

# CRIMINOLOGY

## DEMONSTRATING THE VALIDITY OF TWIN RESEARCH IN CRIMINOLOGY\*

J. C. BARNES,<sup>1</sup> JOHN PAUL WRIGHT,<sup>1,6</sup> BRIAN B. BOUTWELL,<sup>2</sup> JOSEPH A. SCHWARTZ,<sup>3</sup> ERIC J. CONNOLLY,<sup>4</sup> JOSEPH L. NEDELEC,<sup>1</sup> and KEVIN M. BEAVER<sup>5,6</sup>

<sup>1</sup>School of Criminal Justice, University of Cincinnati

<sup>2</sup>School of Social Work, Saint Louis University

<sup>3</sup>School of Criminology and Criminal Justice, University of Nebraska at Omaha

<sup>4</sup>Criminal Justice Department, Pennsylvania State University, Abington

<sup>5</sup>College of Criminology and Criminal Justice, Florida State University

<sup>6</sup>Center for Social and Humanities Research, King Abdulaziz University, Jeddah

**KEYWORDS:** assumptions, behavior genetics, biosocial, empirical, quantitative, twins

*In a recent article published in Criminology, Burt and Simons (2014) claimed that the statistical violations of the classical twin design render heritability studies useless. Claiming quantitative genetics is “fatally flawed” and describing the results generated from these models as “preposterous,” Burt and Simons took the unprecedented step to call for abandoning heritability studies and their constituent findings. We show that their call for an “end to heritability studies” was premature, misleading, and entirely without merit. Specifically, we trace the history of behavioral genetics and show that 1) the Burt and Simons critique dates back 40 years and has been subject to a broad array of empirical investigations, 2) the violation of assumptions in twin models does not invalidate their results, and 3) Burt and Simons created a distorted and highly misleading portrait of behavioral genetics and those who use quantitative genetic approaches.*

“The flaws of twin studies are not fatal, but rather seem no worse (and may be better) than the flaws of the typical causal study that relies on observational data.”  
(Felson, 2012: ii)

Behavioral genetic research has existed for more than 100 years (Maxson, 2007). Since its inception, it has been a lightning rod of criticism, especially by scholars who are inalterably opposed to linking biology with behavior. Over this time, numerous critics (e.g.,

---

\* Additional supporting information can be found in the listing for this article in the Wiley Online Library at <http://onlinelibrary.wiley.com/doi/10.1111/crim.2014.52.issue-4/issuetoc>.

We wish to thank each of the five anonymous reviewers for their scholarly insight. In addition, we would like to acknowledge Mara Brendgen, Francis Cullen, Lisabeth DiLalla, Christopher Ferguson, Judith Rich Harris, Kenneth Kendler, Terrie Moffitt, Christopher Patrick, Steven Pinker, Stephen Tibbetts, Catherine Tuvblad, and Anthony Walsh for their comments, suggestions, and feedback on previous drafts of this article. Of course, all errors and omissions are ours and ours alone. Direct correspondence to J. C. Barnes, School of Criminal Justice, University of Cincinnati, Cincinnati, OH 45221 (e-mail: [jc.barnes@uc.edu](mailto:jc.barnes@uc.edu)).

Joseph, 2004; Lewontin, Rose, and Kamin, 1984) have leveled various charges against twin designs and the assumptions on which they are based. Beginning in the 1970s, politically motivated critics of behavioral genetics launched an all-out crusade against such methods, the findings emanating from them, and even on the researchers themselves (Segerstråle, 2000). These critics called for an end to the idea that biology had anything to do with behavior, noting that sociobiology was a “dangerous idea.” Relying heavily on anecdotes, opinions, and the occasional mathematical example, critics of behavioral genetics were unrelenting in their attack.

In response, a small but growing force of behavioral geneticists, statisticians, and other scholars launched a prolonged effort to collect larger samples of twins, other genetically related relatives, and adoptees. They used these samples to test, retest, and refine behavioral genetic models. Throughout the 1980s, behavioral geneticists published study after study documenting the robustness of behavioral genetic methods, especially those designed to assess heritability (Floderus-Myrhed, Pederson, and Rasmuson, 1980; Paul, 1980; Pederson et al., 1985; Rice, Cloninger, and Reich, 1980; Rushton et al., 1986; Scarr, Scarf, and Weinberg, 1980). These studies quickly multiplied, eventually leading behavioral geneticists to “lose their identity” because they became so integrated within psychology and other fields (Scarr, 1987). By the end of the 1980s, the war was over with a large body of research findings supporting the general thrust of behavioral genetic modeling (Plomin and Bergeman, 1991). As time went on, isolated critics (e.g., Joseph, 2004) emerged only to be greeted by even more empirical evidence in favor of the validity of the findings stemming from behavioral genetic studies. Today, behavioral genetic studies inform a broad range of fields, including medicine, psychiatry, psychology, education, and even criminology. Summarizing the large body of behavioral genetic findings that had emerged prior to 2002, Pinker (2002: 374) observed:

The results [of heritability studies] come out roughly the same no matter what is measured or how it is measured. . . . All of this translates into substantial heritability values, generally between .25 and .75. A conventional summary is that about half of the variation in intelligence, personality, and life outcomes is heritable.

The collective body of biosocial evidence thus directly aligns with the “reality” of findings from every other discipline. Broadly speaking, biosocial criminologists have found that approximately 50 percent of the variance in antisocial phenotypes can be attributed to genetic influences, followed by unique environmental influences and, to a lesser extent, common (or shared) environmental influences (Beaver, 2013; Ferguson, 2010; Moffitt, 2005).

These findings, and others, however, find themselves once again imperiled. A recent publication in *Criminology* by Burt and Simons (2014; hereafter Burt and Simons) called for criminologists to abandon heritability studies and effectively to jettison them from the discipline. In doing so, Burt and Simons called for a *de facto* form of censorship. Their logic was straightforward: Twin studies are fatally flawed, the findings cannot be trusted, old findings should be placed on the scrapheap of scientific history, and new findings only reify what is known to be untrustworthy. As a result, no bona fide criminological journal should publish twin-based research moving forward. Joining history’s critics (Joseph, 2004), Burt and Simons critiqued heritability studies and those who use behavioral genetics models. Although their arguments were multifaceted, they bear a striking

resemblance to the arguments leveled against behavioral geneticists in the 1970s. Even so, their argument was powerful, emotionally appealing, and seductive, especially for those uninformed about behavioral genetics and for those ideologically opposed to biology.

Burt and Simons argued that behavioral genetic methods are fundamentally flawed because violation of “crucial assumptions” and “technical limitations” inevitably leads to upwardly biased estimates of heritability and to downwardly biased estimates of the common environment. Burt and Simons detailed a series of methodological arguments against heritability studies, arguing, for instance, that violations of the equal environments assumption (EEA; “the environment of MZ co-twins is no more similar than that of DZ co-twins”) are “flatly contradicted by both empirical evidence and common sense” (p. 231).<sup>1</sup> Indeed, they even told us “behavioral geneticists acknowledged that the EEA was invalid” (p. 232). Yet the EEA was just one of many behavioral genetic assumptions attacked by Burt and Simons. Because twin research rests on critical assumptions, assumptions Burt and Simons argued are *always* violated, findings from heritability studies are “biologically nonsensical” (p. 225). After purportedly scrutinizing prior studies, Burt and Simons went on to brand entire bodies of carefully collected and meticulously analyzed scientific evidence as useless, and even “preposterous” (p. 236). Burt and Simons then proceeded to tell readers that behavioral genetic findings are “implausible” and that behavioral genetic research rests on a “dubious foundation” (p. 223). In the end, they claimed to have exposed the critical flaws of behavioral genetic research.

But did they expose any fundamental flaws of behavioral genetics? Did they present the readers of *Criminology* a fair and impartial assessment of behavioral genetics research? Did they delve deeply into the vast body of behavioral genetic literature, into the technical aspects of assumption violation, biased parameter estimates, and erroneous conclusions? Or, instead, did they join past critics and reify arguments already shown to be unsubstantiated by empirical evidence? Clearly, as even Burt and Simons recognized, “most of the arguments” in their article “are not original” (p. 225). Does repeating arguments originally proposed in the 1970s, and refuted shortly thereafter, make them relevant today? In short, can we believe Burt and Simons have ascertained unbiased and definitive proof that behavioral genetic models are wrong? Have they done something no other statistician, theorist, or behavioral geneticist in the last four decades has been able to accomplish?

In the following pages, we show that the criticisms of behavioral genetic models advanced by Burt and Simons have not only been answered by dozens of prior studies but also that they are wrong. We show this mathematically, with an in-depth examination of the basis of twin designs, with our own simulation models that directly illustrate the impact of violating assumptions on heritability estimates, and by examining 61 empirical studies that have tested one of the “critical assumptions” pointed out by Burt and Simons: the EEA. In the end, we show that the violations of behavioral genetic assumptions hardly qualify as “fatal flaws” that can be used as a justification to abandon heritability studies. Indeed, the evidence to this fact is overwhelming and should reduce any confidence readers extended to the Burt and Simons article.

If showing mathematically that Burt and Simons grossly exaggerated their claims is not sufficient, we then examine the scholarship underpinning their critique of heritability studies. We show, in detail, where they misquoted scholars, where they misrepresented

---

1. MZ = monozygotic and DZ = dizygotic.

study findings, and where they labeled political ideologues as “experts” in behavioral genetics. We show where they selectively cited some studies, relied heavily on others, and at the same time failed to recognize a voluminous literature that would temper, or even wholly disprove, their claims. Finally, we will show their call for a “post-genomic” criminology to be superfluous, even fanciful, serving only to distract criminologists from the hard work yet to be done in the “genomic” age.

We acknowledge that Burt and Simons critiqued other aspects of behavioral genetic research designs including adoption-based designs and twins reared apart designs. Although the literature has provided ample support for these additional designs and the validity of the findings stemming from them (Bouchard et al., 1990; Pinker, 2002; Plomin, DeFries, et al., 2013), we do not thoroughly discuss this line of research for four reasons. First, and as Burt and Simons acknowledged, the classical twin design is the “main ‘workhorse’ used in behavioral genetics to estimate heritability” (p. 229). Second, comprehensive overviews and empirical assessments of both adoption (Heath et al., 1985; Kendler et al., 2013; Sacerdote, 2004) and twins reared apart studies (Bouchard et al., 1990; Gottfredson, 2010; Segal, 2012) have been completed previously. Third, findings garnered from studies using the classical twin design, the adoption design, and the twins reared apart design largely converge, revealing that genetic influences explain approximately 50 percent of the variance in antisocial phenotypes (Ferguson, 2010; Mason and Frick, 1994; Miles and Carey, 1997; Rhee and Waldman, 2002). Fourth, there is simply insufficient space to provide a thorough discussion of these additional research designs in this article.

## ASSUMPTIONS OF THE CLASSICAL TWIN DESIGN

The classical twin design, *like any statistical model*, rests on a foundation of testable assumptions. If those assumptions fail, then estimates drawn from the classical twin design *may* be upwardly or downwardly biased. For this reason, the popular adage “all models are wrong but some are useful” (Box and Draper, 1987: 424) is important to keep in mind when considering the suitability of the classical twin design or, for that matter, *any* statistical model examining human behavior. In addition to the standard probability theory assumptions that permeate all methods of statistical inference, several assumptions are unique to the classical twin design. The heart of the Burt and Simons attack on behavioral genetic studies rests on the violations of assumptions of the classical twin design, particularly the EEA. Although we provide a discussion of other assumptions in appendix A in the online supporting information, in the following sections, we offer a detailed analysis of the EEA (which will upwardly bias heritability estimates when violated) along with what is perhaps the other most frequently discussed assumption of classical twin designs: random mating (which will downwardly bias heritability estimates when violated).<sup>2</sup> Unlike Burt and Simons, who did not provide any empirical evidence of the consequence(s) of violating the assumptions underlying classical twin designs, we also provide a comprehensive presentation of the available literature including the results of empirical studies that provide direct estimations of the degree to which parameter estimates will be biased when such assumptions are violated.

---

2. Additional supporting information can be found in the listing for this article in the Wiley Online Library at <http://onlinelibrary.wiley.com/doi/10.1111/crim.2014.52.issue-4/issuetoc>.

## MODEL IDENTIFICATION AND THE ACE MODEL

In much the same way that a criminologist might partition variance within individuals and between individuals in a data set with a multilevel model, a behavioral geneticist seeks to partition variance in a measured trait into genetic and environmental components. The total variance in any trait results from five separate influences: 1) additive genetic factors ( $A$ ), 2) dominant genetic factors ( $D$ ), 3) epistatic genetic factors ( $I$ ), 4) common (or “shared”) environmental factors ( $C$ ), and 5) nonshared environmental factors ( $E$ ). Thus, the total *variance* of any trait (referred to as a *phenotype*;  $V_p$ ) can be expressed as (for a detailed discussion presented by Purcell, see Plomin, DeFries, et al., 2013: 373–7):

$$\begin{aligned} V_p = & A + D + I + C + E + 2\text{Cov}(A, D) + 2\text{Cov}(A, I) + 2\text{Cov}(A, C) \\ & + 2\text{Cov}(A, E) + 2\text{Cov}(D, I) + 2\text{Cov}(D, C) + 2\text{Cov}(D, E) + 2\text{Cov}(I, C) \\ & + 2\text{Cov}(I, E) + 2\text{Cov}(C, E) \end{aligned} \quad (1)$$

where  $A$  is the additive genetic effect,  $D$  is the dominant genetic effect,  $I$  is the epistatic genetic effect,  $C$  is the common (shared) environmental effect,  $E$  is the nonshared environmental effect *and* error,  $\text{Cov}(A, D)$  is the covariance between  $A$  and  $D$ ,  $\text{Cov}(A, I)$  is the covariance between  $A$  and  $I$ ,  $\text{Cov}(A, C)$  is the covariance between  $A$  and  $C$ ,  $\text{Cov}(A, E)$  is the covariance between  $A$  and  $E$ ,  $\text{Cov}(D, I)$  is the covariance between  $D$  and  $I$ ,  $\text{Cov}(D, C)$  is the covariance between  $D$  and  $C$ ,  $\text{Cov}(D, E)$  is the covariance between  $D$  and  $E$ ,  $\text{Cov}(I, C)$  is the covariance between  $I$  and  $C$ ,  $\text{Cov}(I, E)$  is the covariance between  $I$  and  $E$ , and  $\text{Cov}(C, E)$  is the covariance between  $C$  and  $E$ .

The solution for  $V_p$  is intuitive and in line with standard probability/counting theory that notes the variance of a sum is calculated as the summation of the unique variance of variable 1, the unique variance of variable 2, and two times their covariance. Several practical issues have prevented researchers from estimating all parts of the equation simultaneously. As noted by Purcell (in Plomin, DeFries, et al., 2013: 377), “by definition, the additive genetic influences are independent of dominance deviations. That is,  $\text{Cov}(A, D)$  will necessarily equal zero.” Thus, this term may be omitted safely. A similar conclusion can be reached for the  $\text{Cov}(A, I)$  parameter and the  $\text{Cov}(D, I)$  parameter. In terms of the environmental parameters, common (shared) environmental factors cannot, by definition, overlap with nonshared environmental factors, allowing us to omit the  $\text{Cov}(C, E)$  parameter. Of the remaining parameters,  $D$  and  $I$  are often omitted by making an additional assumption (see appendix A in the online supporting information). Similarly, the remaining covariance parameters [i.e.,  $\text{Cov}(A, C)$ ,  $\text{Cov}(A, E)$ ,  $\text{Cov}(D, C)$ ,  $\text{Cov}(D, E)$ ,  $\text{Cov}(I, C)$ , and  $\text{Cov}(I, E)$ ] often are omitted by making additional assumptions (see appendix A in the online supporting information). Relying on these assumptions, we are left with the well-known ACE model:

$$V_p = A + C + E$$

Once we have simplified the equation into the  $A$ ,  $C$ , and  $E$  parameters, two additional assumptions are necessary to identify the model. These two assumptions, along with the consequences of violating them, are outlined in the next section.

