

Twin Studies are Still in Trouble:

A Response to Turkheimer

By Jay Joseph, Psy.D.

November 1, 2015

Human behavioral genetics and its allied field of psychiatric genetics are in trouble,¹ as unfulfilled gene discovery expectations during the “euphoria of the 1980s”² have continued to the present day, leading to researchers’ “[nonreplication curse](#)” dysphoria of the 2010s.³ In my recent book [The Trouble with Twin Studies: A Reassessment of Twin Research in the Social and Behavioral Sciences](#),⁴ I presented a detailed argument that genetic interpretations of the common “classical twin method” finding that reared-together MZ twin pairs resemble each other more (correlate higher) for behavioral characteristics than do reared-together same-sex DZ twin pairs are invalid because, among other reasons, [the twin method’s crucial MZ-DZ “equal environment assumption” \(EEA\) is false](#). (MZ pairs are also known as identical or monozygotic, and share a 100% genetic similarity; DZ pairs are also known as fraternal or dizygotic, and share a 50% average genetic similarity.)⁵ I also showed that, [for a different set of reasons](#), genetic interpretations of all “twins reared apart” (TRA) studies published to date have failed to provide scientifically acceptable evidence that human behavioral differences have a genetic basis.

In this context, [behavioral geneticist Eric Turkheimer published a 2015 review](#) of *The Trouble with Twin Studies* (referred to here as “the review”).⁶ Given Turkheimer’s deserved status as a leading (yet at times critical) behavioral genetic twin researcher, a negative review was expected. Nevertheless, I will take this opportunity to address the main issues, and to point to some areas of agreement. In the review I was portrayed as an “arsonist” who aims to “burn” the twin research “cathedral to the ground.” Comparing me to an “arsonist” is an escalation from Daniel Hanson’s characterization of a decade earlier [in his review of my book *The Gene Illusion*](#), where I was a mere “adrenalized boxer,” capable of “throwing some below-the-belt punches.”⁷

Turkheimer objected to my “absolutism,” to my “take no prisoners” position on twin research, and to my rejection of supposedly more “reasonable” middle-ground positions. Here he appealed to conventional wisdom that moderation and sensible compromise positions are good, and that positions labelled as extreme are bad. Commenting on [a blog posting of mine](#) in 2013, he wrote that in the twin study debate, “both sides have to give a little.” I answered, “I don’t think it is necessary for anyone to ‘give a little’ when it relates to what a person believes to be true. Much better, in my view, would be for all sides to state their case and leave it to others to decide who is right.” This is the essence of a debate.

The Equal Environment Assumption (EEA)

Moving on to the issues, the twin method's MZ-DZ equal environment assumption is key to all behavioral genetic and psychiatric genetic positions and theories. In Chapters 7 and 8 of *The Trouble with Twin Studies*, I showed that [MZ pairs experience much more similar environments than experienced by DZ pairs](#), and that MZs experience dramatically higher levels of identity confusion and attachment (as seen in Table 7.1 of my book). I also showed that most twin researchers long ago conceded the point that MZ pairs experience more similar environments than DZ pairs—thereby recognizing that MZ and DZ environments are in fact *unequal*—but that many have upheld the EEA and the twin method through the use of an argument stating that MZ twin pairs' environments are more similar because they “create” or “elicit” more similar environments for themselves because they behave more similarly for genetic reasons. However, this argument, which I called “Argument A,” [is a circular one](#) because the conclusion that genetic factors explain the greater behavioral resemblance of MZ versus DZ twin pairs is based on a *premise* stating the very same thing. Twin researchers invoking Argument A refer to the genetic premise in support of the genetic conclusion, and then refer back to the genetic conclusion in support of the genetic premise, in a continuously circular loop of faulty reasoning.

Turkheimer failed to address or counter my position that Argument A defenses of the EEA are based on faulty circular reasoning, yet he became quite possibly the first prominent behavioral genetic researcher to state clearly that “the EEA is false.” However, instead of concluding that genetic interpretations of MZ-DZ comparisons are invalid, he claimed that “it doesn't really matter,” because “statistical assumptions are always false.” Turkheimer asked, “Who—behavior geneticists, of all people—could imagine that having identical genomes would not influence parents to treat MZ twins more similarly?” Although the meaning of this sentence is not completely clear, he appears to be using the illogical Argument A by suggesting that MZ pairs behave more similarity as a result of their “identical genomes,” which causes parents to treat them more similarly. If he is referring only to similar parental treatment resulting from MZ pairs' greater physical similarity, this is an environmental effect on twin behavioral resemblance.

