

Article

The Garden of Forking Paths; An Evaluation of Joseph's 'A Reevaluation of the 1990 "Minnesota Study of Twins Reared Apart" IQ Study'

Thomas J. Bouchard Jr 

Department of Psychology, University of Minnesota, Minneapolis, Minnesota, USA

The article by Joseph (2022) to which I respond here appeared in the journal *Human Development*. The editor of that journal had not invited a response at the time of publication and rejected a rebuttal, similar to my article below, when it was submitted. In response to an appeal, the editor invited a short rebuttal (3,000 words, 15 references). I submitted a rebuttal, but in order for it to be acceptable the editor required me to imply that Joseph's work was based on "mistaken assumptions, selective reading of research, or errors of omission". In addition, I was told that the term "pseudoscience" was "a form of name calling". I informed the editor I would be lying if I used the phrases suggested (they are not true), and that the term pseudoscience was being used in precisely the manner defined by Feynman in the article cited. I withdrew my submission, refined my manuscript somewhat and submitted it to *Twins Research and Human Genetics*. The final manuscript has been revised following recommendations of a reviewer.

Abstract

Joseph has written what purports to be a refutation of studies of Twins Reared-Apart (TRAs) with a singular focus on the Minnesota Study of Twins Reared-Apart (MISTRA). I show, in detail, that (a) his criticisms of previous TRA studies depend on sources that were discredited prior to MISTRA, as they all failed the test of replicability, (b) the list of biases he uses to invalidate MISTRA do not support his arguments, (c) the accusations of questionable research practices are unsubstantiated, (d) his claim that MISTRA should be evaluated in the context of psychology's replication crisis is refuted. The TRA studies are constructive replications. Like many other scholars, past and present, he has been misled by the variation introduced by small samples (sampling error) and the distortion created by walking in the garden of forking paths. His endeavor is a concatenation of elision and erroneous statistical/scientific reasoning.

Keywords: Minnesota twin study; intelligence; IQ; behavioral genetics; reared-apart twins

(Received 7 April 2023; accepted 12 April 2023; First Published online 5 June 2023)

The scientific study of biological influence on 'human nature' has always been controversial. When asked whether he would discuss man in the *Origin of the Species* (1859), a book he had already withheld for 20 years in deference to the prevailing religious ideology of the day, Darwin replied: 'I think I shall avoid the subject, as so surrounded with prejudices, though I fully admit it is the highest and most interesting problem for the naturalist' (Pearson, 1924, p. 86).

Darwin did eventually publish on the topic (1871), as did his half-cousin Francis Galton (1892/1962). Their theories, when applied to humans, remain controversial among both the public and some intellectuals. Not controversial is the application of their ideas to animals. There is a thriving experimental science of animal behavior genetics (York, 2018) and selection in domestic animals (Grandin, 2022). While familiar to everyone as a pet, the dog has 'emerged as a premier species for the study of morphology behavior, and disease' (Ostrander & Wayne, 2005, p. 1706). There is a

natural affinity between developmental science and genetics. As J. P. Scott pointed out:

We thought that the best time to study the effects of genetics would be soon after birth, when behavior still had little opportunity to be altered by experience. On the contrary, we found that the different dog breeds were most alike as newborns; that is, genetic variation in behavior develops postnatally, in part as a result of the timing of gene action and in part from the interaction of gene action and experience, social, and otherwise. (Scott, 1990, p. vii).

Unless one argues that human beings are not animals, the animal work (true experiments) establishes an overwhelming a priori expectation of finding genetic influences on behavior in humans. It would be astonishing if this were not the case.

'So-Called' Problems with IQ Studies Based on Twins Reared Apart

Joseph (2022) presented numerous disparate criticisms of the early TRA studies and MISTRA.¹ I deal with most of them in the order in which they were presented.

¹Segal (2012, Chapter 14) deals with many of the issues discussed in this article as well as others not raised by Joseph. She also provides details regarding the funding of the study and a list of all grants in Appendix A.

