Genetic Interactions with Prenatal Social Environment:

Effects on Academic and Behavioral Outcomes

Dalton Conley*
New York University & NBER

Emily Rauscher
University of Kansas

December 13, 2012

Abstract word count: 140
Total word count: 9,991
Tables: 6
Figures: 1
Keywords: genetics, birth weight, depression, delinquency, educational achievement

* Address correspondence to Dalton Conley, Faculty of Arts and Sciences,
6 Washington Square North, Room 20; New York, NY 10003 (conley@nyu.edu).

Abstract

Numerous examples of gene by social environment interactions have been reported. In these studies, however, environmental conditions are potentially endogenous to unmeasured genetic characteristics. Thus, gene-gene interactions cannot be ruled out as an alternative explanation. We exploit a natural experiment that randomizes a particular stressor – birth weight discordance within twin pairs – to address this limitation and ask: Do random differences in early environment (prenatal nutrition) moderate genetic effects on depression, delinquency, or GPA? Alternatively, does genotype moderate effects of birth weight? Using Add Health data, the only consistently significant allele-birth weight interaction we reveal works in the opposite direction of Caspi et al.'s classic finding on 5-HTT and maltreatment. Less robust interactions found for DRD2 and MAOA are consistent with this pattern. Results do not necessarily overturn existing research but support our methodological point that gene-environment research must address endogeneity.

Background

Studying genetic-environmental (GE) interactions has long been a goal of social scientists fond of touting the dependence of genetic expression on social structure. The basic GE argument is that genetic effects do not exist in a social vacuum. A specific allele does not have the same effect on individual outcomes; environmental differences determine how the gene manifests. However, how do we get from the adage that "a gene for aggression lands you in prison if you're from the ghetto, but in the boardroom if you're to the manor born" to a serious empirical research agenda on the study of GE interactions?

The basic logic for specifying GE interactions until now has been the following: A certain proportion of a population sample is found to have a variant of a particular allele (an alternative form of a gene). If this allele is shown to be randomly distributed across demographic subgroups (or within a particular subgroup such as an ethnic group), and, likewise, it is found to be associated with a specific social outcome or tendency (such as addictiveness, shyness, schizophrenia) *within* that same population (or subgroup), then researchers may try to find specific environmental conditions which seem to magnify or mitigate its effect—such as family structure, parents' behavior, or simply socioeconomic status.

However, there is a critical flaw with existing GE research. In all cases (e.g., Caspi at el. 2002, 2003; Guo et al. 2008a; Pescosolido et al. 2008; Shanahan et al. 2008), the environmental conditions studied (such as maltreatment or family dinners) are potentially related to the unmeasured genetic characteristics of the subjects and their families. For example, Shanahan and colleagues (2008) find that the "risky" DRD2 genotype reduces the likelihood of post-secondary school attendance for boys and that this risk is moderated by social capital. Yet they

also find that boys with the risky DRD2 genotype are less likely to have high social capital environments. It could be that the alleles are interacting not with differential social environments, but rather with other, non-randomly distributed genes (even if the principal gene in question is indeed randomly distributed).[1] To provide other examples of this challenge, Guo et al. (2008a) indicate that regular family meals eliminate the delinquent tendencies associated with the "risky" DRD2 genotype. Similarly, Pescosolido et al. (2008) find that family support reduces the genetically-influenced risk of alcohol dependence. Although they discuss threats to internal validity and claim only to describe associations, the lurking question of unmeasured environmental or genetic differences that may covary with measured environment remains. The environmental moderators in both cases include family behaviors, which could easily reflect genetic rather than exogenous environmental differences. That is, family support and closeness could be related to the unmeasured genes of the respondent and his/her parents, who to a large extent determine the family environment. For example, other parental genes encouraging social disorganization (and thus fewer planned meals as a family) may be passed on to the adolescent and interact with the "risky" DRD2 allele to produce the observed deleterious outcomes—with the family dinners acting as a proxy.

Furthermore, recent evidence (Fowler et al. 2011) suggests that individuals self-select into social environments based on genotype. Specifically, certain genotypes are correlated within friendship networks and social environment is therefore endogenous to genotype. This evidence casts serious doubt on existing claims of GE interactions in the social sciences, which do not account for endogeneity in the gene-environment association. This is not to say that the interaction effect is not "real"; it suggests it is merely associational, and we should be cautious about drawing causal conclusions about the particular role of family eating habits.

An alternative but neglected perspective – to the GE emphasis on *environment* moderating *genetic* effects – is the potential moderating effect of genotype. For example, while child nutrition and birth weight effects have been well-established, their effects could differ by genotype. Such heterogeneous effects could explain, for example, why low birth weight is not equally detrimental for all children. There is variation in the sequelae of low birth weight by a variety of factors, including poverty, race, and education (Conley et al. 2003), which could reflect underlying GE interaction. In this vein, the present study makes two contributions: 1) methodologically, it advances GE research with an empirical example of a new technique that accounts for endogeneity of the gene-environment association; and 2) substantively, it asks whether the effects of birth weight depend on genotype.

Literature Review

Caspi et al. (2002) claim to have uncovered a GE-interaction among 1,037 white male New Zealand children between the MAOA gene (monoamine oxidase A) and childhood maltreatment. Caspi interacted the short MAOA alleles with the degree of maltreatment the respondents experienced as children to predict an index of anti-social behavior. In a multiple regression context, the interaction effect between the two measures was statistically significant. They argue that this is a true GE interaction effect since the MAOA genotypes were not significantly differently distributed across maltreatment levels—suggesting that this genotype did not itself influence exposure to maltreatment. In a follow-up study (2003) using the same cohort, they find similar evidence of an interaction between stressful life events (between ages 21 and 26) and alleles of the serotonin transporter gene (5-HTT) linked promoter region (5-HTTLPR) in the likelihood of clinical depression at age 26.

However, similar to Guo (2002 & 2008a) or Pescosolido (2008), Caspi et al. (2002; 2003) could actually be uncovering gene-gene interactions, because they did not have an exogenous source of environmental variation. For example, it may be the case that depression was induced by a gene-gene interaction since an underlying unmeasured gene could cause the phenotype of "negative life events" to emerge in one's early 20s: Imagine a gene that promotes excessive thrill-seeking and risk-taking, which, in turn, manifests as negative events during one's early adulthood. As for the MAOA interaction, we face the same issue: While measured maltreatment did not vary by MAOA status, it could have varied by other genes (present in the parents and potentially passed on to the children). Thus, it would not be the maltreatment that interacted with MAOA status but rather the underlying, unmeasured genotype, which, in combination with given MAOA alleles, caused both parents and offspring to act anti-socially.

Thus, existing studies do not address the potential endogeneity of environmental context; in all cases, environmental conditions are potentially related to the unmeasured genetic characteristics of the subjects and their families. Conversely, gene markers may be acting as proxies for social conditions because of population stratification. We argue that GE research must address environmental and genetic endogeneity to be methodologically rigorous. Below, we offer more specific motivations for methodological improvements and an analytical approach which addresses both non-random genes and environment.

Beyond methodology, existing work rarely discusses the alternative interpretation of GE interactions. Namely, the gene could be examined as the moderator rather than social environment. Birth weight is a measure of environment even before birth and is generally conceived as critical for infants, setting their chances for many later outcomes. However, birth

weight may not be equally important for all individuals and effects may vary significantly by genotype.

Endogeneity Concerns

It is difficult to quantify the potential threat of inherited factors to GE findings. (Supplementary Material includes sensitivity analyses that attempt to quantify unobserved variable bias using propensity score matching and Rosenbaum bounds estimation.) However, reported GE findings are likely misestimated. Given that the phenotypes in question are complex behavioral traits (i.e. "quantitative traits"), they are almost certain to be influenced by a wide range of genes and their interactions. In fact, supporting the notional importance of gene-gene interactions is recent genetics research that has shown that among the genes studied in humans (or other [model] organisms such as the fruit fly, *drosophila melanogaster*, or the nematode worm, *Caenorhabditis elegans*), the vast majority of genes are linked in a single network component when measured by either protein-protein interactions, regulatory relationships, or phenotypic co-variation (Jeong et al. 2001; Stelzl et al. 2005). This suggests that one cannot conceptualize the perturbation of one gene as unrelated to the impact of other genes. Unless we measure all genes, we cannot know for sure. (And even if we did, we would not have the statistical power to test for all possible GG interactions, even in a genetic census of all humans on the planet.[2]) Thus, we must leave, for now, GG interactions as a black box. The solution to this conundrum, we argue, is to find a source of E that is orthogonal to G and thus cannot be confounded by unmeasured GG interactions.

