Genotype × Environment Interaction in Psychopathology: Fact or Artifact?

Lindon J. Eaves

Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University School of Medicine, Richmond, Virginia, United States of America

Recent studies have claimed to detect interaction between candidate genes and specific environmental factors (Genotype × Environment interaction, G × E) in susceptibility to psychiatric disorder. The objective of the present study was to examine possible artifacts that could explain widely publicized findings. The additive effects of candidate genes and measured environment on liability to disorder were simulated under a model that allowed for mixture of distributions in liability conditional on genotype and environment. Simulated liabilities were dichotomized at a threshold value to reflect diagnosis of disorder. Multiple blocks of simulated data were analyzed by standard statistical methods to test for the main effects and interactions of genes and environment on outcome. The main outcome of this study was simulated liabilities and diagnoses of major depression and antisocial behavior. Analysis of the dichotomized data by logistic regression frequently detected significant G × E interaction even though none was present for liability. There is therefore reason to question the biological significance of published findings.

Considerable excitement has been generated by recent claims to have detected interaction between the effects of specific genetic markers and specific environmental treatments in susceptibility to psychiatric disorders such as depression (Caspi et al., 2003) and antisocial behavior (Caspi et al., 2002). The appeal of this work has been strengthened by claims to have replicated the original finding (Eley et al., 2004; Foley et al., 2004; Kendler et al., 2005) and by arguments that such interactions are to be expected a priori from what is known, or reasonably conjectured, about the way the human nervous system mediates the impact of salient environmental insults.

If there is widespread Genotype × Environment interaction (G × E) for human behavior, then researchers have both an explanation of past failures to detect the effects of specific genes on behavioral disorders and a paradigm that would direct future efforts and funding.

However, statistical genetic studies of quantitative traits in nonhuman species (Mather & Jinks, 1982) provide significant reason to pause for reflection before embracing such findings as paradigmatic rather than artifactual. Although interactions are widespread in experimental organisms, their contribution is typically smaller than those of main effects. More importantly, even when quantitative traits are considered, the effects of G × E can be simulated by problems of measurement and, in some circumstances, can be generated or removed at will by a simple transformation of scale. In humans, transformations of psychological test scores are routinely conducted to remove apparent G × E that arises as a result of heteroscedasticity. Failure to recognize the sensitivity of genetic findings to scale of measurement can lead to unwarranted complexity in the apparent effects of genes and environment, including G × E, and even result in failure of more simple models for the effects of genes and environment (Eaves et al., 1989).

This note briefly illustrates a possible pathology of scale that could generate apparent specific G × E even though the main effects of specific genes and environments on an underlying quantitative trait are purely additive.

Methods

Model

The effect of measurement and method of analysis on the detection of G × E interaction was investigated through selected simulation studies.

Liability, X, to a given psychiatric disorder was assumed to be continuous. The probability that an individual of marker genotype, i, and measured environment, j, is affected is the probability that X exceeds a threshold of liability, t, that is,

$$P_{ij}(X > t) = \int_{t}^{\infty} \phi_{ij}(X)dX$$

where $\phi_{ij}(X)$ is the probability density function conditional on marker genotype i and measured environment j. For a given threshold, the probability that an individual is affected can be computed for specified $\phi_{ij}(X)$.

Received 1 November, 2005; accepted 7 November, 2005.

Address for correspondence: Lindon Eaves, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University School of Medicine, PO Box 980003, Richmond, VA, 23298, USA. E-mail: eaves@mail2.vcu.edu

