aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Perspectives on Psychological Science
2016, Vol. 11(1) 24–28

DOI: 10.1177/1745691615617442
pps.sagepub.com
SAGE

Weak Genetic Explanation 20 Years Later: Reply to Plomin et al. (2016)

Eric Turkheimer
Department of Psychology, University of Virginia

Abstract

Plomin, DeFries, Knopik, and Neiderhiser (2016, this issue) are correct in their assertion that many discoveries of behavior genetics have proven to be robust and replicable. I note, in contrast, that more specific assertions about the role of genetics in the development of behavior have failed to replicate. Reflecting on why more general findings replicate better than specific ones sheds light on the difficulties of studying complex human development and on the role played by genes in determining its course.

Keywords

behavior genetics, genetic influence, phenotypic null hypothesis, replication, significance testing

I propose a change in the way the major contention of Plomin, DeFries, Knopik, and Neiderhiser (2016; this issue) is framed. The first "replicated finding" of the target article is that all psychological traits show nonzero genetic influence; Irving Gottesman and I coined the First Law of Behavior Genetics (Turkheimer & Gottesman, 1991), stating that all psychological traits are heritable. Both of these statements are a step removed from the empirical observation on which they are based, which is as follows: The degree of similarity between two people on any trait is monotonically related to their degree of genetic relatedness. Putting things this way has several advantages. It avoids the vague causal implications of "genetic influence," about which more below; it removes from the formulation the fraught concept of heritability; and it unifies classical quantitative genetic observations in family members with newer methods based on low-level genetic relatedness in the general population of unrelated persons. Most important, however, it emphasizes that the observation is fundamentally correlational. Once the first law has been stipulated, the task becomes understanding why and how it occurs, as well as delineating its implications for the conduct of behavioral science and human self-understanding.

When I first wrote at length about this subject (Turkheimer, 1998), I proposed a distinction between weak and strong genetic explanation. Weak genetic explanation is the observation that, one way or another, genetic differences among people wind up correlated with phenotypic differences—that on the day we are born, we are not all equally likely to become extraverts or pianists or divorcés. This "finding" would not have surprised our great-grandparents: The apple does not fall far from the tree. Note in particular that the first law does not even require that some people are more capable than others of becoming pianists; it simply says that other things being equal, some people are more likely to do so. Weak genetic explanation of complex individual differences does not imply that those differences have genetic mechanisms for scientists to discover, and as I have already noted, weak genetic explanation extends to everything human. That the first law applies to divorce says nothing in particular about the explanation of divorce, other than that it is subject to the same probabilistic biological constraints as everything else.

Plomin et al. (2016; this issue) use the phrase "genetic influence" as a euphemism for weak genetic explanation. As far as I know, "genetic influence" has never been defined, but here is what I think it means: Suppose you have a database of hundreds of demographic characteristics of American cities. You compute a matrix of pairwise demographic similarities among the cities and find that cities with more similar demographics also have more

Corresponding Author:
Eric Turkheimer, Department of Psychology, University of Virginia, 1715 Mason Lane, Charlottesville, VA 22903
E-mail: ent3c@virginia.edu

similar crime rates. You could then say that crime rates show "demographic influence." To influence means to cause in ways we do not understand, to be correlated in ways that seem plausibly causal. In Turkheimer (1998), I noted that Plomin (1991) used the phrase "genetic influence" 17 times in the first five pages of their article. In the target article, the phrase is used 38 times.

Strong genetic explanation, on the other hand, is the discovery that an observed phenotypic difference is a manifestation of a specific latent genetic mechanism, what Paul Meehl (1977) called a "specific genetic etiology." The co-occurrence of dementia and choreaform movements in Huntington's disease has a strong genetic explanation: Both are the result of a mutation in a single dominant gene. Strong genetic explanations do not have to refer to single gene mechanisms, however. If it turned out that divorce was the result of a network of countable genes with specifiable neurological and then behavioral consequences, eventually compelling people to dissolve their marriages, our conception of divorce would have to change. It would be more than just adding divorce to the long list of characteristics that are genetically influenced; we would be amazed! What had previously seemed to be in the domain of complex human development and self-determination (a domain that we now know, if we did not know it before, comprises pervasive genetic influence) would be seen to be something different and less psychologically complex, more along the lines of a disease or a compulsion.