Another parallel concern is population stratification, which occurs when genetic variants are not randomly distributed within a study population and other environmental or genetic

differences are correlated with this non-random distribution. Hamer and Sirota (2000) illustrate population stratification with a fictitious "chopsticks gene": a scientist sets off in search of a gene for chopstick use and finds a strong association between chopstick use and a specific polymorphism. However, this genotype happens to be more common among Asians, who use chopsticks more often for unrelated cultural reasons, and when the analysis is redone within subgroups, the effect disappears. Thus, non-random distribution of polymorphisms could be correlated with environmental or cultural differences to yield spurious genetic "effects." So while we have so far emphasized the possibility that environment may be acting as a proxy for unmeasured genetic effects (and their interaction with measured alleles), population stratification indicates that measured genes may be acting as proxies for social environments. For example, Thomas and Witte (2002:505) point out that DRD2 is not equally distributed by ethnicity. Therefore, interactions between DRD2 and environment (or the main effect of DRD2 for that matter) may actually be reflecting the social dynamics of ethnicity and not the causal effect of the DRD2 allele. Others (Gelernter and Kranzler 1999; Abdolmaleky et al. 2004; Fan and Sklar 2005; Sabol et al. 1998) show significantly different ethnic distributions for 5-HTT, DRD2, and MAOA. Of course, ethnicity can be controlled for, or we can stratify our sample by ethnic origin, thereby analyzing the effect of DRD2 only within these subpopulations. But we can never be sure whether a given allele is orthogonal to all social environments. In other words, a particular gene variant may be overrepresented in the South, or among the highly devout, or among urban residents. The list of confounders is potentially endless.

      We might anticipate non-random distribution of alleles by chance—due to sampling, migration patterns, and so on—or because those genetic loci have important effects on how we

live (e.g., assortative mating on smoking or drinking behavior). Only through within-family comparisons, can we be sure to address these worries about population stratification.

GE research is not alone in its failure to account for unmeasured genetic differences. Much of the apparent effect of birth weight may reflect inherited differences. While some research has estimated the effect of birth weight using identical twins (Behrman and Rosenzweig 2004; Conley et al. 2003), most studies have relied on sibling (Conley and Bennett 2000; 2001) or even between-family comparisons (Hack et al. 2002; Rich-Edwards 1997; Sorensen et al. 1997), which cannot rule out genetic explanations for differences in outcomes; such an approach may create spurious effects of low birth weight if the same genes driving birth weight also drive educational or developmental outcomes.

Low birth weight is consistently found to have developmental consequences and has been associated with a wide variety of outcomes in later life, including cognitive ability, behavioral problems, and health (Lowe et al. 2009; Hayes and Sharif 2009; Schlotz and Phillips 2009; Cooper et al. 2009). In this way, birth weight (as a proxy for fetal nutrition) is a "stressor," in the same vein as factors studied in existing GE research (e.g., childhood maltreatment and lack of social support). All of these stressors affect the context in which children grow, which research posits to interact with genotype (Caspi et al. 2002; Guo et al. 2008a; Shanahan et al. 2008).

Shifting to an earlier period in child development, we view birth weight as an important indicator of developmental context. Increased prenatal nutrition can act as a nurturing environment, like family support, and dampen genetic risk. For example, if two children have an identical genetic risk for hyperactivity such as two copies of the "risky" DRD2 A1 allele, we might expect birth weight to moderate this risk; higher birth weight could allow one to develop earlier, self-regulate and sit still at a younger age, and gain more from school than the other. The

7

low birth weight child, who develops self-regulation more slowly, may receive negative feedback from school teachers, dislike school, and set off on a less-adaptive trajectory. Similarly, we might expect to see an interaction between birth weight and the long 5-HTT allele, which puts individuals at risk of more depressive symptoms (Caspi et al. 2003, 2010). For instance, low birth weight individuals may be less physically active in later life and miss out on the depression-fighting endorphins of regular exercise. In other words, fetal nutrition could amplify genetic differences in later life. While many intervening factors could be related to later childhood contexts including maltreatment and family closeness, an advantage to studying birth weight is that it occurs very early in life, at the proverbial "starting gate" (Conley et al. 2003).

Methodological Approach

Methods

We deploy a novel approach: We use both MZ (monozygotic or identical) and DZ (dizygotic or fraternal) twin differences in birth weight to predict educational, mental health and behavioral outcomes (all found in previous GE research) using data from the National Longitudinal Survey of Adolescent Health (Add Health). MZ twins hold constant genetic differences that may influence both birth weight and the outcome of interest. Behrman and Rosenzweig (2004) establish that birth weight within identical twin pairs is unrelated to other observed measures in their data and illustrate substantial distribution of birth weight differences between identical twins. We then ask whether the treatment effect of low birth weight varies across twin pairs that are divergent on measured alleles.

While birth weight differences between identical twins are not socially determined as sociologists might traditionally conceive, they are sociologically relevant for at least two

reasons. First, while random in origin due to fetal position, birth weight differences between identical twins are a classic example of cumulative advantage. A slight difference in implantation site can translate to substantial differences in fetal nutrition, which in turn, affects growth and ultimately social outcomes from education to income. The importance of such within-family factors to social stratification has garnered recent attention (Conley 2004). Second, even though this particular etiology of birth weight variation (in utero competition mediated by fetal position) is not due to larger social forces, the mechanism—nutritional deprivation—is the same as those sociologically causal forces that impact fetal nutrition from outside the womb—i.e. the mother's food intake and other health conditions. At least since the Dutch famine during World War II, evidence shows maternal nutrition is affected by social conflict, gender discrimination and the global distribution of food resources (e.g., Roseboom et al. 2001). Caloric deprivation, in turn, affects the fetus with lasting impacts. One estimate asserts that in the developing world, birth weight (i.e. fetal nutrition) influences GDP (Behrman and Rosenzweig 2004). Of course, in the population at large, there are multiple causes of low birth weight. However, in the twin difference context, we hold constant etiologies such as drug, cigarette or alcohol use, maternal preeclampsia, or prematurity. These do not figure into our analyses since they affect both twins. Birth weight has significant sociological antecedents and sequelae; the methodological approach here should be externally valid enough to illuminate those social facts while preserving a quasi-experimental approach (i.e. internal validity). An analogy might be an experiment that adjusts class size "artificially" in order to understand the effects of teacher-student ratios across urban school districts.

However, since the strategy of using MZ twins leaves open the possibility of population stratification, we also present estimates using intra-sibship comparisons among dizygotic

twins. Each DZ twin has an equal chance of inheriting one of two alleles from each parent. In this way, with the exception of other genes that may be linked to the gene in question through linkage disequilibrium (where alleles at certain loci "tag along" with others due to physical proximity on a chromosome), all other genes are orthogonal to the measured genetic difference. However, in the DZ models, we cannot say that birth weight differences are exogenous to unmeasured genetic characteristics that vary between the twins. Thus, we present findings that are robust to both these estimation strategies. (We also show, in auxiliary analysis, that birth weight seems unaffected by these genes.)

To preview our results: While the majority of our findings are null, we find evidence of one GE interaction effect which works in the opposite direction predicted by general understandings in the literature. Birth weight, therefore, does not matter equally by genotype. The fact that we are able to detect a highly robust GE interaction using our restrictive approach belies concerns about our models being under-powered and justifies the use of twins despite smaller sample sizes.

Birth weight is an important difference in childhood context that occurs at the beginning of life. Although it is not the same as measures used in previous GE research, it provides an analytically appealing "stressor" for two reasons. First, it precedes and could even influence measures used in previous GE studies. Second, it is orthogonal to genes for identical twins; that is, genes are shared completely so they cannot affect nutritional advantage relative to one's identical twin in the womb. Even among DZ twins, the uterus is shared and birth weight differences partly reflect proximity to the placenta, which occurs by chance.[3]

We admittedly sacrifice external validity for internal validity. For example, twins are generally lighter at birth than singletons, which means that we cannot generalize to all points in

the distribution of birth weight. Skeptics may fail to see how birth weight, twins, or differences in twin birth weight are relevant to previous GE research or socially interesting variation in general. However, this sacrifice is justified, we believe, because of the causal traction gained. The possibility that previous GE findings are spurious – whether due to gene-gene interactions or genes acting as proxies for environmental differences due to population stratification – warrants the narrow focus of this study. Birth weight varies randomly within MZ twins and alleles are randomly assigned within DZ twins, which make twin pairs an ideal sample to further the investigation of GE interactions. While we do not directly test the environmental measures of the previous studies, our measure of stress—fetal nutrition—complements these and provides an example for future researchers in this area to follow when looking for exogenous environmental shocks.