Corresponding author: Thomas J. Bouchard, Jr.; Email: bouch001@umn.edu

Cite this article: Bouchard T. J., Jr. (2023) The Garden of Forking Paths; An Evaluation of Joseph's 'A Reevaluation of the 1990 "Minnesota Study of Twins Reared Apart" IQ Study'. *Twin Research and Human Genetics* 26: 133–142, <https://doi.org/10.1017/thg.2023.19>

© The Author(s), 2023. Published by Cambridge University Press on behalf of International Society for Twin Studies. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited

Monozygotic Twins are not Genetically Identical

'Recent evidence, however, suggests that the long-running assumption that MZ pairs are genetically identical to each other might not be true' (p. 49). The implication created by this quote is that these new data invalidate the twin method. What Joseph does not tell us is that (a) this assumption was discarded long ago, and (b) this source of influence would cause the twin method to *underestimate* genetic influence. It has long been known that there are chance factors at work that create differences between identical laboratory animals (Gärtner, 1990), even single cell organisms (Koshland, 1984, p. 13). Fisher (1918) called them *random somatic effects*. Darlington (1954) called them *cytoplasmic discordance* and *asymmetry*. Mitchell (2018) calls it *intrinsic randomness*. Behavior geneticists have discussed the issue at length (Molenaar et al., 1993).

The Editor of Science Invited Bouchard to Submit an Article and in Bouchard's View the Publication in Science 'Legitimated the Study'

The implication here is that somehow standard scientific practice was avoided or bypassed. The facts are much simpler. I was asked by Sidney Fox to participate in a conference devoted to the topic of Individuality. I contributed a paper that resulted in a book chapter (Bouchard, 1984). Daniel Koshland, not yet editor of *Science*, but a distinguished professor of biochemistry at the University of California Berkeley, had also been invited. The conference provided an excellent forum to discuss the methodology, strengths and weaknesses of a TRA study. I spent considerable time with Koshland, both during the meeting hours and after, discussing scientific methodology. We never communicated again until he contacted me and requested that I submit a paper to what was to become the first genome issue of *Science*. It is common practice for editors to invite papers for special issues.

'There is a large literature critical of the three studies published before MISTRA. Such studies require randomization and complete separation of the twins into representative homes of the population. None of the three studies nor did MISTRA ... come close to meeting this requirement' (p. 49)

The study of reared-apart twins is a combined experiment of nature — twinning and nurture — adoption. It is a member of a class of studies called 'natural experiments' as opposed to 'planned experiments'. Experimental design mandates the use of randomization and all natural experiments fail to meet that requirement. Joseph credits me with making that distinction and pointing out that conducting a randomized TRA study (an experiment) would be unethical (Bouchard, 1993b, p. 50). He cites, at length, a description of a perfect TRA study by Fancher (1985), which I reproduce here:

A definitive study would have to employ twins who represent a genuinely random sample of the general population, and who have been randomly placed for adoption in a range of homes representative of the entire population. A definitive study would also have to demonstrate that its sample genuinely represents the full population of separated twins, and is not biased toward including only certain kinds of cases. Finally, in an ideal study all twins should have been *completely* separated from each other soon after birth, with no opportunity to communicate with each other or influence each other prior to their testing. (p. 165, emphasis in original)

