

Tilting at Twindmills: Rethinking Sociological Responses to Behavioral Genetics*

JEREMY FREESE

University of Wisconsin-Madison

BRIAN POWELL

Indiana University

Journal of Health and Social Behavior 2003, Vol 44 (June): 130–135

While we commend Horwitz et al. (2003) for speaking to core issues in behavioral genetics, we disagree with many particulars of their article. We are skeptical of their claims regarding the particular contribution offered by both their methods and data. We believe also that the findings they present as challenging the equal environments assumption are, upon closer examination, not persuasive. Most fundamentally, we worry that the way in which Horwitz et al. conceptualize the relationship between genes and environments is not the best means of doing so for sociologists interested in engaging behavioral genetics.

Like Horwitz, Videon, Schmitz, and Davis (this issue), we wish it were more widely recognized that specific heritability estimates rest on numerous assumptions, which together concern social processes at least as much as genetics (Goldberger 1979). As such—and given the seeming widespread interest everywhere else in genes and behavior—the lack of serious sociological engagements with behavioral genetics has been unfortunate. We commend Horwitz et al. for their attempt to speak empirically to the important issue of the equal environments assumption in twin studies. They interpret their results as “suggestive” of its violation—that “some of the greater concordances between monozygotic compared to dizygotic twins that behavior geneticists attribute to genetic factors might stem from the more common social environments that monozygotic twins share” (p. 125). In drawing this conclusion, Horwitz et al. are saying nothing new to those familiar with behavioral genetics. Rutter et al. (2001) report that the equal environments assumption is “likely often to be violated” (p. 304). Maccoby (2000) spec-

ulates that violations of the assumption might explain why twin studies typically produce higher heritability estimates than adoption studies. Even so, behavioral genetics on the whole can be readily faulted for not taking the assumption seriously enough or being too credulous in how they test it (Joseph 1998). Popular portrayals of heritability typically ignore its existence altogether.

While we are therefore pleased Horwitz et al. highlight an assumption that deserves more scrutiny, we believe that the specific empirical analyses Horwitz et al. conduct do not warrant the particular conclusions they draw. Below we briefly consider Horwitz et al.’s analytic strategy and sample in general terms and then focus on their specific results and interpretations. We worry not merely that their analyses will be unpersuasive to behavioral geneticists, but, worse, that they will be persuasive to sociologists only to the extent that sociologists are unfamiliar with behavioral genetics. In some respects, our remarks can be read as “defending” behavioral genetics against Horwitz et al.’s critiques, but, more fundamentally, as we shall later explain, we think sociologists and others need to be working toward ways of thinking about genes and environments that move beyond the familiar conceptualizations drawn from behavioral genetics on which Horwitz et al.’s paper is premised.

*Address correspondence to Jeremy Freese, Department of Sociology, University of Wisconsin-Madison, 1180 Observatory Drive, Madison, WI, 53711, jfreese@ssc.wisc.edu. We would like to thank Simon Cheng, Arthur Goldberger, and Allen Li.

GENERAL CONSIDERATIONS ABOUT THEIR ANALYSIS

How do Horwitz et al. contribute to ongoing debates regarding behavioral genetics? Beyond scrutinizing an underscrutinized assumption, Horwitz et al. assert that they marshal a more appropriate analytic strategy and better data than other work. The actual contributions here, however, are much less than they might seem.

Methodology

Horwitz et al. diverge markedly from standard techniques of behavioral genetics and argue that their analytic approach is superior (pp. 17–18). Horwitz et al. use familiar methods of regression analysis and hypothesis testing to test whether monozygotic twins are more similar on particular traits than dizygotic twins, as would be expected if the trait is heritable. While perhaps adequate for their purposes, *the approach should not be regarded as a technical advance over conventional behavior genetics models, if only because Horwitz et alia's models do not estimate the same parameters that behavior geneticists typically pursue.* One can neither straightforwardly recover heritability estimates nor observe how these estimates change when the “social” variables are added. Nor can one examine whether heritabilities appear stronger among more extreme cases—something for which DeFries-Fulker “extremes” analysis is used.