As many commentators have argued since the 1930s, and as I attempted to show in the book, the results of twin method MZ-DZ behavioral resemblance comparisons are easily explained by environmental (non-genetic) factors.

A New Twin Study “Cathedral”

Turkheimer wrote that while he was reading my book in preparation for the review, “a cathedral was erected in honor of twin studies in the prestigious journal *Nature Genetics*.” He was referring to [a 2015 twin study meta-analysis \(analysis of combined studies\) by Tina Polderman and colleagues](#). This “heritability cathedral” pooled together thousands of twin studies and then calculated heritability estimates, based on MZ-DZ correlational differences, for more than 17,000 physical, medical, and psychological characteristics (traits). The investigators, however, made the same basic interpretive error that twin researchers have been making since the 1920s. In a June 1, 2015 blog posting, I [reviewed the Polderman meta-analysis](#) and concluded:

Because twin method MZ-DZ comparisons are based on the false assumption that MZ and DZ pairs experience equal environments, it doesn't matter whether researchers pool together the results of 5 twin studies, 500 twin studies, 2,748 twin studies, or a million twin studies. Like the individual studies, the pooled results for behavioral characteristics can be *completely* explained by the non-genetic (environmental) influences experienced to a much greater degree by MZ versus DZ twin pairs. Two wrongs don't make a right, and 2,748 environmentally confounded twin studies pooled together don't make a genetic finding, at least as it relates to human behavioral differences.

In addition, although in the review Turkheimer repeated three times that “heritability isn't zero; it isn't one,” the heritability concept is itself controversial, with [several commentators calling for the abandonment of heritability estimates](#) in the behavioral sciences [and medicine](#).⁸ Turkheimer is also critical of some aspects of the practice of heritability estimation.

Earlier, in a May 28th, 2015 posting in his “[Gloomy Prospects Blog](#),” Turkheimer had a different take on the Polderman meta-analysis. Here he was “ambivalent about it,” believing that it constituted “a massive, overwhelming confirmation of what we already knew.” He continued, “The hard question about twin studies is *why* MZ twins are more similar than DZ twins” (emphasis in original). He noted that the critics argue “that the increased similarity of MZs is in fact environmental, the result of violations of the EEA,” an argument that he correctly recognized was not “refuted” by Polderman and colleagues' study.

Comparing twin research and accompanying heritability estimates to a “cathedral” is fitting. A cathedral is an awe-inspiring building with altars, statues, artwork, majestic arches and columns, and stained-glass windows, where worshippers are expected to accept *on faith* doctrines that have no scientific basis whatsoever. Behavioral genetics publications are also impressive looking, and contain awe-inspiring statistics, models, and diagrams—“dazzling statistical pyrotechnics” as one critic put it.⁹

Polderman and colleagues' meta-analysis, Turkheimer wrote in his May 28th blog posting, “represents an inconceivable amount of work. And the meta-analysis itself is beautifully executed. The graphs are striking, the numerical analysis is sophisticated.” These graphs and analyses certainly are spectacular and sophisticated, but they are more akin to a cathedral's beautiful stained-glass windows because the critical assumption underlying twin researchers' genetic interpretations of their pooled MZ-DZ behavioral correlation differences—the assumption that MZ and DZ pairs grow up experiencing equal or similar environments—is utterly false. As the British medical statistician [Lancelot Hogben](#) warned over 80 years ago, almost as if he had this study in mind,

*“There is a danger of concealing assumptions which have no factual basis behind an impressive façade of flawless algebra.”*¹⁰

Due to advances in technology since 1933, we could “upgrade” Hogben's warning by tacking on the phrase, “...and computer-generated online or pdf color graphics, diagrams, and statistical analyses.”