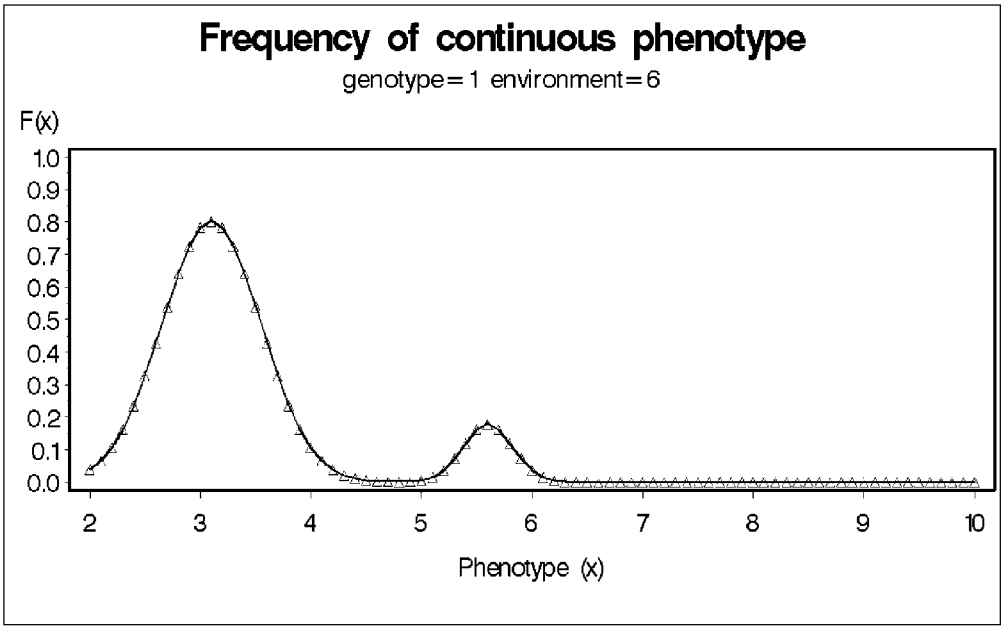

Figure 1

Expected distribution of liability, AA genotype.

A broad range of values for P_{ij} may be generated if $\phi_{ij}(X)$ is a mixture of two distributions for each i and j. For simplicity we assume the component distributions are normal. We let the first distribution, $\phi_{1ij}(X)$, be $N[\mu_{1ij}, o^2_{1ij}]$ and the second, $\phi_{2ij}(X)$, be $N[\mu_{2ij}, o^2_{2ij}]$. In the current application, we assume that the residual variances are the same for all marker genotypes and environments: o^2_1 and o^2_2 for the first and second component distributions respectively. This assumption can be relaxed in a more comprehensive analysis, as can the assumption of normality in the component distributions. Furthermore, we assume initially that the proportions of the components in the mixture depend on measured genotype but not on measured environment. We let p_i be the proportion of individuals of marker genotype i whose liabilities are sampled from the second component, leaving $(1 - p_i)$ as the proportion of individuals of that genotype who are sampled from the first.

In the absence of G × E for liability we may express the means, μ_{1ij} and μ_{2ij} as additive functions of the main effects of the markers, i, and environments, j. We write:

$$\mu_{1ij} = a + Y_i + \beta\varepsilon_j$$

where a is a constant term, Y_i the deviation due to the ith marker genotype, ε_j the value of the jth measured environment and β the regression of liability on measured environment.

In the absence of G × E interaction for the measured environment we may write:

$$\mu_{2ij} = \mu_{1ij} + \delta_i$$

where δ_i is the displacement of the mean of the second component from that of the first.

In concrete terms, we may consider the mean of first component as the principal characteristic of the genotype and the second distribution as yielding a 'bump' in liability due to other factors such as latent environmental insults that may or may not be correlated with the marker genotype, or to characteristics of the assessment process (e.g., subjects not taking the assessment interview equally seriously), the combination of 'gateway' and 'follow-up' items in a diagnostic interview, or even, under some circumstances, the effects of genotyping errors.

Figures 1 and 2 illustrate the expected distribution of liability for two genotypes for a given level of an environmental covariate. The constant is assumed to be zero. In both cases, it is assumed that individuals' environmental effects, ε_j, may take values 0, 1...10. β is assigned an arbitrary value 0.35, generating a range of increasing measured environmental effects from 0 through 3.5. The main effects of three genotypes on the first component of the mixture are assumed to be 1, 2 and 2 for genotypes *AA*, *Aa* and *aa*, respectively, that is, that the allele, a, increasing susceptibility is dominant. The mean of the second component is assumed to be the same for all genotypes in this case, generated by adding values of 3, 2 and 2 for δ_1, δ_2 and δ_3 respectively. The variances are assumed to be $\delta^2_1 = 0.2$ and $\delta^2_2 = 0.05$ for all genotypes. The probabilities that a randomly selected subject will be sampled from the second distribution are assumed to be $p_1, p_2, p_3 = .1$, .02 and .02 respectively.

