rejection of stages, this may be an extreme counterreaction to a previous excess. For decades, at least since Freud's psychosexual theory of the early 20th century, psychology had overemphasized the early years of development. Freud, Piaget, Kohlberg, and others viewed early development as dynamic and dramatic, with little of interest occurring after adolescence. Life span and life course theories rejected such nonsense and instead emphasized development as lifelong. However, in doing so they neglected the special characteristics of early development (infancy through adolescence) that distinguish it from later development.

In our cultural-developmental theory we recognize that development is lifelong and that important changes occur long after adolescence. However, we also recognize that development is especially rapid from infancy through adolescence relative to later life, and that it is densely packed with developmental events. In the first 2 years of life the infant more than doubles in height and weight, and goes from an immobile, gurgling neonate to a walking, talking being with well-developed attachments to persons in the social environment. During adolescence, the dramatic changes of puberty transform persons physically, sexually, and socially in the space of a few years. Nothing in later life results in comparable normative changes in a comparable number of years.

Furthermore, the experiences of early development can have lasting effects on later development, especially if the experiences are adverse. Childhood is a critical period for language development, and a child who is deprived of language stimulation in the early years may find it difficult to make up for the deprivation in adolescence and beyond. If schooling is absent or inadequate during childhood, the child may enter adulthood lacking in the skills necessary for adequate employment and may find the deficiency difficult to remedy as an adult. Although abundant research has shown that many people are remarkably resilient in the face of adverse circumstances in childhood and beyond, it remains true that a variety of physical and emotional deprivations in childhood are predictive of higher risk for problems later in development. Consequently, it is important to account for the special characteristics of early development in a developmental theory of the life course.

Within Cultures, Individual Differences Are Due Substantially to Genetic Variability and Genotype–Environment Interactions

Both life span and life course theories claim to account for the contribution of genes to development. Baltes, in describing life span theory, describes it as involving "developmental biocultural co-constructivism"

(Baltes et al., 2006, p. 586), with "bio" including genetics. Elder acknowledges that "virtually all research on the life course has proceeded without considering the influence of genes and behavior" but seeks to remedy this neglect by describing "mechanisms of gene–environment interactions" (Elder & Shanahan, 2006, p. 702).

However, neither theory accounts adequately for the role genes play in human development. Life span theory emphasizes "plasticity" in development without acknowledging how plasticity is constrained by a person's genotype. Furthermore, life span theory seeks to contribute to the betterment of the world and genetics are sometimes seen as an obstacle to this goal, with one life span theorist even going so far as calling behavior genetics "today's version of the biologizing errors of the past such as eugenics and racial hygiene" (Lerner, 2006, p. 6). Life course theory also shows an inadequate grasp of the implications of behavior genetics, as noted earlier in this chapter.

In our view, the findings of behavior genetics research in recent decades have revolutionary implications for human development that have not been adequately accounted for in either life span or life course theory. These findings show that (a) in the course of early development (infancy through adolescence), the influence of family environment on individual development diminishes whereas the influence of genes increases; and (b) very little of the variability among individuals can be explained by shared family environment. Although these profoundly important discoveries are missing from life span and life course theories, they are explained well by Scarr and McCartney's theory of genotype → environment interactions (hereafter GE interactions), and we seek to incorporate the insights of their theory into our cultural-developmental stage theory of the life course.

Scarr and McCartney (1983) proposed three types of GE interactions: passive, evocative, and active. Passive GE interactions occur because in biological families parents provide both genes and family environment. Genes and environment tend to be mutually reinforcing in biological families, that is, the environment parents provide for their children tends to be consistent with the genetic tendencies they have provided. For example, parents who behave aggressively toward their children, punishing them physically, tend to have children who are more aggressive than other children. This relation is likely to be due not only to the parents' aggressive behavior but also to genes they provided that may have inclined their children to behave aggressively. However, in research on biological families it is difficult to tell how much the relations between family environment and child outcomes are due to genetics and how much are due to environment.

Tsuang, M. T., Lyons, M. J., Eisen, S. A., True, W. T., Goldberg, J., & Henderson, W. (1992). A twin study of drug exposure and initiation of use. *Behavior Genetics, 22,* 756.

Walker, S. O., & Plomin, R. (2006). Nature, nurture, and perceptions of the classroom environment as they relate to teacher assessed academic achievement: A twin study of 9-year-olds. *Educational Psychology, 26,* 541–561.

5

THE SCARR–ROWE INTERACTION BETWEEN MEASURED SOCIOECONOMIC STATUS AND THE HERITABILITY OF COGNITIVE ABILITY

Eric Turkheimer and K. Paige Harden
University of Virginia

Brian D'Onofrio
Indiana University

Irving I. Gottesman
University of Minnesota

The hypothesis that the heritability of cognitive ability might vary with socioeconomic status (SES) was first forwarded by Sandra Scarr in a 1971 paper published in *Science*. Introducing a study of Black and White twins in the Philadelphia school system, Scarr wrote: "The environmental disadvantage hypothesis predicts that IQ scores within advantaged groups will show larger proportions of genetic variance and smaller proportions of environmental variance than IQ scores for disadvantaged groups" (p. 1286).