Twins Reared Apart (TRA) Studies

Turning to TRA studies (also known as “separated twin studies”), which were the main focus of Chapters 2-6 of *The Trouble with Twin Studies*, Turkheimer asked, “But were the twins really perfectly separated, without contact, and raised in absolutely independent family environments?” He answered, “Of course they weren’t.” Framing the issue in this way suggests that while separation may not have been perfect, it was pretty good. In contrast, I constructed three tables in Chapter 2 containing age-at-separation and case-history information, quoted directly from the original publications, for all 75 pairs described in the three “classical” TRA studies of Newman et al., Shields, and Juel-Nielsen. A quick read down the right-hand columns of these tables shows clearly that most pairs were reared *together* for many years, and/or lived nearby to each other and attended school together, and/or had ongoing contact and a close emotional bond. Far from being separated at birth and reared apart in randomly selected homes representing the full range of potential behavior-influencing environments, and meeting each other for the first time when studied, *most studied MZA pairs were only partially reared apart.*¹¹

In addition, most MZA twins were volunteers who knew of the existence of their co-twin when studied, and most grew up in similar behavior-influencing political, cultural, and socioeconomic environments at the same time (known as “cohort effects”). All MZA pairs were the same sex and shared a common prenatal environment, and all received more similar treatment by their social environments based on their very similar physical appearance. In the most famous TRA study to date, the IQ- and personality-centered “Minnesota Study of Twins Reared Apart” (MISTRA) of Bouchard and colleagues (performed between 1979 and 2000), the researchers published very little life-history information on the twins they studied, and then denied independent researchers access to their unpublished information and data. For a number of reasons, it is unlikely that the Minnesota MZA pairs were any more “separated” than were the partially reared-apart pairs described in the earlier studies.

A major issue in the MISTRA, which I discussed in some detail in Chapters 5 and 6, was how the researchers arrived at their conclusions in favor of important genetic influences on IQ, personality, and other behavioral characteristics. The final MISTRA sample consisted of 81 MZA pairs and 56 reared-apart DZ pairs (DZAs). I showed that the researchers had designated DZAs as their control group from the very beginning, and that MISTRA researcher Nancy Segal, in her 2012 book about the study, *Born Together—Reared Apart: The Landmark Minnesota Twin Study*, confirmed this and wrote:

The simple comparison of the MZ (or MZA) and DZ (or DZA) intraclass correlations is *an important first step* in behavioral-genetic analysis because this demonstrates *whether or not there is genetic influence on the trait.* [emphasis added]¹²

Because MZA pairs are more similar to each other genetically than are DZA pairs (100% versus 50%), a mean (average) MZA behavioral trait correlation not significantly higher than the corresponding DZA correlation suggests that non-genetic factors alone are responsible for raising both correlations above zero, since MZAs’ greater genetic resemblance did not lead to their greater behavioral resemblance.

In practice, however, with their data at hand, the Minnesota researchers decided to bypass the “important first step” of determining that the MZA mean correlation is significantly higher

than the control group DZA mean correlation.¹³ Instead, they based their conclusions on heritability estimates produced by analysis of variance “model fitting” procedures, and on the assumption that the MZA correlation “directly estimates heritability” because, as Segal put it, “MZA twins share only their genes.”¹⁴ I showed that MZA pairs share a lot more than their genes, including the environmental factors listed above and other factors. In addition, I showed that model fitting procedures are based on many questionable assumptions about people and genetics. The MISTRA researchers recognized that some of these assumptions are “likely not to hold,”¹⁵ but speculated that “several combinations of violations of assumptions can act to offset each other.”¹⁶

I also documented the MISTRA researchers’ failure to publish important data that might have led to different interpretations of their results. To the best of my knowledge, to this day the researchers have not published their full-sample DZA correlations for the two main IQ tests used in the study, the Wechsler (WAIS) and the Raven’s Progressive Matrices tests, even though they have published their main full-sample DZA *personality* correlations since the 1980s.¹⁷

In Chapter 6 I showed that, based on the MISTRA results that have been published, there does not appear to be a statistically significant difference between the MZA and DZA mean intraclass correlations for either the Wechsler or the Raven test.¹⁸ Contrary to the way the MISTRA results are usually discussed in the scientific literature (including textbooks) and in the popular press, this “important first step” comparison described by Segal and others—a step that “demonstrates whether or not there is genetic influence on the trait”—failed to identify a genetic influence on IQ scores (general intelligence).

