Research in Human Development

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hrhd20

Still Missing

Eric Turkheimer [a]

[a] University of Virginia

Available online: 18 Nov 2011

To cite this article: Eric Turkheimer (2011): Still Missing, Research in Human Development, 8:3-4, 227-241

To link to this article: http://dx.doi.org/10.1080/15427609.2011.625321

Research in Human Development, *8*(3–4), 227–241, 2011

ISSN: 1542-7609 print/1542-7617 online
DOI: 10.1080/15427609.2011.625321

Psychology Press
Taylor & Francis Group

Still Missing

Eric Turkheimer
University of Virginia

The missing heritability problem refers to the gap between heritability estimates for complex human traits based on quantitative genetics and the small magnitude and unreliability of contemporary molecular genetics, especially genome wide association studies. The author reviews the origins of the missing heritability problem and considers research that has attempted to resolve it by quantifying the joint explanatory power of multiple genetic loci, rather than considering their effects one at a time. Although this program has made an important contribution to understanding the role of genetics in the development of complex behaviors, it does not resolve the missing heritability problem.

HERITABILITY FOUND AND LOST

A century of familial studies of twins, siblings, parents and children, adoptees, and whole pedigrees has established beyond a shadow of a doubt that genes play a crucial role in the explanation of all human differences, from the medical to the normal, the biological to the behavioral (Turkheimer, 2000). The era in which these studies were conducted was a heady time for biologically or genetically oriented theorists, as one study after another showed that heritability extended not only to relatively well understood single gene disorders, but also to complex diseases like diabetes and dementia, continuous normal characteristics like height, and continuous and potentially disordered characteristics like weight. Psychiatric disorders and individual differences in normal behavioral characteristics that might once have been thought to be more inherently "psychological," especially intelligence and personality, were shown to be almost as heritable as height and weight. It seemed to many at the time that the world was turning out

Address correspondence to Eric Turkheimer, University of Virginia, Department of Psychology, P. O. Box 400400, Charlottesville, VA 22904-4400. E-mail: ent3c@virginia.edu

to be a much more genetic place than anyone had thought, to the satisfaction of those who were sanguine about the revolution in biological explanation that everyone assumed would follow on the heels of heritability (Plomin & Crabbe, 2000), and to the dismay of those who were not (Collins, Maccoby, Steinberg, Hetherington, & Bornstein, 2000).

At least four methodological concerns ran counter to the prevailing biogenetic mood of the time. One entailed long-standing objections to the foundational assumptions of quantitative genetics (Lewontin, Rose, & Kamin, 1984). Identical twins reared together might experience more environmental similarity than fraternal pairs (the equal environments assumption, e.g., Feldman & Otto, 1997). Epigenetic studies have shown that monozygotic twins may not be quite so genetically identical after all (Petronis, 2006). Adopted children are not assigned to their adoptive environments at random (selective placement), and adoptive environments are strongly selected for minimal environmental quality (Stoolmiller, 1999). Such concerns have been intensely debated elsewhere and will not be my central concern here.

The second concern also involved the statistical assumptions underlying twin and family models, but in a different way. Classical Mendelian genetics is, of course, discrete. Mendel's crossing of smooth and wrinkled peas did not produce moderately wrinkled peas, but a mix of wrinkled and smooth peas in the proportions dictated by the laws of classical genetics. Other than rare single-gene disorders like Huntington's disease, however, the kinds of human characteristics that are the object of nature–nurture discussions deviate from this pattern. Many human characteristics, with height the obvious example, are continuous rather than discrete, and offspring show values intermediate to parents rather than segregating, say, into tall and short varieties. And rarer (although still common compared to single-gene disorders) syndromes like schizophrenia, although they indisputably run in families for genetic reasons, do not appear in pedigrees in a way that would suggest that a single or even a small number of genes is segregating; instead they appear regularly but sporadically.

Squaring phenomena that are continuously distributed and sporadic with Mendelian genetics was accomplished with the Fisher–Wright (FW) model, which is the foundation of modern quantitative genetics. The FW model assumes that complex human traits are the result of a very large number of segregating genes, each with very small, indeed infinitesimal, effects. According to the binomial theorem, when these infinitesimal effects are summed they produce a normally distributed phenotype, with offspring values intermediate to parental values. For traits like schizophrenia that appear as categorical entities, one can add a threshold to the normal distribution resulting from the binomial process, with offspring having scores greater than the threshold displaying the trait. Predictions from such a model match the appearance of complex disorders in pedigrees more closely than models that posit segregating genes of large effect.