Scarr overcame considerable methodological obstacles to demonstrate that her hypothesis appeared to be correct. As is still true today, twins raised in poverty were generally not included in American twin samples, which are most often based on volunteers, or in population-based twin studies conducted in Scandinavia, where poverty itself is

rare. The Philadelphia school system provided a sample that was relatively unselected, but no zygosity data were available, forcing Scarr to resort to comparisons of same-sex and opposite-sex twin pairs, a design with considerably less power than the fully informed twin design (Eaves & Jinks, 1972). Structural equation modeling was not yet the standard for twin analyses, so Scarr simply broke down comparisons of twin correlations for Black and White twins into groups above and below the median for socioeconomic status. Despite the handicaps, heritabilities were found to be consistently higher for Whites than for Blacks, and for higher compared to lower socioeconomic groups, both across races and within them.

Aside from a partial replication in Philadelphia conducted by Scarr (1981b) and a report from Sweden by Fischbein (1980), the hypothesis lay mostly fallow for 25 years. Even as structural equation modeling revolutionized psychological genetics in the '80s and '90s, no studies of variation in heritability as a function of socioeconomic status were undertaken. Along with the continuing absence of appropriate data, an important reason was methodological. Even the sophisticated structural equation programs of the day, most notably LISREL, did not have the ability to estimate models in which latent variables—in this case, the familiar ACE variances of biometric decomposition (A [genetic], C [common or family environment], E [nonshared or unique environment])—entered into interactions with observed variables like socioeconomic status. Conducting such an analysis at the time involved either ingenious computational manipulations that were complex to the point of impracticality (Kenny & Judd, 1984) or required more computational power than was generally available at the time.

Rekindling of interest in the phenomenon almost 30 years after Scarr identified it was sparked by two key factors: availability of suitable data and computational feasibility. In 1999, Rowe, Jacobson, and van den Oord examined data from the National Longitudinal Study of Adolescent Health (Add Health), a large, representative sample of American youth, then in early adolescence, who were administered a version of the Peabody Picture Vocabulary Test; years of education was available for the mothers. The sample included monozygotic (MZ) and dizygotic (DZ) twins, as well as siblings, half-siblings, and unrelated children reared together. The computational advance came in the form of DF analysis, a method introduced by DeFries and Fulker (1985) that allows estimation of latent variable models using ordinary regression programs. In DF analysis, twin data are double entered, and the scores for one twin from a pair are regressed on the other, along with zygosity, and, in the simplest form of the model, the interaction

of zygosity and the cotwin. The latter interaction is an estimate of the extent to which identical twins are more predictable from each other than DZ twins. Moreover, when the model is properly parameterized, the estimated unstandardized regression coefficients can be interpreted directly as standardized variance components in a traditional biometric twin analysis.

Rowe et al. (1999) used an extension of DF analysis proposed by LaBuda and DeFries (1990) in which an observed moderator variable, such as socioeconomic status, is added to the equation and also entered into a three-way interaction with the cotwin score and zygosity. This interaction is a measure of the extent to which the heritability of the trait varies linearly with the moderator. This interaction term was significant in the National Longitudinal Study of Adolescent Health data, producing a crossed interaction, with most of the variance in families with poorly educated mothers explained by the shared environment and most of the variance in families with well-educated families explained by genes. We propose that this effect, hypothesized and first investigated by Scarr in 1971, then revived and established by the late David Rowe, be called the Scarr–Rowe interaction.

In 2003, we (Turkheimer, Haley, Waldron, D'Onofrio, & Gottesman, 2003) conducted an analysis of the SES by heritability analyses in the National Collaborative Perinatal Project (NCPP). The NCPP is particularly well-suited for this purpose. It comprised an unselected sample of twins, many of them raised in poverty. A well-validated measure of socioeconomic status with good psychometric properties is available in the NCPP data set (Myrianthopoulos & French, 1968). In addition, an extensive battery of ability tests was administered to the children in infancy and at ages 4 and 7. Tests included seven subtests of the Wechsler Intelligence Scale for Children (WISC); the three subscales of the Wide Range Achievement Test (WRAT); and a variety of tests of speech, language, and neuropsychological function. Our 2003 paper only reported results for Full Scale, Verbal, and Performance IQ. We employed the computer program Mx (Neale et al., 1999) to estimate the interaction between latent biometric variances and measured socioeconomic status, using methods described by Purcell (2002).