Substituting these parameter values yields the distributions in Figures 1 and 2 for a measured environmental value of 3.0. Changing the value of ε_j will shift the distributions a constant distance to the

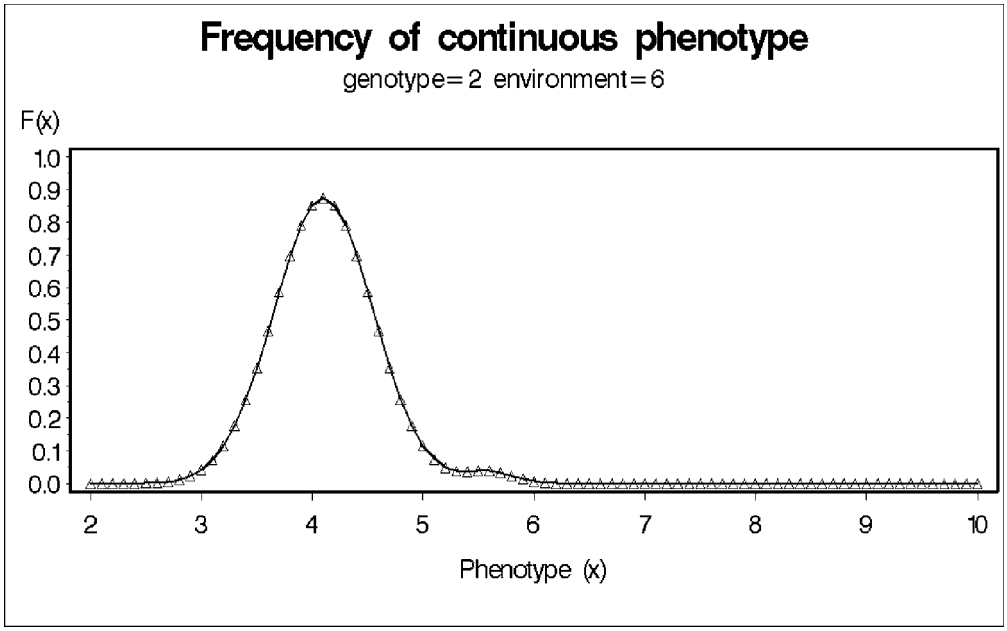

Figure 2

Expected distribution of liability Aa and aa genotypes.

right or left without altering their shape in the absence of G × E. Both distributions show the 'bump' to the right, resulting from the second component. The 'bump' is more distinct for the first genotype because of the higher relative frequency of the second component and the greater difference between the means of the two component distributions.

Figure 3 shows the probabilities of disorder by genotype and measured environment for a threshold value of $t = 3$ across the range of environments. These probabilities are conditional on genotype and environment. In particular cases, only certain restricted values of the environments will be assessed. The patterns of probabilities generated in a particular study, therefore, will depend, for a given threshold, on the specific values realized for the measured environments. Thus, if the measured environments in a particular study correspond to environments 2, 3 and 4 in the figure, the probabilities will increase and fan out as the severity of the environment increases. Environments 0 and 1 would yield probabilities that converge. Environments 0, 1 and 2 would produce a pattern in which the