The history of behavior genetics can be seen as an extended attempt to proceed from weak to strong genetic explanation. The first hope for how this would come about was in the specification of traits that are more or less heritable. "Surely," Plomin et al. state, "it matters if heritabilities were just 5% rather than 50% or perhaps 95%" (p. 5). They do not say why they are so sure about this, but note that 5% and 95% are completely out of the range of heritabilities that actually occur for psychological traits (Polderman et al., 2015). The real discussion is about whether differences between heritabilities of 35% and 65% matter. In any event, trait differences in mid-range heritabilities do not replicate. Researchers investigating the behavior genetics of personality spent 50 years trying to decide which personality traits had the highest heritabilities before finally giving up. And even granting the authors' comparison of 5% and 95%, it matters only in a very specific way. Having two arms has a heritability of 0%. Is growing two arms not genetically influenced? Would it be a mistake to try and understand genetic processes underlying limb generation? Heritability differences mean exactly what they are supposed to mean: At a given place and time, extant genetic variation has a certain correlation with phenotype. Heritability is not a characteristic of a trait that can be discovered, so whatever behavior genetics is or is not, it cannot be just a catalog of heritability coefficients.

The next level of heritability analysis that was supposed to lead to strong genetic explanation was multivariate behavioral genetics, in which elaborate statistical models were used to get beyond simple assertions of heritability to specifications of developmental genetic mechanisms underlying complex human traits. Plomin et al. describe many of these efforts, which have documented that a substantial portion of the reason traits are correlated with each other, across people or within people across time, is that they share some of the genetic background posited by the first law. For scientists unable to randomly assign participants to experimental conditions, this observation has immediate methodological consequences, which represent the most indelible implication of human behavior genetics. If the children of mothers who speak to them in complex sentences do better in the third grade, it is genetically naive to conclude that linguistic complexity causes differences in school performance. Genetic background (and, for that matter, socioeconomic status) is a potential third-variable confound of the relationship. The most important use of twin models in social science is to provide a quasi-experimental grip on situations such as these (Turkheimer & Harden, 2014).

Multivariate quantitative genetics has not, however, produced much in terms of replicable genetic mechanisms that underlie behavior. The reason is telling: Yes, much of the covariance among traits is "genetic" in the weak sense. But the structure of the genetic variance, again and again, has turned out to be no different from the structure of the phenotype it was supposed to explain. For example, the authors bring up the comorbidity of schizophrenia and bipolar disorders. Genetics has not explained this "first fork in the diagnosis of psychosis," because the genetics of psychosis looks just like psychosis itself. Bipolar disorder and schizophrenia are correlated but differentiable, and so is their genetic background. This is the "phenotypic null hypothesis": Beyond assertions of heritability, beyond assertions of genetic background as a source of covariance among traits, most of the multivariate structure of genetic variance is no different than the structure of the phenotype itself (Turkheimer, Pettersson, & Horn, 2014).

Once again, this outcome has been especially clear in the genetics of human personality. Differences in phenotypic personality conform broadly to the familiar five-factor model. For half a century, the finest minds in the genetics of personality attempted to understand the structure of genetic variance in personality and how it might differ from the structure of environmental and/or phenotypic variance. The conclusion: It does not. When factor analyzed, genetic variance, environmental variance, and

phenotypic variance all produce the five-factor model, a phenomenon that McRae, Jang, Livesley, Riemann, and Angleitner (2001) referred to as the "puzzle of parallel structure" (p. 515). Loehlin and Martin (2013) concluded, "the structure of personality is inherent in the evolved phenotype, and is not the immediate consequence of either genetic or environmental organizing factors" (p. 761).

Note that Loehlin and Martin's alternative to genetic was not "environmental" but "phenotypic." The nature-nurture debate was supposed to be about whether differences in behavior were better characterized as genetic or environmental, but that turned out to be the wrong question. Everything is both genetic and environmental. The deeper question involves determining what is to be learned by examining phenotypic phenomena at a genetic level of analysis. The phenotypic co-occurrence of motor and cognitive deficits in Huntington's disease may be of some phenomenological interest, but it makes no sense without reference to the gene that causes both of them. In contrast, the activities of people involved in divorce proceedings can be examined at a genetic level of analysis, but (genetic influence notwithstanding) we do not anticipate a time when people will get genetic testing to help them understand difficulties in their marriages.

All of the foregoing prepares us to ask the key question: What should we expect from the modern genomic era's signature enterprise—the search for covariation between measured DNA and behavior? The distinctions here are subtle. In particular, what do we expect if the phenotypic null hypothesis is true? Under weak genetic assumptions, the phenotypic null hypothesis predicts the existence of small associations between DNA and behavior. Indeed, unless the twin studies were somehow mistaken, covariation between DNA and behavioral differences is inevitable. So any consideration of the contemporary genomics of behavior should forgo surprise at the discovery of small but significant associations: If genes influence behavior and sample sizes are large enough, significant associations between DNA and behavioral differences will be found. The important question is whether the associations will mean anything. The phenotypic null hypothesis predicts that the associations will be tiny and highly contingent on all sorts of other factors and will fail to add up to meaningful developmental theories of the behavioral differences. That is, of course, pretty much what the genomics of complex human behavior has looked like so far.