In this article I will deal with all the issues discussed in the quote. First, I give particular emphasis to the assertion that one needs (a) a sample randomly placed for adoption and (b) that all twins should have been *completely* separated from each other after birth (to preclude their influencing, each other) — in other words, a true experiment. We did not conduct a true experiment. We gathered a sample of convenience and justified our conclusions on the grounds that the sample was reasonable for our purposes. The same is true for all 'natural experiments'. Because of their rarity, a sample of reared-apart twins would necessarily be small. Consequently, requirement (a) is unrealistic. It is also unnecessary. Schmidt and Oh (2016) point out that 'randomization does not work when sample sizes are small' (p. 33). This was also noted earlier by Tversky and Kahneman (1971). Many studies in the behavioral sciences use small samples and, consequently, are not 'true experiments', and these problems bedevil all of them. This problem has been solved in behavior genetics by using replications, multiple corroboration, constructive replications and model fitting, all parts of normal science and discussed below. Requirement (b) is also unnecessary. MISTRA, the TRA studies that came before, and the one that came after (Pedersen et al., 1992), are studies of 'twins reared apart' — they are not studies of 'twins who have never be in contact'. The fact that the twins varied in degree of apartness and contact has made it possible to assess whether those factors were associated (positively or negatively) with degree of similarity on various traits. As I show below, those factors do not explain TRA similarity in IQ.

There is a Sizeable Literature Critical of the TRA Studies

The question is not one of size as the criticisms are largely repetitions of criticisms from three sources: Kamin (1974), Taylor (1980) and Farber (1981). The question is: What is the validity of the criticism? As I demonstrate, they are all invalid.

Leon Kamin — 'The Science and Politics of IQ'

Key criticisms of the TRA literature that are constantly repeated in secondary sources come from Kamin (1974). Reading Joseph's paper, one might think that Kamin is an impeccable source as he cites Kamin's book 13 times. A look at the critical reviews of the book, none of which are mentioned by Joseph, tells an entirely different story. All the reviewers provide numerous examples of statistical and quantitative abuse of the data. I only draw on two reviewers and leave it to the reader to consult the others.

David Fulker (1975)

His book lacks balanced judgment and presents a travesty of the empirical evidence in the field. By exaggerating the importance of what are idiosyncratic details rather than typical features, he totally avoids the necessity to consider the data as a whole. The cumulative picture is overwhelmingly in favor of a substantial heritability. (p. 519)

The evidence for Fulker's claim is quite straightforward. To demonstrate the absurdity of Kamin's causal claims, Fulker created a biometric model based on the claims and it generated absurd results. A simple genetic model fit well.²

²This is an example of the application of parsimony. The use of models and parsimony to increase precision in science is discussed in detail by Gauch (2006).

Kamin sorted cases into small subgroups to show testing bias and to show environmental influences using some of the same cases. As Fulker (1975) pointed out, 'It can hardly be claimed they indicate a striking testing bias for the purpose of one argument but striking environmental influence for another' (p. 510). Fulker devotes considerable space to Kamin's selective use of small samples to draw conclusions consistent with his hypothesis, but which disappear when a slightly different sample is used (text on p. 508 and Table 2). In a textbook on intelligence, Nicholas Mackintosh (1998), discussing Kamin's work, made the same point:

The sure way to guarantee the truth of the adage that one can prove anything with statistics is to trawl through a set of data, performing numerous post hoc analyses on small sub-sets of the data, until one comes up with the desired solution . . . and one's confidence is not increased by evidence of biased reporting. (pp. 98–99)

Douglas Jackson (1975)

Jackson cites Kamin's major conclusion that 'the burden of proof falls upon those who wish to assert the implausible proposition that the way in which a child answers questions devised by a mental tester is determined by an unseen genotype' (p. 176). He ends his review with the claim that 'had the author been equally zealous in evaluating the null hypothesis that such treatments [attempts to increase IQ] make no difference he would have been hard pressed to fail to reject it' (p. 1080).

Kamin reports a number of correlations based on subsets of twins in an effort to establish that there are artifacts. He chooses an incorrect and inflated value for his degrees of freedom, based on individual twins rather than twin pairs, that yields spurious estimates of statistical significance. Worse, he selectively reports correlations chosen post hoc in very small samples. To demonstrate that there is a correlation between IQ and age which might inflate the correlation between twin pairs, Kamin found four high values, but these were based on subsamples selected by him of 7, 3, 9, and 3. He failed to report near-zero age-IQ correlations based on larger samples. (Jackson, 1975, p. 1079)

The problems created by the failure to understand the effects of sample size are legion and not restricted to psychology. The title of a classic paper on the topic by the distinguished statistician Howard Wainer (2007) tells the story: 'The most dangerous equation: Ignorance of how sample size affects statistical variation has created havoc for nearly a millennium'. Reviews by Scarr (1976), Shields (1978) and Bouchard (1982b) also focus on the use of trivially small subsamples to draw false, unreplicable and unparsimonious conclusions.