A very direct way to illustrate the Scarr–Rowe interaction is illustrated in Figure 5.1. In this plot we have avoided parameterizing the results in terms of the familiar but easily misunderstood variance components. Instead, we have computed the absolute value of the difference in Full Scale IQ (FSIQ) for each pair of twins, indicating DZ pairs with a D and MZ pairs with an M. Greater differences are indicative of twin pairs who are less similar to each other. We have then plotted

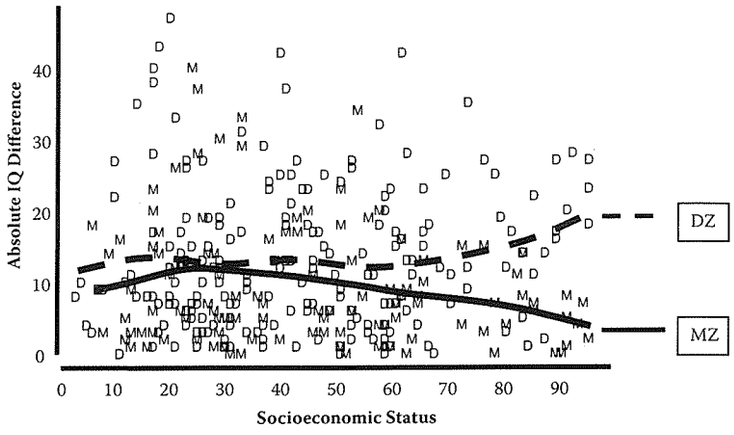

Figure 5.1 Absolute differences between Full Scale IQs of pairs of MZ (M) and DZ (D) twins as a function socioeconomic status.

these differences against SES, and computed a smooth line expressing the difference as a function of SES, separately for MZ and DZ twins. In general, MZ twins are more similar, but it is apparent that in more affluent households, MZ twins become even more similar, whereas the DZ twins become more different. All subsequent analyses of the Scarr–Rowe interaction are simply reparameterizations of this basic phenomenon.

Structural equation modeling demonstrated large crossed interactions for Full Scale and Performance IQ, but not for Verbal IQ. For families at the lowest levels of socioeconomic status, shared environment accounted for almost all of the variation in IQ, with genes accounting for practically none. As socioeconomic status increased, the contribution of shared environment diminished and the contribution of genes increased, crossing in lower middle-class families. Finally, in the most socioeconomically advantaged families (who were by and large not wealthy in the NCPP sample) practically all of the variation in IQ was accounted for by genes, and almost none accounted for by shared environment.

Several studies have attempted to replicate the interaction since our study was published, with mixed results. Kremen et al. (2005) identified a significant interaction between parental education and reading scores on the WRAT among 690 adult twins in the Vietnam Era Twin Registry. Descriptive analyses showed that in the twins with the least educated parents, additive genetics and common environment each accounted for 36% of the variability in reading scores; in the twins with the best educated parents, additive genetics accounted for 56% and common environment 12%. Modeling demonstrated that the phenotypic

variance of reading scores decreased as a function of parental education, but were unable to distinguish statistically two competing models to explain this phenomenon. In one, which they refer to as a "scalar" model, the magnitude of the ACE components all decrease linearly as a function of parental education, so the proportion of variance accounted for by each does not change. In the second model, which the authors preferred, common and nonshared environment decreased as a function of parental education, while additive genetic variance remained constant. The authors also report having reanalyzed the Rowe et al. (1999) data from Add Health, showing that in this study shared environmental variance increased and additive genetic variance decreased as a function of parental education.

Asbury, Wachs, and Plomin (2005) used a sample of 4,446 four-year-old British twins from the Twins Early Development Study in London whose parents were administered a child ability scale over the phone. The only interactions they found were in the opposite direction, that is, children in high-risk impoverished environments showed higher heritabilities. It should be noted that the procedures used in this study differed substantially from those originally reported by Scarr and elaborated on by Rowe et al. (1999) and Turkheimer et al. (2003). Most obviously, the study was conducted in Britain, where the nature of poverty may be very different from that experienced by American children, especially 50 years ago when Scarr's data were collected. In addition, the ability tests were administered over the phone to parents who were asked to report on their children, who were only 4 years old at the time.

Nagoshi and Johnson (2005) found no interaction between paternal education and the familial transmission of cognitive ability from parents to children. We note, however, that although Nagoshi and Johnson describe their study as an attempted confirmation of Turkheimer et al. (2003), the absence of genetically informative data makes the similarity between the two studies quite remote. Nagoshi and Johnson stated, "If, in a general way, heritability decreases, while the effects of shared environment increase in lower income SES groups, when assessing twins, the same results should be attained when assessing parent/child resemblances in cognitive ability, even though such familiality includes both genetic and common family environment effects" (p. 774). Unfortunately this reasoning appears to get the situation exactly backward. Since most instances of the Scarr–Rowe interaction that have been reported include an increase in genetic transmission with high SES coupled with a decrease in shared environmental transmission, the net effect on familiality of transmission could be expected to be a mean of zero, which is exactly what Nagoshi and Johnson report.

Finally, our own lab (Harden, Turkheimer, & Loehlin, 2007) recently re-analyzed Loehlin and Nichols's (1976) National Merit Scholar Qualifying Test (NMSQT) twin data. This sample of 839 twin pairs was drawn from the population of nearly 600,000 adolescents who completed the NMSQT in 1962. The NMSQT is composed of five subtests: English Usage, Mathematics Usage, Social Science Reading, Natural Science Reading, and Word Usage/Vocabulary. Mothers reported on both their own and the fathers' level of education and the annual family income in a written questionnaire. Parental education was classified on a 6-point ordinal scale, from less than an 8th-grade education to a graduate or professional degree. Mid-parent education was calculated as the average of maternal and paternal education (median = 3.5; variance = 1.44). Income was classified on a 7-point ordinal scale (median = 3; variance = 2.37), from less than $5,000 per year to over $25,000 per year (roughly equivalent to less than $31,250 to over $156,250 in 2004 dollars).