A third concern followed from not from the statistical assumptions of the FW model, which has been extraordinarily successful as a statistical method for the modeling of complex genetic effects in families, but rather from the theoretical implications of FW model for the scientific investigation of genetic and environmental developmental processes that actually produce the complex phenomena of interest. Ironically, it is precisely the aspect of the FW model that makes it so successful on a statistical level—the fractioning of genetic variance components into a collection of indistinguishable binomial effects—that makes it problematic as a model for the actual genetic etiology of traits. It is difficult to think of a biomedical disorder or behavioral trait with an etiology that is meaningfully described as an additive accumulation of a large number of very small independent causes, however useful such a model may be for modeling the occurrence of disorders and traits in pedigrees.

It is telling that the theorists who have been most concerned with the etiological limitations of the FW model are not the same environmentalists, antireductionist philosophers or wet-lab developmental biologists who are so suspicious of the statistical assumptions of the classical twin model. Rather, the big worries are expressed by the geneticists and experimental psychopathologists who have set themselves the task of unraveling the etiology of the disorders so clearly shown to be genetic by the FW model. So, for example, Fuller and Thompson (1978), two of the founding fathers of modern behavioral genetics, stated:

> Monogenic theories suggest major biochemical pathways which can be uncovered, whereas polygenic models suggest a complexity of chemical interactions probably intractable to exact study. Thus if most behavior traits must be fit to polygenic models, we may be left only with statistical analyses of such problems as how many genes are involved and the specification of the almost infinite number of interactions between them. Such mathematical exercises seem to us to have only trivial importance and, furthermore, to be of small interest to most biologists and psychologists. (p. 438)

Or, we can turn to the greatest theoretical psychopathologist of the previous century, Paul E. Meehl (1972a):

> One understands fairly clearly what it means to conjecture that a "big-effect monogene" is the specific etiology of a disease . . . But once we have excluded that simple situation, the very meaning of the phrase "specific etiology" begins to "fuzz up." (p. 376)

Meehl's point is particularly interesting in this regard. Meehl knew as much about schizophrenia as anyone on the planet, yet he spent his career chasing a single-gene theory the incorrectness of which was increasingly obvious then, and

decisively so now (Meehl died in 2003). If there were a gene of large effect, of medium effect, of greater-than-tiny effect, for schizophrenia we would have found it by now. Why cling to such a theory in the face of mounting evidence against it? The (admittedly speculative, based on a conversation with the great man shortly before his death) answer is that Meehl was always deeply concerned with scientific epistemology, and he could not bring himself to accept the etiological implications of the FW, which are essentially that schizophrenia doesn't have anything that one would reasonably wish to call an etiology, or what Meehl (1972b) would have called a specific genetic etiology. According to a strict interpretation of the FW model, schizophrenia is just a syndrome that some people are more likely to experience than others, depending on the number of a large and undifferentiated set of genes they happen to have inherited. A single gene theory may have been wrong, but at least it was falsifiable.

A final concern about the success of quantitative genetics is that it has been almost too successful. It was one thing when all the obvious candidates for genetic etiology were demonstrated, one after another, to be heritable: height of course, followed by weight or body mass index, intelligence and personality, schizophrenia and depression. But, with all the usual subjects exhausted, and many dissertations left to write, people conducted twin studies of less likely candidates, and to a troubling degree these all came out to be heritable as well. How much television children watch is heritable (Plomin, Corley, DeFries, & Fulker, 1990). Political attitudes are heritable (Alford, Funk, & Hibbing, 2005). Divorce is heritable (McGue & Lykken, 1992). There has been no end to it, especially as twin studies have expanded into the yet-to-be exhausted areas of economics (risk preference: Zhong et al., 2009) and political science (voting: Fowler, Baker, & Dawes, 2008).

The utter universality of heritability muddies the relationship between the statistical and etiological aspects of the FW model. Presumably, no one expects that a genetic etiology of divorce is about to emerge, notwithstanding its heritability. (Although the reports from economics and political science, having failed to learn the available lessons from a century of twin studies in psychology, still conclude with speculation about undiscovered biological etiologies of voting behavior.) The possibility that so troubles psychopathologists, that an outcome is simply an event that becomes more likely with the accumulation of the tiny effects of a very large number of undifferentiated genes, makes perfect sense for divorce, or at least a lot more sense than the possibility of discovering meaningful "divorce genes" with attendant "divorce pathways" leading to "divorce circuits." The universality of heritability teaches us that heritability is not incompatible with what we have always thought of as the psychological, and conversely that there is no reason to infer from heritability that reduction of a complex characteristic to genetic or neurological structures has become more likely (Turkheimer, 1998). Divorce can be predicted from genes,

but it can't be explained by them. (For a lucid history of discussions of this point, see Griffiths & Tabery, 2008.) I will return to this point at the end of the article.

