et al. (2007) reported evidence for  $G \times$  SES interaction in a positively selected sample of US adolescents that contained very few racial minorities. Kirkpatrick et al. (2015) similarly report evidence for  $G \times$  SES interaction in a sample containing very few Black and Hispanic participants.

### Implications of Meta-Analytic Effect Size Estimates for Reproducibility

Our meta-analytic finding of a  $G \times$  SES interaction in the US, was based on a total sample size of over 10,000 pairs of US twins and siblings. It is instructive to estimate the minimum sample size necessary in order to achieve sufficient power to reproduce the  $G \times$  SES interaction effect in the US. We performed a simulation study in which we generated data based on a model in which all effect sizes ( $s$ ,  $a$ ,  $c$ ,  $e$ ,  $a'$ ,  $c'$ , and  $e'$ ) were set to the meta-analytic estimates for the US (reported in Table 2) and analyzed each dataset with the unconstrained  $G \times$  E model represented in Figure S1 (Purcell, 2002). One third of participants were assumed to be MZ and two-thirds were assumed to be DZ (as would be expected in an unselected sample of same-sex and opposite-sex twins). We increased the sample size in increments of 100 pairs, conducting 100 replications per sample, until at least 80 of the 100 replications produced an  $a'$  estimate that was significant at  $p < .05$  (i.e. power was estimated at a minimum of 80%). Results indicated that a minimum of 3300 twin pairs is needed to achieve at least 80% power. In contrast, when the same procedure was implemented using parameter estimates for FSIQ from Turkheimer et al. (2003), an early and highly publicized study that reported a much larger  $G \times$  SES than was indicated by the current meta-analysis, results indicated that a minimum sample size of only 500 twin pairs is needed to achieve 80% power. It therefore appears that the inconsistency of previous US studies to replicate  $G \times$  SES effects on intelligence may have stemmed from low power associated with overly-optimistic expectations regarding the magnitude of the true interaction effect.

## Discussion

This meta-analysis of published and unpublished data provided clear answers to our three questions. First, studies from the US support a moderately sized  $G \times$  SES interaction on intelligence and academic achievement ( $a' = .074$ ; Figure 1). Second, in studies conducted

outside the US (in Western Europe and Australia), the best estimate for  $G \times \text{SES}$  magnitude was very slightly negative, and not significantly different from zero. Third, the difference between the US and the non-US studies in the estimated magnitude of  $G \times \text{SES}$  effect was itself significant.

Beyond nation, we did not identify any other significant moderators of  $G \times \text{SES}$  effects. We examined whether test performance was measured in childhood or adulthood, childhood age of testing, whether the tests measured achievement/knowledge or intelligence, and whether the tests were of single ability or a composite cognitive measure. None of these additional moderators achieved statistical significance, and the cross-national difference in  $G \times \text{SES}$  effect remained when each of these possible moderators was entered into the meta-analytic model. Thus, the cross-national difference identified does not appear to be an epiphenomenon of cross-national differences in the age ranges examined or the particular intelligence or achievement outcomes measured.

We did not find evidence of publication or reporting bias in our meta-analytic dataset. Both visual and formal tests of funnel plot asymmetry were not significant, both when applied to the overall meta-analytic dataset, and when separately applied to effects sizes from US and non-US samples. Moreover,  $p$ -curve analysis, which is based on the shape of the distribution of  $p$ -values for effect sizes surpassing the  $p < .05$  significance threshold, indicated no evidence for  $p$ -hacking (i.e. questionable research practices). Consistent with the cross national difference identified in our primary meta-analysis,  $p$ -curve analysis further indicated evidential value for a non-zero  $G \times \text{SES}$  effect in US studies, but did not indicate evidential value for a nonzero  $G \times \text{SES}$  effect in the non-US studies. Robustness checks indicated that results were not driven by a small number of studies with either disproportionately large samples or disproportionately large (or small) effect sizes. Variation in  $G \times \text{SES}$  effect sizes in US studies- ranging from near-zero to extremely large- was no greater than would be expected by sampling variability alone.<sup>1</sup>