We are especially puzzled by Horwitz et alia's references to Allison (1990) and Kenny (1996) as supporting their methodological arguments. Space limitations prevent adequate discussion here, but we see deep disanalogies and few transparent connections between the issues and examples discussed in these papers and the twin analyses Horwitz et al. actually do. If behavior genetics models really are flawed in fundamental ways remedied by their comparatively simple approach, we urge Horwitz et al. to provide a more detailed discussion. *Minimally, to stand as a methodological contribution, Horwitz et al. should show that conventional behavioral genetics analyses would have led to substantively different results and then explain why their own results are more credible.* While we are skeptical of the whole exercise of allocating variance to “genes” and “environment,” Horwitz et alia's

modeling arguments do adopt the premises of such an exercise but fail to demonstrate that their approach provides the advantages they claim.

Sample

Horwitz et al. might give the impression that the twin subsample from the National Longitudinal Study of Adolescent Health possesses advantages such that estimates from it should “trump” other twin samples. Although popular portrayals of heritability estimates present them as transcendent biological facts, we agree that they are only potentially meaningful if considered as the property of some population, meaning that estimates can vary over subpopulations, environments, and time. For this reason, basing estimates on a well-drawn sample from a well-defined population is both important and chronically underappreciated. Yet, as Horwitz et al. acknowledge, because of recruiting and weighting issues, the estimation sample of Horwitz et alia's analyses may well *include representatives from various American subpopulations, but it is by no means representative, and the implications of its non-representativeness are unknown.*

Additionally, despite the population-specificity of heritability estimation, Horwitz et al. seem to argue that the National Longitudinal Study of Adolescent Health sample is better for considering heritability and psychological outcomes *generally* than prominent Swedish and other samples, because of its greater “heterogeneity.”¹ Setting aside the seeming US-centrism of their reasoning here, we are concerned that *Horwitz et al. fail to offer a convincing case for why a heterogeneous sample per se is desirable for studying the equal environments assumption.* One might note that Horwitz et al. take as obvious that their estimates should control for (that is, “hold constant”) precisely the kinds of variation that they are pleased their sample possesses. We believe that a specifically heterogeneous sample would provide the most value if either (1) the sample was adequately representative of its heterogeneous population or (2) the sample was large enough to allow one to examine with sufficient power whether estimates vary among subpopulations.² *The National Longitudinal Study of Adolescent Health provides neither of these.*³ We do not wish to dis-

courage informative analyses of the National Longitudinal Study of Adolescent Health twin data, but we urge sociologists not to overestimate its virtues.

SPECIFIC CONSIDERATIONS OF ANALYSIS AND INTERPRETATION

Rather than in its methods or data, the paper's status as a contribution seems to rest squarely on its asserted challenge to the equal environments assumption. Horwitz et al. first provide evidence suggesting greater environmental similarity for monozygotic twins compared to dizygotic twins, but this is unsurprising and, as Horwitz et al. acknowledge, has been demonstrated elsewhere. Horwitz et al.'s contribution seems to rest instead on their apparent demonstration that this similarity in the "social" variables attenuates the excess similarity of monozygotic twins for depression in girls and for having tried alcohol and binge drinking among boys. In all three instances, the credibility of their findings is seriously undermined by weak results, tenuous interpretations, or both.

Consider first their analyses of depressive symptoms among girls. When Horwitz et al.'s social variables are added to the model, the zygosity coefficient indicative of excess monozygotic twin similarity decreases by 57 percent and is no longer significant. Since this is the strongest attenuation, we assume it provides the basis of Horwitz et al.'s dramatic claim in their abstract that their analyses show that measures of environmental similarity sometimes do not just reduce but "eliminate apparent genetic effects" (p. 111). While we think that is reading far too much into the reduction of a coefficient from significance to nonsignificance, especially in a small sample, one must also consider that the excess similarity of monozygotic twins in depressive symptoms is relatively small here to begin with. Moreover, as Horwitz et al. note, the dependent variable is discrepancy in major depression, but discrepancy in depressive symptoms in a sample where the strong majority would not be characterized as "depressed." Because behavioral geneticists have already suggested that genetic inheritance plays a larger role in severe depression than mild depression (see, e.g., Plomin et al. 2001: 217–22), we doubt behavioral geneticists would be much be sur-

prised by the modest initial excess monozygotic similarity observed here.