Plomin et al. rightly emphasize the replicability of genetic influence and promote the robust findings of behavior genetics as a tonic for the so-called replication crisis in the behavioral sciences. The role of behavior genetics in the replication crisis, however, is at once more complex and more interesting. It is true that the most general findings of genetic influence replicate, but attempts to parse the reliable variance components of the ACE model into specific etiologies have not succeeded, and indeed have produced some of the signature failures to replicate that generated the replication crisis in the first place. I have already mentioned that differences in heritabilites among traits do not replicate. Attempts to understand nonshared environmental variance in terms of specific developmental processes have not produced replicable results (Turkheimer & Waldron, 2000), confirming the most pessimistic implications of Plomin and Daniels's (1987) "gloomy prospect":

> One gloomy prospect is that the salient environment might be unsystematic, idiosyncratic, or serendipitous events such as accidents, illnesses, or other traumas. . . . Such capricious events, however, are likely to prove a dead end for research. More interesting heuristically are possible systematic sources of differences between families. (Plomin & Daniels, 1987, p. 8)

By far the most dramatic failure to replicate in behavior genetics has been the collapse of the project to decompose genetic influence into the causal effects of individual alleles via candidate-gene-association studies. Space does not permit a thorough analysis of why candidate-gene studies of complex human behavior did not produce replicable results, but one clear lesson can be drawn. In the domain of human behavior, replicability of results is related to generality of hypothesis. Here is the first law of developmental psychology: Older children perform better than younger children on all tests. And the first law of psychopathology: Individuals with mental disorders perform worse than unaffected controls. Sociology: Poverty is bad for you.[1] One would hope that the outcome of the replication crisis is more than a victory of reliability over validity, a retreat to very general hypotheses that replicate even when they do not inform. Hypotheses do not achieve nullity by being conspicuously false; they achieve it by being obviously true.

Recognition that generalities about human behavior are more replicable than specifics helps us understand something important about the replication crisis. It is not just the science of human behavior that fails to replicate; it is human behavior itself. It is difficult to formulate a replicable theory of how children respond to the divorce of their parents or how carriers of certain monoamine oxidase A (MAOA) alleles respond to environmental deprivation because human behavior is so exquisitely sensitive to the genetic and environmental context in which it occurs. Except in rare cases of outright fraud, failures to replicate do not result from scientific

ignorance or venality but, rather, from a combination of human developmental complexity and the impossibility of establishing experimental control over most of the phenomena of interest (Turkheimer, 2004). If scientists are responsible at all, failure to replicate is the unfortunate consequence of "*p*-hacking," the once widely accepted process in which samples are accumulated with an eye to multiple hypotheses until something finally becomes significant.

This last point sheds an interesting light on the methodologies that have evolved to fill the niche left by the abandonment of candidate-gene association. In genome-wide-association studies, data on hundreds of thousands of individual bits of DNA are collected in large samples and then searched for significant results at highly stringent *p* levels. If (as usually happens) no significant results are discovered the first time around, the process is repeated with even larger samples, continuing until something significant finally emerges. "Hits," as they have come to be known, are now being accumulated for many behavioral characteristics, but the effect sizes for individual SNPs or alleles are vanishingly small (Chabris et al., 2015).

But does this methodology sound familiar? Genome-wide association is unapologetic, high-tech *p*-hacking. In the modern era, when major social science journals discourage null-hypothesis significance tests and replication as opposed to significance has become an obsession, it is nothing short of odd that behavioral science at the bleeding edge of genomic technology has become an extended exercise in stringent but fundamentally old-fashioned significance testing. To assume that the current list of single nucleotide polymorphisms (SNPs) reaching significance for some behavioral trait will be significant again the *next* time someone collects DNA from 100,000 people is to make the most basic of errors about the relationship between statistical significance and replicable science. Some SNPs will replicate. Others will not. It will depend on context.

One of the many intellectual virtues of the first author of the target article is that he often concludes theoretical papers with predictions. Crossing his name with "I [or we] predict" in a bibliographic database provides an intellectually courageous history of the vicissitudes of mainstream behavioral-genetic theorizing over a 40-year period. Perhaps I can focus my differences with this distinguished group of scholars by concluding with a prediction of my own, which I actually made 15 years ago:

> The gloomy prospect looms larger for the genome project than is generally acknowledged. The question is not whether there are correlations to be found between individual genes and complex behavior—of course there are—but instead whether there are domains of genetic causation in which the gloomy prospect does not prevail, allowing the little bits of correlational evidence to cohere into replicable and cumulative genetic models of development. My own prediction is that such domains will prove rare indeed, and that the likelihood of discovering them will be inversely related to the complexity of the behavior under study. (Turkheimer, 2000, p. 164)

The day will soon be here when hundreds of thousands of individuals can be deep sequenced at low cost, and the first law guarantees that associations will be found at whatever level of statistical significance is deemed necessary. It will not make any difference, however, because educational attainment and divorce are not discernible entities at a genetic level of analysis. What we will see instead is a proliferation of small, diverse, contingent findings that do not accumulate into coherent scientific theories. These will not be robust findings with large effect sizes; they will be the signature of a complex problem being addressed at the wrong level of analysis. They will be the keyless sidewalk under the genomic streetlight.