We also replicated the well-established phenomenon of increasing genetic influences, and decreasing shared environmental influences, on intelligence with childhood age (e.g. Briley & Tucker-Drob, 2013; Haworth et al., 2009) in the current meta-analytic dataset. This replication further demonstrates that the studies that met inclusion criteria for our meta-analysis on  $G \times \text{SES}$  effects are representative of the wider body of behavioral genetic research on intelligence. Interestingly, as can be seen by comparing Figures 2 and S2, these  $G \times \text{Age}$  trends closely parallel the US  $G \times \text{SES}$  effect. Genes account for considerably more variation in intelligence at both higher ages and in higher US socioeconomic contexts. Indeed, both phenomena may reflect a process of increased and accumulated effects of gene-environment transactions with the increased opportunity that comes with both social class and age (Tucker-Drob et al., 2013; Turkheimer & Horn, 2014).

The results indicate that  $G \times \text{SES}$  effects are not uniform, but can rather take positive, zero, and even negative values depending on factors that differ at the national-level. The finding

---

<sup>1</sup>The results do, however, illustrate the discoverer's curse: The earliest studies reported much larger effects than those estimated in the current meta-analysis for the US: from 0.13 (Rowe et al., 1999) to as high as .24 (Turkheimer et al., 2003) or .33 (Scarr-Salapatek, 1971; not included in the meta-analysis).

that low SES was associated with attenuated genetic influence on intelligence in the US resolves an important debate. The finding that this interaction is observed only in the US, together with the novel discovery here that the effect may even reverse in sign (Netherlands), suggests that further research on nation-level differences in the effects of family SES on cognitive development is particularly important. Candidate mechanisms that might underlie nation-level differences range from national differences in how concepts of number and letter that underpin literacy and numeracy are imparted (Ramani & Siegler, 2008), to educational quality more broadly (Taylor, Roehrig, Soden-Hensler, Connor, & Schatschneider, 2010), medical and educational access (Bates et al., 2013; Tucker-Drob et al., 2013), and macro-societal characteristics, such as upward social mobility (Ritchie, Bates, & Plomin, 2014) and income support (Duncan, Morris, & Rodrigues, 2011). Our results suggest that large-scale genetically-informed research that incorporates careful measurement and consideration of both proximal and national social factors may provide a unique key to understanding the impact of specific policies on individual differences in intellectual development and academic achievement.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We would like to thank the following individuals for providing us with results of re analyses that we requested: Juliana Gottschling, Kristen Jacobson, Robert Kirkpatrick, William Kremen, Matt Mc Gue, Robert Plomin, Sophie van der Sluis, Marion Spengler, and Maciej Trzaskowski.