## ASSUMPTION 1: HUMANS MATE RANDOMLY (NO ASSORTATIVE MATING)

## ASSUMPTION

The assumption of random mating is required to fit the classical twin design because this assumption defines the variance–covariance matrix for DZ twins. As is shown in appendix B in the online supporting information, the covariance between DZ twins for any trait is expressed as  $\frac{1}{2}A+C$ . The assumption of random mating defines the  $\frac{1}{2}$  portion of the equation. When two humans reproduce, germ cells formed through *meiosis* (which is the process of genetic mixing for sexual reproduction) fuse and form the zygote that will eventually develop into an independent and genetically unique human (Carey, 2003; McConkey, 2004). As a consequence of meiosis and fertilization, a quasi-random 50 percent of the genes from each parent (i.e., a random 50 percent maternally and a random 50 percent paternally) are combined to create the offspring, which will not be genetically identical to either parent at all genetic loci. Thus, one may assume that any offspring produced by two humans will be 50 percent similar to their mother and 50 percent similar to their father at the distinguishing loci. Based on this logic, full siblings and DZ twins are 50 percent similar, on average, within the distinguishing regions of the genome.

Given that the violation of this assumption actually *deflates* heritability estimates, it represents an important counterpoint to the assumptions that are frequently listed as producing inflated heritability estimates. *Against this backdrop, it is somewhat surprising that Burt and Simons did not include any discussion of the consequences of violating the assumption of random mating in their article.*<sup>3</sup> Perhaps Burt and Simons omitted a discussion of assortative mating because they were unaware of the assortative mating literature. In a prior study, however, Simons et al. (2002: 404) noted, “Past research on dating and mate selection has demonstrated strong support for the idea of assortative mating (Collins, 1985).” In the same paper, Simons and colleagues tested for assortative mating, found evidence of mate assortment, and ultimately criticized existing theories for not having incorporated a discussion of assortative mating. Why the issue of assortative mating was not given direct attention in the Burt and Simons critique is, therefore, unclear.

## THE IMPACT OF ASSUMPTION VIOLATION

The process of meiosis ensures that, on average, full siblings and DZ twins will share 50 percent of their distinguishing genotype if mating is random. If mating is not random, then the 50 percent figure may be an underestimate, which would lead to underestimates of the *A* parameter in the ACE model. To see why this is the case, consider the impact of assortative mating (i.e., nonrandom mating) on the DZ variance–covariance matrix. Substituting hypothetical values and solving for *A* clearly indicates that a trait that is completely influenced by additive genetic factors (*A*) will produce a heritability estimate

---

3. Burt and Simons provided an indirect acknowledgment of the assumption of random mating by stating that classical twin designs assume “[t]he genes of MZ twins are 100 percent identical and are approximately 50 percent identical for DZ twins” (Burt and Simons: 230). Although this statement necessitates the assumption of random mating, this point is not made clear by Burt and Simons and is perhaps even obfuscated by lumping the assumption of 100 percent genetic similarity of MZ twins together with the assumption of 50 percent (on average) genetic similarity for DZ twins. The former relies on principles of molecular biology that define the DNA structure of MZ twins.

for the  $A$  parameter that is below 1.00 because of a violation of this assumption. Specifically, if a hypothetical trait were completely the result of additive genetic factors, then we should expect estimates of  $A$  to, on average, hover around 1.00. But, if the assumption of random mating is violated, then the  $\frac{1}{2}A$  value in the DZ correlation matrix will be too low, producing an overall estimate of  $A$  that is below 1.00 because the correlation for MZ twins will be 1.00 but the correlation for DZ twins will be above the expected .50; the value will reflect the amount of genetic correlation that is *actually* present. When this occurs in practice, the ACE model (which uses the  $\frac{1}{2}A$ ) attributes any portion of the DZ correlation that is above .50 to the shared environment ( $C$ ). Thus, violation of the random mating assumption leads to inflated estimates of the shared environment effect and deflated estimates of heritability.

#### EMPIRICAL EVIDENCE OF ASSUMPTION VIOLATION

An impressive body of research exists regarding mate similarity across a variety of demographic factors (Vandenburg, 1972) and other traits/behaviors such as level of education (Domingue et al., 2014; Mare, 1991) and political affiliation (Alford et al., 2011). One may have expected such correlations among these factors, but the focus for this discussion is on whether mates tend to correlate in their antisocial behavior (or across important correlates of antisocial behavior, such as self-control). Mental illness, addiction, and drug use display a pattern of similarity between mates that suggests the presence of assortative (or nonrandom) mating for antisocial behavior to some degree (Jacob and Bremer, 1986; Merikangas and Spiker, 1982; Rhule-Louie and McMahon, 2007). Findings from a diverse line of scholarship, moreover, suggest humans select mates who display similar levels of antisocial and aggressive behavior (e.g., Boutwell and Beaver, 2010; Boutwell, Beaver, and Barnes, 2012; Capaldi, Kim, and Owen, 2008; Haynie et al., 2005; Krueger et al., 1998; Rhule-Louie and McMahon, 2007; Rowe and Farrington, 1997). Despite some variation, each of these analyses has reported a positive and statistically significant correlation between mates for antisocial outcomes, and there is relative consistency in the strength of the observed associations across studies. Krueger and colleagues (1998) reported mating assortment for antisocial behavior in couples that was around  $r = .50$ . This correlation is similar in magnitude (depending on the trait in question) to the correlations reported in other studies (Boutwell and Beaver, 2010; Boutwell, Beaver, and Barnes, 2012; Rowe and Farrington, 1997). In short, empirical evidence shows that sexual partners do *not* mate randomly, and thus, the assumption of random mating is likely consistently violated in the classical twin design on many behavioral phenotypes (Alford et al., 2011).

#### CALCULATING AND SIMULATING THE IMPACT OF ASSUMPTION VIOLATION

To demonstrate the consequences of violating the assumption of random mating in the ACE model, we followed a two-pronged approach. First, we created a simple script in the computer program  $R$  that would estimate  $A$ , which will be referred to with the heritability estimate notation of  $h^2$  to reduce confusion (i.e.,  $A$  will be used to refer to the “actual” or “true” level of additive genetic influence and  $h^2$  is used to refer to the “estimated” level of additive genetic variance that is retrieved from the ACE model). The coefficient for  $h^2$  was estimated for different levels of “true”  $A$  and at different values of the degree to which DZ twins *actually* overlap in distinguishing genes that



**Table 1. Impact of Violating the Assumption of Random Mating on  $h^2$  Estimates**

Calculation Results	"True" Parameter Values			
	A = .25, C = .50		A = .50, C = .25	
	$h^2$ Estimate	Difference from "True" Parameter	$h^2$ Estimate	Difference from "True" Parameter
Genotypic similarity of DZs = .50	.250	.000	.500	.000
Genotypic similarity of DZs = .52	.240	-.010	.480	-.020
Genotypic similarity of DZs = .54	.230	-.020	.460	-.040
Genotypic similarity of DZs = .56	.220	-.030	.440	-.060
Genotypic similarity of DZs = .58	.210	-.040	.420	-.080
Genotypic similarity of DZs = .60	.200	-.050	.400	-.100
Simulation Results	Average $h^2$ Estimate	Average $c^2$ Estimate	Average $h^2$ Estimate	Average $c^2$ Estimate
Genotypic similarity of DZs = .50	.252	.497	.503	.246
90 percent range	(.162-.339)	(.417-.577)	(.395-.610)	(.143-.344)
Genotypic similarity of DZs = .55	.222	.526	.447	.301
90 percent range	(.131-.318)	(.441-.610)	(.333-.561)	(.195-.410)
Genotypic similarity of DZs = .60	.200	.548	.401	.347
90 percent range	(.116-.293)	(.468-.623)	(.305-.513)	(.249-.439)

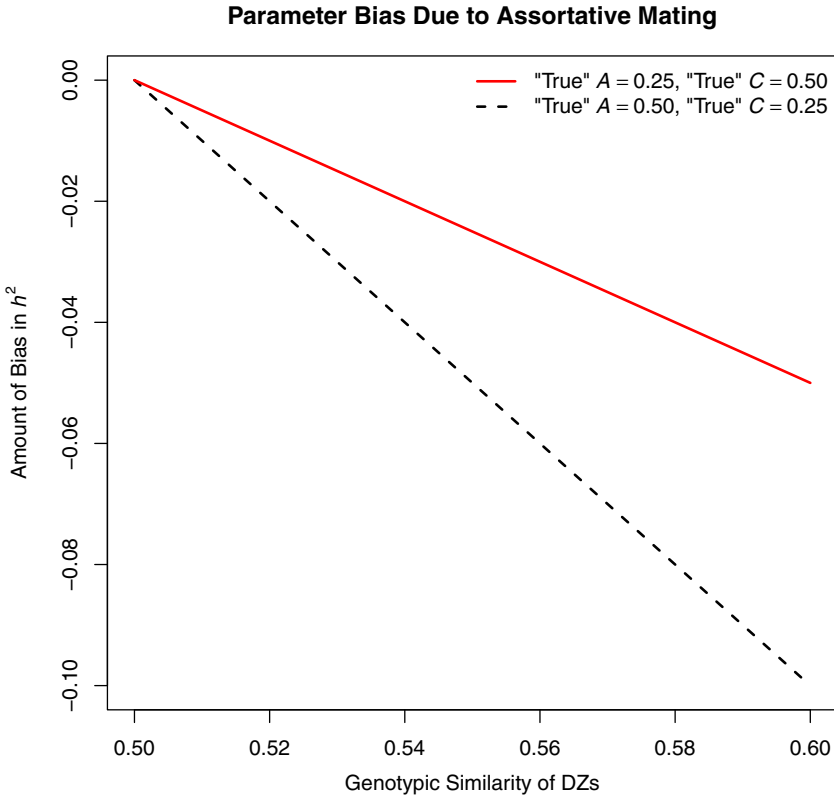
influence the trait in question. The latter element—the *actual* level of genetic overlap for DZ twins—was set to range between .50 (i.e., no violation of the assumption) and .60 (a fairly substantial departure from the assumption). These values were chosen because the evidence presented previously suggests the level of assortative mating for antisocial outcomes is likely to range between  $r = .25$  and  $r = .50$ , which may translate to a DZ genotypic relatedness score that ranges between .01 and .10 higher than the assumed level of .50 (Lynch and Walsh, 1998: 158). The R script is provided in appendix E in the online supporting information. The results from the calculations are presented in table 1 under the panel labeled "Calculation Results."

As shown in table 1, the  $h^2$  estimate decreases as the level of genotypic similarity for DZ twins increases above .50 (i.e., as the assumption begins to fail). Interestingly, the degree of bias is greater under conditions where additive genetic factors account for more variance in the trait; bias is greater when "true"  $A = .50$  compared with when "true"  $A = .25$ . When "true"  $A$  is set to .50, then a .01 increase in the genotypic similarity of DZs translates to a reduction in the  $h^2$  estimate of 1 percentage point. The same increase in the genotypic similarity of DZs amounts to a reduction in the  $h^2$  estimate of .5 percentage points when "true"  $A = .25$ . The results from the calculations also are presented graphically in figure 1.

Our second approach to demonstrating the impact of violating the assumption of no assortative mating was to produce two series of computer simulations: once where the "true" parameters were set to  $A = .25$ ,  $C = .50$ , and  $E = .25$  and a second time where the "true" parameters were set to  $A = .50$ ,  $C = .25$ , and  $E = .25$ . In both simulations, the *actual* level of genetic overlap for DZ twins was set to vary among .50, .55, and .60. Variance-covariance matrices from 500 simulated data sets—each with 500 MZ twin pairs



**Figure 1. Heritability Bias at Different Levels of Random Mating Violation and Different Levels of “True” A and “True” C**



and 500 DZ twin pairs—for each condition were analyzed with the latent variable ACE modeling program provided in the *OpenMx* package (Neale and Maes, 2004) available in *R*. As shown in the bottom panel of table 1, the simulations confirmed the results from the calculations discussed previously by revealing that the *A* parameter is *underestimated* and that the *C* parameter is *overestimated* when the random mating assumption is violated. When the level of genotypic similarity among DZ twins is set to .55 (i.e., a violation of the assumption), the ACE model underestimates the *A* parameter by roughly 3 percentage points, on average, when “true” *A* = .25. The simulations revealed that heritability estimates are approximately 5 percentage points lower than the “true” value when the genotypic similarity of DZ twins is .55 and “true” *A* = .50. As the genotypic similarity among DZ twins is set to higher values, the degree of underestimation of *A* increases. Inversely, the *C* parameter is consistently overestimated as the level of genotypic similarity among DZ twins increases.

**Conclusion:** *When the assumption of random mating fails, a portion of the variance that should be attributed to A is instead attributed to C, and this bias is more substantial for traits with higher levels of “true” additive genetic variance. Assortative mating therefore downwardly biases heritability estimates (i.e.,  $h^2$ ) and upwardly biases the estimate of the*

*common/shared environment (i.e.,  $c^2$ ). These results are contrary to the general thrust of the critique offered by Burt and Simons.*

## ASSUMPTION 2: THE EQUAL ENVIRONMENTS ASSUMPTION

### ASSUMPTION

Genetic factors are solely responsible for the increased similarity between MZ twins relative to DZ twins. This assumption is well known and often is referred to as the *equal environments assumption (EEA)*. In terms of the equations presented in appendix B in the online supporting information, the EEA allows one to solve for  $A$ ,  $C$ , and  $E$  by assuming  $C$  (i.e., the shared environment) carries a similar influence across all sibling pairings (MZ and DZ twins in the current example). Given that the crux of the Burt and Simons critique rests on the violation of this assumption, we pay close attention to the issue. We also should note that should this assumption prove valid or if violations of the assumption have a trivial influence on heritability estimates, then the Burt and Simons argument against the validity of heritability studies will be drawn into serious question.

### THE IMPACT OF ASSUMPTION VIOLATION

If the EEA fails, then the ACE model equations along with the variance–covariance equations will be biased. The most likely direction of bias vis-à-vis an EEA violation is that  $A$  will be overestimated,  $C$  will be underestimated, and  $E$  should be unaffected. The logic behind this conclusion is simple: If certain types of siblings/twins receive more  $C$  than other types of siblings/twins, then those siblings will be more similar to one another as a result of having greater levels/impacts of  $C$ . This becomes problematic for the classical twin design because critics often argue that MZ twins will receive greater levels of  $C$  relative to DZ twins because they tend to look more similar to one another. In addition, MZ twins are the same sex but roughly half of all DZ twins are opposite sex. For these reasons, critics claim the greater similarity observed for MZ twins may simply be because they receive more similar treatment from the environment rather than their greater level of genetic similarity, effectively violating the EEA. Directly related to these observations, critics of twin research have correctly pointed out that MZ twins tend to have more environments in common relative to DZ twins, including parental treatment (Kendler et al., 1994), closeness with one another (Horwitz et al., 2003; Lykken et al., 1990), belonging to the same peer networks (McGuire and Segal, 2013), being enrolled in the same classes (Cronk et al., 2002), and being dressed similarly (Cronk et al., 2002; Loehlin and Nichols, 1976).