Likewise, we also sacrifice some power by identifying off within-twin pair differences. Thus, effects would have to be (almost) twice as large as models that use all individuals with genetic data in Add Health (these approaches also suffer from inflation of standard errors due to the non-independence of observations). So, while post-hoc power tests are discouraged in the literature (Levine and Ensom 2001; Hoenig and Heisey 2001), the fact that previous studies report powerful impacts of these genes (and interaction effects) means that we should still be able to detect them with our reduced power, even if previous estimates are partially spurious. Finally we report results with an alpha <.1 (or .05 for a one-tailed test which is reasonable given priors about the valence of effects). That said, acting as if we were designing an experiment to detect an effect size that is $\beta >= 0.15$, at $\alpha < .05$, with three predictors (other than the fixed effects), our minimum required sample size for a study with 0.80 power would be 76. All models meet this

minimum. Meanwhile, our targeted, minimally detectable effect size of 0.15 is well within early

reports for population studies of these genes on these phenotypes.[4]

MZ twin pair fixed effects exploit random variation in birth weight to identify a gene-

environment interaction. Out of concern for population stratification, we also show results for

dizygotic twins. Fixed effects model the differences in outcomes within twin pairs (i.e. between

twins):

$$Y_{ij} = a + bBirthweight_{ij} + cRiskyAllele_{ij} + dBirthweight_{ij} * RiskyAllele_{ij} + eSex_{ij} + fTWINSET_j + \varepsilon_{ij}$$

where Y_{ij} is the outcome for a given twin i in pair j, b is the effect of birth weight within twin

pair j that has the so-called normal allele, c is the main effect of differences in the risky gene, b

$+ d$ is the effect of birth weight (within twin pairs) for those with the risky allele, e represents

differences due to sex (for dizygotic twin pairs only), and ε_{ij} is the sibling-specific (i.e.

idiosyncratic) error, assumed to be unrelated to genes, birth weight, and control factors.

Monozygotic twins share the same genes and sex, so c and e drop out. Individuals are

compared to their twin who is the same age, obviating the issue of age differences.

Before presenting interaction effects, we show descriptive statistics and estimated main

effects of birth weight and genotype within twin pairs. Assuming random variation, main

effects of birth weight can be identified within MZ and DZ twin pairs. The main effect of a

risky genotype can be identified within DZ twin pairs. However, fixed effect models cannot

identify the main effect of a genotype within identical twin pairs because there is no genetic

variation. While we present fixed effect models in all other cases, random effect models

estimate main effects of the alleles in question among MZ twins. Random effects assume that

genotype is randomly assigned to MZ twin pairs. Caution is therefore warranted in interpreting

these coefficients, because they could reflect a spurious relationship due to population stratification.

Data and Measures

The National Longitudinal Study of Adolescent Health (Harris 2009) provides birth weight and sequenced genotype data for three genes putatively related to behavioral and health outcomes conditional on environment (5-HTT, DRD2, and MAOA). These genes are involved in neurotransmitter (e.g., dopamine and serotonin) transport, receiving and recycling, vital for cognition and behavior. Previous research, including animal-based genetic manipulation, has identified these three loci as influential on a variety of outcomes (Cases et al. 1995; Shih and Thompson 1999).

As in previous GE research, we focus on outcomes in young adulthood. Wave 3 of the Add Health was collected in 2001-2 when respondents were ages 18-26. Siblings of individuals identified as twins in the stratified sample were added, yielding 64% of sibling pairs from the probability sample and 36% from convenience sampling. Buccal swabs were collected in Wave 3 from 2,612 of the 3,139 eligible siblings from Wave 1 (a compliance rate of 83%; Harris et al. 2006) for DNA sequencing at the Institute for Behavioral Genetics. Monozygosity was genetically confirmed (Harris et al. 2006). Our sample includes over 200 twin pairs not missing birth weight, genetic data, or outcome data for either twin (sample size for each model depends on the number of pairs with complete outcome data).

Research links polymorphisms in the human genes DRD2, 5-HTT, and MAOA with behavior and health outcomes. At the D2 dopamine receptor gene locus (DRD2), the A1 allele is related to fewer receptor binding sites (Pohjalainen et al. 1998). Compared to the A2 allele,

possessing the A1 allele has been associated with anxiety, depression, novelty seeking, impulsiveness, lack of inhibition, and substance use (Lawford et al. 2006; Noble et al. 1998; Wiers et al. 1994; Blum et al. 1991). Consistent with previous research, individuals possessing an A1 allele are considered to have the risky genotype.[5]

Previous research suggests that individuals with a short allele in the promoter region of the serotonin transporter gene locus (5-HTT) have stronger depressive reactions to stressful life experiences (Caspi et al. 2003).[6] Both men and women carry two copies of DRD2 and 5-HTT. For both of these genes, those with no copies of the risky allele are specified in models below, but results are similar with alternative specifications.

The MAOA gene codes for monoamine oxidase A, which chaperones and breaks down neurotransmitters (e.g., serotonin, dopamine) and for which variation has been linked to disposition toward aggression in both animals and humans (Guo et al. 2008b; Rowe 2001; Cases et al. 1995; Shih and Thompson 1999; Brunner et al. 1993). Sabol et al. (1998) and Deckert et al. (1999) found lower activity and less efficient transcription among shorter MAOA alleles with 2 or 3 repeats. Given the debate about how best to specify MAOA genotype (Sabol et al. 1998; Deckert et al. 1999), we specify it in a variety of ways: any 3.5 or 4R vs. others; any 3.5, 4, or 5R vs. others; and 4R vs. others. Results are similar using different specifications and we present results using the distinction recommended by Sabol et al. (1998) – 3.5 or 4R alleles compared to 2, 3, or 5R. Men carry one copy of the MAOA gene because it is on the X chromosome. Women with two copies and men with one copy of a 3.5 or 4 repeat are included in the "non-risky" category below.

Birth weight is reported by parents, measured in ounces, and logged. Results using raw birth weight in ounces and an indicator for low birth weight are also examined (low birth

weight includes those 1500-2500g; no one in this sample was <1500g); results are similar and discussed below but not shown. The average birth weight difference between twins is non-trivial 8 ounces. Although this measure is retrospective, when children are teens, parents typically remember birth weight well (e.g., Walton et al. 2000 report an 85% accurate recall rate when children are teenagers). Nevertheless, errors are possible and birth weight is missing for 20% of the individual sample (30% of pairs). We address the possibility that birth weight is not missing at random in multiple ways (comparing descriptive statistics for those with and without missing data; the missing indicator method; and assigning the midpoint of 7 ounces for those with valid information for pounds at birth and missing only ounces). Results are similar.

Depression is measured using nine items of the Center for Epidemiologic Studies-Depression Scale (CES-D). CES-D normally includes more items that were omitted from Wave 3. Therefore we also include the other six questions about the frequency of depressive symptoms in Wave 3. The sum of responses for all items (listed in the supplemental section) indicates the frequency of depressive symptoms. Results are also investigated using an indicator for "any symptoms" and logged scores (after adding one to avoid excluding those with no symptoms). Following Fletcher and Lehrer (2009) and Roberts et al. (1991), age and gender-specific threshold measures of depression are also investigated. Results are largely the same in these specifications and are not presented. The Caspi study measured depression using the Diagnostic Interview Schedule, which may capture a different level of depression and make results less comparable.

Delinquency is measured using 12 questions from the Add Health Wave 3 survey that ask about deviant behavior in the past 12 months. For example, questions include how often you used someone else's credit/bank card without their permission, deliberately wrote a bad

check, stole something, or used a weapon (see supplemental material for a full list). Sensitivity analyses use an indicator for the presence of any delinquent behaviors and logged scores (after adding one)—results are similar to a linear specification. We show results from raw delinquency and depression measures in the main analysis because results are easier to interpret; we show logged measures in an appendix.

Educational achievement is measured using cumulative high school GPA gathered from high school transcripts. An indicator of college attendance (as well as a continuous measure for highest grade completed) is also tested, for comparability with Shanahan et al. (2008). In general, results do not differ from those for GPA and thus discussion concentrates on analysis of the continuous measure of achievement (i.e. GPA). A supplemental table presents results for college attendance.

Results

Table 1 provides descriptive statistics. Twins in the Add Health sample with sibling and genetic data have an average birth weight of 90 ounces (5.6 pounds; 4.5 in logged ounces). Average twin birth weight in the US was similar in the early 1990s: approximately 84 ounces (Alexander et al. 1998). We find a higher GPA among MZ than DZ twins; other averages are similar by twin type.