### Funding

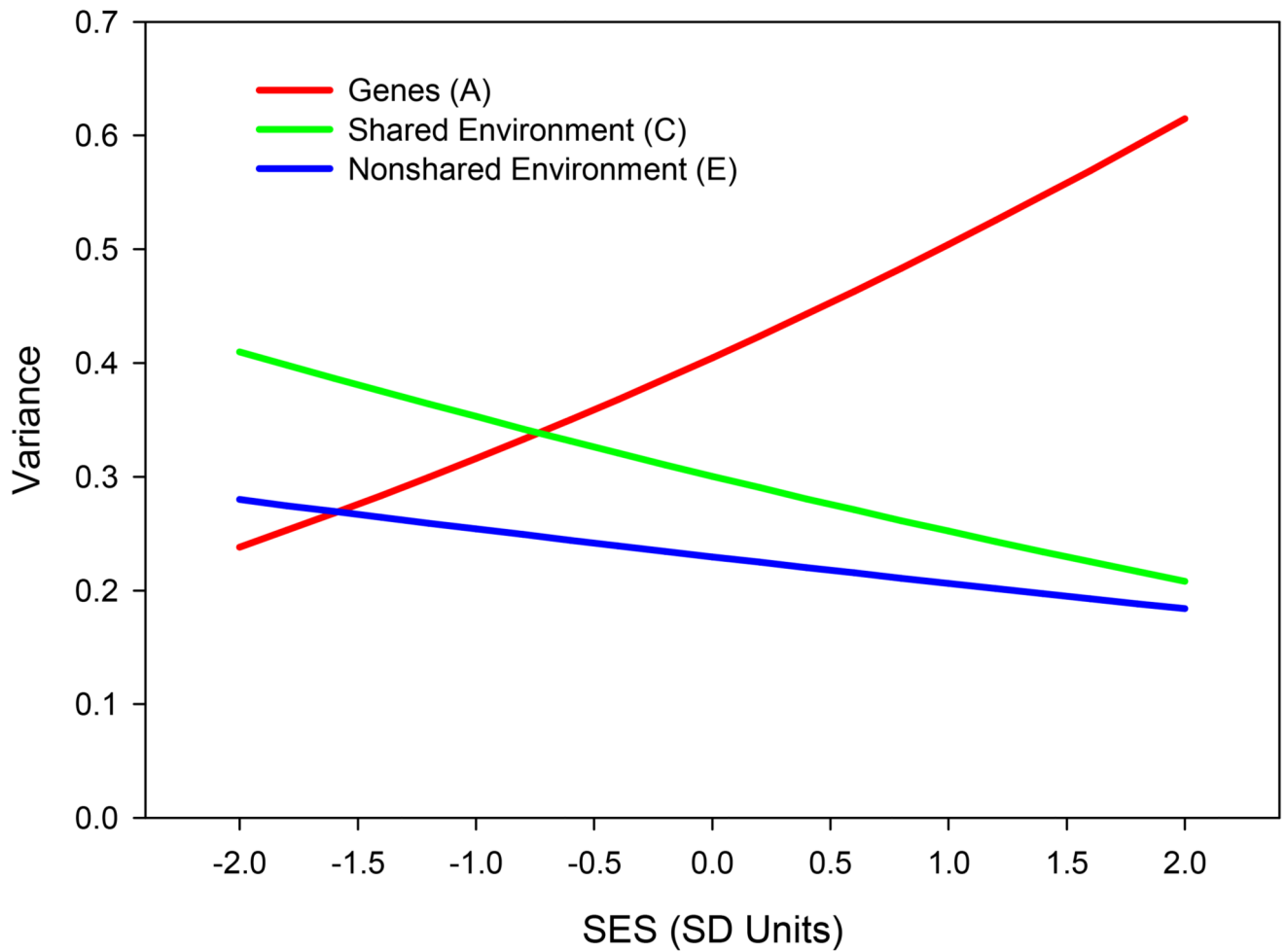
The Population Research Center at the University of Texas is supported by NIH grant R24HD042849.

## References

- Adler NE, Newman K. Socioeconomic disparities in health: pathways and policies. *Health Aff (Millwood)*. 2002; 21(2):60–76. [PubMed: 11900187]
- Asbury K, Wachs TD, Plomin R. Environmental moderators of genetic influence on verbal and nonverbal abilities in early childhood. *Intelligence*. 2005; 33(6):643–661.
- Bartels M, van Beijsterveldt CE, Boomsma DI. Breastfeeding, maternal education and cognitive function: a prospective study in twins. *Behavior Genetics*. 2009; 39(6):616–622. [PubMed: 19653092]
- Bates TC, Hansell NK, Martin NG, Wright MJ. When does socioeconomic status moderate the heritability of IQ?: Evidence from an Australian adolescent sample. *Child Development*. under review.
- Bates TC, Lewis GJ, Weiss A. Childhood Socioeconomic Status Amplifies Genetic Effects on Adult Intelligence. *Psychological Science*. 2013; 24(10):2111–2116. [PubMed: 24002887]
- Bouchard TJ Jr. McGue M. Familial studies of intelligence: a review. *Science*. 1981; 212(4498):1055–1059. [PubMed: 7195071]
- Briley DA, Tucker-Drob EM. Explaining the increasing heritability of cognitive ability across development: a meta-analysis of longitudinal twin and adoption studies. *Psychol Sci*. 2013; 24(9):1704–1713. [PubMed: 23818655]
- Bronfenbrenner U, Ceci SJ. Nature-nurture reconceptualized in developmental perspective: a bioecological model. *Psychological Review*. 1994; 101(4):568–586. [PubMed: 7984707]