In light of these observations, the EEA has been the subject of much debate and has sparked the production of a large literature that spans several decades and cuts across multiple fields of study (e.g., Allison et al., 1996; Bulik, Sullivan, and Kendler, 1998; Conley et al., 2013; Cronk et al., 2002; Derks, Dolan, and Boomsma, 2006; Eaves, Foley, Silberg, 2003; Felson, 2014; Hannagan and Hatemi, 2008; Hatemi et al., 2009; Kendler and Gardner, 1998; Kendler et al., 2000; Littvay, 2012; Rose et al., 1988; Scarr and Carter-Saltzman, 1979). Certain critics have cited violations of the EEA as a damning limitation for twin research in sociology (Horwitz et al., 2003), political science (Beckwith and Morris, 2008; Charney, 2008; Suhay and Kalmoe, 2010), educational psychology (Richardson and Norgate, 2005), and social psychology (Simons, Beach, and Barr, 2012).

Perhaps the most aggressive critic of twin studies has been Joseph (2004, 2006, 2010) who questioned the findings of heritability studies because of the presence of unequal environments (i.e., a violated EEA). Burt and Simons simply echoed Joseph's (2004) sentiments, stating: "[W]e *think* it is unquestionably the case that violations of the EEA are inflating heritability and decreasing shared environmental effects to a substantial degree" (p. 236, emphasis added). Like many before them, Burt and Simons failed to acknowledge that their discussion of the potential effects of violating the EEA was an empirically testable issue (Littvay, 2012).

#### EMPIRICAL EVIDENCE OF ASSUMPTION VIOLATION

To assess the current empirical reality, we performed an exhaustive search of the literature bearing directly on the EEA. Appendix D in the online supporting information displays all of the studies that have examined the EEA and that were located through a systematic search of the literature using ProQuest, Web of Science, and PsycINFO. Key search words and terms used to locate such studies included "EEA," "equal environments assumption," and "unequal environments." Any matches that provided an empirical assessment or comprehensive overview of the EEA were selected. In total, this search process, along with the inclusion of publications that were cited by Burt and Simons, generated 61 pieces of scholarship.

The bolded studies at the top of appendix D in the online supporting information represent the studies that were cited by Burt and Simons. As shown, Burt and Simons cited a total of nine studies, only two of which included an empirical analysis.<sup>4</sup> One of these studies (Horwitz et al., 2003) examined whether the EEA was violated but did not examine the effect of such violations on heritability estimates. The other empirical study (Cronk et al., 2002) included by Burt and Simons was actually miscited. Burt and Simons cited this study as evidence of a violated EEA, but the primary conclusions of the study clearly indicated (even within the abstract) that controlling for the presence of unequal environments *did not* result in a significant change in the heritability estimates for any of the examined outcomes. Indeed, the average change in heritability estimates was only .02 (or 2 percentage points). Based on the consistency of their findings, Cronk et al. (2002) concluded that "[o]ur results *support the validity of the assumptions of equal environments*, upon which conclusions from these twin studies are based" (p. 836; emphasis added).

A close inspection of appendix D in the online supporting information reveals Burt and Simons cherry-picked studies that align directly with their argument that violations of the EEA are pervasive and undermine heritability estimates. The studies included in appendix D in the online supporting information tested for violations of the EEA across 1,233 environments and violations were detected in only 112 of them (9 percent). Of the 61 studies available, only 13 concluded that the EEA was invalid (21 percent), but of these only 6 performed any empirical analysis (10 percent), and none of these studies actually estimated the impact of the presence of unequal environments on heritability estimates. However, several studies examined directly the effect of violating the EEA on heritability estimates. Appendix D in the online supporting information includes 11 studies that

---

4. One of the cited articles (Richardson, 2011) was a summary piece published in an online, non-peer-reviewed newsletter, *GeneWatch*. Another study is unpublished and three others are books that provide no new data. After taking these citations into account, Burt and Simons cited a total of two empirical pieces of literature, only one of which actually tested the impact of unequal environments on heritability or shared environmental estimates.

estimated the impact of unequal environments on heritability estimates, with the average effect being an upward bias of about .012 (or about 1 percentage point) in the heritability estimate. What this necessarily means is that the widely cited heritability estimate of .50 for antisocial behaviors may be upwardly biased by .012 and the “true”  $A$  is actually closer to .488. However, we should note that these estimates do not take into account violations of other assumptions (e.g., assortative mating; the presence of evocative gene-environment correlation) that may downwardly bias heritability estimates. Nonetheless, as appendix D in the online supporting information indicates, Burt and Simons did not provide readers with a systematic and unbiased account of the EEA literature.

The results presented in appendix D in the online supporting information are revealing and show convincingly that the EEA likely has little-to-no influence on heritability estimates. These conclusions are echoed by Felson (2014; see also Felson, 2012: ii), who performed “the most comprehensive evaluation of the equal environments assumption to date.” Using the Midlife Development in the United States (MIDUS) survey, Felson examined 32 outcomes that tapped a range of psychological and sociological domains. In addition, he examined a diverse set of environmental similarity measures that included similarity of childhood environment, proportion of lives lived together, frequency of contact, level of psychological intimacy, and how often each twin shared advice with his or her co-twin. Importantly, these environmental similarity measures contain several experiences that are cited often as evidence of unequal environments. The results revealed that only 1 of the 58 estimated models led to a significant change in  $h^2$  estimates after accounting for the similarity measures. Although the remaining models revealed nonsignificant differences, such differences were still modest averaging around .10 (or 10 percentage points). Importantly, the changes in  $h^2$  after accounting for the environmental similarity measures did not follow any discernable pattern and were not consistent over time, indicating that any potential bias introduced is likely *random* and does not *systematically* bias  $h^2$  or  $c^2$  estimates. Felson (2012: ii) concluded that “[t]he flaws of twin studies are not fatal, but rather seem no worse (and may be better) than the flaws of the typical causal study that relies on observational data.”<sup>5</sup>

Although numerous studies examining the potential moderating effects of environmental similarity on  $h^2$  and  $c^2$  estimates have found that violations of the EEA result in statistically nonsignificant parameter deviations (e.g., Allison et al., 1996; Borkenau et al., 2002; Bulik, Sullivan, and Kendler, 1998; Cronk et al., 2002; Felson, 2014; Hettema, Neale, and Kendler, 1995; Kendler et al., 1994; Kendler and Gardner, 1998; Klump et al., 2000; Littvay, 2012; Loehlin and Nichols, 1976; Morris-Yates et al., 1990; Plomin, Willerman, and Loehlin, 1976; Scarr and Carter-Saltzman, 1979), additional methodologies also have been employed. Perhaps the most empirically rigorous method of assessing the validity of the EEA is through the use of misclassified twin samples. More specifically, in samples where zygosity is determined via responses to self-report questionnaires tapping confusability, misclassifications can occur where MZ twins are initially classified as DZ twins and vice versa. Once genotyping tests are conducted, however, these classifications are corrected. Several studies have drawn on this unique situation to employ a more robust test of the EEA (Conley et al., 2013; Gunderson et al., 2006; Kendler et al., 1993; Xian

---

5. As with the empirical assessments listed in table 2, Felson (2012) came to this conclusion without assessing the impact of violating other assumptions (e.g., assortative mating) that may downwardly bias heritability estimates.

et al., 2000). This situation presents an ideal way to test whether the EEA is violated and whether such violations result in meaningful changes in estimates of  $h^2$  and  $c^2$ . Assuming that unequal environments result in biased  $h^2$  estimates, DZ twins that are mistaken as MZ twins should more closely resemble one another across the phenotypes of interest relative to correctly identified DZ twins. Similarly, if critics of twin studies are correct, then MZ twins incorrectly classified as DZ twins should be less similar to one another across the examined phenotypes relative to correctly classified MZ twins.

In the most recently performed misclassification study, Conley et al. (2013) used three distinct samples—the National Longitudinal Study of Adolescent Health (Add Health), the Child and Adolescent Twin Study in Sweden, and the Minnesota Twin Family Study (MTFS)—to examine whether unequal environments (measured as zygosity misclassification) produce biased  $h^2$  estimates. Importantly, Conley et al. examined a wide range of outcome measures, many of which are commonly used by criminologists, including height, weight, body mass index, depression, attention deficit hyperactivity disorder, delinquency, and high-school GPA. The results revealed that heritability estimates are not significantly inflated when the EEA is violated (such as when twins are misclassified). In addition, violating the EEA actually resulted in artificially *deflated* heritability estimates in some of the estimated models, leading the authors to conclude that “it seems reasonable to take results from an ACE model more or less at face value” (p. 425). Once again, these findings indicate that any potential bias stemming from violations of the EEA is likely random, as evidenced by the significant *deflation* of  $h^2$  estimates in some models. Importantly, these results directly align with previous studies that analyzed misclassified twin pairs (Gunderson et al., 2006; Kendler et al., 1993; Xian et al., 2000), revealing a consistent overall pattern of findings.<sup>6</sup>

Germane to the discussion of random bias because of a violation of the EEA is the premise by Burt and Simons that including opposite-sex DZ twins is a methodologically unsound practice of twin research. To justify their argument, they relied on two points. First, they referred to a “voluminous” literature regarding differences in experiences between sexes. However, they offered no citations to contextualize the type of experiences to which they refer nor did they discuss how those differences in experiences are relevant to their argument against twin studies in terms of the *effect* on heritability estimates. This oversight is important because it is left to the reader’s imagination to figure out those experiences and the ways in which such experiences could affect heritability estimates.<sup>7</sup>

Second, Burt and Simons suggested that heritability estimates are always inflated by including opposite-sex twins in statistical analyses. This, however, is not the case. Including opposite-sex twins in genetic analysis is a conventional practice as it allows for the testing of sex differences in heritability and environmental influences. Statistical

- 
6. Another interesting manner by which the EEA has been assessed was recently completed in two studies by Segal and colleagues (Segal, 2013; Segal, Graham, and Ettinger, 2013). In these studies, the authors employed a sample of genetically unrelated look-alikes (i.e., non-kin doppelgänger pairs). Critics of twin studies have noted that a violation of the EEA is likely to result in part because of the degree to which people (MZs) who look more alike are treated more similarly (compared with DZs). Segal and her colleagues illustrated that there was virtually no concordance across a wide variety of personality characteristics among the unrelated look-alike pairs, suggesting the EEA is upheld.
  7. Although it may seem intuitive that including different-sex twins may produce biased results, science advances based on empirical evidence and not on intuition or common sense. Therefore, such a claim requires evidence of the effect of including different-sex twins on estimates of heritability.

modeling strategies have been developed and others have been modified to handle samples containing opposite-sex twin pairs (e.g., Purcell and Sham, 2003). Even more revealing is that there is not consistent evidence showing significant sex differences when it comes to heritability estimates (Meier et al., 2011; Viding et al., 2004), which indicates that opposite-sex twin correlations are not significantly different from same-sex twin correlations. The findings generated from studies that analyze samples containing opposite-sex twin pairs showed exactly this, with opposite-sex twin correlations commonly not being significantly different from same-sex twins (Meier et al., 2011). Interestingly, and in direct contradiction to what Burt and Simons claimed, sometimes the opposite-sex twin correlations are greater in magnitude when compared with same-sex twin correlations (Saudino, Ronald, and Plomin, 2005).

Burt and Simons cited two studies (Meier et al., 2011; Saudino, Ronald, and Plomin, 2005) to support their claim that including opposite-sex twins in studies results in a violation of the EEA. Interestingly, the findings and conclusions drawn from these two studies provide evidence that runs counter to the Burt and Simons claim. More specifically, although Meier et al. (2011) found significant differences in cross-twin correlations between same-sex and opposite-sex DZ twins wherein same-sex DZ twins (both males and females) possessed greater levels of concordance relative to opposite-sex DZ twins across some of the examined outcomes, they also reported that the correlations were not significantly different between opposite-sex and same-sex twins for some outcomes. The other study cited by Burt and Simons (Saudino, Ronald, and Plomin, 2005) reported findings wherein opposite-sex twin correlations were *greater* than same-sex twin correlations for some of the examined outcomes. The key takeaway points are that 1) the inclusion of opposite-sex twins is not an unconventional practice in behavioral genetic studies, 2) these types of twin pairs can provide information as to potential etiological differences between males and females, and 3) including opposite-sex twins in statistical analyses does not seem to have any consistent effect on biasing variance component estimates in any systematic direction.

#### CALCULATING AND SIMULATING THE IMPACT OF ASSUMPTION VIOLATION

In an effort to provide a more direct example of the biasing effects of EEA violations on estimates of  $h^2$  and  $c^2$ , Burt and Simons argued that “if the shared environmental effect is .3 for MZ twins and .2 for DZ twins, then heritability estimates will be inflated by 20%” (p. 232). These values were garnered from an unpublished manuscript by Suhay and Kalmoe (2010), and even though they represent arbitrary values from a hypothetical example, Burt and Simons presented them as empirical evidence favoring an inflated  $h^2$  estimate and deflated  $c^2$  estimate stemming from a violation of the EEA. However, this example is overly simplistic and highly problematic in two important ways. First, the degree to which an EEA violation inflates  $h^2$  is directly tied to the degree to which shared environmental influences *actually* affect the trait (i.e., “true”  $C$ ). As “true”  $C$  decreases, the biasing effect of the EEA *decreases*. This information has bearing on the Suhay and Kalmoe calculation because they arbitrarily chose values for the MZ and DZ correlations but did not specify the “true” value for  $C$ . So, the reader is left without a baseline value in which to determine how much  $h^2$  has been inflated and  $c^2$  deflated.

In addition, the amount of error introduced by the EEA in Suhay and Kalmoe’s (2010) calculations is understated, erroneously leading to the conclusion that a small violation



amounts to large biases. The authors stated: “[t]o see more clearly why this is the case, imagine a small amount of environmental error has crept into the relevant MZ and DZ trait correlations” (Suhay and Kalmoe, 2010: 5). They then provided a hypothetical example where  $c^2_{MZ} = .3$  and  $c^2_{DZ} = .2$ . The “small amount of environmental error” is reflected in the MZ correlation being 50 percent higher than the DZ correlation: hardly a “small amount of environmental error.” Moreover, their mathematical discussion of the biasing impact of the EEA is too simplistic. Violations of the EEA will operate by “down weighting” the  $C$  parameter for DZ twins. To see why this is the case, recall that the EEA assumes the amount of  $C$  is equivalent across MZ and DZ twins. If the assumption fails, then the only logical outcome (speaking both conceptually and mathematically) is that MZ twins will receive 100 percent of  $C$  but DZ twins will receive less. Thus, we must “down weight” the proportion of  $C$  that is received by the DZ twins to test the violations from the EEA. In this way, Suhay and Kalmoe’s example is misleading because the values assigned for the cross-twin correlations correspond to DZ twins sharing only 67 percent of  $C$  compared with the amount of  $C$  shared by MZ twins. This represents a significant violation of the EEA and begs the question of whether the seemingly innocuous example is realistic.