One concern is potentially selective data on birth weight. Approximately 20% of the sample is missing parent-reported birth weight. An additional 10% is missing birth weight data for their twin, excluding about a third of all identical twins. Table 1 compares twins with and without complete birth weight data. Both identical and fraternal twins missing birth weight

information for their twin sibling were born significantly lighter on average than others. This suggests excluded twins may have weighed less at birth than those included in the study.

Low weight babies experienced the strongest environmental insult and their outcomes may be most sensitive to genotype. Under-representation of low weight babies may therefore cause attenuation bias in GE interaction estimates. An alternative story further reduces concern. Parents may better remember the birth weight of the lighter (more at risk) twin since there may be more drama associated with her perinatal period as compared to the heavier, healthier twin. In this scenario, those with missing birth weight data would be nearer the twin average.

A related concern is the potential relationship between missing birth weight data and outcomes or alleles of interest. MZ twins with complete birth weight info have significantly higher delinquency rates than those without – amounting to about ½ an additional delinquent act such as stealing or damaging property. This difference remains significant with binary (indicating any symptoms) and logged measures of delinquency. Twins with and without birth weight data show no differences in the specific alleles investigated here. These birth weight and delinquency differences, though slight, suggest missing birth weight data could be non-random. Nevertheless, sensitivity analyses – using both the missing indicator method and assigning the midpoint for those missing only ounces – yield similar results.

A further concern is that the genes in question could influence birth weight differences within twin pairs. If the risky MAOA gene, for example, is associated with smaller birth weight differences between DZ twins, the reduced birth weight variation could yield insignificant interactions. However, regressions checking for such an association suggest it is not a concern (Table 2). Among MZ twin pairs, no genes are associated with twin pair birth

weight difference or average. Looking at 5-HTT alone does not change results. Among DZ twins, pairs in which one twin has a risky DRD2 allele have slightly more similar birth weights. This effect holds when including twin pairs missing birth weight for one twin. However, the DRD2 effect is not found when specifying those with two risky DRD2 alleles and is only marginal when predicting difference in raw ounces at birth. At the individual level, these genes are all unrelated to birth weight. Including those with missing birth weight does not change the results.

Table 3 shows main effects for each outcome. In Panel 1, models predicting each outcome include only the gene which previous research predicts should have effects. Panel 2 includes all three genes. Despite previous evidence that the short 5-HTT allele increases depression symptoms, DZ twin fixed effect regressions—controlling for a variety of potential confounders—show no significant effect of this genotype on depression, whether including all three genes or only 5-HTT. Results are similar using the natural log of depression symptoms (Table S3).

Although prior research predicts an effect of MAOA and DRD2 on delinquency and school continuation, our analyses of twins find insignificant main effects of variation at these genetic loci (controlling for gender and birth weight). While rare significant relationships emerge, the overarching pattern is that main genetic effects are insignificant regardless of the specifications tested. Results shown in Table 3 are similar when including twins missing birth weight, specifying the genes differently, or using alternative measures of birth weight or the outcome variable. The absence of main genetic effects suggests previous results may be biased by population stratification.

Similarly, main effects of birth weight are insignificant in all of the identical and

fraternal twin models shown (when an interaction with genotype is not included). This absence of birth weight effects contradicts Conley and Bennett (2000), who use singleton comparisons, which fail to address underlying genetic differences or experiences *in utero*. These results also contradict Behrman and Rosenzweig (2004), who use twin pairs. The difference could reflect sample differences (e.g., younger cohorts in this study) or controls for genotype included here.

While previous research suggests those with risky alleles should benefit most from positive environment, this analysis reveals that additional fetal nourishment actually *increases* depression among those homozygous for the risky allele. Weighing an additional pound (above the MZ mean of 90 oz.) increases depression symptoms by over 4 points (about 0.8 standard deviations) for those with two copies of the short allele, but has no effect for others. (With a mean difference of ½ a pound at birth, few twins have a one-pound difference, but we present results with this large difference for ease of interpretation.) Results for depression (Table 4) reveal that birth weight effects depend on 5-HTT genotype when including all or identical twins, but not fraternal twins alone. Model 3, limited to identical twins, suggests that an additional pound, compared to one's twin, reduces depressive symptoms by 2 points for those with two long alleles (nearly half a standard deviation). That same pound increases depressive symptoms by 2.8 points for those with one or two short alleles (about 0.5 standard deviations). Figure 1 illustrates this GE interaction, which remains significant with a Bonferroni correction for 10 hypotheses (more than the 3 outcomes by 3 genes or 9 hypotheses investigated).

[Figure 1 about here]

Regressions limited to individuals with two long alleles find no effect, while birth weight increases depression among those with any short allele. Results (not shown) are null using an indicator for any depressive symptoms, but are similar (though less precise) using a

logged measure of depression or including twins with missing birth weight. See Table S4.

Meanwhile, birth weight effects on depression do not depend on DRD2 genotype. An interaction between the long MAOA allele and birth weight is found among all twins (Model 4 in Table 4). But this disappears among fraternal twins and is only marginal among identical twins. Supporting the counterintuitive GE interaction with 5-HTT, the interaction with MAOA is also different than prior literature would predict. That is, birth weight seems to reduce depression among those with the "good" allele.

Within identical twin pairs, birth weight does not moderate effects of MAOA genotype on delinquency (Table 5). There is a significant interaction with MAOA among all twins. However, similar to our results for depression, the environmental advantage of fetal nutrition accrues to those with the "good" rather than the "risky" genotype. In other words, contrary to evidence that a short MAOA genotype moderates environmental stress (Caspi et al. 2002; Kim-Cohen et al. 2006), additional prenatal stress for a twin with the risky MAOA genotype actually *decreases* delinquency propensities.[7] Effects disappear within identical twins, which suggests omitted differences could drive results from previous research. Thus, although GE research would predict early childhood environment (fetal nutrition) to dampen effects of risky MAOA genotypes even among twins, we do not find consistent evidence of this. Rather, in some models environmental advantage amplifies the benefits of holding a "good" MAOA allele. While others find insignificant MAOA interaction effects (Haberstick et al. 2005), our null finding is limited to identical twins, which suggests previous GE interactions could reflect population stratification.

Finally, results from twin comparisons shown in Table 6 suggest previous evidence of a DRD2-social capital interaction may have been biased by omitted differences. Among all

twins, results suggest that those with no copies of the risky A1 allele have significantly *lower* high school GPAs and experience a boost from fetal nutrition (i.e. are more sensitive to environment). This interaction is marginally significant among fraternal twins and disappears within identical twins (Model 3).[8] Thus, genetic differences between fraternal twins appear to account for apparent environmental interaction effects with DRD2.

To summarize, neither birth weight nor the alleles we measure appear to have any direct, main effects on the outcomes we study. However, birth weight interacts in both DZ and MZ twin models such that decreased birth weight (previously considered a risk factor) results in lower risk of depression—but only for those who have the "risky" serotonin transporter promoter region allele. Furthermore, results suggest that birth weight effects depend on genotype.

Discussion and Conclusions

Research claiming GE interaction fails to address a potential relationship between genes and the environmental context in question. We motivate the need to address this shortcoming and offer a method for assessing GE interaction effects: by deploying both MZ and DZ comparisons, which each complement the other's inferential weakness. Our results do not necessarily overturn previous findings because: results could differ by age; the treatment effect of twin birth weight differences is unclear; and external validity is limited. Nevertheless, our analysis should encourage future GE research to address endogeneity.

Reviewing our findings in light of previous studies, Caspi et al. (2003) presented evidence that sensitivity to environmental insults increases with each short 5-HTT allele – i.e., those with one short and one long allele fall between those with two copies of either. We, too, find a linear interaction between number of long-alleles and sensitivity to environment;

however, our finding works in the opposite direction. Compared to those with two copies of the long ("good") allele, heterozygotes show significantly higher depression with increased birth weight – an additional 3.7 points with each pound. Those with two copies of the short allele show the strongest interaction with this specification, increasing 5.7 points on the depression scale with an additional pound at birth (about 1.1 standard deviations). Those with short 5-HTT promoters are indeed more sensitive to their environment; however, they respond in the opposite way as we would predict.[9]

Although the measures used here are not strictly comparable (and findings warrant replication in other samples as well as research that attempts to uncover the mechanism), the pattern of null findings for the other gene and GE interaction effects within twin pairs suggests that previous significant findings *may* be biased due to population stratification or omitted environmental factors (i.e. a failure to account for endogeneity). Equally intriguing, the only GE interactions we found suggest that genes and environment work together to amplify existing benefits. Interactions which disappear among MZ twins could reflect population stratification. However, the pattern of greater advantages accruing to those with "good" genes suggests future research may gain clarity by focusing on (random) environmental advantage rather than disadvantage.