Turkheimer did not discuss or dispute these findings, which, leaving aside [other numerous potentially invalidating problems and biases found in TRA studies](#), by themselves overturn MISTRA claims in favor of genetic influences on intelligence. He also did not discuss or dispute my contention that the researchers failed to publish their full-sample DZA IQ correlations, nor did he try to explain why these correlations were not published. Instead, he discussed MZA-DZA correlational patterns in 42 individual ability measures, which he believed show that “MISTRA comprises both genetic and environmental effects.”

Molecular Genetic Research

Commenting on attempts to identify genes at the molecular genetic level, Turkheimer recognized “to a quite remarkable extent, it has proven impossible to find” DNA variants that influence behavioral variation, and that “scientists have not identified a single gene that would meet any reasonable standard as a ‘gene for’ schizophrenia, intelligence, depression, or extraversion.” He then claimed that I ignored “the most recent developments, such as the now numerous statistical ‘hits’ in large-scale genome-wide association studies of height, schizophrenia, and even educational achievement.” I actually did mention some of these,¹⁹ but I didn’t emphasize them because, as I showed, [literally thousands of previous sensationalized “hits” of this type](#) did not survive subsequent attempts to replicate them. Virtually all [turned out to be false positives](#). This certainly [has been the case in psychiatry](#), and in behavioral genetics as well. As I concluded, experience demands that we treat all behavioral gene finding claims as false positive results, until proven otherwise.

Turkheimer agreed with me to a certain extent when he recognized that “the EEA is false,” and that “it has proven impossible to find” genes for the “entire standard roster of heritable human traits.” Since these were two of the three main interrelated themes of my book (the third was that TRA studies are massively flawed), there seems to be some common ground between us. The main area of disagreement is interpretation. Turkheimer concludes that the twin method remains valid because twins create their own (admittedly) unequal environments, and that genes for behavior exist and await discovery (the “[missing heritability problem](#)”), possibly through the use of recently developed techniques. For me, problems with genetically interpreted twin (and adoption) study findings, in combination with decades of negative molecular genetic results, lead to the conclusion that genes for human behavioral differences are unlikely to exist, and that genetic interpretations of twin data are wrong.

Other Issues

Chapter 11 of *The Trouble with Twin Studies* contained a parable about arsonists who burn down buildings in a large city. Turkheimer wrote that the parable “completely undermines Joseph’s own argument” because some houses in the parable were more arson resistant than others (analogous to differing genetic predispositions for psychiatric disorders). Although in retrospect I should have explained this better, the main purpose of the parable was to show that *even if* there are genetic differences among people that contribute to different behavioral outcomes, an emphasis on genetics is still a wrong and even harmful approach. Behavioral genetics, despite its leaders’ frequent statements that environmental factors are important, promotes an approach that emphasizes genetics, largely absolving society and political leaders from [the need to improve people’s social, familial, physical, and political environments](#). The parable drives this point home, and in no way undermines any position I argued in the book.

I was criticized in the review for “not reasoning forward from a known set of facts,” but instead, of confabulating “backwards from a fixed conclusion, eliding any segments of the evidence that don’t lead to the preordained destination.” Behavioral genetics, in fact, is where “confabulating backwards from a fixed conclusion” is the norm, and this is seen in the review itself. Another example is a 1998 behavioral genetic adoption study of personality, where Plomin and colleagues concluded in favor of genetics and the “heritability” of personality traits despite finding no personality test score correlation (.01) between birthparents and their 240 adopted-away biological offspring—a correlation that in the researchers’ own words “directly indexes genetic influence.”²⁰ I discussed this study in detail in Appendix B, but it was not mentioned in the review.²¹

Turkheimer wrote that, in my book, “consideration of any kind of genetic influence on behavior is inseparable from its worst possible consequences in victim blaming and eugenics.” This is a false characterization of my work, although it is true that I and many other critics over the years have told the disturbing story of eugenics, and warned of its possible revival. Interestingly, in 2015, leading behavioral genetic twin researchers Matt McGue (of MISTRA fame) and Irving Gottesman wrote, “Perhaps no area within psychology has received as much ethical scrutiny as genetics research on intelligence, and we agree that such research bears the burden of its early association with the eugenics movement.”²² As this passage from a pair of esteemed senior behavioral genetic researchers shows, the legitimate search for genetic influences on behavioral differences is compatible with the continuing need to remind scientists and society of the “worst possible consequences” of conducting this research.