Because of the problematic and overly simplistic nature of the example offered by Suhay and Kalmoe (2010), we provided a more realistic and comprehensive test of the EEA using a series of calculations and computer simulations. Mathematically, violations of the EEA were introduced to the model by “down weighting” the  $C$  parameter for DZ twins. Thus, the level of  $C$  present in the MZ variance/covariance matrix is the “true” level of  $C$  and the amount of  $C$  present in the DZ variance/covariance matrix must be “down weighted” to demonstrate an EEA violation (see appendix B in the online supporting information for the variance/covariance matrices).

When this strategy was applied to the calculations introduced previously (see appendix E in the online supporting information for the  $R$  script), evidence was produced to support the notion that EEA violations will inflate  $h^2$  as a function of the degree to which the EEA is violated. Table 2, which is structured similarly to table 1, presents the results from a series of calculations where the EEA was purposely violated under two conditions: 1) where “true”  $A = .25$  and “true”  $C = .50$  and 2) where “true”  $A = .50$  and “true”  $C = .25$ . Also, consistent with the previous analysis, the results from the calculations were plotted and are presented in figure 2. As shown in both the table and the figure, departures from the EEA consistently lead to overestimates of  $h^2$  and underestimates of  $c^2$ . Note, however, that the bias in  $h^2$  was weaker when “true”  $C$  was set to the lower value.

The results from the computer simulations—again, 500 simulated data sets were generated for each condition and the ACE model was estimated using *OpenMx* in  $R$ —corroborated the results presented in the calculations portion of the table. Both the calculations and the simulation results indicate that departures from the EEA lead to overestimates for  $h^2$  and underestimates of  $c^2$ , but these biases are weaker as the “true” impact of shared environmental factors decreases.

**Conclusion:** *Taken together, this body of findings provides clear evidence suggesting that the EEA is typically not violated and that estimates of  $h^2$  and  $c^2$  garnered from behavioral genetic models are relatively unbiased. Even in situations where the EEA is almost certainly violated (e.g., misclassified twins), such studies have indicated that estimates of  $h^2$  and  $c^2$  are highly robust and experience only substantively minor changes (Carey,*



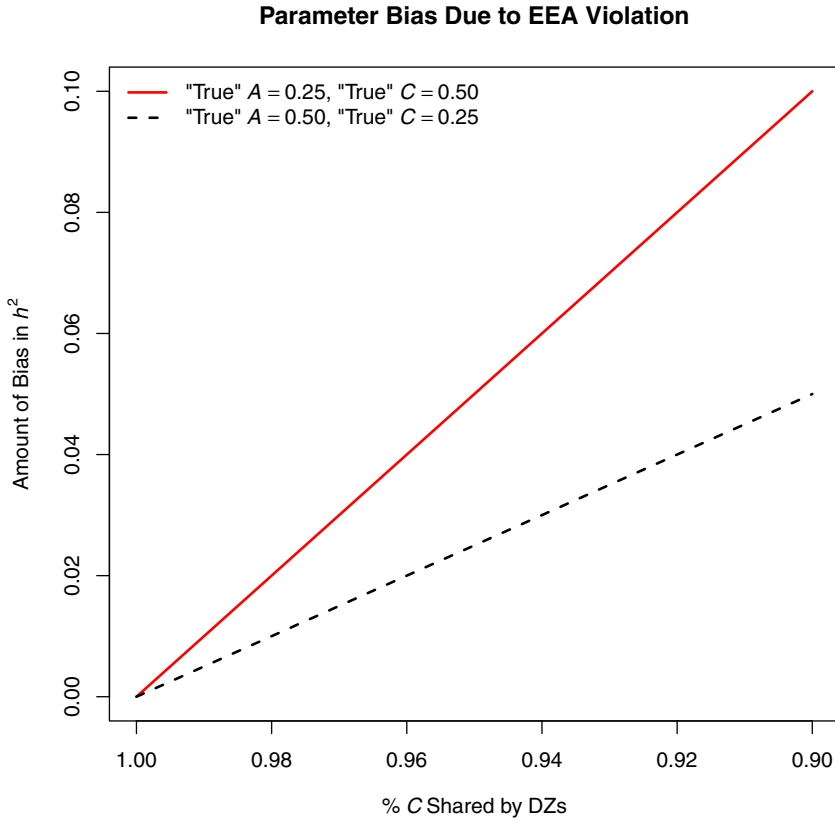
**Table 2. Impact of Violating the Equal Environments Assumption on  $h^2$  Estimates**

Calculation Results	"True" Parameter Values			
	A = .25, C = .50		A = .50, C = .25	
	$h^2$ Estimate	Difference from "True" Parameter	$h^2$ Estimate	Difference from "True" Parameter
% C shared by DZ = 100	.250	.000	.500	.000
% C shared by DZ = 98	.270	.020	.510	.010
% C shared by DZ = 96	.290	.040	.520	.020
% C shared by DZ = 94	.310	.060	.530	.030
% C shared by DZ = 92	.330	.080	.540	.040
% C shared by DZ = 90	.350	.100	.550	.050
Simulation Results	Average $h^2$ Estimate	Average $c^2$ Estimate	Average $h^2$ Estimate	Average $c^2$ Estimate
% C shared by DZ = 100	.246	.502	.496	.252
90 percent range	(.163-.331)	(.419-.582)	(.391-.599)	(.147-.355)
% C shared by DZ = 95	.302	.448	.528	.222
90 percent range	(.206-.397)	(.362-.538)	(.413-.639)	(.118-.333)
% C shared by DZ = 90	.348	.401	.548	.201
90 percent range	(.251-.460)	(.297-.483)	(.439-.676)	(.081-.302)

2003). When differences in  $h^2$  were detected, no discernable pattern emerged in such differences indicating that any potential bias introduced was likely random (Cronk et al., 2002; Eaves et al., 2003; Felson, 2012, 2014). Additionally, in instances where Burt and Simons argued there is an obvious violation of the EEA (i.e., inclusion of opposite-sex DZ twins), the effect on heritability estimates also seems to be random. Simulations and an analysis of all of the existing data on the EEA converge to reveal that when the EEA is violated, estimates are inflated by between 1 and 5 percentage points.

This conclusive set of findings is perhaps one potential explanation for the minimal discussion of the EEA within the criminological literature. The debate has been settled and the EEA is not a sufficient reason to dismiss heritability studies or the estimates gleaned from these studies. Similar to how a large literature stretching back several decades has clearly defined the robustness of the assumptions that accompany simple linear regression and the minimal biasing effects that may result when such assumptions are violated, a similar line of research has demonstrated that the EEA does not result in any systematic bias within  $h^2$  and  $c^2$  estimates. We are certain that by using the vague and subjective inclusion criteria Burt and Simons relied on to accumulate biosocial studies critiqued in their article we could identify hundreds of criminological studies that have actively violated the basic assumptions of linear regression (e.g., the assumption that errors are identically [normally] and independently distributed across the sample space) and have not formally acknowledged such assumptions. It is important to note, however, that all the studies included in the Burt and Simons table (pp. 234-5) referenced scholarship bearing on the classical twin design assumptions and that the authors of those studies have written extensively about these assumptions (Beaver, 2009, 2013). Although Burt and Simons painted

**Figure 2. Heritability Bias at Different Levels of EEA Violation and Different Levels of “True” A and “True” C**



a picture of deception at the hands of biosocial criminologists, the truth is that the EEA is frequently not discussed because it has no consistent or meaningful impact on heritability estimates.

Despite the presence of a cohesive set of empirical findings that span multiple literatures and decades, indicating that the EEA does not systematically bias heritability estimates, Burt and Simons failed to acknowledge the existence of this literature. Rather, they cited sources that did not include analyses capable of backing their claims. In the interest of scholarly transparency, we have brought the full body of literature that empirically assesses the EEA to light. After assessing this body of evidence, our sentiments align closely with those of Carey (2003: 301, emphasis in original)<sup>8</sup>:

Taken together, all these lines of evidence suggest that the equal environments assumption meets the definition of a robust assumption. A *robust assumption* is one that might actually be violated, but the effect of violating the assumption is so small

8. Our reading of the available evidence also aligns with that of Rowe and Osgood (1984: 534) who stated, “Although [the EEA] is often questioned, the empirical evidence is largely supportive.”

that the estimates and substantive conclusions are not altered. For example, Newtonian physics is incorrect, but one can use Newtonian principles to build a bridge or design a skyscraper. In these situations, the assumptions of Newtonian physics are robust even though they are technically wrong.

#### JOINT VIOLATION OF THE RANDOM MATING ASSUMPTION AND THE EEA

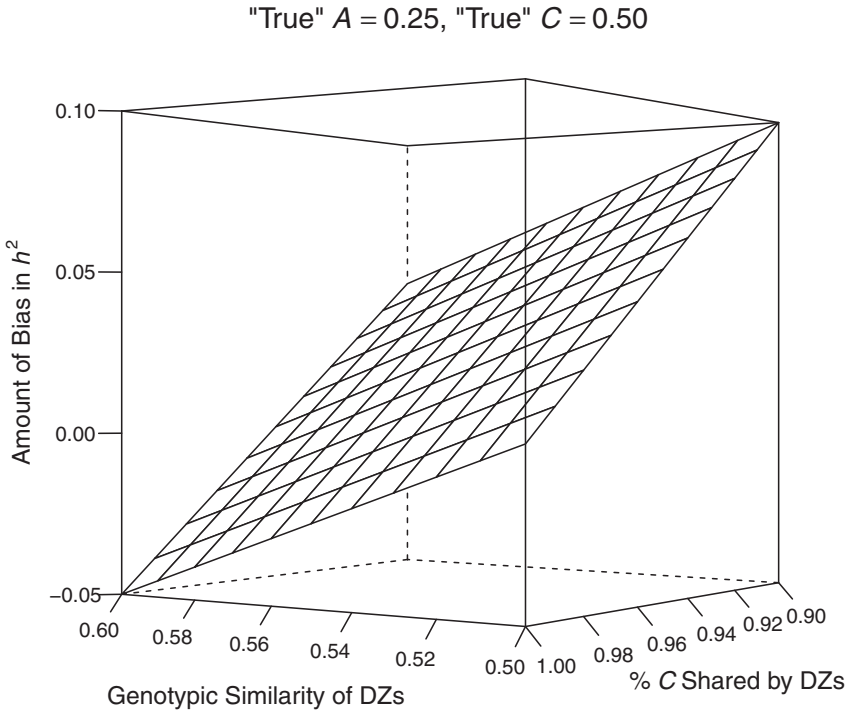
As revealed in the calculations and simulation results presented previously, violations of the random mating assumption and violations of the EEA lead to *opposite-sign effects* on heritability estimates. This result is not surprising. Indeed, it is both intuitive and has been anticipated by scholars for more than two decades. Raine (1993: 58–9) noted:

It is concluded that methodological problems of twin studies are just as likely to decrease heritability estimates as they are to artificially inflate them. Rutter et al. (1990) have suggested that in all probability these effects will tend to cancel each other out. It is important, therefore, not to ignore research findings from twin studies. In spite of its limitations, the study of twins remains a key method in the field of behavior genetics, and future twin studies of crime using up-to-date biometric modeling are likely to make increasingly important contributions to genetic research on crime.

The finding of opposite-sign effects leads to at least two important questions or concerns that warrant close consideration. First is the question of whether one assumption is more likely to be violated than the other, and if so, what are the “downstream” consequences of this reality? Given the evidence presented in this analysis, a conservative answer is one that acknowledges the violation of both assumptions in everyday practice. This leads to a second question or concern. Specifically, if both assumptions are violated in practice, then how much and in which direction are heritability estimates biased? This question is empirical, meaning it can be tested using the calculations above and using the same simulations described previously. To be brief, we augmented the calculations and simulations to include violations of *both* assumptions simultaneously. We estimated the calculations and simulations twice: 1) where “true”  $A = .25$  and “true”  $C = .50$ , and 2) where “true”  $A = .50$  and “true”  $C = .25$ . Figures 3 and 4 reveal the findings from these analyses in three-dimensional space. First, figure 3 shows the joint impact of violating the assumption of random mating and violating the EEA when “true”  $A = .25$  and “true”  $C = .50$ . As was noted, when the two assumptions were analyzed separately, the EEA tends to have a greater biasing impact at lower values of “true”  $A$  and higher values of “true”  $C$  compared with violations of the assumption of random mating under these same conditions. The three-dimensional plot displays this same finding.

The plane presented in the box in figure 3 represents all of the  $h^2$  estimates gleaned from calculations under the different permutations of the two conditions. Note that the plane touches the vertical axis closest to the reader, and this intersection represents the “true”  $A$  value (i.e.,  $h^2 = .25$ ). If we follow the horizontal axis to the right, the axis labeled “%  $C$  Shared by DZs,” we see that  $h^2$  estimates tend to increase as the EEA Bias increases (which is reflected by lower values of  $C$  for DZs). If we follow the horizontal axis to the left, the axis labeled “Genotypic Similarity of DZs,” we see that  $h^2$  estimates tend to decrease as the assumption of random mating is violated. The overestimates of

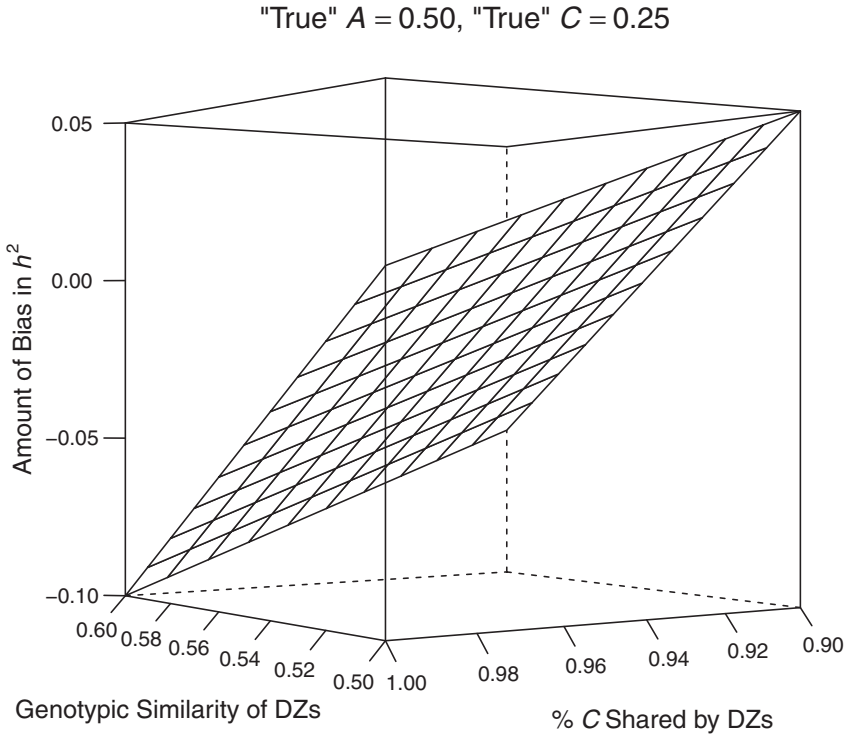
**Figure 3. Three-Dimensional Plot of the Joint Bias of EEA Violation and Random Mating Violation**



$h^2$  that are presented on the right side of the box are more substantial at the right edge of the plane, representing a case where the EEA is violated but the assumption of random mating is *not* violated. As we move toward the center of the plane, we see the  $h^2$  estimates are not as inflated, representing the “cancelling out effect” of violating the assumption of random mating. At the back corner of the box, where the plane touches the vertical axis,  $h^2$  is estimated to be .30 (.05 points higher than the “true”  $A$ ). This point represents the case when *both* assumptions are violated in their most extreme form (at least in terms of the parameters set for this study). Thus, the evidence presented in figure 3 suggests that violations of the EEA, even when the assumption of random mating is violated, may lead to slightly overestimated values of  $h^2$  when “true”  $A$  is in the moderate-to-low range and “true”  $C$  is in the moderate-to-high range.