I too am a behavior geneticist, so it is important to conclude this response with a "lest I be misunderstood" paragraph. It is remarkable that in this day and age there continues to be a school of thought maintaining that behavior genetics is fundamentally mistaken about even weak genetic influence, that the nearly universal findings of quantitative genetics can be dismissed because of methodological assumptions of twin studies (Joseph, 2014) or contemporary findings in epigenetics (Charney, 2012). Those arguments can be evaluated on their own terms, but my point of view must not be cited in their support. Genetic influence is real and has profound methodological implications for how human behavior is studied.

Even more important, when properly understood, behavior genetics frames a deep scientific question about the explanation of human behavior. Readers may or may not agree that molecular genetics will never, in principle, have anything useful to say about marital status. But molecular genetics already has had important things to say about macular degeneration, cystic fibrosis, and senile dementia. If any of us are to make useful predictions, they should be about what the future holds for schizophrenia and extraversion. Where will such ambiguously psychophysical entities end up on an axis of developmental complexity running from Huntington's disease to divorce? This, not genes versus environment, is the real question posed by behavior genetics. I am more skeptical than most of my colleagues about the reductive power of genetics to explain such things, but I recognize

that the scientific jury is still out. In the meantime, all I ask is that inevitable findings of weak genetic influence not be accepted as strong genetic explanations of complex human behavior while we wait for the progress of science to take its inevitable course.

Declaration of Conflicting Interests

The author declared no conflicts of interest with respect to the authorship or the publication of this article.

Funding

The author received support from National Institute of Aging Grant R03 AG048850.

Note

1. The sociologist Jeremy Freese has informed me that the first law of sociology is actually, "Some people do, and some people do not," but that is another story.

References

Chabris, C. F., Lee, J. J., Cesarini, D., Benjamin, D. J., & Laibson, D. I. (2015). The fourth law of behavior genetics. *Current Directions in Psychological Science*, *24*, 304–312.

Charney, E. (2012). Behavior genetics and post-genomics. *Behavioral & Brain Sciences*, *35*, 331–358.

Joseph, J. (2014). *The trouble with twin studies: A reassessment of twin research in the social and behavioral sciences*. New York, NY: Routledge.

Loehlin, J. C., & Martin, N. G. (2013). General and supplementary factors of personality in genetic and environmental correlation matrices. *Personality and Individual Differences*, *54*, 761–766.

McRae, R. R., Jang, K. L., Livesley, W. J., Riemann, R., & Angleitner, A. (2001). Sources of structure: Genetic, environmental, and artifactual influences on the covariation of personality traits. *Journal of Personality*, *69*, 511–535.

Meehl, P. E. (1977). Specific etiology and other forms of strong influence: Some quantitative meanings. *The Journal of Medicine & Philosophy*, *2*, 33–53.

Plomin, R. (1991). Human behavioral genetics. *Annual Review of Psychology*, *42*, 161–190.

Plomin, R., & Daniels, D. (1987). Why are children in the same family so different from one another? *Behavioral & Brain Sciences*, *10*, 1–16.

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2016). Top 10 replicated findings from behavioral genetics *Perspectives on Psychological Science*, *11*, 3–23.

Polderman, T. J., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, *47*, 702–709.

Turkheimer, E. (1998). Heritability and biological explanation. *Psychological Review*, *105*, 782–791.

Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science*, *9*, 160–164.

Turkheimer, E. (2004). Spinach and ice cream: Why social science is so difficult. In L. DiLalla (Ed.), *Behavior genetics principles: Perspectives in development, personality, and psychopathology* (pp. 161–189). Washington, DC: American Psychological Association.

Turkheimer, E., & Gottesman, I. I. (1991). Is H 2= 0 a null hypothesis anymore? *Behavioral & Brain Sciences*, *14*, 410–411.

Turkheimer, E., & Harden, K. P. (2014). Behavior genetic research methods. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 159–187). Cambridge, United Kingdom: Cambridge University Press.

Turkheimer, E., Pettersson, E., & Horn, E. E. (2014). A phenotypic null hypothesis for the genetics of personality. *Annual Review of Psychology*, *65*, 515–540.

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