Our analysis also contributes to the research literature on birth weight. Not only do we find little to no main effects of birth weight on important behavioral and academic outcomes, we also find that when interacted with certain alleles, birth weight works in the opposite direction as previously supposed. As with our GE findings, it could be the case that pre- and postnatal effects work in opposite directions, averaging to zero (endnote 3), but that still begs the question of why previous scholars have found significant average treatment effects for

related outcomes. Alternatively, results could reflect the lower average birth weight among twins, with environmental treatment clustered toward the lower end of the distribution. Thus, our results are certainly not conclusive and call for further investigation. More broadly, our results suggest that birth weight effects are heterogeneous by genotype. If this finding is replicated, it has important public health implications. For example, health officials could work to boost nutrition among fetuses most at risk of low birth weight effects.

Beyond the specific results presented here, we purport to have developed a careful method for assessing GE interaction effects, which we encourage other researchers to use: deploying both MZ and DZ comparisons, which each complement the other's inferential weakness. These methods and findings have important implications for future social scientific research involving genetic data. As more surveys collect genetic information from respondents, opportunities to investigate GE interactions increase. However, without random sources of genetic and environmental variation, GE results even from quality surveys will remain questionable.

Notes

[1] Recent genome-wide association studies are particularly subject to this criticism of population stratification. For example, Beauchamp et al. (2009) use principle components analysis (PCA) to account for as much variability as possible, but it is impossible to know whether it successfully identifies all important subpopulations. Within-family studies avoid population stratification.

[2] The discovery of about 21,000 genes—a figure much lower than originally hypothesized—means a tractable number of alleles for geneticists to study. However, if this lowly number of genes explains the development of human beings in all their forms, then gene-gene interactions are probably quite important.

[3] Of course, birth weight differences between twins may proxy an entirely social effect rather than fetal nutrition. Namely, perhaps the smaller twin is perceived as weak and thus stigmatized (or lavished with attention and resources). We are indifferent to what exactly the causal mechanism is for birth weight to produce an effect on our measured phenotype as long as the mechanism is not contingent on genes. If, however, those families with allele A tend to overinvest in their lower weight twin while those with allele B tend to stigmatize the lighter sibling, then we would detect a treatment effect that would average to zero. The use of DZ twin estimates, which identify the allelic-birth weight interaction effect on within-family differences, could somewhat mitigate these concerns. We identify a local average treatment effect (LATE) that cannot identify the mechanism by which birth weight interacts with genetic predisposition.

[4] For example, Caspi et al. (2002: 853) report an MAOA-maltreatment interaction effect size of -0.36. Others report interaction effects ranging from -0.11 to -0.89 (Caspi et al. 2003:388) and from -0.58 to -0.72 (Guo et al. 2008a:599), nearly all larger than our targeted effect size.

5 Guo et al. (2008a) present an interaction for heterozygotes – with exactly one A1 allele – not for either homozygous type, which makes it difficult to interpret their results. It is unclear whether heterozygosity or having a short allele is driving results.

6 Recently scholars have determined that this locus is tri-allelic. However, our data only contain the bi-allelic measure, which makes null findings more likely through measurement error.

7 Results in Table 5 are similar when addressing missing birth weight. Using logged delinquency, the MAOA interaction is only marginally significant among all twins (see Table S5) and is generally not robust to other specifications of the gene.

8 Including twin pairs missing complete birth weight, the interaction is insignificant. No significant interaction effects with DRD2 emerge for years of education or college attendance (Table S2).

9 It could, of course, be the case that social treatment of twins differs by birth weight in a way that concurs with Caspi et al.'s findings. Namely, parents may differentially invest in twins by birth weight, favoring the "at risk," lower birth weight twin and thus "neglecting" the heavier one, leading to a post-natal interaction effect between 5-HTT and parental investment (as proxied by birth weight).

References

Abdolmaleky, Hamid Mostafavi, Stephen V. Faraone, Stephen J. Glatt, and Ming T. Tsuang. 2004. Meta-Analysis of Association between the T102C Polymorphism of the 5HT2A Receptor Gene and Schizophrenia. *Schizophrenia Research* 67(1):53–62.

Alexander, G. R., M. Kogan, J. Martin, and E. Papiernik. 1998. "What Are the Fetal Growth Patterns of Singletons, Twins, and Triplets in the United States?" *Clinical Obstetrics and Gynecology* 41(1):115-25.

Behrman, Jere R. and Mark R. Rosenzweig. 2004. "Returns to Birthweight." *The Review of Economics and Statistics* 86(2):586–601

Beauchamp, Jonathan, David Cesarini, J. Niels Rosenquist, James H. Fowler, and Nicholas A. Christakis. 2009. "A Genome Wide Association Study of Educational Attainment." A paper presented at the IZA (Institute for the Study of Labor) Workshop: Genes, Brains, and the Labor Market in Bonn, Switzerland. http://econ.as.nyu.edu/docs/IO/14167/Cesarini1.pdf.

Blum, Kenneth, Ernest P. Noble, Peter J. Sheridan, Olivia Finley, Anne Montgomery, Terry Ritchie, Tulin Ozkaragoz, Robert J. Fitch, Frank Sadlack, Donald Sheffield, Tommie Dahlmann, Sheryl Halbardier, and Harou Nogami. 1991. "Association of the A1 Allele of the D2 Dopamine Receptor Gene with Severe Alcoholism." *Alcohol* 8(5): 409-16.

Brunner, H.G., M. Nelen, X.O. Breakefield, H.H. Ropers, B.A. van Oost. 1993. "Abnormal Behavior Associated with a Point Mutation in the Structural Gene for Monoamine Oxidase A." *Science* 262(5133):578-80.

Cases, Olivier, Isabelle Seif, Joseph Grimsby, Patricia Gaspar, Kevin Chen, Sandrine Pournin, Ulrike Muller, Michel Aguet, Charles Babinet, Jean Chen Shih, and Edward De Maeyer. 1995. "Aggressive Behavior and Altered Amounts of Brain Serotonin and Norepinephrine in Mice Lacking MAOA." *Science* 268:1763-66.

Caspi, Avshalom, Joseph McClay, Terrie E. Moffitt, Jonathan Mill, Judy Martin, Ian W. Craig, Alan Taylor, and Richie Poulton. 2002. "Role of Genotype in the Cycle of Violence in Maltreated Children." *Science* 297:851-54.

Caspi, Avshalom, Karen Sugden, Terrie E. Moffitt, Alan Taylor, Ian W. Craig, HonaLee Harrington, Joseph McClay, Jonathan Mill, Judy Martin, Antony Braithwaite, and Richie Poulton. 2003. "Influence of Life Stress on Depression: Moderation by a Polymorphism in the 5-HTT Gene." *Science* 301:386-89.

Caspi, Avshalom, Ahmad R. Hariri, Andrew Holmes, Rudolf Uher, and Terrie E. Moffitt. 2010. "Genetic Sensitivity to the Environment: The Case of the Serotonin Transporter Gene and Its Implications for Studying Complex Diseases and Traits." *American Journal of Psychiatry* 167:509-27.

Conley, Dalton. 2004. *The Pecking Order: Which Siblings Succeed and Why*. New York: Pantheon Books.

Conley, Dalton and Neil G. Bennett. 2000. "Is Biology Destiny? Birth Weight and Life Chances." *American Sociological Review* 65:458–467.

Conley, Dalton and Neil G. Bennett. 2001. "Birthweight and Income: Interactions Across Generations." *Journal of Health and Social Behavior* 42(4):450-465.

Conley, Dalton, Kate W. Strully, N.G. Bennett. 2003. *The Starting Gate: Birth Weight and Life Chances*. Berkeley: University of California Press.

Cooper, Cyrus, Nicholas Harvey, Zoe Cole, Mark Hanson, and Elaine Dennison. 2009. "Developmental origins of osteoporosis: the role of maternal nutrition." *Advances in Experimental Medicine and Biology* 646:31-9.

Deckert, J., M. Catalano, Y.V. Syagailo, M. Bosi, O. Okladnova, D. Di Bella, et al. 1999. "Excess of High Activity Monoamine Oxidase A Gene Promoter Alleles in Female Patients with Panic Disorder." *Human Molecular Genetics* 8(4):621-4.

Dehejia, Rajeev H. and Sadek Wahba. 1999. "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs." *Journal of the American Statistical Association* 94: 1053-1062.

DiPrete, Thomas A. and Markus Gangl. 2004. "Assessing Bias in the Estimation of Causal Effects: Rosenbaum Bounds on Matching Estimators and Instrumental Variables Estimation with Imperfect Instruments." *Sociological Methodology* 34(1):271-310.