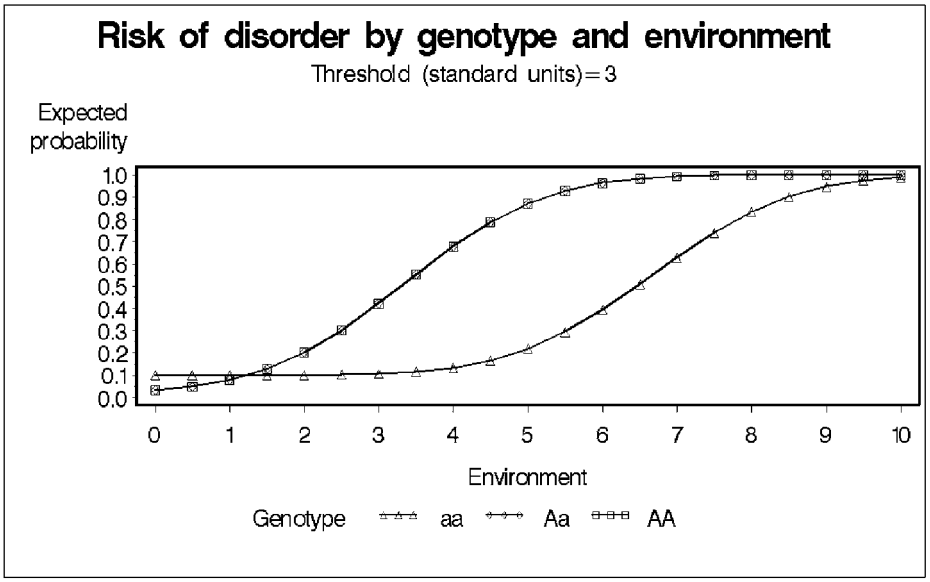

Figure 3

Expected frequency of disorder by genotype and environment.

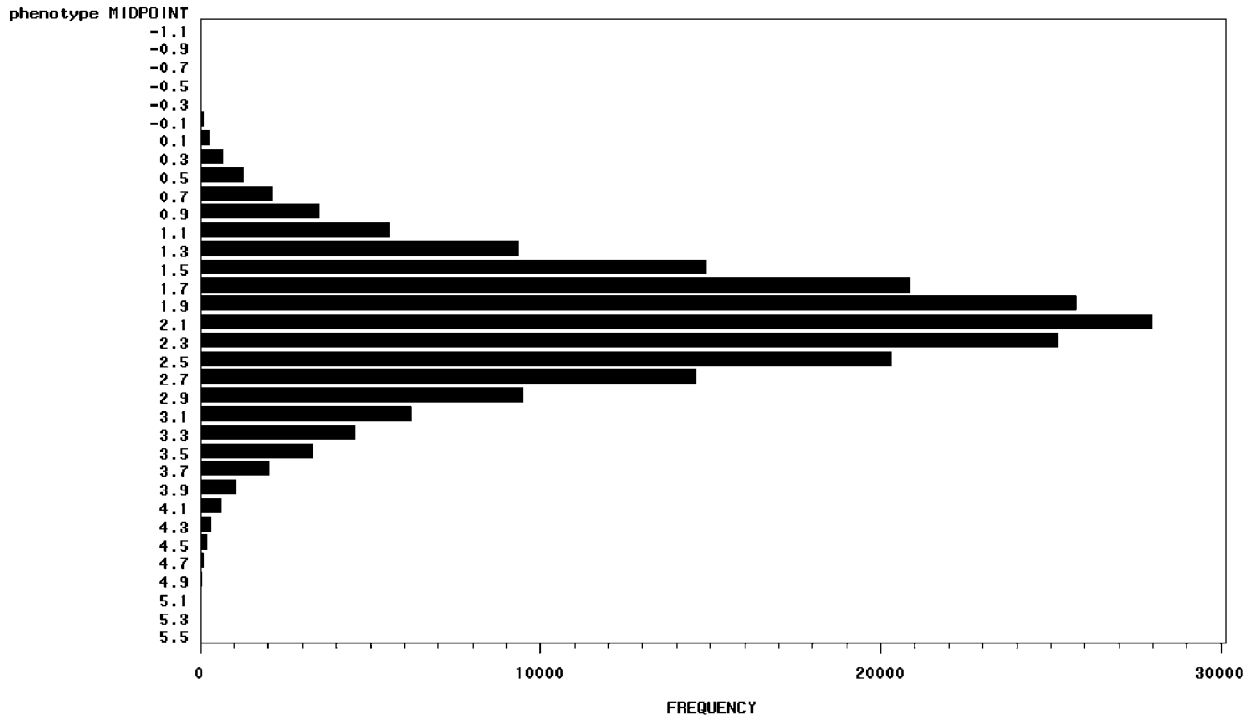

Figure 4

Distribution of 1,000,000 simulated phenotypes (4 environments).