Kamin claimed 'we are entitled to conclude that today, as in the past, untrue facts and fallacious conclusions tend to reflect the social and ideological biases of the theorist' (Eysenck & Kamin, 1981, p. 349). *The Science and Politics of IQ* is an unequivocal example of that claim.

Kamin Redux

In their attempts to discredit the TRA studies, the two individuals, discussed below, applied Kamin's use of selected samples in their own idiosyncratic manner. Farber used larger (but still small) samples than Kamin, but her analyses were equally misleading. Taylor, like Kamin, used small samples. Farber did not cite Taylor and Taylor could not have cited Farber as she had not yet published. They appear to have carried out their analyses independently. Both sources are repeatedly cited by Joseph, Taylor is cited 7 times and Farber 10 times.

Howard Taylor (1980) — 'The IQ Game: A Methodological Inquiry Into the Heredity Environment Controversy'

According to Joseph, citing Taylor, the MISTRA study is *bad science*:

In 1980, sociologist Howard Taylor described what he called 'The IQ game,' by which he meant IQ-genetic researchers' 'use of assumptions that are implausible as well as arbitrary to arrive at some numerical value for the genetic heritability of human IQ scores on the grounds that no heritability calculations could be made without benefit of such assumptions' (Taylor, 1980). The MISTRA IQ study can be seen as an exemplar of 'IQ game' bad science. (p. 62)

The necessity, usefulness, and validity of assumptions used for model fitting kinship data were addressed by Heath (1982) in his critical review of Taylor's book. The logic is simple, parsimonious, and standard scientific strategy:

If we can show that such different assumptions lead to different predictions, and then test some of these predictions, this will further our understanding of the inheritance of IQ. The sterility of Taylor's approach is that, if we assume that for every relationship there is a special environment, we can never make any testable predictions. If we avoid this extreme assumption, we can at least fit models and use the parameter estimates thus obtained to make quantitative predictions for new sets of relationships. The failure of these predictions would then lead us to reexamine our original assumptions. (p. 214)

Taylor claimed to show that much of the similarity in IQ between monozygotic twins reared apart (MZAs) in the three classic studies, was due to similarity in their environments. He documented this claim by classifying twins into groups of high and low environmental similarity based on four different measures: age of separation, reunion in childhood, rearing by relatives, and similarity in social environments. Two of the three studies made use of more than one IQ test, so it was possible for me to return to the original publications and ask, 'Do his results constructively replicate?'

I present only one example of my analysis of his work. Taylor argued that twins reared by related relatives are more similar (.75) than those reared by families that are not related to the adopted twin (.56). My analysis using the alternate IQ measure refutes this finding as the correlations not only do not replicate, they reverse, .66 versus .77. Taylor's analysis is almost identical to one done by Kamin (1974), whose results were repeated by Lewontin et al. (1984, p. 107) and elsewhere.

The title of my paper was 'Do Environmental Similarities Explain the Similarity in Intelligence of Identical Twins Reared Apart?' I found that his 'conclusions regarding the MZA data are simply erroneous and cannot be substantiated from the evidence at hand. The answer to the question posed in the title of this paper is NO!'