This study developed some new models that are relevant to those we will present later in this chapter. For the first time, rather than modeling interactions between socioeconomic variables and observed ability scores, we modeled interactions with the biometric components of a latent variable representing the common variance shared among the five observed ability scores. Separate models were fit examining interactions with parental education and parental income. Results were similar for the two socioeconomic indicators, with around 40% of the variance accounted for by both additive genetics and the common environment in the families with the lowest incomes and levels of education, whereas in the richer and better educated families the preponderance of the variance was accounted for by genetics and about 30% accounted for by the common environment. The effect was somewhat stronger, and statistically significant, for income as opposed to education, although we could not reject a model in which the interactions for the two variables were identical.

Many questions about the Scarr–Rowe interaction remain to be answered. The first, of course, is the extent to which the effect is generalizable across populations, abilities, and environments. In this regard the interaction faces a difficult standard, because the effect on which it is based—the heritability of ability itself—is one of the most replicable findings in the social sciences. But given the variety of intelligence tests and the cultures in which they may be administered, the well-established differences between children and adults in the biometrics of ability, and the wide range of environmental circumstances in which people around the world are raised, one would not necessarily expect

the interaction to occur in every population or test that might be studied. Nevertheless, we feel it has been reasonably well established that the effect can be expected to occur in populations containing substantial numbers of participants reared in poverty. In all cases to date, such samples have been American.

Another question is how impoverished environments must be before the effect becomes observable. We are inclined to think that some degree of actual poverty is required in the form of deficient schools and chaotic home environments. All of the studies that have been successful in finding the Scarr–Rowe interaction have used samples that included the lower end of the American socioeconomic distribution, whereas the two studies that have nominally failed to find it, in Hawaii and Britain, did not. The National Merit study is an exception to this pattern, showing the effect even though the sample was strongly selected for at least middle-class status. One might speculate that the National Merit Test is more difficult than a standardized IQ test, allowing socioeconomic effects to appear at relatively lower levels of deprivation.

A related question is whether the moderation of heritability by SES is linear or curvilinear. Scarr (1981a), among others, has hypothesized that the effect of environment on ability might be nonlinear or even a threshold, with differences among poor environments exerting a stronger effect than differences among adequate ones. Although our 2003 report of the NCPP described the results as "nonlinear," as time has gone by we have become more circumspect about the possibility of detecting nonlinearity in the effect. Socioeconomic status is measured on a weak scale, probably no more than ordinal, variance components are often a quadratic function of estimated parameters, and the most recent statistical methods used to study the interaction often involve models of the log of the variance components rather than the variances themselves. All of these considerations call into question any specification or speculation about the shape of the functions for SES by variance components.

An even more difficult question about the environmental component of the Scarr–Rowe interaction involves its composition. Environment and socioeconomic status are extremely coarse measures. From the point of view of establishing social policy it would be very useful to know exactly what it is about impoverished environments that explain their potency. Is it poor schools, or disturbed parent–child relations, or exposure to environmental toxins? Evans' (2004) comprehensive review of impoverished environments makes it abundantly clear that children raised in poverty are disadvantaged in every way imaginable. Evans writes in the abstract:

> Poor children confront widespread environmental inequities. Compared with their economically advantaged counterparts, they are exposed to more family turmoil, violence, separation from their families, instability, and chaotic households. Poor children experience less social support, and their parents are less responsive and more authoritarian. Low-income children are read to relatively infrequently, watch more TV, and have less access to books and computers. Low-income parents are less involved in their children's school activities. The air and water poor children consume are more polluted. Their homes are more crowded, noisier, and of lower quality. Low-income neighborhoods are more dangerous, offer poorer municipal services, and suffer greater physical deterioration. Predominantly low-income schools and day care are inferior.

Unfortunately, as we have argued elsewhere (Turkheimer, 2005), this is another topic on which it may prove difficult to make meaningful progress using existing large-scale data sets. In our view, omnibus environmental variables comprise a multitude of tiny environmental effects, which are effectively summed by either gross variables like SES or latent variance components like the shared environment (see McCartney & Berry, this volume, for some evidence of proximal environmental moderators of genetic effects). Indeed, Evans (2004) hypothesizes that exposure to multiple environmental stressors may be the key to understanding the cumulative effects of impoverished environments on development.

Past attempts to decompose the nonshared environmental variance component into individual environmental effects have been largely unsuccessful (Turkheimer & Waldron, 2000), but Evans and English (2002) have demonstrated that the accumulation of childhood risk exposures may mediate the effects of any individual risk factor.