regression lines cross over in a way that is remarkably similar to that shown by the published figures of Caspi et al. (2002) and Foley et al. (2004). The figures have an important implication at the outset, namely, that the same underlying distributions of residual differences in liability can yield a wide range of patterns for the probability of disorder simply as a function of the range of environmental influences chosen for study, including several different patters of G × E or no G × E at all. The 'cross-over' pattern will necessitate the inclusion of G × E interaction terms in any logistic regression model for the effects of genes and environments on liability to the disorder if the underlying distribution of residual differences in liability resembles that in Figures 1 and 2. However, such interactions are expected to be nonspecific and may even involve heterogeneity in the contribution of residual *genetic* effects and have nothing whatever to do with interaction between the candidate gene and specific measured environment.

Simulation

The same parameters used to generate Figures 1 to 3 were used for the simulation of liability and diagnostic data under two scenarios: (1) where environments 0, 1, 2 and 3 were represented; (2) where only environments 0, 1 and 2 were included. The frequency of the decreasing allele was assumed to be .3, yielding frequencies of .09, .42 and .49 for the AA, Aa and aa respectively. In the 'four-level environment' case, the environments were assumed to be present in frequencies of .70, .15, .10 and .05. In the 'three-level environment' model the frequencies were assumed to be .75, .15 and .10. Marker genotypes and measured environments were assumed to be independent, that is, there is no genotype–environment correlation.

Two kinds of study design were considered. In the first, 1000 random subjects were simulated and analyzed without selection. This corresponds roughly to the approach adopted by Caspi et al. (2002, 2003), Foley et al. (2004), and Kendler et al. (2005) in their reports of positive findings. A second study design was also considered in which analysis focused on the upper and lower 15% of the distribution of 2000 random subjects. This is the strategy employed by Eley et al. (2004) in their report of significant G × E. In both cases the simulations were repeated 100 times to assess the frequency of significant main effects of genes, environment and G × E interaction. The environments were considered to be ordered and the genotypes treated as nominal categories. Simulations and analysis were conducted in SAS (SAS Institute Inc., 2001).

Statistical Analysis

The first series of 100 simulations were analyzed in two ways. First, a linear model for the continuous liabilities was fitted and Type III (partial) sums of squares computed, together with regression coefficients and tests of significance for the linear model including the effects of genes, environments and G × E interaction. These analyses were followed by a logistic