Probing further, we find that 85 percent of the attenuation in the zygosity coefficient can be attributed to the inclusion of the "time together" variable; twins who both report that they spend "a lot" of time together are less discrepant on depressive symptoms, and more monozygotic twins report spending a lot of time together than dizygotic twins.⁴ That this result challenges the equal environments assumption is based on the interpretation that more time together means similar environmental experiences (i.e., "exposure to stressors"), which, in turn, yields similar levels of depressive symptoms.

Importantly, however, the distribution of Horwitz et al.'s depressive symptoms measure (a CES-D variant) is effectively left-censored at 0, with a substantial rightward skew. Consequently, discordance on depressive symptoms between twins is correlated with their average level of depression ($r = .36$). The time twins spend together is correlated about the same with their average CES-D score as with the difference between their scores ($r \approx .24$), and Horwitz et al.'s full model actually fits average depression as an outcome better than twin discrepancy in depression ($R^2 = .15$ vs. $R^2 = .10$). Horwitz et al.'s analytic approach thus offers no leverage against the alternative interpretation that spending more time together has salutary consequences for female twins' psychological well-being, which, because of the distribution of the CES-D, implies that they will also be less discordant on this measure. Even worse, if depressive states often involve social withdrawal, then some pairs might spend (or report spending) less time together precisely because one or both of them is depressed. *In other words, for their main finding challenging the equal environments assumption among females, Horwitz et al.'s result may reflect either a reversal of cause and effect or a misattribution of an effect on differences between twin pairs (their average depression) as an effect on differences within pairs.*

For males, significant excess similarity of monozygotic twins was observed for the three alcohol-related outcomes: having tried alcohol, frequency of drinking, and frequency of binge drinking. The first and third of these outcomes are attenuated by the inclusion of the social variables in their model. From their

analyses of trying alcohol, Horwitz et al. conclude that “apparent genetically determined similarities are mediated by social sources of concordance” (p. 125). However, as before, the evidence of excess monozygotic similarity—apparently genetically determined or otherwise—is weak to begin with ($p = .04$, one-tailed). This would probably again not surprise behavioral geneticists, for the meaning of the zygosity coefficient here is not obvious. Just under 90 percent (22./25) of the 13-year-old male twins concordant on this measure are concordant because neither reports having tried alcohol; meanwhile, 72 percent of concordant 18-year old male twins are concordant because both have tried alcohol. Observed discordance might be thus better thought of as twins happening to be sampled during their period of discordance, as all were once concordant (both 0) and most will be concordant again (both 1). We are surprised that Horwitz et al. evince no recognition of this crucial analytic issue, as it is most certainly not resolved by “controlling for age” with a linear term. In any event, while including the social variables does reduce the marginally significant zygosity coefficient to nonsignificance, *the change between models is not significant by Horwitz et alia’s calculations, indicating that the results do not provide evidence against the equal environments assumption by the conventional standards of hypothesis testing.*

For boys’ binge drinking, while a model for continuous outcomes is used, 68 percent of pairs have no discrepancy, and over 90 percent of the pairs with no discrepancy report no binge drinking. Given such a distribution, one might instead have estimated a model for the dichotomous outcome of whether the twins are discrepant in their reported binge drinking. We computed this model, and what was already a modest 23 percent attenuation in the untransformed zygosity coefficient, with the addition of the social variables, becomes only an 11 percent decrease. Thus, the substantive influence of adding the social variables on the association between zygosity and binge drinking is slight at best.

Moreover, the “number of three best friends who drink” variable is responsible for 80 percent of this small attenuation in the zygosity coefficient when discrepancy in binge drinking is measured “continuously,” and it is responsible for *all* of it when the dichotomous measure is used. However, we believe Horwitz et al.

should not have included this independent variable in their model. Direction of causality is again at issue: To what extent does having friends who drink influence one’s drinking, to what extent does one’s drinking influence one’s friends’ drinking behavior, and to what extent does drinking behavior influence who one considers one’s best friends? Horwitz et al. provide little theoretical reason for understanding this variable as a cause of monozygotic twins experiencing more similar environments than dizygotic twins, *net* of monozygotic twins spending more time together and having more overlap of friends (already included as controls). In short, although we are supposed to interpret the attenuation of the zygosity coefficient here as empirical evidence of violation of the equal environments assumption, *the attenuation is small, and the variable responsible for it likely should have been excluded anyway.*