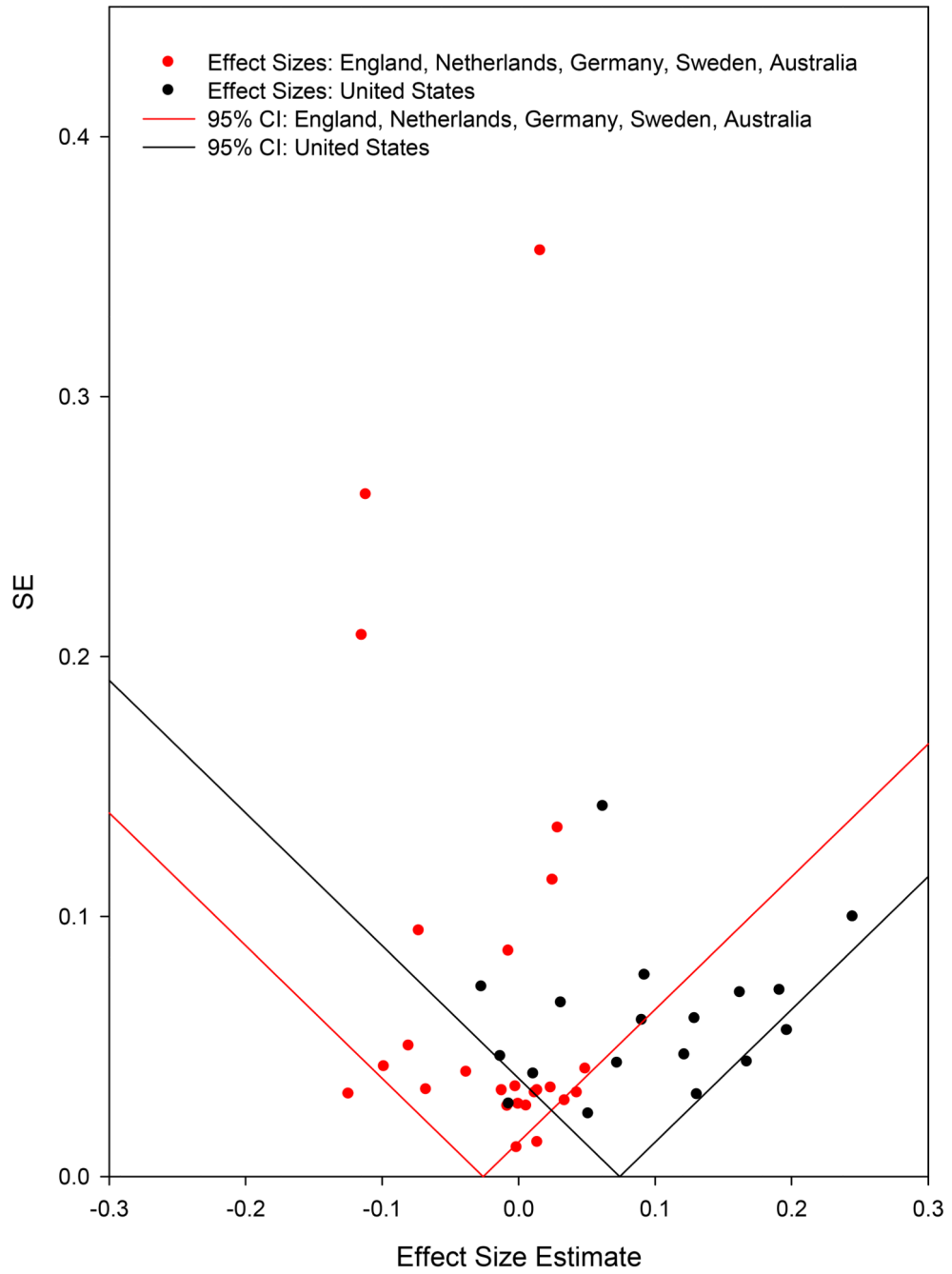
- Cheung AK, Harden KP, Tucker-Drob EM. GenexEnvironment interactions in early externalizing behaviors: parental emotional support and socioeconomic context as moderators of genetic influences? *Behavior Genetics*. 2014; 44(5):468–486. [PubMed: 24980660]
- Cheung MWL. A model for integrating fixed-, random-, and mixed-effects meta-analyses into structural equation modeling. *Psychological Methods*. 2008; 13(3):182–202. [PubMed: 18778151]
- Davies G, Tenesa A, Payton A, Yang J, Harris SE, Liewald D, et al. Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular Psychiatry*. 2011; 16(10):996–1005. [PubMed: 21826061]
- DeNavas-Walt, C.; Proctor, BD.; Bureau, USC. *Current Population Reports, P60-249, Income and Poverty in the United-States: 2013*. US Department of Commerce Economics and Statistics Administration US CENSUS BUREAU; Washington, DC: 2014.
- Duncan GJ, Morris PA, Rodrigues C. Does money really matter? Estimating impacts of family income on young children’s achievement with data from random-assignment experiments. *Developmental psychology*. 2011; 47(5):1263. [PubMed: 21688900]
- Eaves LJ, Jinks JL. Insignificance of evidence for differences in heritability of IQ between races and social classes. *Nature*. 1972; 240:84–88. [PubMed: 4564817]
- Fischbein S. IQ and social class. *Intelligence*. 1980; 4(1):51–63.
- Friend A, DeFries JC, Olson RK. Parental education moderates genetic influences on reading disability. *Psychological Science*. 2008; 19(11):1124–1130. [PubMed: 19076484]
- Friend A, DeFries JC, Olson RK, Pennington B, Harlaar N, Byrne B, et al. Heritability of high reading ability and its interaction with parental education. *Behavior Genetics*. 2009; 39(4):427–436. [PubMed: 19296213]
- Galton, F. *Hereditary Genius: An inquiry into its laws and consequences*. Macmillan and Co.; London: 1869.
- Grant MD, Kremen WS, Jacobson KC, Franz C, Xian H, Eisen SA, Lyons MJ. Does parental education have a moderating effect on the genetic and environmental influences of general cognitive ability in early adulthood? *Behavior genetics*. 2010; 40:438–446. [PubMed: 20300818]