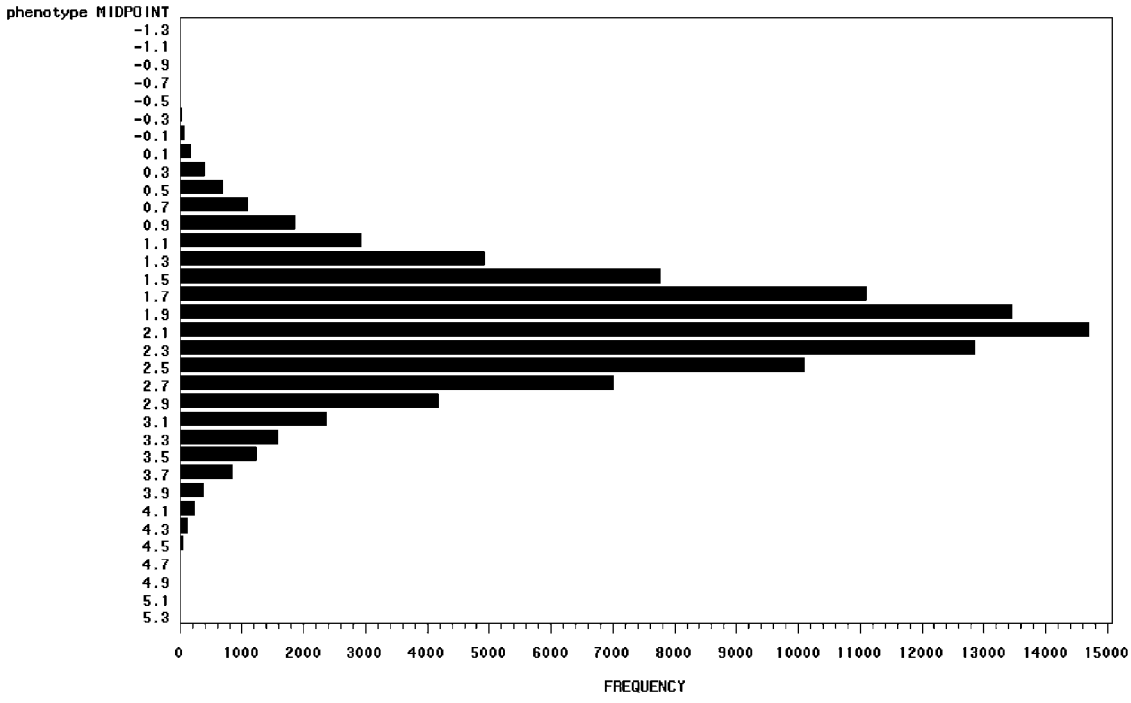

Figure 5

Distribution of 1,000,000 simulated phenotypes (3 environments).

regression analysis of the dichotomous outcome fitting the same model. The approach of logistic regression applied to a random sample was adopted by Caspi et al. (2002, 2003) and Foley et al. (2004). The second series of simulations selected the top and bottom 15% after inspection of the liability distributions and used a log-linear model to predict membership of the upper or lower groups as a function of the main effects and interactions of genes and environment. This the approach used by Eley et al. (2004) in their analysis of adolescent data selected on the basis of screening with the Mood and Feelings Questionnaire.

Results

Summary Statistics

Figures 4 and 5 illustrate the distributions of liability in 1,000,000 subjects simulated under the four-level and three-level environment scenarios. In both cases there is little hint of any underlying pathology of scale. The distributions are unimodal and much more symmetric than is usually found with symptom counts in psychiatric interview data.

Tables 1 and 2 summarize the preliminary statistics for the four- and three-environment models pooling over independent blocks of trials. The observed frequencies of the genotype–environment contributions are close to those expected from the marginal frequencies of the three genotypes and multiple levels of environment given genetic and environmental effects are independent. A cursory examination of the mean liabilities by genotype and environment confirm that the differences between environments are the same across genotypes and consistent with the absence of G × E at the level of liability, in spite of the marked bimodality of the residual genetic and/or environmental effects.

Detection of G × E Interaction

Tables 3 and 4 summarize the outcomes of fitting linear and logistic regressions to the liabilities and dichotomized outcomes for the 100 blocks of samples. When continuous liability is chosen as dependent variable and a linear model fitted by GLM on the assumption of normal errors, the power for the detection of the main effects of genes and environment is 100% in both sets of simulations. Tests of G × E are significant at the 5% level in 15% and 19% of samples, which is a significant excess over what is expected by chance alone. Thus, the pathology of measurement leads to a slight but significant excess in the frequency of significant tests of G × E, even though the underlying model is additive. However, the false positive rate for tests of G × E under the linear model for the raw scores pales into insignificance compared with what can happen when testing for G × E in logistic regression using diagnosis as the dependent variable. For the case of four environmental levels (Table 3), with admittedly relative high prevalence rates in the most severe environment (see Table 1), significant 'G × E' is detected in 70% of samples even though the raw data are simulated without any effects

Table 1

Summary Statistics for Simulated Data (*N* = 1,000,000, 4 Environments)