We are more hopeful about the possibilities for understanding the domains of ability in which the Scarr–Rowe effect is manifest. Compared to global environmental constructs, ability is relatively straightforward to measure and has a robust and well-understood multivariate structure. The NCPP, in contrast to most other studies that have been conducted, included a broad array of cognitive abilities to form the basis of a multivariate analysis of the interaction. In this chapter we report a first step in this direction, in which we conduct a factor analysis of 10 measures of cognitive ability, extract a common factor, and decompose both the common factor and unique variances into biometric components that interact with observed socioeconomic status. In addition, we conduct the analysis using a relatively new statistical procedure, BUGS (Bayesian inference using Gibbs sampling).

METHODS

Participants

In the current study, we used data from the National Collaborative Perinatal Project, which included a large national sample of American mothers, who were enrolled into the study during pregnancy (n = 48,197), and their children (n = 59,397), who were followed from birth until age 7 (Nichols & Chen, 1981). Participants were recruited from 12 urban hospitals around the country and included a high proportion of racial minorities and impoverished families. Extensive medical, psychological, and socioeconomic data were obtained for the mothers during pregnancy, and for the children at birth and at ages 8 months, 1 year, 4 years, and 7 years. Socioeconomic scores were obtained at mother's registration in the study and at the 7-year evaluation. These scores were based on the 100-point system of Myrianthopoulos and French (1968) and computed from a linear combination of parental education, occupational status, and income.

The sample included 623 twin births. Of these, 320 pairs with complete data regarding IQ, SES, and zygosity remained at the 7-year follow-up. Twins remaining in the sample at 7 years of age did not differ from twins lost to the sample in terms of birth order, mother's marital status at birth, family SES at birth, race, or family income at birth. One additional DZ pair was eliminated as an outlier, because of an 81-point difference between the IQ scores of the twins; the twin with the lower IQ was identified as brain damaged at birth. Of the remaining 319 pairs, 114 were monozygotic and 205 were dizygotic. Of the DZ pairs, 81 were same-sex pairs and 124 were opposite-sex pairs. There were no significant mean differences for any analysis variables between same- and opposite-sex DZ pairs, and no differences in the twin correlations, so the opposite-sex pairs were combined with the same-sex pairs in all analyses.

The twins were classified as 43% White, 54% Black, and 3% "other." The sample included a high proportion of impoverished families. The median number of years of education of the head of household was between 10 and 11 years, and 25% of household heads were not educated past the ninth grade. The median occupation was "service worker," and 25% of the household heads received occupational ratings of "laborer" or lower, including 14% with no occupation. The median family income was between $6,000 and $7,000 annually, equivalent to $22,100 in 1997

dollars, the most recent year for which an equivalent scale is available. Twenty-five percent of the families had incomes below the 1973 poverty level for a family of four (U.S. Census Bureau, 2002).

Measures

We selected 10 tests for multivariate analysis, including the seven WISC subtests included in the NCPP (Block Design, Coding, Digit Span, Vocabulary, Comprehension, Information, and Picture Arrangement) and the three subscales of the WRAT (Math, Reading, and Spelling). All tests were administered within a few months of the child's seventh birthday, so raw scores were used.

Statistical Analysis

The goal of our analysis was to identify a common factor among the 10 test scores, simultaneously performing a biometric twin analysis on the common factor and 10 unique variances, including interaction terms of the ACE components of the common and unique variances with measured socioeconomic status. It should be noted at the outset that this model, including 33 latent variable interactions (three on the common variance and three on each of the unique variances), presents formidable computational challenges and is probably beyond the range of available computers and likelihood-based software like Mplus or Mx.

The model included 10 λ_k loadings of the items onto a common factor η. Each item was also regressed on SES (transformed to have a mean of zero), with unstandardized regression weight β_k. The common variance was decomposed into ACE components, with the log variance of each component a linear function of SES, with intercepts A_0, C_0, and E_0, and unstandardized regression coefficients A_1, C_1, and E_1. The E_0 parameter was fixed to zero to identify the scale of the latent variances. The unique variances, μ_k, of each of the items were also decomposed into ACE components, with variances modeled as a log linear function of SES, with intercepts A_{k0}, C_{k0}, and E_{k0}, and slopes A_{k1}, C_{k1}, and E_{k1}.

This model was estimated using the Markov chain Monte Carlo (MCMC) method via the software WinBUGS by the MRC Biostatistics Unit. WinBUGS implements the Gibbs sampling algorithm (Geman & Geman, 1984) to iteratively simulate values for model parameters, given a specified prior distribution and an initial value for each parameter. The output of the Gibbs sampler constitutes a Markov chain. Under a wide set of conditions, the distribution of the Markov chain converges on the posterior distribution of parameters, that is, on the distribution of parameters given the data (Gelman, Carlin, Stern, & Rubin, 2003).

The primary advantage of applying MCMC to the analysis of interactive behavior genetic models is that it makes the estimation of such a model computationally feasible; using traditional maximum likelihood methods, the models are almost always intractable (Eaves & Erkanli, 2003). In addition, the Gibbs sampler simulates values for all unknowns about the data, not only those traditionally thought of as model parameters, but also scores on latent genetic and environmental factors and missing data values. Whether the distribution of the Markov chain was stationary (i.e., whether the model had "converged") was evaluated using basic diagnostic plots (trace plots, autocorrelation plots, and density plots). These diagnostic plots are available from the first author upon request.