The same exercise was repeated for the case when “true”  $A = .50$  and “true”  $C = .25$ . The results from this set of calculations are presented in three-dimensional space in figure 4. In a general sense, the results reported in figure 4 mirror those from figure 3. The vertical axis closest to the reader reflects the “true”  $A$  (i.e.,  $h^2 = .50$ ). The horizontal axis to the right reveals the impact of EEA bias, and the horizontal axis to the left reveals the impact of assortative mating (i.e., violating the assumption of random mating). The primary difference between figure 3 and figure 4 is the latter indicates that violations of the assumption of random mating are more consequential for  $h^2$  estimates compared with violations of the EEA. Figure 4 clearly displays a steeper downward slope on the

**Figure 4. Three-Dimensional Plot of the Joint Bias of EEA Violation and Random Mating Violation**



left horizontal axis compared with the upward slope on the right horizontal axis. Finally, the point at which the plane touches the vertical axis in the far corner of the box shows the impact of violating both assumptions in their most extreme form (at least in terms of the parameters set for this study). Here, the  $h^2$  estimate ( $h^2 = .45$ ) is lower than the “true”  $A$  by .05 points. Thus, when “true”  $A$  is in the moderate-to-high range and “true”  $C$  is in the moderate-to-low range, the cumulative effect of jointly violating the EEA and the assumption of random mating is that  $h^2$  will tend to be *underestimated*.

**Conclusion:** *When the assumptions of random mating and the EEA are considered in tandem, calculation and simulation results reveal that violations of one assumption tend to counterbalance violations of the other. Based on these results, we cannot conclude that violations of the EEA will overstate heritability because violations of the assumption of random mating lead to underestimates. Scholars should, therefore, be skeptical of the unilateral dismissal of heritability studies by Burt and Simons because of, among other reasons, their lack of a discussion of the assumption of random mating. Our conclusions therefore reinforce the comments of Rowe and Osgood (1984: 537) who stated that, “Although individual studies can be faulted on one ground or another, the overall pattern of results is so regular that to ignore genetic factors requires either outlandish assumptions or a very selective reading of the literature.”*

## CERTAIN ASSUMPTIONS NEED NOT APPLY: ALTERNATIVE METHODS OF ESTIMATING $h^2$

The Burt and Simons claim that EEA violations bias  $h^2$  estimates was dramatically overstated, especially in light of random mating assumption violations and the finding that the level of bias depends on the values of the “true” parameters. The available empirical literature, along with the results from our calculations and simulations, support the current position of the classical twin design representing a robust method. Nonetheless, staunch critics may still argue that the EEA limits the viability of the classical twin design. As a result, we offer one final point. Allegations of an inflated heritability estimate in twin studies can be assessed using methodologies that *do not* rely on the assumptions of the twin method. If these methodologies produce results that align closely with those produced by the classical twin design, then previous concerns over violations of the EEA and violations of the assumption of random mating can be alleviated. One such alternative method for generating heritability estimates for phenotypic variance is known as genome-wide complex trait analysis (GCTA; Yang et al., 2011). In an oversimplified statement, GCTA “scans” the entire genome for every individual in the sample (sample sizes often number in the thousands) and runs thousands of statistical analyses (correcting the  $p$  values for multiple testing bias). Next, the association between the measured genes and the total trait variance can be used to estimate the portion of variance in the trait that is a result of genetic factors. Importantly, GCTA does not employ kinship pairs as it exploits chance genetic similarity in single-nucleotide polymorphisms (SNPs) between unrelated individuals to calculate heritability estimates for phenotypic variance (Plomin, DeFries, et al., 2013; Yang et al., 2010). Given that GCTA does not employ the types of kinship pairs that prevail in the classical twin design, the assumptions of the classical twin design, including the EEA, need not apply.<sup>9</sup>

Because GCTA is a cutting-edge technique, the number of published studies employing the method to estimate the heritability of various phenotypes is limited. In the first study using GCTA, Yang et al. (2010) generated an estimate of the proportion of variance in height explained by more than 300,000 SNPs garnered from more than 3,900 *unrelated* respondents. The results of the GCTA analysis indicated that these 300,000+ SNPs explained 45 percent of the variance in height. To be clear, unlike twin-based studies, the GCTA assesses the collective additive influence of the measured SNPs to estimate heritability. In this way, the heritability estimates gleaned from a GCTA can be compared with those provided by twin-based studies to establish the amount of “missing heritability” between the two methods.<sup>10</sup> Thus, the 45 percent figure represents the amount of variance in height that is caused by the SNPs included in the GCTA. Of crucial importance is a comment about the study made by the authors in a subsequent publication.

---

9. Except, of course, the assumption that non-kin doppelgangers will only be as similar as their genetic profiles (see footnote 6).

10. Therefore, unlike what Burt and Simons claimed, researchers in behavioral and molecular genetics have such strong confidence in the results of twin-based studies that they create novel methods to assess the difference between the broad heritability estimates produced by twin-based studies and the minimal effects observed in genome-wide association studies (GWAS) and single-gene association studies. In other words, far from calling for an end to twin-based studies that produce estimates of heritability, these experts employed such estimates as a *benchmark* from which to assess sophisticated molecular genetic analyses.

Visscher, Yang, and Goddard (2010) provided their rationale for excluding close relatives by indicating their desire to ensure that any observed effect was a result of genetic factors rather than of nongenetic factors captured by the shared environment. The authors noted, however, that “leaving these few pairs [of close relatives] in or out made very little difference to the results” (Visscher, Yang, and Goddard, 2010: 521). Consequently, the GCTA study by Yang et al. (2010) supports two conclusions of twin-based research on the variance in height: 1) Variance is primarily caused by genetic effects and 2) shared environmental effects have little influence on the variance. Although this study is limited to assessing variance in height, it illustrates that the twin-based method is not hopelessly biased as there is convergence in illustrating the effect of genetic factors.

GCTA also has been employed in assessments of schizophrenia (Lee et al., 2012) and other psychiatric disorders (Lee et al., 2013), components of personality (Vinkhuyzen et al., 2012), behavioral inhibition (McGue et al., 2013), and intelligence in both childhood (Benyamin et al., 2013) and adulthood (Davies et al., 2011). The findings across these studies illustrate that variance in the phenotypes under examination is influenced by variance in the SNPs included in the analyses. As noted, these types of studies are not susceptible to the assumptions of the twin-based research; yet they produce estimates of  $h^2$  that often are within the confidence intervals produced by the classical twin design. In an effort to assess directly the relationship between GCTA studies and the twin-based methodology, Plomin, Haworth, et al. (2013) compared estimates of heritability produced by the two methodologies for a variety of phenotypes (weight, height, general cognitive ability, nonverbal cognitive ability, verbal cognitive ability, and language ability) within the same sample. The results of the analyses indicated that at least 40 percent of the  $h^2$  estimates gleaned from the classical twin design were accounted for by the genetic effects captured by the GCTA (see table 1 in Plomin, Haworth, et al., 2013). In terms of general cognitive ability, the authors found that approximately two thirds of the  $h^2$  estimate produced by the classical twin design also was identified by the GCTA. Overall, the study provided strong support for the methodologies employed in twin-based studies. Indeed, the authors (p. 566) concluded by noting:

Although GCTA analysis and other DNA-based methods are exciting additions to behavioral genetic research, we suggest that traditional quantitative-genetic methods, such as twin studies and adoption studies, will continue to make important contributions to understanding how genotypes become phenotypes, in part because twin and adoption studies are as much studies of environmental influence as they are of genetic influence.

Specific to the study of criminal behavior, an examination by Bentley and colleagues (2013) employed a technique similar to GCTA called “candidate systems of genes (CSG)” (p. 1074). Using data from the Nurse Family Partnership Program, Bentley and colleagues examined the variance in a latent antisocial variable that was a function of variance in a latent genetic variable (based on eight different polymorphisms). The results of the analyses indicated that the latent genetic variable was significantly predictive of antisocial behavior ( $\beta = .413, p = .006$ ) and that the latent genetic variable explained 16 percent of the variance in antisocial behavior. When compared with genome-wide association studies on antisocial behaviors where single SNPs have rarely reached statistical significance (e.g., Tielbeek et al., 2012), these results are remarkable. Selecting



just eight polymorphisms (out of a potential pool numbering in the thousands), the authors were able to account for a considerable proportion of the variance in antisocial behavior.

**Conclusion:** *When multiple divergent methods converge to illustrate a consistent finding, it is logically reasonable and empirically sound to accept as valid the results of the divergent methods. The primary reason for such a conclusion is the nonoverlapping assumptions underlying the various methods. This dynamic is illustrated in the convergence in findings between classical twin design studies and various molecular genetics methodologies. GCTA and CSG studies are examples of methodologies that are not subject to the same assumptions as classical twin designs and yet provide convergence in terms of the differential influence of genetic and environmental factors on a variety of behavioral phenotypes, including antisocial behavior. In direct contrast to the appraisal provided by Burt and Simons, the most cutting-edge research emanating from molecular genetics relies heavily on the findings of classical twin design studies and is continually providing empirical support for the validity of such studies.*

## SOCIAL CONSTRUCTION OF BIOLOGICAL REALITY

We have shown empirically that violations of the assumptions of behavioral genetics studies do not invalidate heritability estimates. This is not a matter of opinion but a matter of mathematical evidence. Under certain conditions, our calculations and simulations revealed that heritability estimates will be slightly upwardly biased (probably no more than 5–10 percentage points). Under other conditions, heritability estimates will be downwardly biased (probably no more than 5–10 percentage points). Under the most likely condition, where multiple violations occur simultaneously, the biasing influences of assumption violations wash out, with upwardly biasing factors canceling downwardly biasing factors. Moreover, the overall pattern of findings flowing from the 61 studies examining the EEA revealed the same conclusions offered by our calculations and simulation data. Needless to say, no “fatal flaw” in behavioral genetic methodologies or assumptions has been discovered and the conclusion that “all of these models are biased toward *inflating heritability and underestimating shared environmental influences*” (Burt and Simons, 2014: 226, emphasis in original) is unequivocally incorrect.

We turn now to secondary critiques of the Burt and Simons article. Next, we show where Burt and Simons presented a misleading portrait of the behavioral genetics literature. We examine their oversights, beginning with their misrepresentation of the literature and their misunderstanding of the value of heritability in modern behavioral genetics research (i.e., epigenetics). Although less important than the presented mathematical evidence validating heritability studies, these problems highlight the various distortions of evidence and prior scholarly work endemic in the Burt and Simons article.

## SELECTIVE CITING AND MISREPRESENTING RESEARCH

The questions raised against heritability studies by Burt and Simons are empirical issues that can be addressed with available data. Burt and Simons, however, did not empirically assess the veracity of their criticisms. Instead, and as we have already shown, Burt and Simons supported their criticisms by selectively citing from the literature, pulling snippets of information from a study that seem to support their beliefs, or citing heavily from

scholarship that was dismantled long ago in other fields of study, practices that some may refer to as the “social construction of reality.”

For example, the selection criteria for the studies included in table 1 in the Burt and Simons (pp. 234–5) article were highly subjective. After a closer examination, we identified three primary issues with their table and the search criteria used to create it. First, when we attempted to replicate the vague search criteria offered by Burt and Simons, Google Scholar revealed dozens of relevant studies, all of which were excluded from their table 1. Unsurprisingly, many of the omitted studies do not necessarily conform to the primary thrust of the Burt and Simons article (e.g., Barnes, Beaver, and Boutwell, 2011; Beaver et al., 2008) and were not identified by Burt and Simons. For example, despite identifying numerous studies conducted by Beaver et al., they overlooked at least two of his studies that directly estimated the heritability of crime and/or delinquency and that included multiple measures of the environment (Beaver, 2011a; Beaver, DeLisi, et al., 2009). One such paper, for example, directly examined gene–environment interactions with 13 environmental measures drawn from the Add Health data (Beaver, 2011a).<sup>11</sup> In addition, at least two omitted studies were published in *Criminology* (Barnes, Beaver, and Boutwell, 2011; Beaver et al., 2008). Second, Burt and Simons gave no empirical reason why they used a 2008 cut-off. Had they searched *Criminology* prior to 2008, they would have had to include additional studies, studies that also found heritability estimates on delinquency of .50 (e.g., Haynie and McHugh, 2003).<sup>12</sup> Third, some of the information included in their table 1 is highly misleading or even completely incorrect. More specifically, in their discussion of studies that produce heritability estimates, Burt and Simons said the following: “[These studies] compare individual phenotypes across varying degrees of genetic relationships and use these comparisons to estimate genetic and environmental influences *without actually measuring either*” (p. 229, emphasis in original). We point out that 50 percent of the studies they chose to include in the table actually did measure an environmental variable.

In addition to arguing against the replication of twin studies, Burt and Simons singled out the Add Health data set as being problematic because it is used in most heritability studies by biosocial criminologists. They pointed out that the entire twin sample nested within the larger probability sample includes only 289 MZ twin pairs and 452 DZ twin pairs for a total of  $N = 1,482$  twins. We are somewhat perplexed by this as Simons has built his career on the FACHS data that include a little more than 800 respondents. In terms of sample size, the twin subsample of the Add Health data dovetails nicely with the FACHS data. We should further note that it is not uncommon for entire perspectives, theories, and disciplines (at certain times) to be guided by only one data set. For example, Farrington’s work in developmental/life-course research has been based on data from a little more than 400 males from England (Piquero, Farrington, and Blumstein, 2007), Sampson

- 
11. In addition, Burt and Simons failed to acknowledge a special issue in *Journal of Criminal Justice* that includes multiple heritability studies and that was available online in early 2013 (Tuvblad and Beaver, 2013).
  12. Although Haynie and McHugh (2003) claimed to have completely accounted for heritability in subsequent models, these subsequent models were estimated incorrectly. As a result, the conclusion of this study should have been that the heritability of delinquency is approximately .50, not .00, and that the loss of statistical significance was caused by inflation in the standard error relative to the drop in coefficient resulting from the incorrect application of the DeFries–Fulker analysis.

and Laub's (1993) life-course work has been based on 1,000 White males from Boston, and the National Youth Survey guided much of what was known about criminological theory for 10–15 years. To our knowledge, there has not been any serious attempt to limit publications on these samples or to isolate the scholars who have used such data. Burt and Simons continued to disparage the Add Health data by pointing out that the measurement of key constructs is less than ideal. Perhaps this is partially true, but fallibility in the measures would simply deflate heritability estimates because measurement error is captured by the nonshared environmental estimate (i.e.,  $e^2$ ). What makes these criticisms all the more surprising is that Simons has pointed out in no less than five separate publications that the Add Health data represent an ideal data set to examine genotypic influences on social behaviors (Simons, Beach, and Barr, 2012; Simons and Lei, 2013; Simons et al., 2011, 2012, 2013).