- Hanscombe KB, Trzaskowski M, Haworth CM, Davis OS, Dale PS, Plomin R. Socioeconomic Status (SES) and Children’s Intelligence (IQ): In a UK-Representative Sample SES Moderates the Environmental, Not Genetic, Effect on IQ. *PLoS One*. 2012; 7(2):e30320. [PubMed: 22312423]
- Harden KP, Turkheimer E, Loehlin JC. Genotype by environment interaction in adolescents’ cognitive aptitude. *Behavior Genetics*. 2007; 37(2):273–283. [PubMed: 16977503]
- Hauser RM. Educational Stratification in United-States. *Sociological Inquiry*. 1970; 40(2):102. &
- Haworth CM, Wright MJ, Martin NW, Martin NG, Boomsma DI, Bartels M, et al. A twin study of the genetics of high cognitive ability selected from 11,000 twin pairs in six studies from four countries. *Behav Genet*. 2009; 39(4):359–370. [PubMed: 19381794]
- Ioannidis JP. Why most discovered true associations are inflated. *Epidemiology*. 2008; 19(5):640–648. [PubMed: 18633328]
- Jensen AR. How much can we boost IQ and scholastic achievement. *Harvard educational review*. 1969; 39(1):1–123.
- Kirkpatrick RM, McGue M, Iacono WG. Replication of a gene-environment interaction Via Multimodel inference: additive-genetic variance in adolescents’ general cognitive ability increases with family-of-origin socioeconomic status. *Behavior Genetics*. 2015; 45(2):200–214. [PubMed: 25539975]
- Kremen WS, Jacobson KC, Xian H, Eisen SA, Waterman B, Toomey R, et al. Heritability of word recognition in middle-aged men varies as a function of parental education. *Behavior Genetics*. 2005; 35(4):417–433. [PubMed: 15971023]
- Muthén, LK.; Muthén, BO. *Mplus User’s Guide*. 7 ed.. Muthén & Muthén; Los Angeles, CA: 1998-2012.
- Nagoshi CT, Johnson RC. Socioeconomic status does not moderate the familiarity of cognitive abilities in the Hawaii Family Study of Cognition. *Journal of biosocial science*. 2005; 37(06):773–781. [PubMed: 16221325]