THE MISSING HERITABILITY PROBLEM

It was widely thought that the Human Genome Project would deliver the vindication of quantitative genetics, especially as it applied to human behavior (Plomin & Crabbe, 2000). Everyone assumed that once the human genome was sequenced the "genes for" the phenomena that had been demonstrated to be heritable would be just around the corner, but it hasn't happened. Early findings of linkage for the major heritable psychological disorders and traits—schizophrenia and depression, intelligence and personality—with specific genetic markers were small, unfocused genetically in that they identified regions of chromosomes rather than specific genetic variants, and unreliable in replication. The "candidate gene studies" that followed targeted particular genes, but the results were still small and even harder to replicate.

Most recently, genome wide association studies, or GWAS, have become the dominant paradigm for gene finding. GWAS use single nucleotide polymorphisms, or SNPs, individual segments of DNA nucleotides for which variation among individuals only includes two alleles from the available four: A, T, C, or G. SNPs are indicators of genetic variation, in the sense that variation in SNPs is an imperfect marker for variation in functional genes. To the extent variation in SNPs is associated with variation in genetic variants that in turn covary with a phenotype (such genetic covariation is referred to as linkage disequilibrium, or LD), SNPs provide a comprehensive set of indicators of relations between genes and an outcome of interest. Currently available technology can place upward of a million SNPs on a single chip. The availability of these inexpensive chips renewed expectations that the discovery of the allelic molecules underlying genetic variation, and thus the establishment of the genetic etiology of complex traits, was finally at hand, but once again it hasn't come to pass. What has happened instead is that for any given characteristic we have discovered a handful of SNPs that appear to be in LD with an unknown but certainly very large number of genes, which are more or less predictably correlated with an outcome of interest. For any given SNP, effect sizes are generally less than 1% of the phenotypic variance, even for something as uncontroversially heritable as height. The genetic mechanisms of the major heritable syndromes and traits have not been found.

This gap between the predictive and explanatory power of genes has come to be known as "missing heritability," and it is the central conundrum of modern genomics. The phrase, apparently coined by Maher (2008), refers to the gap between the enormous success of population genetics in establishing that

all human differences are heritable, and the disappointing paucity of findings to have subsequently emerged from the sequencing of the human genome about the same human traits. The important question is: What exactly is missing? It is not really heritability per se. That is to say, there are no longer legitimate concerns about whether schizophrenia, depression, and intelligence, let alone height, are heritable. No one expects that once all of the GWAS are done, it will turn out that the critics were right all along, and the apparent heritability of height was merely the result of violations of the assumptions of twin studies. Point estimates of heritability coefficients may not mean very much and have few implications for the existence of meaningful biological etiology, but as to the general question of whether height runs in families for genetic reasons, only a methodological pedant or a die-hard environmentalist would make much of a fuss. Of course height is heritable, in the simple sense that it is possible to predict the heights of children from the heights of their biological parents, and, in the absence of widespread disease or malnutrition, to do so as well for adopted-away children as for children raised in their biological homes. What is missing from contemporary genomics is not heritability, but a meaningful link between statistical and etiological models of the transmission of complex traits.

Height is the poster child of the missing heritability problem, as I discussed in detail elsewhere (Turkheimer, 2011). Adopted almost as a proof of concept by the genomic community, a 2008 issue of *Nature Genetics* published three reports of GWAS for height (Gudbjartsson et al., 2008; Lettre et al., 2008;Weedon et al., 2008), along with a summary article by Visscher (2008), who went on to play a key role in almost all of the studies that followed. The consortium of studies included 65,000 participants, each with a half million SNPs. None of the individual SNPs accounted for as much as 1% of the variation in height, and the handful of SNPs that reached statistical significance in more than one of the studies accounted for around 3%. Whether this result should be regarded as a failure or as a first step toward greater successes was the topic of my earlier paper on this topic; in either case, it was emblematic of the missing heritability problem. In the remainder of this article I address one aspect of height genomics that has been the topic of intensive investigation since the original reports were published: How can one combine the effects of multiple SNPs to obtain an omnibus estimate of their effect?