Beyond the omission of relevant research and curious selective derision of the Add Health data, Burt and Simons also selectively quoted scholars, resulting in an overall distortion of such scholars' original intentions. This is particularly salient when Burt and Simons quoted Moffitt's (2005) seminal review. Burt and Simons (p. 226, emphasis in original) wrote:

Although evidence from the different methods are used to “provide convergent findings,” given that “each of the primary designs used by behavioral geneticists has its own Achilles heel(s)” (Moffitt, 2005: 57), we show that all of these models are biased toward *inflating heritability* and *underestimating shared environmental influences*.”

Burt and Simons used Moffitt's words to make the argument that heritability estimates are biased and, as a result, should be abandoned. However, when reading Moffitt's words from the original article, we see that the *actual* quote was (Moffitt, 2005: 57):

A fundamental assumption guiding this review is that sturdy inferences ought to be drawn from a cumulative body of studies whose methods differ as much as possible, but provide convergent findings about the same construct. As we have seen, each of the primary designs used by behavioral geneticists has its own Achilles heel(s), but fortunately, each design's idiosyncratic flaws are offset by compensatory strengths of the other designs. As a consequence, although particular studies and particular designs may be subject to critique, this does not invalidate inferences derived from the entire cumulative evidence base.

A second example of misrepresentation of previous scholarship by Burt and Simons is evident when they argued that biosocial criminologists frequently ignore the assumptions of the classical twin design, particularly the EEA. To reinforce this, they quoted Beaver (2011b: 86 [sic]) as saying “the only reason that MZ twins should be [sic] more similar than DZ twin pairs [sic] is because they share twice as much genetic material” (Burt and Simons: 236). Clearly, Burt and Simons are highlighting that Beaver (2011b), and other biosocial criminologists, strategically failed to inform readers about twin-based assumptions that, if violated, could be causing MZ twins to be more similar to each other than DZ twins. Again, however, if we go to the original Beaver (2011b: 87) article, we see that the full, unedited quote reads: “As a result, *if the assumptions of twin-based research are met*, the only reason that MZ twins should be phenotypically more similar than DZ twins is because they share twice as much genetic material” (emphasis added). Burt and Simons

therefore edited the Beaver (2011b) quote and misrepresented his words in such a way as to provide support for their argument. When the actual unedited quote from Beaver (2011b) is read, the statement flatly contradicts their claim that biosocial criminologists ignore the assumptions of twin-based research.

Where Burt and Simons selectively cited some studies and incorrectly cited and quoted others, they also relied heavily on highly questionable sources. They cited Joseph (1998, 2001, 2004, 2006, and 2010), for example, an amazing *70 times* in their article and online supporting information. This averages out to one citation of Joseph per page. Relying so heavily on a single source makes it difficult to see how Burt and Simons introduced anything that Joseph had not already discussed. Moreover, Burt and Simons cited as evidence an unpublished manuscript (Suhay and Kalmoe, 2010) and a newsletter (Richardson, 2011) from a website constructed and maintained by individuals politically opposed to gene-behavior research. Their selection of just a few sources of information becomes more suspect when juxtaposed against appendix D in the online supporting information, which indicates that Burt and Simons could have easily selected from the 60+ scholarly studies that empirically assessed the EEA.

As we have shown, the call by Burt and Simons for a biosocial criminology “rooted in reality” rests on a social construction of reality (p. 14). The oversights, misrepresentations, anecdotes, distortions, and misquotes paint a carnival-mirror-like picture of heritability studies, their uses, and their findings. Studies that would complicate their narrative were left out of their discussion. Scholars that support their narrative were given tremendous weight. Quotes taken from others were highly edited, even to the point that a reading of the original quote goes on to provide conclusions directly counter to their position.

#### FAILURE TO APPRECIATE THE BENEFITS OF HERITABILITY STUDIES FOR MODERN BEHAVIORAL RESEARCH

It was not until biosocial criminologists challenged the discipline to take seriously genetic and biological influences that the sociological tide turned to a more integrated focus (Beaver and Wright, 2013). With the use of behavioral genetic approaches, including heritability studies, biosocial criminologists brought empirical evidence to a field uninitiated in the use of twin and extended-family designs. Most criminologists have simply ignored the findings flowing from such studies, but notable exceptions exist (Benson, 2013; Cullen, 2011). This lack of acknowledgment, occurring against a longstanding backdrop of disciplinary bias, likely encourages criminologists to accept the Burt and Simons assertions despite the evidence we provide (Walsh and Ellis, 2004). Before scholars make this jump, however, we encourage readers to consider the following benefits of heritability studies.

First, heritability studies are the “first step” in a long march toward a true biosocial criminology, a criminology that is disinterested in whether biology or the environment gets the credit for causing crime but instead cares about getting the puzzle pieced together correctly. Biosocial criminologists have published an array of studies on gene  $\times$  environment interactions (GxEs) and gene–environment correlations (*r*GEs) (Barnes, Beaver, and Boutwell, 2013; Beaver, Wright, and DeLisi, 2008; Schwartz and Beaver, 2011), they have published studies using extended family designs with alternative behavioral genetic assessments (Beaver et al., 2009; Connolly and Beaver, 2014), and they have published studies using molecular markers in search of direct and interactive effects (Beaver et al., 2013; Schwartz and Beaver, 2014; Wright et al., 2012). Moreover,

they have written books on the interlocking qualities of nature and nurture over the life course (Benson, 2013; Wright, Tibbetts, and Daigle, 2008). Heritability studies thus represent only one part of the arsenal of biosocial criminologists, an arsenal that is growing in size and complexity.

Second, all statistical models have evolved and all have assumptions that are, more or less, meaningful if violated. This applies equally to social statistics, which has undergone evolution over time and is subject to a range of important assumptions. Hypothesis testing, for example, encouraged the development of the general linear model, which morphed into truncated variation models, which helped form the basis for complex trajectory and other latent grouping analyses. Unsatisfied with the statistical limits of each approach, scholars developed new approaches, such as hierarchical linear modeling and latent class analysis. Not only did scholars develop each approach, we note, but also they empirically tested what happens when each underlying statistical assumption is violated. Ordinary least-squares (OLS) regression is undoubtedly the most used statistical technique in the social sciences. The assumptions of OLS are known and have been tested thoroughly. They are so well known that almost no scholar now discusses the assumptions of OLS in their studies, even though most scholars continually violate those assumptions, such as homoskedasticity and normally distributed errors. Should OLS regression be abandoned? Of course not.

In a similar way, ACE models have evolved over time. They now can be used to examine the genetic and environmental covariation between traits (Loehlin, 1996) and can test for sex differences (Boisvert et al., 2013). Highly complex behavioral genetic designs, including genetic growth curves (McArdle and Plassman, 2009), Bayesian integration to assess GxEs (Eaves, Foley, and Silberg, 2003), and “children of twins” designs (D’Onofrio et al., 2007) emerged out of basic heritability studies. Our point is as follows: Scholars should not abandon research or research methods; instead, they should work to revise their methodologies and statistical models to address known problems.

Third, heritability studies and their partitioning of variance into genetic and environmental sources have led scholars to a better, more nuanced understanding of *environmental* factors. For example, family processes and parenting behaviors tied to offspring conduct are clearly tangled in the complex web of biology and environment (Harris, 1998). As McGue (2010) and others (Pinker, 2002) have noted, by ignoring biological influences scholars were led to erroneous conclusions about the effects of parents and families on crime—conclusions that brought harm to the lives of parents and children. Scholars were so locked into their standard social science paradigm, however, that it took the work of Rowe (1994) and subsequently Harris (1998) to show how the elements of the nonshared environment were important to understanding why some children were influenced by family processes while other children in the same household were not. Heritability studies provided these insights, and they led to more refined studies into parenting and families (Beaver, 2008; Harris, 1998; Rowe, 1994; Wright and Beaver, 2005). If that example is insufficient, then consider that heritability studies provided the first evidence that drug addiction and alcoholism were not the products of “bad morals” but of genetically influenced sensitivity to substances. Heritability studies also provided the first evidence that ADHD, conduct disorder, obsessive-compulsive disorder, autism spectrum disorders, and other psychiatric problems were not caused by faulty environments or poor mothering but instead were strongly influenced by genetics. We could go on, but the point should be clear: Behavioral genetics research has led to important

discoveries and to a more refined understanding of the types of environmental factors that matter, the types that do not, and the types that matter for some people or in some conditions but not in others. In short, lives have been improved by the results of behavioral genetic studies. They have humanized domains of behavior, have led to more effective social and pharmaceutical interventions, and have been instrumental in destigmatizing complex social behaviors like homosexuality.<sup>13</sup>

Yet, Burt and Simons criticized the classical twin design by drawing on examples where the logic of classical twin studies seems to break down. For example, they tell us that “eyedness” shows the limited utility of heritability. They argue that our genome equips all humans with two eyes and yet heritability estimates of “eyedness” based on standard twin equations would be zero (.00). This is correct. Heritability estimates would be zero (.00). However, counting eyes does not amount to quantifying and explaining *variance*; in fact, in their example, “eyedness” is a constant and mathematical constants cannot be explained by variables, including genetic variables, regardless of the methodological design being used.<sup>14</sup> The example is logically flawed.

They also argued that the partitioning of variance into genetic and environmental sources is nearly impossible because of the constant interaction between genes and the environment. Both Plomin, DeFries, et al. (2013) and Harris (2006) have responded to this criticism and have dismantled the often-cited example of trying to quantify the area of a rectangle (phenotype) by discussing the relative contributions of the width (genes) and the height (environment). Clearly, the area of a rectangle is the product of width and height. As Plomin, DeFries, et al. (2013: 89–92) noted, though, “if we ask not about a single rectangle but about a population of rectangles, the variance in areas could be due entirely to length, entirely to width, or both.” In the “damned rectangle” example, as Harris (2006) called it, width and height (similar to genes and environment) are being applied to a single rectangle. In reality, behavior geneticists examine samples containing numerous “rectangles” of many different sizes. When viewed in this way, it makes sense to quantify variance in the area of rectangles based on differences in length and height. What this necessarily means is that for a single individual, his or her genes and environment in interaction contribute to his or her phenotypic score, but when examining phenotypic variance in a sample of individuals, it is possible to partition variance into genetic and environmental components.

Instead of partitioning variance, biosocial criminologists, according to Burt and Simons, should explore GxEs, epigenetic processes, and social neuroscience. The problem, however, is that much of the GxE literature is under heavy criticism and epigenetics is in its infancy. Unfortunately, Burt and Simons couched their discussion of epigenetics and GxEs in language that makes it seem as if this body of research is generally accepted and easily

---

13. We thank Steven Pinker (2002, and personal communication) for pointing out that behavioral genetic research helped to remove the blame from parents for every “pathology” and helped to dispel archaic ideas such as the belief that homosexuality is a contagious choice.

14. Unless, of course, the sample included Oedipus Rex and Moshe Dayan (thanks to Steven Pinker [personal communication] for this tongue-in-cheek example). In this case, eyedness would vary as a result of environmental factors. The most important point is that *heritability* explains variance, not innateness or species-wide traits. Both variance and species-wide traits can be genetically influenced, but only the former is addressed by the classical twin design and behavior geneticists.

replicated by neuroscientists, geneticists, and epigeneticists. In reality, nothing could be further from the truth. Yet, Burt and Simons followed the lead of many before them who, as Smith (2011: 539) stated, use epigenetics as “the currently fashionable response to any question to which you do not know the answer.” Contrary to the picture painted by Burt and Simons, epigeneticists are urging social scientists to be more cautious when discussing epigenetic influences on social behavior. In the words of two preeminent epigeneticists, Heijmans and Mill (2012: 4): “[E]pigenetics will not be able to deliver the miracles it is sometimes claimed it will.” Perhaps unknown to sociologists who have hung their future of the field on epigenetics, epigeneticists are confronted with the same problems genomic and biosocial scientists are encountering. In addition, the GxE literature has been plagued by failures to replicate, especially for novel GxEs (for an overview of the literature, see Duncan, Pollastri, and Smoller, 2014), with some recent studies indicating that greater than 90 percent of detected GxE effects are likely false positives (Duncan and Keller, 2011). As a result, some top journals have adopted screening criteria for GxE studies that require replication of a novel GxE before the paper is considered for publication (Hewitt, 2012).

These points draw attention to one final issue worth considering. Specifically, we argue that heritability studies are not biased and that scholars reconsider the call by Burt and Simons for an “end” to heritability studies. There is still much to be gained from heritability studies and the classical twin design. For instance, recent heritability studies have shown that genetic factors underlie the etiology of criminological variables that may have otherwise been assumed to be purely social in origin (e.g., Beaver, 2011b). Additionally, twin studies provide an avenue by which scholars can more accurately estimate the impact of environmental factors on antisocial behavior. Twin studies can be used to control for genetic influences so that the impact of an environmental variable on antisocial behavior can be analyzed without the confounding influence of genetic factors (e.g., Burt et al., 2010). Thus, the value of the classical twin study has not depreciated.

## REFERENCES

- Alford, John R., Peter K. Hatemi, John R. Hibbing, Nicholas G. Martin, and Lindon J. Eaves. 2011. The politics of mate choice. *The Journal of Politics* 73:362–79.
- Allison, David B., Jaakko Kaprio, Maarit Korkeila, Markku Koskenvuo, Michael C. Neale, and Kazuo Hayakawa. 1996. The heritability of body mass index among an international sample of monozygotic twins reared apart. *International Journal of Obesity* 20:501–6.
- Barnes, J. C., Kevin M. Beaver, and Brian B. Boutwell. 2011. Examining the genetic underpinnings to Moffitt’s developmental taxonomy: A behavioral genetics analysis. *Criminology* 49:923–54.
- Barnes, J. C., Kevin M. Beaver, and Brian B. Boutwell. 2013. A functional polymorphism in a serotonin transporter gene (5-HTTLPR) interacts with 9/11 to predict gun-carrying behavior. *PLoS ONE* 8:e70807.
- Beaver, Kevin M. 2008. Nonshared environmental influences on adolescent delinquency involvement and adult criminal behavior. *Criminology* 46:341–69.
- Beaver, Kevin M. 2009. *Biosocial Criminology: A Primer*. Dubuque, IA: Kendall-Hunt.
- Beaver, Kevin M. 2011a. Environmental moderators of genetic influences on adolescent delinquent involvement and victimization. *Journal of Adolescent Research* 26:84–114.