- Nelson CA 3rd, Zeanah CH, Fox NA, Marshall PJ, Smyke AT, Guthrie D. Cognitive recovery in socially deprived young children: the Bucharest Early Intervention Project. *Science*. 2007; 318(5858):1937–1940. [PubMed: 18096809]
- Nisbett RE, Aronson J, Blair C, Dickens W, Flynn J, Halpern DF, et al. Intelligence: new findings and theoretical developments. *American Psychologist*. 2012; 67(2):130–159. [PubMed: 22233090]
- Plomin R, Haworth CM, Meaburn EL, Price TS, Wellcome Trust Case Control, C. Davis OS. Common DNA markers can account for more than half of the genetic influence on cognitive abilities. *Psychological Science*. 2013; 24(4):562–568. [PubMed: 23501967]
- Prescott CA, McArdle JJ, Achorn DL, Kaiser A, Lapham SJ. Estimating Genetic and Family Environment Effects when Zygosity is Unknown by using Data from Twin and Sibling Pairs. *Behavior Genetics*. 2014
- Purcell S. Variance components models for gene environment interaction in twin analysis. *Twin Res*. 2002; 5(6):554–571. [PubMed: 12573187]
- Ramani GB, Siegler RS. Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Dev*. 2008; 79(2):375–394. [PubMed: 18366429]
- Rhemtulla M, Tucker-Drob EM. Gene-by-socioeconomic status interaction on school readiness. *Behavior genetics*. 2012; 42(4):549–558. [PubMed: 22350185]
- Ritchie SJ, Bates TC, Plomin R. Does Learning to Read Improve Intelligence? A Longitudinal Multivariate Analysis in Identical Twins From Age 7 to 16. *Child development*. 2014
- Rowe DC, Jacobson KC, Van den Oord EJ. Genetic and environmental influences on vocabulary IQ: parental education level as moderator. *Child Development*. 1999; 70(5):1151–1162. [PubMed: 10546338]
- Simonsohn U, Nelson LD, Simmons JP. P curve: A key to the file-drawer. *Journal of Experimental Psychology: General*. 2014; 143(2):534. [PubMed: 23855496]
- Scarr-Salapatek S. Race, social class, and IQ. *Science*. 1971; 174(4016):1285–1295. [PubMed: 5167501]
- Soden-Hensler, B. An examination of gene  $\times$  socioeconomic status interactions for reading achievement. Florida State university; Florida; 2012.
- Taylor J, Roehrig AD, Soden-Hensler B, Connor CM, Schatschneider C. Teacher quality moderates the genetic effects on early reading. *Science*. 2010; 328(5977):512–514. [PubMed: 20413504]
- Tucker-Drob EM, Briley D, Harden KP. Genetic and Environmental Influences on Cognition across Development and Context. *Current directions in psychological science*. 2013; 22(5):349–355. [PubMed: 24799770]
- Tucker-Drob EM, Rhemtulla M, Harden KP, Turkheimer E, Fask D. Emergence of a Gene  $\times$  Socioeconomic Status Interaction on Infant Mental Ability Between 10 Months and 2 Years. *Psychological Science*. 2011; 22(1):125–133. [PubMed: 21169524]
- Turkheimer E, Haley A, Waldron M, D'Onofrio B, Gottesman II. Socioeconomic status modifies heritability of IQ in young children. *Psychological Science*. 2003; 14(6):623–628. [PubMed: 14629696]
- Turkheimer E, Harden KP, D'Onofrio B, Gottesman II. The Scarr–Rowe interaction between measured socioeconomic status and the heritability of cognitive ability. *Experience and development: A festschrift in honor of Sandra Wood Scarr*. 2009:81–97.
- Turkheimer, E.; Horn, EE. Interactions Between Socioeconomic Status and Components of Variation in Cognitive Ability. In: Finkel, D.; Reynolds, CA., editors. *Behavior Genetics of Cognition Across the Lifespan*. Vol. 1. Springer; New York: 2014. p. 41-68.
- Van Den Oord EJ, Rowe DC. An examination of genotype-environment interactions for academic achievement in an US national longitudinal survey. *Intelligence*. 1997; 25(3):205–228.
- van der Sluis S, Willemsen G, de Geus EJ, Boomsma DI, Posthuma D. Gene-environment interaction in adults' IQ scores: measures of past and present environment. *Behavior Genetics*. 2008; 38(4): 348–360. [PubMed: 18535898]