PARTITIONING HERITABILITY

Although theories like Meehl's that posited the genes of large effect for complex syndromes and traits were once more fashionable than they are today, we have noted above that the FW model that is used to compute heritability specifically

assumes radically polygenic effects. FW models have always fit family data better than models that involve a segregating gene of large effect, or even of a countable set of "oligogenic" effects, as they were briefly and optimistically called as a stopgap against the rising tide of infinitesimal polygenic effects that eventually became the missing heritability problem (Gershon, 2000). It should come as no surprise, then, that the effect of individual genes or SNPs, are small, and this realization quickly gives rise to another question: How can one compute the joint effect of multiple SNPs on a phenotype? Once again, height has become the prototype for understanding the effects of multiple SNPs. A single investigator, Peter Visscher of the University of Melbourne, has played a leading role in the development of the statistical and genetic procedures to address the problem. His name appears in one role or another on practically every paper that has been published on the topic. For convenience I refer to the overall research paradigm as Visscher's.

There are a number of ways to quantify the collective effect of multiple SNPs, from the very simple to extremely complex. The most basic is to count them, computing a summed index of the number of height-associated SNPs (each SNP can have a score of 0, 1, or 2, depending on how many of the "tall" values each participant inherits from his or her mother and father) that achieved statistical significance in a particular study. In Weedon et al. (2008) the sum of 20 SNPs that had reached the rigorous levels of significance employed in such studies accounted for 2.9% of the variation in height, or a difference of about 2 inches in the heights of individuals falling at the lowest and highest 5% of the SNP scale. In a similar analysis in Lettre et al. (2008), 12 SNPs explained 2% of the variation and about $1\frac{1}{2}$ inches between the extremes.

The difficulty with such methods is their complex relationship with the notion of statistical significance as it is employed in GWAS. Because the number of potential associations in GWAS is so vast, investigators have had to adopt herculean levels of significance for each one (10^{-8} is the current standard), resulting in very low statistical power for all but the largest associations, which are not large to begin with. So it is not surprising that the set of SNPs clearing such a high hurdle does not account for much of the genetic variation that twin studies show to be there. Goldstein (2009) estimated the overall distribution of effect sizes of SNPs for height and concluded that 93,000 would be required to account for the full 80% heritability of height.

A more comprehensive method was employed in another extensive meta-analysis of height GWAS conducted by Lango-Allen et al. (2010). Lango-Allen et al. combined results from 46 studies of height, including a total of more than 133,000 participants. A series of successively more stringent filters was applied to identify 180 SNPs reaching genome-wide significance at $p < 5 \times 10^{-8}$. These SNPs were weighted and combined to predict height using least squares fixed effect regression and accounted for 10.5% of the variation in height. Relaxing the

significance threshold to $p < 0.05$ increased the proportion of variance explained to 13.3%. Inferential methods were employed to estimate that a total of 697 loci would collectively account for 15.7% of the variation, and that a half million participants would be required to have 99% power to detect them.

An even more sophisticated method for understanding the relationship between multiple SNPs and complex outcomes has recently been demonstrated by Yang et al. (2010). This method offers the clearest view of the possibilities of accumulating effects over multiple SNPs and the implications of doing so. In particular, the article by Yang et al. (2010) represents the most complete attempt to date to account for the heritability of height at the level of SNPs. Moreover, it was accompanied by an invaluable explanatory paper (Visscher, Yang, & Goddard, 2010) that clarifies the methods that were employed and the authors' views about their implications.

Yang et al. (2010) began by stating that there are only two possible explanations of the missing heritability problem when it is construed literally as the difference between the heritability of height when computed from family studies and the percentage of variation accounted for by an appropriately weighted set of multiple SNPs. Either some of the relevant SNPs have not been included in the analysis because of the statistical power problems discussed above, or the set of SNPs on the chip are not in complete linkage disequilibrium with the "causal variants," that is, the genes themselves, which are the actual genetic basis of height. The meaning of the word *causal* in this phrase is not entirely clear and will be important in what follows. Nevertheless, it is important to remember that SNPs are mere indicators of the actual genetic sequence that is the ultimate object of interest, and like any observed indicator of a latent construct, they are imperfect.

Yang et al.'s (2010) methods have more in common with quantitative genetic twin and family studies than the count- or regression-based methods of summing SNPs that had been previously employed. Yang et al. showed that the complete set of SNPs on a chip, not selected at all for the magnitude of the individual associations with height, can be used to estimate a pairwise matrix of genetic relatedness among all participants in the study. The individuals in the study are not, in fact, genetically related in the usual sense, and Yang et al. removed any pairs that appeared to be related to each other. The relatedness they analyzed is the "background" relatedness of all humans. However, the estimate of relatedness among individuals based on similarity of their SNPs is limited by the extent of the LD between the SNPs and the "causal variants." This allows the researchers to partition heritability into two parts, one that is reproduced by the LD between SNPs and the causal variants, and another that is not. Yang et al. found that 45% of the phenotypic variation in height can be "captured" by pairwise similarity based on the SNPs. The remainder of the 80% heritability of height, they suggested, is consistent with a model in which the rest of the variation arises from a large

number of causal variants that have a minor allele frequency (MAF, the proportion of the population with the less common variant) of less than .1.