- Beaver, Kevin M. 2011b. The effects of genetics, the environment, and low self-control on perceptions of maternal and paternal socialization: Results from a longitudinal sample of twins. *Journal of Quantitative Criminology* 27:85–105.
- Beaver, Kevin M. 2013. *Biosocial Criminology: A Primer*, 2nd ed. Dubuque, IA: Kendall-Hunt.
- Beaver, Kevin M., Matt DeLisi, Michael G. Vaughn, John Paul Wright, and Brian B. Boutwell. 2008. The relationship between self-control and language: Evidence of a shared etiological pathway. *Criminology* 46:939–70.
- Beaver, Kevin M., Matt DeLisi, John Paul Wright, and Michael G. Vaughn. 2009. Gene-environment interplay and delinquent involvement: Evidence of direct, indirect, and interactive effects. *Journal of Adolescent Research* 24:147–68.
- Beaver, Kevin M., J. Eagle Schutt, Brian B. Boutwell, Marie Ratchford, Kathleen Roberts, and J. C. Barnes. 2009. Genetic and environmental influences on levels of self-control and delinquent peer affiliation: Results from a longitudinal sample of adolescent twins. *Criminal Justice and Behavior* 36:41–60.
- Beaver, Kevin M., and John Paul Wright. 2013. Reflections on the ascendancy of biosocial criminology. *ACJS Today* 38:1–11.
- Beaver, Kevin M., John Paul Wright, Brian B. Boutwell, J. C. Barnes, Matt DeLisi, and Michael G. Vaughn. 2013. Exploring the association between the 2-repeat allele of the MAOA gene promoter polymorphism and psychopathic personality traits, arrests, incarceration, and lifetime antisocial behavior. *Personality and Individual Differences* 54:164–8.
- Beaver, Kevin M., John Paul Wright, and Matt DeLisi. 2008. Delinquent peer group formation: Evidence of a gene x environment correlation. *The Journal of Genetic Psychology* 169:227–44.
- Beckwith, Jon, and Corey A. Morris. 2008. Twin studies of political behavior: Untenable assumptions? *Perspectives on Politics* 6:785–91.
- Benson, Michael L. 2013. *Crime and the Life Course: An Introduction*, 2nd ed. New York: Roxbury.
- Bentley, Mary Jane, Haiqun Lin, Thomas V. Fernandez, Maria Lee, Carolyn M. Yrigollen, Andrew J. Pakstis, Lillya Katsovich, David L. Olds, Elena L. Grigorenko, and James F. Leckman. 2013. Gene variants associated with antisocial behavior: A latent variable approach. *Journal of Child Psychology and Psychiatry* 54:1074–85.
- Benyamin, Beben, Beate St Pourcain, Oliver S. P. Davis, Gail Davies, Narelle K. Hansell, Marie-Jo A. Brion, Robert M. Kirkpartick, et al. 2013. Childhood intelligence is heritable, highly polygenic and associated with *FNBP1L*. *Molecular Psychiatry* 19:253–8.
- Boisvert, Danielle, John Paul Wright, Valerie Knopik, and Jamie Vaske. 2013. A twin study of sex differences in self-control. *Justice Quarterly* 30:529–59.
- Borkenau, Peter, Rainer Riemann, Alois Angleitner, and Frank M. Spinath. 2002. Similarity of childhood experiences and personality resemblance in monozygotic and dizygotic twins: A test of the equal environments assumption. *Personality and Individual Differences* 33:261–9.
- Bouchard, Thomas J., Jr., David T. Lykken, Matthew McGue, Nancy L. Segal, and Auke Tellegen. 1990. Sources of human psychological differences: The Minnesota Study of Twins Reared Apart. *Science* 250:223–8.
- Boutwell, Brian B., and Kevin M. Beaver. 2010. The intergenerational transmission of low self-control. *Journal of Research in Crime and Delinquency* 47:174–209.

- Boutwell, Brian B., Kevin M. Beaver, and J. C. Barnes. 2012. More alike than different: Assortative mating and antisocial propensity in adulthood. *Criminal Justice and Behavior* 39:1240–54.
- Box, George E. P., and Norman R. Draper. 1987. *Empirical Model-Building and Response Surfaces*. New York: Wiley.
- Bulik, Cynthia M., Patrick F. Sullivan, and Kenneth S. Kendler. 1998. Heritability of binge-eating and broadly defined bulimia nervosa. *Biological Psychiatry* 44:1210–8.
- Burt, Callie H., and Ronald L. Simons. 2014. Pulling back the curtain on heritability studies: Biosocial criminology in the postgenomic era. *Criminology* 52:223–62.
- Burt, S. Alexandra, Brent Donnellan, Mikhila N. Humbad, Brian M. Hicks, Matt McGue, and William G. Iacono. 2010. Does marriage inhibit antisocial behavior? An examination of selection vs. causation via a longitudinal twin design. *Archives of General Psychiatry* 67:1309–15.
- Capaldi, Deborah M., Hyoun K. Kim, and Lee D. Owen. 2008. Romantic partners' influence on men's likelihood of arrest in early adulthood. *Criminology* 46:267–99.
- Carey, Gregory. 2003. *Human Genetics for the Social Sciences*. Thousand Oaks, CA: Sage.
- Charney, Evan. 2008. Politics, genetics, and “greedy reductionism.” *Perspectives on Politics* 6:337–43.
- Collins, Randall. 1985. *Sociology of Marriage and the Family: Gender, Love, and Property*. Chicago, IL: Nelson-Hall.
- Conley, Dalton, Emily Rauscher, Christopher Dawes, Patrik K. E. Magnusson, and Mark L. Siegal. 2013. Heritability and the equal environments assumption: Evidence from multiple samples of misclassified twins. *Behavior Genetics* 43:415–26.
- Connolly, Eric J., and Kevin M. Beaver. 2014. Examining the genetic and environmental influences on self-control and delinquency: Results from a genetically informative analysis of sibling pairs. *Journal of Interpersonal Violence* 29:707–35.
- Cronk, Nikole J., Wendy S. Slutske, Pamela A.F. Madden, Kathleen K. Bucholz, Wendy Reich, and Andrew C. Heath. 2002. Emotional and behavioral problems among female twins: An evaluation of the equal environments assumption. *Journal of the American Academy of Child & Adolescent Psychiatry* 41:829–37.
- Cullen, Francis T. 2011. Beyond adolescence-limited criminology: Choosing our future—The American Society of Criminology 2010 Sutherland Address. *Criminology* 49:287–330.
- Davies, Gail, Albert Tenesa, Anthony Payton, Jian Yang, Sarah E. Harris, David Liewald, X. Ke, et al. 2011. Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular Psychiatry* 16:996–1005.
- Derks, Eske M., Conor V. Dolan, and Dorret I. Boomsma. 2006. A test of the equal environment assumption (EEA) in multivariate twin studies. *Twin Research and Human Genetics* 9:403–11.
- Domingue, Benjamin W., Jason Fletcher, Dalton Conley, and Jason D. Boardman. 2014. Genetic and educational assortative mating among U.S. adults. *Proceedings of the National Academy of Sciences*. E-pub ahead of print.
- D’Onofrio, Brian M., Eric Turkheimer, Robert E. Emery, Hermine H. Maes, Judy Silberg, and Lindon J. Eaves. 2007. A children of twins of parental divorce and offspring psychopathology. *Journal of Child Psychology and Psychiatry* 48:667–75.

- Duncan, Laramie E., and Matthew C. Keller. 2011. A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *The American Journal of Psychiatry* 168:1041–9.
- Duncan, Laramie E., Alisha R. Pollastri, and Jordan W. Smoller. 2014. Mind the gap: Why many geneticists and psychological scientists have discrepant views about gene-environment interaction (GxE) research. *American Psychologist* 69:249–68.
- Eaves, Lindon, Debra Foley, and Judy Silberg. 2003. Has the “equal environments” assumption been tested in twin studies? *Twin Research* 6:486–9.
- Felson, Jacob. 2012. *What Can We Learn From the Twin Studies? An Evaluation of the Equal Environments Assumption*. New York: Mellen Press.
- Felson, Jacob. 2014. What can we learn from twin studies? A comprehensive evaluation of the equal environments assumption. *Social Science Research* 43:184–99.
- Ferguson, Christopher J. 2010. Genetic contributions to antisocial personality and behavior: A meta-analytic review from an evolutionary perspective. *The Journal of Social Psychology* 150:160–80.
- Floderus-Myrhed, Birgitta, Nancy Pedersen, and Ingrid Rasmuson. 1980. Assessment of heritability for personality, based on a short-form of the Eysenck Personality Inventory: A study of 12,898 twin pairs. *Behavior Genetics* 10:153–62.
- Gottfredson, Linda S. 2010. Lessons in academic freedom as lived experience. *Personality and Individual Differences* 49:272–80.
- Gunderson, Erica P., Ai-Lin Tsai, Joe V. Selby, Bette Caan, Elizabeth J. Mayer-Davis, and Neil Risch. 2006. Twins of mistaken zygosity (TOMZ): Evidence for genetic contributions to dietary patterns and physiologic traits. *Twin Research and Human Genetics* 9:540–9.
- Hannagan, Rebecca J., and Peter K. Hatemi. 2008. The threat of genes: A comment on Evan Charney’s “Genes and Ideologies.” *Perspectives on Politics* 6:329–35.
- Harris, Judith Rich. 1998. *The Nurture Assumption: Why Children Turn Out the Way They Do*. New York: Free Press.
- Harris, Judith Rich. 2006. *No Two Alike: Human Nature and Human Individuality*. New York: W. W. Norton.
- Hatemi, Peter K., John R. Alford, John R. Hibbing, Nicholas G. Martin, and Lindon J. Eaves. 2009. Is there a “party” in your genes? *Political Research Quarterly* 62:584–600.
- Haynie, Dana L., Peggy C. Giordano, Wendy D. Manning, and Monica Longmore. 2005. Adolescent romantic relationships and delinquency involvement. *Criminology* 43:177–210.
- Haynie, Dana L., and Suzanne McHugh. 2003. Sibling deviance: In the shadows of mutual and unique friendship effects? *Criminology* 41:355–92.
- Heath, Andrew C., Kenneth S. Kendler, Lindon J. Eaves, and David Markell. 1985. The resolution of cultural and biological inheritance: Informativeness of different relationships. *Behavior Genetics* 15:439–65.
- Heijmans, Bastiaan T., and Jonathan Mill. 2011. Commentary: The seven plagues of epigenetic epidemiology. *International Journal of Epidemiology* 41:1–5.
- Hettema, John M., Michael C. Neale, and Kenneth S. Kendler. 1995. Physical similarity and the equal-environment assumption in twin studies of psychiatric disorders. *Behavior Genetics* 25:327–35.

- Hewitt, John K. 2012. Editorial policy on candidate gene association and candidate gene-by-environment interaction studies of complex traits. *Behavior Genetics* 42:1–2.
- Horwitz, Allan V., Tami M. Videon, Mark F. Schmitz, and Diane Davis. 2003. Rethinking twins and environments: Possible social sources for assumed genetic influences in twin research. *Journal of Health and Social Behavior* 44:111–29.
- Jacob, Theodore, and Deborah A. Bremer. 1986. Assortative mating among men and women alcoholics. *Journal of Studies on Alcohol and Drugs* 47:219–22.
- Joseph, Jay. 1998. The equal environments assumption of the classical twin method: A critical analysis. *The Journal of Mind and Behavior* 19:325–58.
- Joseph, Jay. 2001. Separated twins and the genetics of personality differences: A critique. *The American Journal of Psychology* 114:1–30.
- Joseph, Jay. 2004. *The Gene Illusion: Genetic Research in Psychiatry and Psychology Under the Microscope*. New York: Algora.
- Joseph, Jay. 2006. *The Missing Gene: Psychiatry, Heredity, and the Fruitless Search for Genes*. New York: Algora.
- Joseph, Jay. 2010. Genetic research in psychiatry and psychology: A critical overview. In *Handbook of Developmental Science, Behavior, and Genetics*, eds. K. E. Hood, C. T. Halpern, G. Greenberg, and R. M. Learner. Malden, MA: Wiley-Blackwell.
- Kendler, Kenneth S., and Charles O. Gardner. 1998. Twin studies of adult psychiatric and substance dependence disorders: Are they biased by differences in the environmental experiences of monozygotic and dizygotic twins in childhood and adolescence? *Psychological Medicine* 28:625–33.
- Kendler, Kenneth S., Laura M. Karkowski, Michael C. Neale, and Carol A. Prescott. 2000. Illicit psychoactive substance use, heavy use, abuse, and dependence in a US population-based sample of male twins. *Archives of General Psychiatry* 57:261–9.
- Kendler, Kenneth S., S. Larsson Lönn, Nancy A. Morris, Jan Sundquist, Niklas Långström, and Kristina Sundquist. 2013. A Swedish national adoption study of criminality. *Psychological Medicine* 4:1–13.
- Kendler, Kenneth S., Michael C. Neale, Ronald C. Kessler, Andrew C. Heath, and Lindon J. Eaves. 1993. A test of the equal-environment assumption in twin studies of psychiatric illness. *Behavior Genetics* 23:21–7.
- Kendler, Kenneth S., Michael C. Neale, Ronald C. Kessler, Andrew C. Heath, and Lindon J. Eaves. 1994. Parental treatment and the equal environment assumption in twin studies of psychiatric illness. *Psychological Medicine* 24:579–90.
- Klump, Kelly L., Amanda Holly, William G. Iacono, Matt McGue, and Laura E. Willson. 2000. Physical similarity and twin resemblance for eating attitudes and behaviors: A test of the equal environments assumption. *Behavior Genetics* 30:51–8.
- Krueger, Robert F., Terrie E. Moffitt, Avshalom Caspi, April Bleske, and Phil A. Silva. 1998. Assortative mating for antisocial behavior: Developmental and methodological implications. *Behavior Genetics* 28:173–86.
- Lee, S. Hong, Teresa R. DeCandia, Stephan Ripke, Jian Yang, The Schizophrenia Psychiatric Genome-Wide Association Study Consortium (PGC-SCZ), The International Schizophrenia Consortium (ISC), The Molecular Genetics of Schizophrenia Collaboration (MGS), et al. 2012. Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nature Genetics* 44:247–50.

- Lee, S. Hong, Stephan Ripke, Benjamin M. Neale, Stephen V. Faraone, Shaun M. Purcell, Roy H. Perlis, Bryan J. Mowry, et al. 2013. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics* 45:984–94.
- Lewontin, Richard C., Steven Rose, and Leon J. Kamin. 1984. *Not in Our Genes: Biology, Ideology, and Human Nature*. New York: Pantheon Books.
- Littvay, Levente. 2012. Do heritability estimates of political phenotypes suffer from an equal environment assumption violation? Evidence from an empirical study. *Twin Research and Human Genetics* 15:6–14.
- Loehlin, John C. 1996. The Cholesky approach: A cautionary note. *Behavior Genetics* 26:65–9.
- Loehlin, John C., and Robert C. Nichols. 1976. *Heridity, Environment, and Personality: A Study of 850 Sets of Twins*. Austin: University of Texas Press.
- Lykken, David T., Matt McGue, Thomas J. Bouchard, Jr., and Auke Tellegen. 1990. Does contact lead to similarity or similarity to contact? *Behavior Genetics* 20:547–61.
- Lynch, Michael, and Bruce Walsh. 1998. *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates, Inc.
- Mare, Robert D. 1991. Five decades of educational assortative mating. *American Sociological Review* 56:15–32.
- Mason, Dehryl A., and Paul J. Frick. 1994. The heritability of antisocial behavior: A meta-analysis of twin and adoption studies. *Journal of Psychopathology and Behavioral Assessment* 16:301–23.
- Maxson, Stephen C. 2007. A history of behavior genetics. In *Neurobehavioral Genetics: Methods and Applications*, 2nd ed., eds. B. C. Jones and P. Mormède. Boca Raton, FL: CRC Press.
- McArdle, John J., and Brenda L. Plassman. 2009. A biometric latent curve analysis of memory decline in older men of the NAS-NRC Twin Registry. *Behavior Genetics* 39:472–95.
- McConkey, Edwin H. 2004. *How the Human Genome Works*. Boston, MA: Jones and Bartlett Learning.
- McGue, Matt. 2010. The end of behavioral genetics? *Behavior Genetics* 40:284–96.
- McGue, Matt, Yiwei Zhang, Michael B. Miller, Saonli Basu, Scott Vrieze, Brian Hicks, Steve Malone, William S. Oetting, and William G. Iacono. 2013. A genome-wide association study of behavioral disinhibition. *Behavior Genetics* 43:363–73.
- McGuire, Shirley, and Nancy L. Segal. 2013. Peer network overlap in twin, sibling, and friend dyads. *Child Development* 84:500–11.
- Meier, Madeline H., Wendy S. Slutske, Andrew C. Heath, and Nicholas G. Martin. 2011. Sex differences in the genetic and environmental influences on childhood conduct disorder and adult antisocial behavior. *Journal of Abnormal Psychology* 120:377–88.
- Merikangas, Kathleen Ries, and Duane G. Spiker. 1982. Assortative mating among inpatients with primary affective disorder. *Psychological Medicine* 12:753–64.
- Miles, Donna R., and Gregory Carey. 1997. Genetic and environmental architecture on human aggression. *Journal of Personality and Social Psychology* 72:207–17.
- Moffitt, Terrie E. 2005. Genetic and environmental influences on antisocial behaviors: Evidence from behavioral-genetic research. *Advances in Genetics* 55:41–104.
- Morris-Yates, Allen, Gavin Andrews, Peter W. Howie, and Scott Henderson. 1990. Twins: A test of the equal environments assumption. *Acta Psychiatrica Scandinavica* 81:322–6.