Genotype	Environment	*N*	%	Mean	*SD*	% affected
AA	0	62,933	6.30	1.26	0.87	10.0
AA	1	13,524	1.25	1.59	0.85	9.5
AA	2	8737	0.90	1.94	0.86	9.9
AA	3	4456	0.45	2.31	0.88	11.9
Aa	0	294,550	29.40	2.03	0.49	3.2
Aa	1	62,916	6.30	2.38	0.49	9.2
Aa	2	41,710	4.20	2.73	0.49	26.7
Aa	3	20,958	2.10	3.08	0.49	55.6
Aa	0	342,694	34.30	2.03	0.49	3.2
Aa	1	73,581	7.35	2.38	0.49	9.2
Aa	2	49,381	4.90	2.73	0.49	26.2
Aa	3	24,560	2.45	3.08	0.49	55.2
Sample		1,000,000	100.00	2.13	0.65	9.1

of G × E interaction on underlying liability. The apparent effect of G × E is still marked, but less serious under the model that omits the more severe environmental category (Table 4). In this case more than 25% of tests of G × E are still significant at the 5% level. If investigators succumbed to the temptation of reporting outcomes of borderline significance ($\alpha = 0.10$) the number of 'reportable' findings would increase to more than 40%.

The situation is worse when logistic regression is used to test for G × E in selected samples. In this case, 100% of the tests yield significant G × E when none was included at the outset. That is, treating selected samples as if they were random is grossly misleading with respect to the detection of G × E.

Comment

The simulation studies show that, contrary to expectations, logistic regression does not rescue investigators from false conclusions about G × E when none is present on the underlying liability of the disorder. Excluding the possibility of genotype–environment correlation does not remove the possibility of detecting spurious G × E interaction. Surprisingly, when there is reason to think the distribution of liability might be 'bumpy', using logistic regression to test for G × E might actually lead to a higher false positive rate than simply fitting a linear model to the raw scores on the assumption of normal errors. Given the way in which psychiatric diagnoses are generated from interview data and the various possibilities for genotyping error, there may be several ways of generating data of the type we have simulated that have little to do with the neurobiology of genes or environment.

The above considerations do not disprove published claims to have detected and, even, to have replicated specific G × E interactions in the etiology of psychiatric disorder. However, they do counsel critical reflection before such findings are used to justify a new paradigm for research in psychiatric genetics. Artifact may be as replicable as fact, perhaps even more so. When highly significant and apparently novel

Table 2

Summary Statistics for Simulated Data (*N* = 1,000,000, 3 Environments)

Genotype	Environment	*N*	%	Mean	*SD*	% affected
AA	0	67,388	6.75	1.26	0.87	10.0
AA	1	13,394	1.35	1.59	0.85	9.5
AA	2	8868	0.90	1.94	0.86	9.9
Aa	0	315,424	31.50	2.03	0.49	3.2
Aa	1	63,149	6.10	2.38	0.49	9.2
Aa	2	41,561	4.20	2.73	0.49	26.7
Aa	0	367,335	36.75	2.03	0.49	3.2
Aa	1	73,404	7.35	2.38	0.49	9.2
Aa	2	49,477	4.90	2.73	0.49	26.2
Sample		1,000,000	100.00	2.08	0.62	6.7

Table 3

Percentage of Significant Main Effects and Interactions Under 3 Analytical Scenarios, 4 Environments ($\alpha = 0.05$)

Sample	Phenotype	Genes (G)	Environment (E)	G × E	G × E. No G
Random $N = 1000$	Continuous	100%	100%	15%	0%
Random $N = 1000$	Dichotomous	55%	92%	70%	24%
Selected $N = 2000$	Dichotomous	99%	100%	100%	1%

findings are replicated in a discipline beset by ambiguity there may be good reason to consider alternative, less dramatic, explanations. The well-known possibility that pathologies of scale may masquerade as G × E interaction and/or epistasis has still to be considered in a critical evaluation of published findings.

Typically, it has been assumed that treating continuous data as categorical and using statistical methods such as logistic regression minimizes the chances of detecting G × E that is an artifact of scale. This may be the case when nonadditive effects on the latent trait reflect the heteroscedasticity bedeviling many efforts to measure behavior and its disorders. However, the above example shows that this is not necessarily the case. We have simulated data that generate figures remarkably similar to some of those published and replicated. The simulated data often show even the published pattern of significant main effect of the environment, no significant main effect of genotype, and significant G × E interaction. Furthermore, if the phenotype is analyzed as a continuous trait, using standard methods for linear modeling, there is typically no strong indication of G × E interaction. It is only when the outcome is dichotomized and analyzed with logistic regression that (spurious) interactions emerge. This is counter to what is commonly assumed, but cannot be ignored when so much is at stake. The situation is compounded when selected samples are treated as if they were random. In this case interactions are more likely to be found rather than not simply as an artifact of working with selected samples without correction for ascertainment.