HAVE WE FOUND THE LOST HERITABILITY?

So is the missing heritability problem solved, at least for height? The answer to the question depends on exactly what the missing heritability problem consisted of in the first place. From the outset, it was known that identical twins, who share among other things 100% of their SNPs, were correlated at greater than $r = .8$ for height; fraternal twins who share on average 50% of their SNPs, are correlated only slightly more than one half that amount, and so forth down the quantitative genetic scale of relatedness. The Visscher program has demonstrated how the statistical relationship between genotype and height is manifested at the level of genetic variants and SNPs: less than 1% attributable to any single SNP, around 10% attributable to linear combinations of statistically significant SNPs, around 45% attributable to all genetic variants in LD with "hundreds or even thousands" of SNPs on currently available chips, and the remainder attributable (somewhat more speculatively) to rare genetic variants. Visscher et al. (2010) state that their study is "the first to show that at least half the heritability for height . . . is captured by common SNPs."

Of the three reservations about quantitative genetic heritability that were outlined at the outset—the assumptions of twin and family studies, the universality of heritability, and the absence of mechanism—the new paradigm has put the first to rest, and before continuing to explain my skepticism about whether the most important problems have been solved, it is worth appreciating what a significant accomplishment this is. Thanks to the Visscher program of research, it should now be impossible to argue that the whole body of quantitative genetic research showing the universal importance of genes for human development was somehow based on a sanguine view of the equal environments assumption in twin studies, putting an end to an entire misguided school of thought among traditional opponents of classical quantitative (and by association behavioral) genetics (e.g., Joseph, 2010; Kamin & Goldberger, 2002).

The twin study critics may have been wrong to assert that the heritability of voting behavior and divorce was illusory, but the intuition that motivated their objections—that the quantitative genetics of voting does not tell us much about why people vote, once you get past the general surprise that everything is heritable—remains valid and is not significantly assuaged by the Visscher program. Return to the use of the word *captured* in the sentence from Visscher et al. (2010) quoted above. What does it mean for a variance component like heritability to be "captured by" common SNPs? Presumably, it denotes something like the ambiguous language about variance "accounted for" or "explained" in analysis of

variance, referring to a noncausal accounting of variability in terms of statistical correlations, as critiqued by the most important paper of the antitwin study era, Lewontin's (1974) classic, "The Analysis of Variance and the Analysis of Causes" (see also Turkheimer, 2000). The Visscher program accounts for the variability of height but does not identify its causes.

As noted above, the word *causal* appears in the Yang et al. (2010) and Visscher et al. (2010) articles, but in a very particular context, in the phrase *causal variants*, which Visscher et al. used to refer to variation in the alleles for which SNPs are imperfect indicators. Visscher et al. did not define what they mean by the word *causal*, but from the context it is apparent that the causal variants for height are just the alleles of genes that are statistically associated with height. That is, there are no noncausal variants, in the sense of genes that are associated with an outcome spuriously, in a way that would not be regarded as causal.

The Visscher et al. (2010) paradigm thus circumvents the core inferential problem in the science of complex human characteristics, which is how to identify true causal relations in the inchoate network of small associations that typifies phenomena that are highly complex in their genesis and not amenable to randomized experimental control. Visscher et al. estimated the total genetic causal effect in the limited sense that all genetic associations are considered biologically foundational and thus causal in general, but that is exactly the same sense as established by twin studies: somehow, a vast number of genes with very small individual statistical effects are inputs into a developmental process that eventuates in height. In the heyday of twin studies, investigators used to state that twin-based heritability demonstrated that genes "influence" heritable phenotypes (Turkheimer, 1998), as a way of emphasizing the undeniable statistical relation between genetic and phenotypic variation while fudging the vastness of the causal gap between them. Divorce is heritable, but how does variation in genes produce variation in marital status? Twin studies proved incapable of providing a meaningful answer, and it was this shortcoming that molecular genetics was supposed to remedy.

The Visscher et al. (2010) program should drive a stake through the heart of a classical line of argument against classical behavioral genetics and its attendant statistical assumptions. Nevertheless, it is difficult to see how their will make much of an impact on the more contemporary problem, which is that quantitative genetics, despite demonstrating the universality of heritability, has failed to offer much in the way of etiological insight into complex behaviors, and moreover that the very ubiquity of heritability has made it problematic to differentiate between heritable phenotypes that have genetic mechanisms and those that do not.