The review ended with the conclusion, “*The Trouble with Twin Studies* is science denial.” I argue that criticism of behavioral genetic research *strengthens* science, since good science is greatly served by rejecting and casting out bad science—such as the classical twin method and all TRA studies published to date—from its ranks. On the other hand, leading behavioral geneticists and psychiatric geneticists refuse to recognize that decades of failed behavioral gene discovery attempts constitute a scientific finding that such genes are unlikely to exist.²³ *This*, one could argue, is real science denial.

Conclusion

Eric Turkheimer is a thoughtful and articulate spokesperson for the field of behavioral genetics. He is a past president of the [Behavioral Genetics Association](#), and has conducted numerous research projects, including a widely cited 2003 study “Socioeconomic Status Modifies Heritability of IQ in Young Children.”²⁴ He is to be commended for participating in a dialogue that his colleagues often avoid, even though he believes, as he put it in the review, that the EEA and the MISTRA are “the two most thoroughly worked-over topics in the history of the nature-nurture debate,” and that these “topics and their attendant arguments were complete in their current form 20 years ago.” [This far-from-settled debate](#), which in the past was mostly one-sided, is now intensifying in the context of decades of paradigm-threatening gene discovery failures.

Future public and scientific examination of the validity and usefulness of twin research, as well as other behavioral genetic and psychiatric genetic methods and concepts, can only be welcomed—even better as a joint examination by all sides of the issue working together as equal partners towards the common goal of uncovering the truth about the world and its people. If my book helps ignite such an examination, it will have served its purpose well.

Notes

¹ I thank Claudia Chaufan, Roar Fosse, M.C. Jones, and Ken Richardson for providing helpful feedback on an earlier draft of this article.

² Plomin et al., (2013), *Behavioral Genetics* (6th ed.), New York: Worth Publishers, p. 240.

³ Faraone, S., (2013), [Real Progress in Molecular Psychiatric Genetics](#), *Journal of the American Academy of Child and Adolescent Psychiatry*, 52, 1006-1008, p. 1007.

⁴ Joseph, J., (2015), [The Trouble with Twin Studies: A Reassessment of Twin Research in the Social and Behavioral Sciences](#), New York: Routledge.

⁵ Many previously accepted biological and genetic assumptions underlying twin research may not be true, including the assumption that MZ pairs are 100% genetically identical throughout their lives. See Charney, E., (2012), [Behavior Genetics and Postgenomics](#), *Behavioral and Brain Sciences*, 35, 331-358

⁶ Turkheimer, E., (2015), [Arsonists at the Cathedral](#), *PsycCRITIQUES*, 60 (40), 1-4. DOI: <http://dx.doi.org/10.1037/a0039763>