Fan, J.B. and P. Sklar. 2005. "Meta-Analysis Reveals Association between Serotonin Transporter Gene Stin2 VNTR Polymorphism and Schizophrenia." *Molecular Psychiatry* 10:928-38.

Fletcher, Jason M. and Steven F. Lehrer. 2009. "Using Genetic Lotteries within Families to Examine the Causal Impact of Poor Health on Academic Achievement." NBER Working Paper 15148. http://www.nber.org/papers/w15148.

Fowler, James H., Jaime E. Settle, Nicholas A. Christakis. 2011. "Correlated Genotypes in Friendship Networks." *PNAS* 108(5):1993-1997.

Guo, Guang, Michael E. Roettger, and Tianji Cai. 2008a. "The Integration of Genetic Propensities into Social-Control Models of Delinquency and Violence among Male Youths." *American Sociological Review* 73(4):543-68.

Guo, Guang, Xiao-Ming Ou, Michael Roettger, Jean C. Shih. 2008b. "The VNTR 2 Repeat in MAOA and Delinquent Behavior in Adolescence and Young Adulthood: Associations and MAOA Promoter Activity." *European Journal of Human Genetics* 16(5):626-34.

Haberstick, Brett C., Jeffrey M. Lessem, Christian J. Hopfer, Andrew Smolen, Marissa A. Ehringer, David Timberlake, and John K. Hewitt. 2005. "Monoamine Oxidase A (MAOA) and Antisocial Behaviors in the Presence of Childhood and Adolescent Maltreatment." *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* 135B(1):59-64.

Hack, Maureen, Daniel J. Flannery, Mark Schluchter, Lydia Cartar, Elaine Borawski, and Nancy Klein. 2002. "Outcomes in Young Adulthood for Very-Low-Birth-Weight Infants." *The New England Journal of Medicine* 346(2):149-57.

Hamer, D. and L. Sirota. 2000. "Beware the Chopsticks Gene." *Molecular Psychiatry* 5:11-13.

Harris, Kathleen Mullan, Carolyn Tucker Halpern, Andrew Smolen, and Brett C. Haberstick. 2006. "The National Longitudinal Study of Adolescent Health (Add Health) Twin Data." *Twin Research and Human Genetics* 9(6):988-97.

Harris, Kathleen Mullan. 2009. The National Longitudinal Study of Adolescent Health (Add Health), Waves I & II, 1994–1996; Wave III, 2001–2002 [machine-readable data file and documentation]. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill.

Hayes, Breda and Farhana Sharif. 2009. "Behavioural and Emotional Outcome of Very Low Birth Weight Infants – Literature Review." *Journal of Maternal-Fetal and Neonatal Medicine* 22(10):849-56.

Hoenig, John M. and Dennis M. Heisey. 2001. "The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis." *The American Statistician* 55(1):19-24.

Jeong, H., S. P. Mason, A.-L. Barabási, Z. N. Oltvai. 2001. Lethality and Centrality in Protein Networks. *Nature* 411:41-2.

Kim-Cohen, J., A. Caspi, A. Taylor, B. Williams, R. Newcombe, I.W. Craig, T.E. Moffitt. 2006. MAOA, Maltreatment, and Gene-Environment Interaction Predicting Children's Mental Health: New Evidence and Meta-Analysis. *Molecular Psychiatry* 11(10):903–13.

Lawford, Bruce R., Ross Young, Ernest P. Noble, Burnett Kann, Terry Ritchie. 2006. "The D-2 Dopamine Receptor (DRD2) Gene Is Associated with Co-Morbid Depression, Anxiety and Social Dysfunction in Untreated Veterans with Post-Traumatic Stress Disorder." *European Psychiatry* 21(3):180-5.

Levine, Marc and Mary H.H. Ensom. 2001 "Post Hoc Power Analysis: An Idea Whose Time Has Passed?" *Pharmacotherapy* 21:405-9.

Lowe, Jean, Sarah J. Erickson, Peggy Maclean, and Susanne W. Duvall. 2009. "Early Working Memory and Maternal Communication in Toddlers Born Very Low Birth Weight." *Acta Paediatrica* 98(4):660-3.

Noble, Ernest P., Tulin Z. Ozkaragoz, Terry L. Ritchie, Xuxian Zhang, Thomas R. Belin, and Robert S. Sparkes. 1998. "D2 and D4 Dopamine Receptor Polymorphisms and Personality." *American Journal of Medical Genetics: Neuropsychiatric Genetics* 81:254-67.

Pescosolido, Bernice A., Brea L. Perry, J. Scott Long, Jack K. Martin, John I. Nurnberger, Jr., and Victor Hesselbrock. 2008. "Under the Influence of Genetics: How Transdisciplinarity Leads Us to Rethink Social Pathways to Illness." *American Journal of Sociology* 114(S1):S171–S201

Pohjalainen, T., Rinne, J. O., Någren, K., Lehikoinen, P., Anttila, K., Syvälahti, E., Hietala, J. 1998. "The A1 Allele of the Human D_2 Dopamine Receptor Gene Predicts Low D_2 Receptor Availability in Healthy Volunteers." *Molecular Psychiatry* 3:256-60.

Rich-Edwards, Janet W., Meir J. Stampfer, JoAnn E. Manson, Bernard Rosner, Susan E. Hankinson, Graham A. Colditz, Walter C. Willett, and Charles H. Hennekens. 1997. "Birth Weight and Risk of Cardiovascular Disease in a Cohort of Women Followed up Since 1976." *British Medical Journal* 35:396-400.

Roberts, Robert E., Peter M. Lewinsohn, and John R. Seeley. 1991. "Screening for Adolescent Depression: A Comparison of Depression Scales." *Journal of the American Academy of Child & Adolescent Psychiatry* 30(1):58-66

Roseboom, Tessa J., Jan H.P. van der Meulen, Anita C.J. Ravelli, Clive Osmond, David J.P. Barker, and Otto P. Bleker. 2001. "Effects of Prenatal Exposure to the Dutch Famine on Adult Disease in Later Life: An Overview." *Molecular and Cellular Endocrinology* 185:93-8.

Rosenbaum, Paul R. 2002. *Observational Studies, Second Edition.* New York: Springer.

Rosenbaum, Paul R. and Donald B. Rubin. 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score." *The American Statistician* 39:33-8.

Rosenbaum, Paul R. and Donald B. Rubin. 1984. "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score" *Journal of the American Statistical Association* 79:516-24.

Rowe, David C. 2001. *Biology and Crime*. Los Angeles: Roxbury.

Sabol, Sue Z., Stella Hu, and Dean Hamer. 1998. "A Functional Polymorphism in the Monoamine Oxidase A Gene Promoter." *Human Genetics* 103(3):273-9.

Schlotz, Wolff and David I.W. Phillips. 2009. "Fetal Origins of Mental Health: Evidence and Mechanisms." *Brain, Behavior, and Immunity* 23(7):905-16.

Shanahan, Michael J., Stephen Vaisey, Lance D. Erickson, and Andrew Smolen. 2008. "Environmental Contingencies and Genetic Propensities: Social Capital, Educational Continuation, and Dopamine Receptor Gene DRD2." *American Journal of Sociology* 114(S1):S260-S86.

Shih, J. C. and R. F. Thompson. 1999. "Monoamine Oxidase in Neuropsychiatry and Behavior." *American Journal of Human Genetics* 65:593-8.

Sorensen, Henrik Toft, Svend Sabroe, Jorn Olsen, Kenneth J. Rothman, Matthew W. Gillman, and Peer Fischer. 1997. "Birth Weight and Cognitive Function in Young Adult Life: Historical Cohort Study." *British Medical Journal* 315:401-3.

Stelzl, Ulrich, Uwe Worm, Maciej Lalowski, Christian Haenig, Felix H. Brembeck, Heike Goehler, et al. 2005. "A Human Protein-Protein Interaction Network: A Resource for Annotating the Proteome." *Cell.* 122(6):957-68.

Thomas, Duncan C. and John S. Witte. 2002. "Point: Population Stratification: A Problem for Case-Control Studies of Candidate-Gene Associations?" *Cancer Epidemiology, Biomarkers & Prevention* 11:505-12.

Walton, K.A., L.J. Murray, A.M. Gallagher, G.W. Cran, M.J. Savage, C. Boreham. 2000. "Parental Recall of Birthweight: A Good Proxy for Recorded Birthweight?" *European Journal of Epidemiology* 16(9):793-6.

Wiers, R.W., J.A. Sergeant, and W.B. Gunning. 1994. "Psychological Mechanisms of Enhanced Risk of Addiction in Children of Alcoholics: A Dual Pathway?" *Acta Paediatrica Supplement* 404:9-13.