FUTURE PROSPECTS

As this article went to press, the importance of understanding the theoretical implications of this research program was reinforced, as another Visscher-led

group followed in the footsteps of twin researchers by extending the paradigm from the domain of intuitively genetic physical characteristics like height to an even more complicated and certainly more controversial phenotype: human intelligence (Davies et al., 2011). It is not difficult to summarize what they found. It is heritable, and its heritability can be established using SNP-based genetic similarity just as it can with identical and fraternal twins. But there is no evidence of anything even resembling a "gene for" intelligence, and no promising signs that all the infinitesimal genetic associations are about to produce a meaningful genetic account of the development of intelligence. The heritability of intelligence is detectible, and it is a little less deniable on methodological grounds than it used to be, but it is still "missing" in the sense I outline here.

Technically advanced though it may be, the Visscher program is still in its early stages, at roughly where quantitative genetics was during the 20th century, estimating the heritability of something like height, for which some kind of underlying genetic architecture is uncontroversial, however difficult it has proved to specify. What if we followed the twin researchers to the next step and conducted the exact same research program for marital status? Twin studies have shown that marital status is perhaps half as heritable as height, in the range of 40%. If we collected SNPs on a very large sample of divorced and nondivorced individuals, why wouldn't we expect the same result Visscher et al. (2010) obtain for height or intelligence? Roughly one half of the 40% would be "captured by" the SNPs, and the rest would be attributable to imperfect LD between the SNPs and the "causal" variants. We could then conclude with Visscher et al.

> Our results show that half the genetic variance is tracked by common SNPs and this variance is split among hundreds or even thousands of SNPs. The remaining half of the genetic variance could well be split among a large number of causal variants with MAF < 0.1 but there is no reason to believe that this missing variance is explained by only a few variants and therefore most must explain a small amount of the variance. (p. 521)

We could then anticipate the truly enormous population studies of marital status that would be required to power discovery of the individual genetic effects underlying divorce.

Further consideration of the genomics of divorce elucidates the fuzzy boundary between genetic prediction and explanation. It is clear, for starters, that there are no "genes for" divorce, and Visscher and his colleagues (2010) are sophisticated enough not to use that kind of language, even for height. But are there "causal variants" for divorce? If "causal variants" means nothing more than variation in the genome that covaries with marital status, then they have to exist, as the twin studies have already shown, and as the Visscher program would reiterate if anyone bothered to conduct it. Not only are there certain to be thousands of alleles related to marital status (polygenicity), but there is no reason to think that the

consequences of variation at any particular locus would be limited to something like marital status (pleiotropy). So there are thousands of genes, each with myriad effects, that act as inputs into a developmental system that sometimes eventuates in divorce. So for divorce as (maybe) opposed to height, it is clear that the causal gap between the activity of individual alleles and eventual outcomes in marital status is so great as to render the attribution of cause to particular alleles all but meaningless.

The causal networks underlying complex human outcomes might turn out to be so broken up into infinitesimal components that it is impossible to use them as building blocks for recognizable scientific explanation (Turkheimer, 2000). I have elsewhere referred to this possibility as the "gloomy prospect." In the context of the extraordinary genomic and mathematical tools brought to bear in the Visscher program, it may seem obscurantist to contend that despite it all, marital status and perhaps even height may never yield to explanation at the level of alleles, and to that extent heritability will remain missing. But it is important to recognize that in the same sense, environmental explanation has been missing much longer than heritability has. Divorce is 60% environmental, shared and mostly nonshared in families. Environmental social science has already engaged in a long effort to specify the individual environmental events that add up to marital disruption, without success. For the most part I refer the reader elsewhere for an extended discussion of the relationship between environmental biometric variance components and individual units of environmental causation (Turkheimer, 2011), but it is worth pausing to consider how clear it is in the environmental realm that statistical prediction from comprehensive but very general environmental composites (like socioeconomic status or the nonshared environment) does not necessarily translate into causal accounts of mechanism at the level of the infinitesimal environmental atoms which compose them (see Turkheimer, 2005).

For all the difficulties of inferring human developmental causation when randomized experimentation is not possible, the outlook is not entirely gloomy. In the years since I first outlined the daunting problems facing environmental social scientists (Turkheimer, 2004), my lab and many others have endeavored to show that twin studies are valuable for much more than just computing heritabilities (Moffitt, 2005; Rutter, Pickles, Murray, & Eaves, 2001; Turkheimer, 2008).