**Figure 1.**

Meta analytic estimate of Gene $\times$ SES interaction on cognitive test performance for the United States. Note that cognitive test scores were standardized to a z scale within each dataset prior to model fitting. This plot is very close to (but not identical) to a plot in which the y axis represents proportions of variance, where proportions are instantaneous for each level of SES.



**Figure 2.**  
Funnel plot of Gene x SES interaction effect sizes for US and Non-US Studies.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



Table 1

Studies of Gene $\times$ SES Interaction included in Meta-Analysis.

Study	Study Name	Originally Reported in	N Pairs	k Effect Sizes	Country	Age Range	Data acquisition
1	Brisbane Adolescent Twin Study (BATS)	Bates, Hansel, Martin & Wright (Bates et al., under review)	1176	1	Australia	Adolescence	Raw data access
2	Cognitive Ability, Self-Perceived Motivation, and School Achievement (CoSMoS)	Spengler, Gottschling, & Spinath, (2011, unpublished)	542	1	Germany	Adolescence	Provided by authors
3	Early Childhood Longitudinal Study-Birth Cohort (ECLS-B)	Tucker-Drob et al. (2011), Rhemulla and Tucker-Drob (2012), Tucker-Drob (unpublished)	750	6	US	Early Childhood	Raw data access
4	Florida State Twin Registry (FSTR)	Soden-Hensler (2012)	466	1	US	Middle Childhood	Reported in article
5	Midlife in the United States (MIDUS)	Bates et al. (2013)	851	1	US	Adulthood	Raw data access
6	Minnesota Center for Twin and Family Research (MTFS)	Kirkpatrick, McGue, & Iacono (Kirkpatrick, McGue, & Iacono, 2015)	2494	1	US	Adolescence	Provided by authors
7	National Collaborative Perinatal Project (NCPP)	Turkheimer et al. (2003)	319	3	US	Early Childhood	Raw data access
8	National Longitudinal Study of Adolescent Health (Add Health)	Rowe et al. (1999), Jacobson (unpublished)	1909	2	US	Adolescence and Early Adulthood	Provided by authors
9	National Merit Twin Study (NMSQT)	Harden et al. (2007)	839	2	US	Adolescence	Raw data access
10	Netherlands Twin Registry - Child (NTR-C)	Bartels et al. (2009)	3132	1	Netherlands	Middle Childhood	Re-analysis of data reported in article
11	Swedish Twins (SWT)	Fischbein (1980)	215	2	Sweden	Middle Childhood	Re-analysis of data reported in article
12	Twins Early Development Study (TEDS)	Hanscombe et al. (2012), Asbury, Wachs, and Plomin (2005)	8716	16	England	Childhood and Adolescence	Results from authors