- Neale, Michael C., and Hermine H. M. Maes. 2004. *Methodology for Genetic Studies of Twins and Families*. Dordrecht, the Netherlands: Kluwer Academic.
- Paul, Steven M. 1980. Sibling resemblance in mental ability: A review. *Behavior Genetics* 10:277–90.
- Pederson, Nancy L., Gerald E. McClearn, Robert Plomin, and Lars Friberg. 1985. Separated fraternal twins: Resemblance for cognitive abilities. *Behavior Genetics* 15:407–19.
- Pinker, Steven P. 2002. *The Blank Slate: The Modern Denial of Human Nature*. New York: Viking Press.
- Piquero, Alex R., David P. Farrington, and Alfred Blumstein. 2007. *Key Issues in Criminal Career Research: New Analyses of the Cambridge Study in Delinquent Development*. New York: Cambridge University Press.
- Plomin, Robert, and Cindy S. Bergeman. 1991. The nature of nurture: Genetic influence on “environmental” measures. *Behavioral and Brain Sciences* 14:373–86.
- Plomin, Robert, John C. DeFries, Valerie S. Knopik, and Jenae M. Neiderhiser. 2013. *Behavioral Genetics*, 6th ed. New York: Worth.
- Plomin, Robert, Claire M. A. Haworth, Emma L. Meaburn, Thomas S. Price, Wellcome Trust Case Control Consortium, and Oliver S. P. Davis. 2013. Common DNA markers can account for more than half of the genetic influence on cognitive abilities. *Psychological Science* 24:562–8.
- Plomin, Robert, Lee Willerman, and John C. Loehlin. 1976. Resemblance in appearance and the equal environments assumption in twin studies of personality traits. *Behavior Genetics* 6:43–52.
- Purcell, Shaun, and Pak C. Sham. 2003. A model-fitting implementation of the DeFries-Fulker model for selected twin data. *Behavior Genetics* 33:271–8.
- Raine, Adrian. 1993. *The Psychopathology of Crime: Criminal Behavior as a Clinical Disorder*. New York: Academic Press.
- Rhee, Soo Hyun, and Irvin D. Waldman. 2002. Genetic and environmental influences on antisocial behavior: A meta-analysis of twin and adoption studies. *Psychological Bulletin* 128:490–529.
- Rhule-Louie, Dana M., and Robert J. McMahon. 2007. Problem behavior and romantic relationships: Assortative mating, behavior contagion, and desistance. *Clinical Child and Family Psychology Review* 10:53–100.
- Rice, John C., Robert Cloninger, and Theodore Reich. 1980. Analysis of behavioral traits in the presence of cultural transmission and assortative mating: Applications to IQ and SES. *Behavior Genetics* 10:73–92.
- Richardson, Ken. 2011. Wising up on the heritability of intelligence. *GeneWatch* 24:15–8.
- Richardson, Ken, and Sarah Norgate. 2005. The equal environments assumption of classical twin studies may not hold. *British Journal of Educational Psychology* 75:339–50.
- Rose, Richard J., Markku Koskenvuo, Jaakko Kaprio, Seppo Sarna, and Heimo Langinvainio. 1988. Shared genes, shared experiences, and similarity of personality: Data from 14,288 adult Finnish co-twins. *Journal of Personality and Social Psychology* 54:161–71.
- Rowe, David C. 1994. *The Limits of Family Influence: Genes, Experience, and Behavior*. New York: Guilford Press.
- Rowe, David C., and David P. Farrington. 1997. The familial transmission of criminal convictions. *Criminology* 35:177–202.

- Rowe, David C., and D. Wayne Osgood. 1984. Heredity and sociological theories of delinquency: A reconsideration. *American Sociological Review* 49:526–40.
- Rushton, J. Philippe, David W. Walker, Michael C. Neale, David K. B. Nias, and Hans J. Eysenck. 1986. Altruism and aggression: The heritability of individual differences. *Journal of Personality and Social Psychology* 50:1192–1198.
- Sacerdote, Bruce. 2004. *What happens when we randomly assign children to families?* Cambridge, MA: National Bureau of Economic Research, Working Paper No. 10894.
- Sampson, Robert J., and John H. Laub. 1993. *Crime in the Making: Pathways and Turning Points Through Life*. Cambridge, MA: Harvard University Press.
- Saudino, Kimberly J., Angelica Ronald, and Robert Plomin. 2005. The etiology of behavior problems in 7-year-old twins: Substantial genetic influence and negligible shared environmental influence for parent ratings and ratings by same and different teachers. *Journal of Abnormal Child Psychology* 33:113–30.
- Scarr, Sandra 1987. Three cheers for behavior genetics: Winning the war and losing our identity. *Behavior Genetics* 17:219–28.
- Scarr, Sandra, and Louise Carter-Saltzman. 1979. Twin method: Defense of a critical assumption. *Behavior Genetics* 9:527–42.
- Scarr, Sandra, E. Scarf, and R. A. Weinberg. 1980. Perceived and actual similarities in biological and adoptive families: Does perceived similarity bias genetic estimates? *Behavior Genetics* 10:445–58.
- Schwartz, Joseph A., and Kevin M. Beaver. 2011. Evidence of a gene  $\times$  environment interaction between perceived prejudice and MAOA genotype in the prediction of criminal arrests. *Journal of Criminal Justice* 39:378–84.
- Schwartz, Joseph A., and Kevin M. Beaver. 2014. Exploring whether genetic differences between siblings explain sibling differences in criminal justice outcomes. *Comprehensive Psychiatry* 55:93–103.
- Segal, Nancy L. 2012. *Born Together—Reared Apart: The Landmark Minnesota Twin Study*. Cambridge, MA: Harvard University Press.
- Segal, Nancy L. 2013. Personality similarity in unrelated look-alike pairs: Addressing a twin study challenge. *Personality and Individual Differences* 54:23–8.
- Segal, Nancy L., Jamie L. Graham, and Ulrich Ettinger. 2013. Unrelated look-alikes: Replicated study of personality similarity and qualitative findings on social relatedness. *Personality and Individual Differences* 55:169–74.
- Segerstråle, Ullica. 2000. *Defenders of the Truth: The Battle for Science in the Sociobiology Debate and Beyond*. New York: Oxford University Press.
- Simons, Ronald L., Steven R. H. Beach, and Ashley B. Barr. 2012. Differential susceptibility to context: A promising model of the interplay of genes and the social environment. *Advances in Group Processes* 29:139–63.
- Simons, Ronald L., and Man Kit Lei. 2013. Enhanced susceptibility to context: A promising perspective on the interplay of genes and the social environment. In *Handbook of Life-Course Criminology: Emerging Trends and Directions for Future Research*, eds. Chris L. Gibson and Marvin D. Krohn. New York: Springer.
- Simons, Ronald L., Man Kit Lei, Steven R. H. Beach, Gene H. Brody, Robert A. Philibert, and Frederick X. Gibbons. 2011. Social environment, genes, and aggression: Evidence supporting the differential susceptibility perspective. *American Sociological Review* 76:883–912.



- Simons, Ronald L., Man Kit Lei, Eric A. Stewart, Steven R. H. Beach, Gene H. Brody, Robert A. Philibert, and Frederick X. Gibbons. 2012. Social adversity, genetic variation, street code, and aggression: A genetically informed model of violent behavior. *Youth Violence and Juvenile Justice* 10:3–24.
- Simons, Ronald L., Leslie G. Simons, Man Kit Lei, Steven R. H. Beach, Gene H. Brody, Frederick X. Gibbons, and Robert A. Philibert. 2013. Genetic moderation of the impact of parenting on hostility toward romantic partners. *Journal of Marriage and the Family* 75:325–41.
- Simons, Ronald L., Eric Stewart, Leslie C. Gordon, Rand D. Conger, and Glen H. Elder, Jr. 2002. A test of life-course explanations for stability and change in antisocial behavior from adolescence to young adulthood. *Criminology* 40:401–34.
- Smith, George D. 2011. Epidemiology, epigenetics and the “gloomy prospect”: Embracing randomness in population health research and practice. *International Journal of Epidemiology* 40:537–62.
- Suhay, Elizabeth, and Nathan P. Kalmoe. 2010. The equal environment assumption in twin studies of political traits: Social confounds and suggested remedies. Unpublished manuscript. <http://sites.lafayette.edu/suhaye/files/2011/01/Violations-of-the-EEA-Suhay-Kalmoe.pdf>.
- Tielbeek, Jorim J., Sarah E. Medland, Beben Benyamin, Enda M. Byrne, Andrew C. Heath, Pamela A. F. Madden, Nicholas G. Martin, Naomi R. Wray, and Karin J. H. Verweij. 2012. Unraveling the genetic etiology of adult antisocial behavior: A genome-wide association study. *PLoS ONE* 7:e45086.
- Tuvblad, C., and Kevin M. Beaver. 2013. Genetic and environmental influences on antisocial behavior. *Journal of Criminal Justice* 41:273–6.
- Vandenberg, Steven G. 1972. Assortative mating, or who marries whom? *Behavior Genetics* 2:127–57.
- Viding, Essi, Frank M. Spinath, Thomas C. Price, Dorothy V. M. Bishop, Philip S. Dale, and Robert Plomin. 2004. Genetic and environmental influence on language impairment in 4-year-old same-sex and opposite-sex twins. *Journal of Child Psychology and Psychiatry* 45:315–25.
- Vinkhuyzen, Anna A. E., Nancy L. Pederson, Jian Yang, Sun H. Lee, Patrick K. E. Magnusson, William G. Iacono, Matt McGue, et al. 2012. Common SNPs explain some of the variation in the personality dimensions of neuroticism and extraversion. *Translational Psychiatry* 2:e102.
- Visscher, Peter M., Jian Yang, and Michael E. Goddard. 2010. A commentary on “Common SNPs explain a large portion of the heritability for human height” by Yang et al. (2010). *Twin Research and Human Genetics* 13:517–24.
- Walsh, Anthony, and Lee Ellis. 2004. Ideology: Criminology’s Achilles’ heel? *Quarterly Journal of Ideology* 27:1–25.
- Wright, John Paul, and Kevin M. Beaver. 2005. Do parents matter in creating self-control in their children? A genetically informed test of Gottfredson and Hirschi’s theory of low self-control. *Criminology* 43:1169–202.
- Wright, John Paul, Rebecca Schnupp, Kevin M. Beaver, Matt DeLisi, and Michael G. Vaughn. 2012. Genes, maternal negativity, and self-control: Evidence of a gene x environment interaction. *Youth Violence and Juvenile Justice* 10:245–60.
- Wright, John Paul, Stephen G. Tibbetts, and Leah E. Daigle. 2008. *Criminals in the Making: Criminality Across the Life Course*. Thousand Oaks, CA: Sage.

- Xian, Hong, Jeffrey F. Scherrer, Seth A. Eisen, William R. True, Andrew C. Heath, Jack Goldberg, Michael J. Lyons, and Ming T. Tsuang. 2000. Self-reported zygosity and the equal-environments assumption for psychiatric disorders in the Vietnam Era Twin Registry. *Behavior Genetics* 30:303–10.
- Yang, Jian, Beben Benyamin, Brian P. McEvoy, Scott Gordon, Anjali K. Henders, Dale R. Nyholt, Pamela A. Madden, Andrew C. Heath, Nicholas G. Martin, Grant W. Montgomery, Michael E. Goddard, and Peter M. Visscher. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* 42: 565–9.
- Yang, Jian, S. Hong Lee, Michael E. Goddard, and Peter M. Visscher. 2011. GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics* 88:76–82.

J. C. Barnes is an associate professor in the School of Criminal Justice at the University of Cincinnati. He is a biosocial criminologist whose research seeks to understand how genetic and environmental factors combine to impact criminological phenomena. Recent works have attempted to reconcile behavioral genetic findings with theoretical developments in criminology.

John Paul Wright is a professor of criminal justice in the School of Criminal Justice at the University of Cincinnati. He is also a scholar in the Center for Social and Humanities Research, King Abdulaziz University, Jeddah, Saudi Arabia. His work examines the biological connections to violence.

Brian B. Boutwell is an associate professor of criminology and criminal justice in the School of Social Work at Saint Louis University. His research interests include the evolution of complex outcomes including intelligence, violence, and chronic criminality.

Joseph A. Schwartz is an assistant professor in the School of Criminology and Criminal Justice at the University of Nebraska at Omaha. His research interests include behavior genetics, biosocial criminology, the association between intelligence and behavior, and additional factors involved in the etiology of criminal behavior.

Eric J. Connolly is an assistant professor in the Department of Criminal Justice at Pennsylvania State University, Abington. He received his Ph.D. in criminology and criminal justice from Florida State University. His research interests include biosocial criminology, life-course/developmental criminology, and quantitative behavior genetics.

Joseph L. Nedelec is an assistant professor in the School of Criminal Justice at the University of Cincinnati. He received his Ph.D. in criminology from Florida State University. His research interests include biosocial criminology, evolutionary psychology, intelligence, quantitative behavior genetics, and cybercrime.

Kevin M. Beaver is a professor in the College of Criminology and Criminal Justice at Florida State University and Visiting Distinguished Professor in the Center for Social and Humanities Research at King Abdulaziz University. His research focuses on the biosocial underpinnings to antisocial behaviors.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web site:

**Appendix A.** Discussion of Additional Assumptions of Twin-Based Research

**Appendix B.** Mathematical Foundations of Behavioral Genetics

**Appendix C.** Latent Variable ACE Model

**Appendix D.** Comprehensive List of Studies Examining the Equal Environments Assumption

**Appendix E.** *R* Script for Carrying Out Calculations of Assumption Violations