Suppose one is interested in the effects of parental divorce on children (e.g., D'Onofrio et al., 2007). Simple correlations between parental marital status and child outcomes are causally ambiguous, both genetically (alcoholic parents are more likely to get divorced, and then to pass the associated genetic background to their children) and environmentally (poor people have higher rates of divorce); random assignment is out of the question. But if one can find pairs of identical twin parents, one of whom is divorced and the other continuously married, the children of the nondivorced parents become a potent control group for the children of divorce. Is the control perfect, equivalent to what would be available

in a world where children were randomly assigned to parenting conditions? Of course not. Those twin parents have spouses, for one thing. Nevertheless, working with improved yet still imperfect causal control is the daily business of social science. Properly understood, behavioral genetics has always been about using family designs to make incrementally better causal inferences than can be made without genetically informative data. Heritability is a distraction.

One interesting possibility is that the innovative quantitative genetic methodology developed by Visscher and colleagues (2010) will also find applications in mainstream developmental social science after the heritability of everything has been re-computed. If molecular-based genetic similarity based on SNPs could be used to control for genetic confounds in social scientific studies of natural variation in phenomena like marital status, it would expose much larger data sets to genetically informed analysis, because the samples would not have to be limited to twins or adoptees. Such humble scientific progress will not occur until geneticists abandon the chimera of "genes for" depression or intelligence, and until those who have opposed behavior genetics concede that genetic influence on behavior is not an illusion created by the statistical assumptions of twin studies.

CONCLUSION

In the first paragraph of the explanatory article (Visscher et al., 2010) that followed the publication of Yang et al. (2010), Visscher et al. (2010) took the unusual step of letting the reader know that Yang et al. was rejected for publication at two journals before being published in *Nature Genetics*. My theoretical concerns notwithstanding, it is obvious to this reader (I was not a reviewer) that the article is an important contribution, so it was apparently misunderstood for a variety of reasons. In my case, I did not fully understand the statistical genetics until Visscher et al. explained it more simply. In the last paragraph of the explanatory paper, however, Visscher et al. identified a deeper reason:

> Why have we encountered so much apparent misunderstanding of the methods and results in the human genetics community? The core of our method is heavily steeped in the tradition of prediction of random effects and the estimation of variance due to random (latent) effects. While estimation and partitioning of variance has a long history in human genetics, in particular in twin research, the prediction of random effects is alien to many human geneticists . . . . Another reason could be the simultaneous use of population genetics and quantitative genetics concepts and theory in our paper, since these are usually applied in different applications, e.g., gene mapping or estimation of heritability. (p. 522)

As a social scientist and twin researcher, I had to struggle with the biological and statistical genetics underlying the Yang et al. analyses, but the analysis of

variance, the acausal "capturing" and "tracking" of one domain of variance with another came naturally to me. The situation was reversed for the geneticists who were the target audience of the paper: biologically based scientists, accustomed to genes that have an actual causal pathway to their outcomes. Over and above its technical brilliance, the real contribution of the Yang et al. article is to bring into focus this conceptual chasm between biological and quantitative genetics, and thus between the physical sciences and social science. Genomics is only now learning a hard lesson that social scientists had to learn a long time ago: sometimes prediction is just prediction. That is what the missing heritability problem is really about, and why it has not yet been solved.

REFERENCES

Alford, J. R., Funk, C. L., & Hibbing, J. R. (1978). Are political orientations genetically transmitted? *American Political Science Review*, *99*, 153–157.

Collins, W. A., Maccoby, E. E., Steinberg, L., Hetherington, E. M., & Bornstein, M. H. (2010). Contemporary research on parenting: The case for nature and nurture. *American Psychologist*, *55*(2), 218–232.

Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S. E., Liewald, D., . . . Luciano, M. (2011). Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular Psychiatry*, *16*(10), 996–1005.

D'Onofrio, B. M., Turkheimer, E., Emery, R. E., Maes, H. H., Silberg, J., & Eaves, L. J. (2007). A children of twins study of parental divorce and offspring psychopathology. *Journal of Child Psychology and Psychiatry*, *48*, 667–675.

Feldman, M. W., & Otto, S. P. (1997). Twin studies, heritability, and intelligence. *Science*, *278*, 1383–1387.

Fowler, J. H., Baker, L. A., & Dawes, C. T. (2008). Genetic variation in political participation. *American Political Science Review*, *102*, 233–248.

Fuller, J. L., & Thompson, W. R. (1978). *Foundation of behavior genetics*. St. Louis, MO: Mosby.

Gershon, E. (2000). Bipolar illness and schizophrenia as oligogenic diseases: Implications for the future. *Biological Psychiatry*, *47*, 240–244.

Goldstein, D. B. (2009). Common genetic variation and human traits. *New England Journal of Medicine*, *360*, 1696–1698.