Figure 1: Effect of an Additional Pound at Birth on Depression Symptoms by 5-HTT Genotype
MZ Twins, N=206 p<.01, from 8 oz. below to 8 oz. above mean birth weight of 90 oz.
Any s indicates any short allele; ll indicates two copies of the long allele.
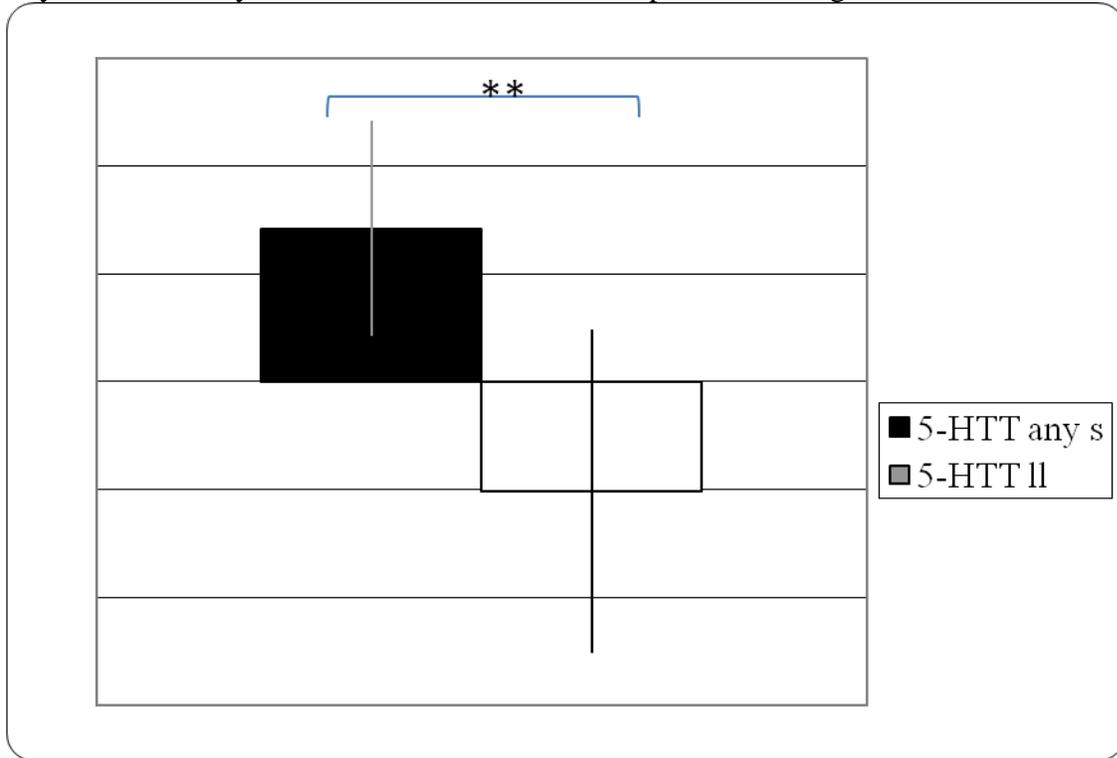
Table 1: Individual Twins with and without Complete Birth Weight Data for their Twin Sibling

	MZ w/ twin BW			MZ w/o twin BW			DZ w/ twin BW			DZ w/o twin BW		
	Mean	Std Error	N	Mean	Std Error	N	Mean	Std Error	N	Mean	Std Error	N
Birth Weight (log oz)	4.5*	0.01	208	4.37*	0.05	15	4.52*	0.01	285	4.42*	0.04	28
Cumulative GPA	2.82*	0.05	181	2.62*	0.08	89	2.61+	0.06	252	2.43+	0.08	104
Highest Grade Comp	13.41	0.13	208	13.08	0.18	100	13.31+	0.12	285	12.96+	.16	129
Depression	5.71	0.35	208	5.77	0.46	98	5.54	0.30	284	5.80	0.49	127
Log Depression	1.63	0.05	208	1.68	0.07	98	1.58	0.05	284	1.61	0.07	127
Any Depression	0.93	0.02	208	0.95	0.02	100	0.89	0.02	285	0.93	0.02	129
Delinquency	0.78**	0.15	202	0.29**	0.09	100	0.69	0.11	283	0.77	0.16	129
Log Delinquency	0.32**	0.04	202	0.14**	0.04	100	0.29	0.03	283	0.32	0.05	129
Any Delinquency	0.30**	0.03	208	0.15**	0.04	100	0.25	0.03	285	0.27	0.04	129
DRD2 – no A1	0.53	0.04	203	0.58	0.05	100	0.61	0.03	285	0.54	0.04	127
MAOA – any 3.5/4R	0.68	0.03	200	0.63	0.05	97	0.71	0.03	280	0.69	0.04	127
5-HTT – ll	0.31	0.03	206	0.33	0.05	98	0.33	0.03	284	0.32	0.04	128

Indicates significant mean difference (two-tailed t-test, unequal variance) ** $p<.01$, * $p<0.05$, + $p<0.10$
With twin BW indicates both twins have complete birth weight information. Without twin BW indicates that only one twin in a pair has complete birth weight information and the other is excluded.

Table 2: Predicting Birth Weight: Within-Twin BW Difference and Individual Twin BW

VARIABLES	(1) Birth Weight Difference	(2) Birth Weight Difference	(3) Birth Weight	(4) Birth Weight
Twin pair	MZ	DZ	MZ – RE	DZ
DRD2 both no A1	-0.021	0.000		
	(0.021)	(0.023)		
DRD2 one no A1	0	-0.054*		
	(0)	(0.027)		
5-HTT both ll	0.024	-0.015		
	(0.022)	(0.024)		
5-HTT one ll	0	-0.022		
	(0)	(0.024)		
MAOA both any 3.5/4R	0.020	-0.009		
	(0.023)	(0.027)		
MAOA one any 3.5/4R	0	0.028		
	(0)	(0.030)		
Male	0.020	-0.021	0.061+	0.033
	(0.021)	(0.019)	(0.034)	(0.022)
Individual				
DRD2 - no A1			-0.019	0.016
			(0.033)	(0.028)
MAOA - any 3.5/4R			0.028	0.037
			(0.037)	(0.030)
5-HTT - ll			-0.053	-0.017
			(0.035)	(0.029)
Constant	0.075**	0.146**	4.482**	4.475**
	(0.025)	(0.033)	(0.040)	(0.034)
Observations	95	136	190	272
R-squared	0.033	0.066		0.799
Number of pairs	95	136	95	136

Standard errors in parentheses
** $p<0.01$, * $p<0.05$, + $p<0.1$

BW = Birth Weight

Table 3: Main Effects of Genes and Birth Weight: Twin Pairs

Panel 1 – Including Hypothesized Gene

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Depression				Delinquency				HS GPA			
	Twins	DZ	MZ	MZ-RE	Twins	DZ	MZ	MZ-RE	Twins	DZ	MZ	MZ-RE
5-HTT ll	-0.31	-0.32	0	0.08								
	(1.10)	(1.14)	(0)	(0.88)								
MAOA any 3.5/4R					-0.10	-0.09	0	-0.64*				
					(0.40)	(0.45)	(0)	(0.29)				
DRD2 no A1									-0.05	-0.05	0	0.17
									(0.13)	(0.15)	(0)	(0.15)
Birth Weight (log oz)	1.86	0.86	3.78	-0.16	-0.53	-0.92	0.24	0.64	-0.02	0.13	-0.30	-0.19
	(2.73)	(3.49)	(4.45)	(2.30)	(0.95)	(1.31)	(1.29)	(0.68)	(0.30)	(0.45)	(0.33)	(0.28)
Male	-0.93	-0.91	0	-0.30	0.76*	0.77*	0	0.19	-0.45**	-0.46**	0	-0.34*
	(0.80)	(0.83)	(0)	(0.82)	(0.30)	(0.34)	(0)	(0.28)	(0.09)	(0.11)	(0)	(0.15)
Constant	-2.17	2.28	-11.27	6.54	2.73	4.49	-0.46	-1.90	3.07*	2.31	4.16**	3.77**
	(12.32)	(15.76)	(20.04)	(10.33)	(4.25)	(5.88)	(5.81)	(3.05)	(1.37)	(2.03)	(1.49)	(1.24)
Observations	486	280	206	206	456	270	186	186	392	232	160	160
R-squared	0.618	0.593	0.654		0.634	0.609	0.692		0.847	0.815	0.912	
No. of pairs	243	140	103	103	228	135	93	93	196	116	80	80