-
- ⁷ Hanson, D., (2005), [The Gene Illusion Confusion](#), *PsycCRITIQUES*, 50 (52), 1-4, p. 1. DOI: <http://dx.doi.org/10.1037/04131512>. For another review of *The Gene Illusion* by a behavioral genetic researcher, see Spinath, F. M., (2004), [Review of the book *The Gene Illusion: Genetic Research in Psychiatry and Psychology under the Microscope*], *Intelligence*, 32, 425-427. See [Richard Holdsworth's review](#) for a more positive evaluation of *The Gene Illusion*.
- ⁸ Chaufan, C., (2008), Unpacking the Heritability of Diabetes: The Problem of Attempting to Quantify the Relative Contributions of Nature and Nurture, *DataCrítica: International Journal of Critical Statistics*, 2, 23-38.
- ⁹ Lerner, R. M., (1995), The Limits of Biological Influence: Behavioral Genetics as the Emperor's New Clothes [Review of the Book *The Limits of Family Influence*], *Psychological Inquiry*, 6, 145-156, p.148.
- ¹⁰ Hogben, L., (1933), *Nature and Nurture*, London: George Allen & Unwin, p. 121, emphasis added.
- ¹¹ Farber, S. L., (1981), *Identical Twins Reared Apart: A Reanalysis*, New York: Basic Books.
- ¹² Segal, N. L., (2012), *Born Together—Reared Apart: The Landmark Minnesota Twin Study*, Cambridge, MA: Harvard University Press, p. 62.
- ¹³ The MISTRA researchers' decision to bypass the step of directly comparing correlations, and to instead focus on analyzing and partitioning variances, was described in a 1988 MISTRA publication. See Tellegen et al., (1988), [Personality Similarity in Twins Reared Apart and Together](#), *Journal of Personality and Social Psychology*, 54, 1031-1039, p. 1034.
- ¹⁴ Segal, 2012, p. 334.
- ¹⁵ McGue, M., & Bouchard, T. J., Jr., (1989), "Genetic and Environmental Determinants of Information Processing and Special Mental Abilities: A Twin Analysis," in R. Sternberg (Ed.), *Advances in the Psychology of Human Intelligence* (Vol. 5, pp. 7-45), Hillsdale, NJ: Erlbaum, p. 23.
- ¹⁶ Johnson et al., (2007), [Genetic and Environmental Influences on the Verbal-Perceptual-Image Rotation \(VPR\) Model of the Structure of Mental Abilities in the Minnesota Study of Twins Reared Apart](#), *Intelligence*, 35, 542-562, pp. 548–549.
- ¹⁷ See also Kamin, L. J., & Goldberger, A. S., (2002), Twin Studies in Behavioral Research: A Skeptical View, *Theoretical Population Biology*, 61, 83-95. Although the MISTRA researchers have published their full-sample California Personality Inventory DZA correlations for decades, I am not aware of any publication reporting MZA or DZA correlations for another major personality test they administered, the "Sixteen Personality Factor" test (16PF). A list of MISTRA publications (through 2010), tests, inventories, and questionnaires [can be downloaded from Nancy Segal's website](#).
- ¹⁸ The MISTRA Wechsler (WAIS) full-scale IQ test score correlations were MZA = .62, versus DZA = .50. The Raven's Progressive Matrices IQ test score correlations were MZA = .55, versus DZA = .42. The information is incomplete because, as I showed in Chapter 6, the MISTRA researchers did not publish or analyze their full-sample DZA IQ correlations. The Wechsler IQ correlations were taken from Segal, 2012, p. 286, based on unpublished figures given to her by Bouchard. Segal did not state the number of MZA and DZA twin pairs, but the final MISTRA sample consisted of 81 MZA and 56 DZA pairs. The Raven IQ correlations were taken from Johnson et al., 2007, p. 552, Table 3, and were based on 74 MZA and 52 DZA pairs. The [VassarStats website](#) provides a test of statistical significance between two independent sample correlation coefficients. This test shows that both the MISTRA Wechsler and Raven MZA versus DZA correlations fail to differ below the conventional .05 level of statistical significance, meaning that the difference between these correlations is assumed to have occurred by chance (the null hypothesis stating that the correlations do not differ is not rejected). See also Joseph, 2015, Chapters 5 and 6.

¹⁹ For example, see Joseph, 2015, p. 182.

²⁰ Plomin et al., (1998), [Adoption Results for Self-Reported Personality: Evidence for Nonadditive Genetic Effects?](#), *Journal of Personality and Social Psychology*, 75, 211-218, p. 211.

²¹ See also Joseph, J., (2013), “The Lost Study: A 1998 Adoption Study of Personality that Found No Genetic Relationship between Birthparents and Their 240 Adopted-Away Biological Offspring,” in R. Lerner & J. Benson (Eds.), *Advances in Child Development and Behavior*, 45, 93-124, San Diego: Elsevier.

²² McGue, M., & Gottesman, I. I., (2015), “[Classical and Molecular Genetic Research on General Cognitive Ability](#),” *The Genetics of Intelligence: Ethics and the Conduct of Trustworthy Research*, special report, *Hastings Center Report* 45, no. 5 (2015): S25-S31, p. S30. DOI: 10.1002/hast.495

²³ Latham, J., & Wilson, A., (2010), [The Great DNA Data Deficit: Are Genes for Disease a Mirage?](#), *The Bioscience Research Project*.

²⁴ Turkheimer et al., (2003), [Socioeconomic Status Modifies Heritability of IQ in Young Children](#), *Psychological Science*, 14, 623-628.