Code and initial values for the full model is available from the first author upon request. Prior distributions of parameters were noninformative. WinBUGS was used to simulate a chain of 30,000 updates of the Gibbs sampler for each model. The first 20,000 iterations were discarded as "burn-in" (i.e., overly influenced by initial values), and the remaining 10,000 were used to characterize the posterior distribution of the parameters of interest. The posterior distribution mean is reported as the point estimate for a given parameter; the median, 2.5 percentile and 97.5 percentile of distributions (95% credible interval) are reported to describe the point estimate and uncertainty about parameter values. Models estimated using WinBUGS are compared using the deviance information criterion (DIC; Spiegelhalter, Best, Carlin, & van der Linde, 2002). The DIC reflects a trade-off between the fit of the data to the model and corresponding model complexity. Lower values indicate better model fit, with differences greater than 10 DIC ruling out the model with higher DIC.

RESULTS

We began by estimating a baseline model in which the common variance and the unique variance in each of the 10 items were decomposed into ACE components, that is, all interaction terms in the model were set to zero. The standardized A, C, and E parameters for the common variance were .58, .35, and .07, respectively. Table 5.1 gives the factor loadings, commonalities, and ACE decompositions of the unique variance of the 10 items.

Next, we estimated a model that included interactions between the ACE components of the common variance and observed SES. The fit of this model was superior to that of the baseline model, a reduced model that did not include interactions on either the common factor or on

Table 5.1 Estimated Parameters from Reduced Model with No SES Interactions

			Variance (%)		
Test	**Loading**	**Communality**	h^2	c^2	e^2
Common Variance	—	—	0.58	0.35	0.07
WRAT Math	1.08	0.71	0.10	0.24	0.66
WRAT Reading	2.59	0.76	0.44	0.24	0.32
WRAT Spelling	1.44	0.78	0.59	0.03	0.38
Block Design	0.47	0.18	0.42	0.00	0.58
Coding	1.25	0.21	0.01	0.38	0.61
Digit Span	0.38	0.49	0.00	0.00	1.00
Vocabulary	0.35	0.45	0.50	0.00	0.50
Comprehension	0.94	0.30	0.39	0.00	0.61
Information	0.88	0.42	0.12	0.40	0.48
Picture Arrangement	0.26	0.16	0.00	0.26	0.73

Note: Proportions of variance for individual tests are for unique variance in each test.

the uniquenesses (DIC = 35665.4, ΔDIC = 340.6), suggesting that the interaction parameters are required for adequate fit. The magnitude of the A variance had a positive relationship with SES, while the magnitude of the C variance was negatively related to SES. The E variance was not related to SES. Of the three ACE components of the common variance, only the A component had a confidence interval around the interaction with SES that did not contain zero, although the C component was in the predicted direction and barely contained zero. In Figure 5.2 we have plotted the total common variance against SES, with the total variance decomposed into ACE components. Genetic variance increases as a function of SES, whereas the C component decreases. The increase in the total common variance with increasing SES is attributable to increases in common genetic variance.

Finally, we fit a full model in which all 11 sets of ACE parameters, one for the common variance and one each of the unique variances of the items, were decomposed into ACE parameters that interacted with observed SES. Although the confidence intervals for 10 of the 30 interaction parameters excluded zero, compared to the fit of the full model (DIC = 35321.2), the reduced model that did not include any interactions on the uniqueness fit equally well (ΔDIC = 3.6). Thus we did not interpret them individually and selected the previous model, with interactions on the components of the common but not the unique variance, as the final model.

DISCUSSION

Since Sandra Scarr first identified more than 30 years ago what we refer to here as the Scarr–Rowe interaction, subsequent investigators have endeavored to sharpen her seminal scientific insight. To be sure, data have become available that are superior to what Scarr managed to assemble in the Philadelphia school system, and computational power to analyze the effect has increased by several orders of magnitude. The two most important questions, we have suggested, are the specification of the most important proximal environments contributing to the effect, analysis of the components of intelligence that are most sensitive to it. The data we have reported here are a contribution to the latter question.

In our 2003 report of the Scarr–Rowe interaction in the NCPP data, we showed that the Scarr–Rowe interaction was observable for Performance IQ but not for Verbal IQ. At the subtest level, the interaction was present for some tests (mostly educational and achievement tests, in addition to the Performance IQ subtests) but not for others. There are two possible explanations for such a diversity of results: either they are a reflection of systematic differences among the abilities tapped by the subtests, or they are simply random variations around an overall mean interaction that is manifest in the common variance. The analysis presented in the current chapter suggests that these differences may not reflect important differences among the subtests, since the interaction is robustly estimated in the common variance in almost exactly the same form that was observed for Performance IQ, and is not detectable in the unique variances contributing to the differences among the subtests.