Panel 2 – Including All Three Genes

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Depression				Delinquency				HS GPA			
	Twins	DZ	MZ	MZ-RE	Twins	DZ	MZ	MZ-RE	Twins	DZ	MZ	MZ-RE
5-HTT ll	-0.44	-0.46	0	0.52	-0.70+	-0.71	0	0.29	0.18	0.18	0	-0.23
	(1.14)	(1.22)	(0)	(0.97)	(0.39)	(0.43)	(0)	(0.29)	(0.13)	(0.16)	(0)	(0.17)
MAOA any 3.5/4R	0.75	0.78	0	-1.34	-0.04	-0.02	0	-0.74*	-0.29*	-0.29+	0	-.16
	(1.18)	(1.22)	(0)	(0.99)	(0.40)	(0.45)	(0)	(0.30)	(0.13)	(0.16)	(0)	(0.17)
DRD2 no A1	0.81	0.82	0	-0.38	0.003	0.01	0	-0.04	-0.07	-0.07	0	0.28
	(1.12)	(1.16)	(0)	(0.89)	(0.38)	(0.42)	(0)	(0.27)	(0.13)	(0.15)	(0)	(0.15)
Birth Weight (log oz)	1.60	0.80	3.16	-0.68	-0.61	-1.04	0.25	0.64	0.05	0.18	-0.19	-0.17
	(2.81)	(3.57)	(4.63)	(2.43)	(0.96)	(1.31)	(1.33)	(0.72)	(0.30)	(0.45)	(0.32)	(0.27)
Male	-0.94	-0.91	0	-0.88	0.75*	0.77*	0	0.09	-0.52**	-0.52**	0	-0.43**
	(0.90)	(0.93)	(0)	(0.93)	(0.30)	(0.34)	(0)	(0.29)	(0.10)	(0.12)	(0)	(0.16)
Constant	-1.90	1.51	-8.35	10.32	3.30	5.22	-0.49	-1.86	2.94*	2.28	3.68*	3.86**
	(12.65)	(16.05)	(20.84)	(10.94)	(4.32)	(5.90)	(6.01)	(3.24)	(1.36)	(2.03)	(1.43)	(1.22)
Observations	460	270	190	190	446	268	178	178	376	222	154	154
R-squared	0.615	0.593	0.646		0.635	0.616	0.682		0.856	0.824	0.922	
No. of pairs	230	135	95	95	223	134	89	89	188	111	77	77

Standard errors in parentheses
** p<0.01, * p<0.05, + p<0.1

All models include twin pair fixed effects, except the MZ-RE models, which indicate random effects (because genotype is the same within identical twin pairs).

Table 4: Interaction Effects – Depression, with twin pair fixed effects

VARIABLES	(1)	(2)	(3)	(4)	(5) Depression	(6)	(7)	(8)	(9)
	Twins	DZ	MZ	Twins	DZ	MZ	Twins	DZ	MZ
Birth Weight * 5-HTT-ll	-12.32**	-8.40+	-27.33**						
	(4.27)	(5.03)	(8.59)						
5-HTT-ll	55.38**	37.62	0						
	(19.31)	(22.74)	(0)						
BW * MAOA any 3.5/4R				-10.90*	-9.04	-18.96+			
				(4.95)	(5.80)	(10.33)			
MAOA any 3.5/4R				49.98*	41.59	0			
				(22.41)	(26.27)	(0)			
BW * DRD2 no A1							-5.36	-7.45	2.48
							(4.49)	(5.26)	(9.19)
DRD2 no A1							25.00	34.45	0
							(20.32)	(23.77)	(0)
BW	6.90*	4.13	15.95**	9.44*	7.17	17.48+	4.78	5.23	2.61
	(3.21)	(3.98)	(5.73)	(4.42)	(5.25)	(8.98)	(3.66)	(4.60)	(6.25)
Male	-0.95	-0.92	0	-0.77	-0.77	0	-0.91	-0.89	0
	(0.79)	(0.83)	(0)	(0.89)	(0.93)	(0)	(0.81)	(0.83)	(0)
Constant	-24.94+	-12.49	-28.03	-37.12+	-26.96	-14.38	-15.90	-18.08	-11.81
	(14.47)	(17.98)	(19.91)	(19.96)	(23.70)	(20.22)	(16.51)	(20.75)	(20.81)
Observations	486	280	206	470	272	198	482	282	200
R-squared	0.631	0.601	0.685	0.625	0.596	0.667	0.616	0.597	0.645

Standard errors in parentheses
** p<0.01, * p<0.05, + p<0.1

Table 5: Interaction Effects – Delinquency, with twin pair fixed effects

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				Delinquency Score					
	Twins	DZ	MZ	Twins	DZ	MZ	Twins	DZ	MZ
BW * MAOA any 3.5/4R	-3.60*	-3.55+	-4.36						
	(1.69)	(2.14)	(3.04)						
MAOA any 3.5/4R	16.22*	-16.00	0						
	(7.68)	(9.69)	(0)						
BW * 5-HTT-ll				-2.05	-1.82	-3.27			
				(1.50)	(1.89)	(2.64)			
5-HTT-ll				8.75	7.70	0			
				(6.78)	(8.55)	(0)			
BW * DRD2 no A1							0.26	0.17	0.89
							(1.81)	(1.99)	(4.10)
DRD2 no A1							-1.14	-0.71	0
							(8.20)	(8.98)	(0)
Birth Weight (log oz)	2.01	1.47	3.59	0.27	-0.48	2.13	-0.91	-1.22	-0.46
	(1.52)	(1.94)	(2.66)	(1.13)	(1.50)	(1.78)	(1.48)	(1.74)	(2.82)
Male	0.82**	0.84*	0	0.78**	0.80*	0	0.80*	0.81**	0
	(0.30)	(0.34)	(0)	(0.28)	(0.31)	(0)	(0.32)	(0.31)	(0)
Constant	-8.76	-6.31	-2.24	-0.76	2.64	-4.25	4.39	5.77	0.72
	(6.86)	(8.74)	(5.91)	(5.11)	(6.77)	(6.13)	(6.68)	(7.82)	(9.28)
Observations	456	270	186	472	278	194	468	280	188
R-squared	0.641	0.617	0.699	0.631	0.610	0.682	0.628	0.604	0.654

Standard errors in parentheses
** $p < 0.01$, * $p < 0.05$, + $p < 0.1$

Table 6: Interaction Effects – Cumulative High School GPA, with twin pair fixed effects

VARIABLES	(1)	(2)	(3)	(4) GPA	(5)	(6)	(7)	(8)	(9)
	Twins	DZ	MZ	Twins	DZ	MZ	Twins	DZ	MZ
BW * DRD2 no A1	1.11*	1.41+	0.17						
	(0.51)	(0.71)	(0.67)						
DRD2 no A1	-5.07*	-6.41*	0						
	(2.33)	(3.23)	(0)						
BW * MAOA any 3.5/4R				0.42	0.35	0.75			
				(0.57)	(0.80)	(0.70)			
MAOA any 3.5/4R				-2.20	-1.89	0			
				(2.59)	(3.67)	(0)			
BW * 5-HTT-ll							-0.29	-0.61	0.74
							(0.50)	(0.69)	(0.65)
5-HTT-ll							1.49	2.91	0
							(2.25)	(3.11)	(0)
Birth Weight (log oz)	-0.64	-0.72	-0.38	-0.26	-0.08	-0.73	0.11	0.38	-0.63
	(0.42)	(0.62)	(0.45)	(0.48)	(0.68)	(0.60)	(0.37)	(0.52)	(0.44)
Male	-0.45**	-0.46**	0	-0.53**	-0.53**	0	-0.45**	-0.45**	0
	(0.09)	(0.11)	(0)	(0.10)	(0.12)	(0)	(0.09)	(0.11)	(0)
Constant	5.85**	6.16*	4.14**	4.35*	3.48	3.86**	2.38	1.07	4.62**
	(1.87)	(2.80)	(1.50)	(2.19)	(3.07)	(1.39)	(1.65)	(2.37)	(1.51)
Observations	392	232	160	386	224	162	396	230	166
R-squared	0.851	0.821	0.912	0.857	0.824	0.923	0.848	0.816	0.913

Standard errors in parentheses
** $p<0.01$, * $p<0.05$, + $p<0.1$

Biosketches

Dalton Conley is University Professor at New York University. He also holds appointments at NYU's School of Medicine, the Wagner School of Public Service, and the National Bureau of Economic Research. His research focuses on the determinants of economic opportunity within and across generations.

Emily Rauscher is Assistant Professor of Sociology at the University of Kansas. She studies intergenerational inequality, particularly as it relates to education, occupation, and health.