It may be premature to claim detection of highly specific G × E interactions. Replication across studies means little because the same artifacts of scale or sample selection may apply to multiple studies. Likewise, a failure to replicate may imply nothing more than difference in the choice of measurement, different threshold for diagnosis, or even the selection of covariates that discriminate most effectively at different points on the scale of measurement.

How will the truth become clear? If current G × E is an artifact of measurement, it is predicted that interactions will be general rather than specific, that is involve other covariates of liability apart from those chosen for study. Other covariates that show significant main effects should also enter into interactions. Different patterns of interaction, or no interaction, may be detected with different measurements of the same construct. Interactions will not just involve G × E but also interactions between genes (epistasis) or even between environmental covariates themselves. Furthermore, the direction of interactions may be inconsistent across covariates (e.g., different life events) within a study or between studies as a function of severity and threshold. It is only after a more detailed and self-critical consideration of large and varied data sets, many of which are already available, that we will be sure that current findings cannot be explained by mechanisms that have little to do with the neurobiology of psychiatric disorder.

Whether we conclude ultimately that apparent G × E interaction reveals anything specific about neuorgenetics, widespread replication of recent findings would point to the fact that, if environmental factors are known to affect susceptibility to a psychiatric disorder, it makes sense to take these into account when attempting to identify the effects of specific genetic markers through association studies. However, depending on how the disorder is assessed and the specific choice of environments, the effects of a candidate gene and environment may appear as purely main effect, purely G × E interaction, or a mixture of both.

Acknowledgments

This work is supported by grants MH45268 (LJE) MH68521 (LJE) and MH55557 (JL Silberg) from the National Institute of Mental Health, and an award

Table 4

Percentage of Significant Main Effects and Interactions Under 3 Analytical Scenarios, 3 Environments ($\alpha = 0.05$)

Sample	Phenotype	Genes (G)	Environment (E)	G × E	G × E. No G
Random $N = 1000$	Continuous	100%	100%	19%	0%
Random $N = 1000$	Dichotomous	56%	97%	27%	1%
Selected $N = 2000$	Dichotomous	99%	100%	100%	0%

from NARSAD (JL Silberg). We thank Dr. Jonathan Kuhn for insightful discussion and Dr. Thalia Eley for her helpful response to an inquiry.

References

Caspi, A., McClay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., Taylor, A., & Poulton, R. (2002). Role of genotype in the cycle of violence in maltreated children. *Science, 297,* 851–854.

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., McClay, J., Mill, J., Martin, J., Braithwaite, A., & Poulton, R. (2003). Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene. *Science, 301*, 386–389.

Eaves, L. J., Eysenck, H. J., & Martin, N. G. (1989). *Genes, culture and personality: An empirical approach*. New York: Academic Press.

Eley, T. C., Sugden, K., Corsico, A., Gregory, A. M., Sham, P., McGuffin, P., Plomin, R., & Craig, I. W. (2004). Gene–environment interaction analysis of serotonin system markers with adolescent depression. *Molecular Psychiatry*, *9,* 908–915.

Foley, D. L., Eaves, L. J., Wormley, B., Silberg, J. L., Maes, H. H., Kuhn, J., & Riley, B. (2004). Childhood adversity, monoamine oxidase a genotype, and risk for conduct disorder. *Archives of General Psychiatry, 61,* 738–744.

Kendler, K. S., Kuhn, J. W., Vittum, B. S., Prescott, C. A., & Riley, B. (2005). The interaction of stressful life events and a serotonin transporter polymorphism in the prediction of episodes of major depression: A replication. *Archives of General Psychiatry, 62,* 529–535.

Mather, K., & Jinks, J. L. (1982). *Biometrical genetics: The study of continuous variation* (2nd ed.). London: Chapman Hall.

SAS Institute Inc. (2001). Statistical analysis system (Release 8.2) [Computer software]. Cary, NC: Author.