Griffiths, P. E., & Tabery, J. (2008). Behavioral genetics and development: Historical and conceptual causes of controversy. *New Ideas in Psychology*, *26*, 332–352.

Gudbjartsson, D. F., Walters, D. F., Thorleifsson, H. S., Halldorsson, B. V., Zusmanovich, P., . . . Stefansson, K. (2008). Many sequence variants affecting diversity of adult human height. *Nature Genetics*, *40*, 609–615.

Joseph, J. (2010). The genetics of political attitudes and behavior: Claims and refutations. *Ethical Human Psychology and Psychiatry*, *12*, 200–217.

Kamin, L. J., & Goldberger, A. S. (2002). Twin studies in behavioral research: A skeptical view. *Theoretical Population Biology*, *61*, 83–95.

Lettre, G., Jackson, A. U., Gieger, C., Schumacher, F. R., Berndt, S. I., . . . Guiducci, C. (2008). Identification of ten loci associated with height highlights new biological pathways in human growth. *Nature Genetics*, *5*, 584–591.

Lewontin, R. C. (1974). The analysis of variance and the analysis of causes. *American Journal of Human Genetics*, *26*, 400–411.

Lango-Allen, H., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., . . . Raychaudhuri, S. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, *467*, 832–838.

Lewontin, R. C., Rose, S. P., & Kamin, L. J. (1984). *Not in our genes: Biology, ideology and human nature*. New York, NY: Pantheon Books.

Maher, B. (2008). Personal genomes: The case of the missing heritability. *Nature*, *456*, 16–21.

McGue, M., & Lykken, D. T. (1992). Genetic influence on risk of divorce. *Psychological Science*, *3*, 368–373.

Meehl, P. E. (1972a). A critical afterward. In I. I. Gottesman & J. Shields (Eds.), *Schizophrenia and genetics: A twin study vantage point* (pp. 367–415). New York, NY: Academic Press.

Meehl, P. E. (1972b). Specific genetic etiology, psychodynamics, and therapeutic nihilism. *International Journal of Mental Health*, *1*, 10–27.

Moffitt, T. E. (2005). The new look of behavioral genetics in developmental psychopathology: gene-environment interplay in antisocial behaviors. *Psychological Bulletin*, *131*, 533–543.

Petronis, A. (2006). Epigenetics and twins: three variations on the theme. *Trends in Genetics*, *22*, 347–350.

Plomin, R., Corley, R., DeFries, J. C., & Fulker, D. W. (1990). Individual differences in television viewing in early childhood: Nature as well as nurture. *Psychological Science*, *6*, 371–377.

Plomin, R., & Crabbe, J. (2000). DNA. *Psychological Bulletin*, *126*(6), 806–828.

Rutter, M., Pickles, A., Murray, R., & Eaves, L. (2001). Testing hypotheses on specific environmental causal effects on behavior. *Psychological Bulletin*, *127*, 291–299.

Stoolmiller, M. (1999). Implications of the restricted range of family environments for estimates of heritability and nonshared environment in behavior-genetic adoption studies. *Psychological Bulletin*, *125*, 392–409.

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. (2005). Mobiles: A gloomy view of the prospects for developmental behavioral genetics. In E. Perense (Ed.), *Wrestling with behavioral genetics* (pp. 100–108). Baltimore, MD: Johns Hopkins University Press.

Turkheimer, E. (2008, Spring). A better way to use twins for developmental research. *LIFE Newsletter*, *2*, 1–5.

Turkheimer, E. (2011). GWAS and EWAS. In K. S. Plaisance & T. A. C. Reydon (Eds.), *Philosophy of behavioral biology (Boston Studies in the Philosophy of Science)* (pp. 43–64). Dordrecht, The Netherlands: Springer.

Visscher, P. M. (2008). Sizing up human height variation. *Nature Genetics*, *40*, 489–490.

Visscher, P. M., Yang, J., & Goddard, M. E. (2010). A commentary on 'common SNPs explain a large proportion of the heritability for human height.' *Twin Research and Human Genetics*, *13*, 517–524.

Weedon, M. N., Lango, H., Lindgren, C. M., Wallace, C., Evans, D. M., . . . Hall, A. S. (2008). Genome-wide association analysis identifies 20 loci that influence adult height. *Nature Genetics*, *40*, 575–583.

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., . . . Montgomery, G. W. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, *42*(7), 565–569.

Zhong, S., Chew, S. H., Set, E., Zhang, J., Xue, H., Sham, P. C., . . . Israel, S. (2009). The heritability of attitude toward economic risk. *Twin Research and Human Genetics*, *12*, 103–107.