We have still further doubts about Horwitz et alia’s interpretations of patterns, in part because the patterns seem severely undertheorized. If Horwitz et al. seek to demonstrate that results behavioral geneticists attribute to genes might be the product of environmental causes, then one might be dismayed by how little serious contemplation they give to how particular environmental factors are supposed to influence the outcomes studied. Behavioral geneticists regularly criticize many discussions of environmental effects as being largely hand-waving, and Horwitz et al. may inadvertently provide another example that fits this characterization well. Their interpretations of the results that suit the general finding seem unfortunately both vague and ad hoc. One looks at Tables 3 and 4 and wonders why time spent together predicts similarity in depression and trying alcohol for females but not at all for males, while similar attractiveness significantly predicts similarity in trying alcohol for males but not for females. If “the greater amount of time monozygotic twins spend together. . . increases their exposure to common stressors” (p. 123) for females, why not for males? If male twins similar in attractiveness select “into peer groups that have common attitudes toward using alcohol” (p. 123) why don’t female twins, especially since the article cited to support their assertion (Kennedy 1990) reports no relevant sex differences? *Without specific, well-articulated theory, it is hard to see why results so ultimately*

modest and fragile should be granted much credibility.

CONCEPTUALIZING THE CAUSALITY OF GENES AND ENVIRONMENTS

Taken together, the problems with Horwitz et alia's analyses are sufficiently serious that their results offer little toward understanding the excess monozygotic similarity on the psychobehavioral traits they examine. Indeed, if Horwitz et al. have intended to put forth analyses that challenge core propositions of behavioral genetics, we worry that they may have instead provided more fodder that "avid proponents" of genetic zealotry can use to challenge the worthiness of sociological participation in these debates. Yet, even if their analyses were more compelling, there are two reasons why we would still be uncomfortable with how Horwitz et al. conceptualize the issues they pursue.

First, we have deep unease with Horwitz et alia's assertions that "ultimately, the issue of whether social or genetic causes are primary is theoretical" and that "theoretical assumptions, not empirical findings, determine where to end the chain of causation between social and genetic factors" (p. 125). These statements seem to deploy the more respectable "theoretical" where "practical" or "aesthetic" would be more accurate, and they seem to encourage stopping inquiry precisely where we think it might become more interesting. We find ultimately unproductive the notion of drawing a division between "genetic" and "social" causes and then saying that theoretical assumptions allow one to assign as primary and ultimate whichever one prefers.

Second, we question the exercise of trying to figure out what percentage of depression or alcohol usage is really "genetic" and what percentage is "environmental." Horwitz et al. rightly recognize that labeling attributes such as attractiveness as "social" is complicated if genes are presumed to explain why monozygotic twins tend to be more similarly attractive than dizygotic twins. Any consequences of attractiveness would seem to represent an interaction of genes and environments. Horwitz et al. use attractiveness to assert that "the separation of genetic and environmental influences is always complex; allocating variance to one or the other source is not as straightforward" as it seems (p. 125). Here is

how our positions differ: While Horwitz et al. seem to see allocating variance between genes and environments as "complex," and that the central analytic matter is determining whether the allocation to genes is "overstated" or not because of violations of the equal environments assumption, we regard *the very idea that this kind of binary allocation is even possible for many outcomes as being precisely what sociology needs to move beyond* if the discipline is to contribute more than just a loyal opposition that adopts the same oversimplifying conceptualizations as much behavioral genetics. Causal interactions defeat contingent attributions of potency to the individual causes; sociologists readily recognize this elsewhere when thinking about "interaction effects," but fall back into thinking otherwise when confronted with the duolith of "genetic" and "social" causes.

Sociologists who believe genes have something to do with various outcomes and wish to better understand what that something is face many challenges. Foremost is strong resistance within the discipline, much of which is based on the deeply serious but ultimately unfounded concern that acknowledging the consequentiality of genes necessarily sets one down one unsavory slippery slope or another. Another obstacle, we are convinced, is provided by commonplace conceptualizations that circumscribe thinking in unhelpful ways. Among the worst is that there exists 100 percentage points of credit for how an outcome is determined and that an exercise of central importance is figuring out how to divvy up these points between "genes" and "the environment." When thinking is restricted in this way, one will indeed reach an impasse where the same empirical results can lead one to give the credit to one side or the other. But this need not mean that one has reached the point where "theory" should enter as a *deus ex machina* that determines what one chooses to believe, but rather perhaps that one has reached the limitations of a tired and impoverished way of thinking about these issues.