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Study	Study Name	Originally Reported in	N Pairs	k Effect Sizes	Country	Age Range	Data acquisition
13	Vietnam Era Twin Registry (VET)	Grant et al. (2010) Kremen et al. (2005)	3203	2	US	Early and Middle Adulthood	Results from authors
14	Netherlands Twin Registry-Adult (NTR-A)	van der Sluis et al. (2008)	314	4	Netherlands	Early and Middle Adulthood	Results from authors

Note: The number of pairs for van der Sluis et al. (2008) was approximated by dividing the reported number of individual participants by 2. The number of pairs for the Early Childhood Longitudinal Study was rounded to the nearest 50 in accordance with data security regulations. All parameter estimates for individual studies are provided in the archived datafile, located at <https://osf.io/vkmpc>.

Meta-analytic results for full Structural Equation Model parameter estimates for US and non-US studies.

**Table 2**

Meta-analysis parameter	Effect Size													
	$s$	$s$ SE	$a$	$a$ SE	$a'$	$a'$ SE	$c$	$c$ SE	$c'$	$c'$ SE	$e$	$e$ SE	$e'$	$e'$ SE
US	0.300	0.027	0.636	0.044	0.074	0.020	0.548	0.07	-0.046	0.032	0.479	0.033	-0.025	0.016
Europe	0.280	0.022	0.672	0.045	-0.027	0.022	0.507	0.052	-0.029	0.012	0.471	0.034	0.063	0.065
Difference	0.020	0.033	-0.036	0.065	0.101	0.032	0.042	0.086	-0.017	0.028	0.008	0.047	-0.088	0.068
$\tau$	0.073	0.016	0.097	0.013	0.029	0.012	0.158	0.018	0.007	0.195	0.115	0.018	0.137	0.069

Note. Parameters listed in the rows labeled "US" and "Europe" are the meta-analytic expectations for the US and Europe, respectively. Parameters listed in the row labeled "Difference" represent the difference between the meta-analytic expectation for the US and that of Europe.  $\tau$  represents the standard deviation of the random effect, representing residual heterogeneity in effect sizes.

**Table 3**

Additional moderation tests of the  $a'$  effect size representing the  $G \times \text{SES}$  interaction.

Meta-analysis parameter	Moderation Model 1 (Full Sample)		Baseline Childhood Model (Age < 20 years)		Moderation Model 2 (Age < 20 years)		Moderation Model 3 (Age < 20 years)		Moderation Model 4 (Age < 20 years)		Moderation Model 5 (Age < 20 years)	
	$a'$	SE	$a'$	SE	$a'$	SE	$a'$	SE	$a'$	SE	$a'$	SE
US	0.076	0.021	0.078	0.022	0.123	0.034	0.092	0.026	0.085	0.038	0.060	0.025
Europe	-0.026	0.022	-0.026	0.023	0.008	0.024	-0.009	0.017	-0.023	0.017	-0.028	0.023
Diff	0.102	0.032	0.104	0.035	0.114	0.034	0.101	0.034	0.108	0.043	0.087	0.037
Older than 20 years	-0.011	0.016										
Childhood Age					-0.003	0.002						
SES Single Indicator (1) vs. Composite (0)							-0.031	0.031				
Achievement/ Knowledge (1) vs. Intelligence (0)									-0.015	0.045		
Single Ability (1) vs. Composite (0) Cognitive Measure											0.046	0.033
$\tau$	0.029	0.012	0.033	0.013	0.030	0.013	0.031	0.012	0.032	0.012	0.027	0.016

Note. Parameters listed in the rows labeled "US" and "Europe" are the meta-analytic expectations for the US and Europe, respectively. Parameters listed in the row labeled "Difference" represent the difference between the meta-analytic expectation for the US and that of Europe.  $\tau$  represents the standard deviation of the random effect, representing residual heterogeneity in effect sizes.