Further analysis will be required before this conclusion can be reached definitively, however. More than one factor can be extracted from the NCPP battery, and simple analyses of these factors suggest that some may manifest the Scarr–Rowe interaction more strongly than others (D'Onofrio, Turkheimer, Hamagami, Harden, & Gottesman, 2005). Our earlier investigations of this question were hampered by the computational limitations we have discussed here, and we intend to return to them using the winBUGS methods that we have now developed.

Several of the other questions we identified at the outset remain to be addressed. Most important, the range of environmental circumstances in which the effect can be expected to occur is still unknown. The difficulty in addressing this question is that the data required to answer it—ability data from large samples of twins who have not been selected for at least a middle-class upbringing—are extremely difficult to come by. It is interesting to note, however, that a massive source of relevant data remains untapped. We are referring to the statewide achievement

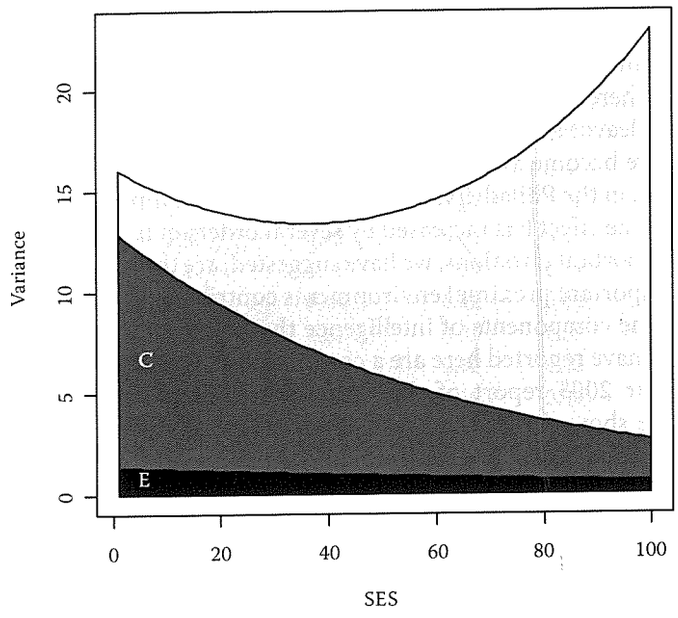

Figure 5.2 Proportions of variance of common variance among tests attributable to genetics (A), shared environment (C), and nonshared environment (E), plotted as a function of measured socioeconomic status.

testing that is now universally administered in public schools. Tens of thousands of unselected twins are being given cognitive testing on an annual basis, but so far it has not been possible to overcome the daunting ethical and political obstacles that stand in the way of analyzing the data.

Another intriguing question that awaits further analysis is whether the Scarr–Rowe interaction would continue to be observed in the ability scores of adults. It is well-known that familial effects on cognitive ability are observed in childhood but decline quickly after adolescence, and it would be important to determine whether the heightened environmental effects associated with poverty in childhood survive into adulthood. To answer this question, it will be necessary to identify adult twins for whom data are available concerning their environments during their childhood. We are currently investigating some possible sources of such data.

One lesson to be learned from the Scarr–Rowe interaction is methodological rather than substantive. As we have noted, twins from impoverished families are usually not recruited into American volunteer

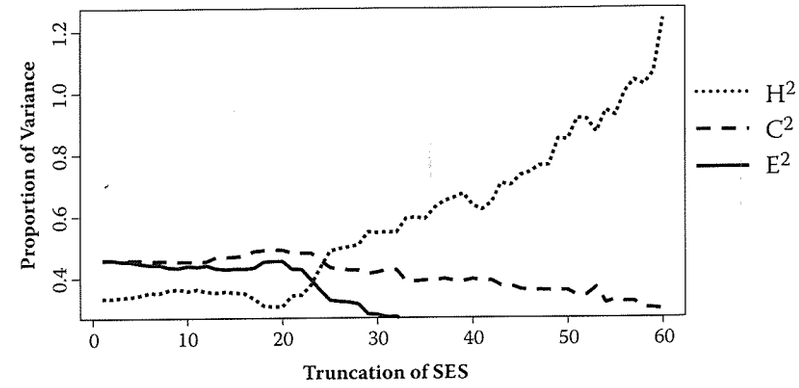

Figure 5.3 Proportions of variance of Full Scale IQ when the poorest twins are removed from the sample. The sample is truncated below SES value on the *x* axis.

twin samples. If, as the analyses presented here suggest, the biometric parameters underlying variation in important psychological constructs change as a function of socioeconomic status, then many conclusions that have been reached on the basis of volunteer twin samples may be seriously biased. Figure 5.3 is an illustration of this problem taken from the NCPP analysis. In this figure, instead of computing the ACE parameters as a function of SES, we systematically biased the sample by truncating it one twin pair at a time, starting with the poorest pair and moving up. So at the extreme left of the graph, one sees the ACE parameters for the entire sample. The next point to the right shows the results with the poorest pair omitted, then with the two poorest pairs omitted, and so forth. As more and more of the lower end of the distribution is truncated, additive genetics becomes more and more dominant in the results, until it eventually violates the assumptions of the model and is estimated to explain more than 100% of the variance in IQ. This result should be cause for some caution among those who wish to accept very high heritabilities of intelligence based on samples that have not been carefully constructed to be representative.