CONCLUSION

Despite our various criticisms, we commend Horwitz et al. for highlighting that heritability estimates might not have the meaning often attributed to them. Their kind of critique, done properly, may well be useful for public debates. Yet serious consideration of gene-

environment interactions undermines the purported meaning of heritability estimates even more fundamentally. Horwitz et al. are exactly right that genes and environments interact to produce outcomes in ways that are “complex,” but for too long too many people have taken the mere recognition of this complexity as being intellectually sufficient. We believe that the details of this complexity can and will increasingly become an empirical object of study in a theoretical language that fully appreciates the co-implication of genes and environments, regardless of whether sociology participates. Sociologists may find this project and language appealing and even strengthening to their enterprise, but we suspect that this might be so only to whatever extent sociologists interject their voices into discussions. While we do not wish to discourage work that engages the “strongest assertions” of behavioral genetics on its own terms, we want also to make sure that sociologists recognize that these terms are not the only way in which the underlying issues can be conceived, and the search for more satisfying terms may be an endeavor ripe for genuinely meaningful sociological contribution.

NOTES

1. Although Horwitz et al. also trumpet the large size of the National Longitudinal Survey of Adolescent Health twin sample, it is considerably smaller than the samples that can be drawn from the twin registries cited in comparison.
2. Heterogeneous samples may also be particularly valuable in comparison to homogeneous samples that exclude environmental conditions considered particularly likely to have substantial psychological consequences (i.e., “high-risk” environments, see Rutter et al. 2001), although, in this regard, results from a specifically homogeneous subpopulation of adolescents in high-risk environments could also be quite interesting.
3. Our separate analyses of blacks and whites provided little evidence of divergent results, but the small sample sizes are inadequate to stake much confidence in these results.
4. Our re-analysis does not repeat two of Horwitz et alia’s analytic decisions with which we disagree: They treat a set of triplets as three independent dizygotic pairs, and they impute a value of 0 on parental education to missing values without including a separate dummy variable indicating imputation.

REFERENCES

- Allison, Paul D. 1990. “Change Scores as Dependent Variables in Regression Analysis.” *Sociological Methodology* 20:93–114.
- Goldberger, Arthur S. 1979. “Heritability.” *Economica* 46:327–47.
- Horwitz, Allan V., Tami M. Vidcon, 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.
- Joseph, Jay. 1998. “The Equal Environment Assumption of the Classical Twin Method: A Critical Analysis.” *Journal of Mind and Behavior* 19:325–58.
- Kennedy, Janice H. 1990. “Determinants of Peer Social Status: Contributions of Physical Appearance, Reputation, and Behavior.” *Journal of Youth and Adolescence* 19:233–44.
- Kenny, David A. 1996. “Models of Non-Independence in Dyadic Research.” *Journal of Personal and Social Relationships* 13:279–94.
- Maccoby, Eleanor E. 2000. “Parenting and Its Effects on Children: On Reading and Misreading Behavior Genetics.” *Annual Review of Psychology* 51:1–27.
- Plomin, Robert, John C. DeFries, Gerald E. McClearn, and Peter McGuffin. 2001. *Behavioral Genetics*, 4th ed. New York: Worth Publishers and W. H. Freeman.
- Rutter, Michael, Andrew Pickles, Robin Murray, and Lindon Eaves. 2001. “Testing Hypotheses on Specific Environmental Causal Effects on Behavior.” *Psychological Bulletin* 127:291–324.

Jeremy Freese is an Assistant Professor in Sociology at the University of Wisconsin-Madison. His research includes work on the relationship between biology and social outcomes, engaging primarily evolutionary psychology.

Brian Powell is the Allen D. and Polly S. Grimshaw Professor of Sociology of Indiana University. His areas of interest include sociology of education, social psychology, sociology of mental health and sociology of the family. Current work examines the educational and mental health implications of differential family structure (with Lala Carr Steelman and Jun Xu), the relationship between perceptions of equity in the family and emotional states (with Kathryn Lively and Lala Carr Steelman), the role that schools play in students’ weight gain (with Doug Downey) and the role that self-enhancing tendencies play in professors’ assessments of grade inflation (with Janice McCabe).