Finally, there is also a need to return to the clear theoretical focus that Scarr brought to her early work on this subject in 1971. Now that software is readily available, it would be possible to re-analyze practically every twin analysis that has ever been conducted, with the familiar variance components moderated by socioeconomic status, or by age or gender or race. One would not want the field to wind up in the atheoretical tabulation of moderated variance components, without explicit reference to the developmental processes that underlie them.

REFERENCES

Asbury, K., Wachs, T. D., & Plomin, R. (2005). Environmental moderators of genetic influence on verbal and nonverbal abilities in early childhood. *Intelligence, 33*, 643–661.

DeFries, J. C., & Fulker, D. W. (1985). Multiple regression analysis of twin data. *Behavior Genetics, 15*, 467–473.

D'Onofrio, B. M., Turkheimer, E., Hamagami, F., Harden, K. P., & Gottesman, I. I. (2005). *Gene by environment interaction in multiple factors of intellectual ability and academic achievement.* Unpublished manuscript.

Eaves, L., & Erkanli, A. (2003). Markov chain Monte Carlo approaches to analysis of genetic and environmental components of human developmental change and GxE interaction. *Behavior Genetics, 33*, 279–299.

Eaves, L. J., & Jinks, J. L. (1972). Insignificance of evidence for differences in heritability of IQ between races and social classes. *Nature, 240*, 84–88.

Evans, G. W. (2004). The environment of childhood poverty. *American Psychologist, 59*, 77–92.

Evans, G. W., & English, K. (2002). The environment of poverty: Multiple stressor exposure, psychophysiological stress, and socioemotional adjustment. *Child Development, 73*, 1238–1248.

Fischbein, S. (1980). IQ and social class. *Intelligence, 4*, 51–63.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003). *Bayesian data analysis* (2nd ed.). London: CRC Press.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6*, 721–741.

Harden, K. P., Turkheimer, E. & Loehlin, J. C. (2007). Genotype by environment interaction in adolescents' cognitive aptitude. *Behavior Genetics, 37*, 273–283.

Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin, 96*, 201–210.

Kremen, W. S., Jacobson, K. C., & Xian, H. (2005). Heritability of word recognition in middle-aged men varies as a function of parental education. *Behavior Genetics, 35*, 417–433.

LaBuda, M. C., & DeFries, J. C. (1990). Genetic etiology of reading disability: Evidence from a twin study. In G. V. Pavlidis (Ed.), *Perspectives on dyslexia* (Vol. 1, pp. 47–76). New York: Wiley.

Loehlin, J. C., & Nichols, R. C. (1976). *Heredity, environment, and personality: A study of 850 sets of twins.* Austin, TX: University of Texas Press.

Myrianthopoulos, N. C., & French, K. S. (1968). An application of the U.S. Bureau of Census socioeconomic index to a large diversified population. *Social Science and Medicine, 2*, 283–299.

Nagoshi, C. T., & Johnson, R. C. (2005). Socioeconomic status does not moderate the familiality of cognitive abilities in the Hawaii Family Study of Cognition. *Journal of Biosocial Science, 37*, 773–781.

Neale, M., Boker, S., Xie, G., & Maes, H. (1999). *Mx: Statistical modeling.* (Available from University of Virginia, Department of Psychiatry, Box 126 MCV, Richmond, VA, 23298)

Nichols, P., & Chen, T. (1981). *Minimal brain dysfunction: A prospective study.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Plomin, R. (1999). Genetics and general cognitive ability. *Nature, 402*, C25–C29.

Purcell, S. (2002). Variance components models for gene-environment interaction in twin analysis. *Twin Research, 5*, 554–771.

Rowe, D. C., Jacobson, K. E., & van den Oord, E. (1999). Genetic and environmental influences on vocabulary IQ: Parental education level as a moderator. *Child Development, 70*, 1151–1162.

Scarr, S. (1981a). *Race, social class and individual differences in I.Q.: New studies of old issues.* Hillsdale, NJ: Erlbaum.

Scarr, S. (1981b). The effects of family background: A study of cognitive differences among black and white twins. In S. Scarr (Ed.), *Race, social class and individual differences in I.Q.: New studies of old issues* (pp. 261–315). Hillsdale, NJ: Erlbaum.

Scarr-Salapatek, S. (1971). Race, social class, and IQ. *Science, 4*, 1285–1295.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B, 64*, 583–640.

Turkheimer, E. (2005). Mobiles: A gloomy view of the prospects for developmental behavioral genetics. In E. Perense (Ed.), *Wrestling with behavioral genetics.* Baltimore: Johns Hopkins University Press.

Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B. M., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science, 14*, 623–628.

Turkheimer, E., & Waldron, M. C. (2000). Nonshared environment: A theoretical, methodological and quantitative review. *Psychological Bulletin, 126*, 78–108.