Susan Farber (1981) — 'Identical Twins Reared Apart: A Reanalysis'

Farber brought together the data available on all the twins reared apart in the extant literature. Twenty-five percent of the book deals with IQ. I reviewed the book and draw from that source (Bouchard, 1982a) and only report some of the flaws here. Farber complained:

My own evaluation, particularly of the allegedly scientific analyses made of the IQ data, is more caustic. Suffice it to say that it seems that there has been a great deal of action with numbers but not much progress — or sometimes not much common sense. (p. 22)

Despite this complaint the book contains an appendix that is 44 pages long with 54 tables dealing with the IQ data. The one figure that I expected to find was not there — the correlation for what she classified as the Highly Separated Group, a curious omission. Consequently, I carried out the computation:

The results were surprising! For the entire group: $n = 39$, $r_i = .76$, mean = 97.42, SD = 14.28. For the females: $n = 26$, $r_i = .76$, mean = 97.96, SD = 14.29. For the males: $n = 13$, $r_j = .76$, mean = 96.35, SD = 14.20. The three arrays show the slight depression in IQ characteristic of most older twin samples, a standard deviation comparable to the normative population, and identical intraclass r 's that are indistinguishable from the full sample for which separation is ignored. (Bouchard, 1982a, p. 191)

The findings for the full sample where degree of separation is ignored are males, $n = 32$, $r_i = .74$, females, $n = 50$, $r_i = .76$, all cases, $n = 82$, $r_i = .75$.

The tables also contain a surprise regarding the influence of amount of contact:

A second aspect of the book is an elaborate statistical treatment of the IQ data from the separated MZ twin studies. Some interesting analyses are provided, but readers are hereby cautioned to watch out for the graphs and summaries in Chapter 7. *These suggest that the amount of contact between separated MZ twins accounts for some 20–30% of the IQ variance. Perhaps, but only if one assumes that the mechanisms involved work in the opposite directions in males and females (emphasis added)* (see Appendix E, p. 350). For the sexes combined, the amount of contact between the twins does not predict their resemblance. (Loehlin, 1981, p. 297)

As noted, one must go to the appendix to discover this new 'complex pattern of environmental effects'. This 'complex pattern' is known as a disordinal interaction (Bouchard, 1993b, Figure 2.13). Given that the authors recognized that 'far too many parameters were being estimated and tested for the number of observations available', it may seem gratuitous to point out that interactions are even more unlikely to replicate than main effects when using small samples (Border et al., 2019).

I called Kamin's, Taylor's, and Farber's approach to the data 'pseudoanalysis', but the term never took hold. Other names have been invented and are in widespread use; JARKING (justifying after results are known), HARKING (hypothesizing after results are known) and *p*-hacking. An older term is 'data dredging'. A newer, and more comprehensive term is '*the garden of forking paths*' (Gelman & Loken, 2014). The statistical abuses characterized by these terms explain why many scientific findings (true experiments) turned out to be false (Ioannidis, 2005; Szucs & Ioannidis, 2017, 2021). As Ritchie (2020) has pointed out:

Scientists who knowingly run low powered research, and the reviewers and editors who wave through tiny studies for publication, are introducing a subtle poison into the scientific literature, weakening the evidence that it needs to progress. (p. 143)

The key critics of older TRA studies, repeatedly cited by Joseph, have failed to show that, age of separation, reunion in childhood, degree of separation, amount of contact, rearing by relatives or similarity in social environments had *any* influence on TRA IQ similarity.

Dizygotic Twins Reared Apart (DZA)

Joseph implies that we concealed data gathered from the dizygotic twins reared apart (DZAs). That we were studying both MZA and DZA twins is mentioned in the first sentence of the abstract and the first sentence of the article itself. We did this to make it clear that MISTRA was an ongoing research program (Urbach, 1974a,

1974b; Zwaan et al., 2018), that additional twins were being recruited and, in the future, we would publish additional analyses. The DZAs were not included because the sample was small; the purpose of the paper was to report a constructive replication of previous studies of MZA twins in the brief format provided by *Science* and explain the methodology underlying the study of MZA twins. We included the findings for IQ from the three previous MZA studies (Table 2, p. 225), all of which had smaller samples than ours. Based on (a) our previous review of the relevant literature (Bouchard & McGue, 1981, cited as reference 9), (b) the previous TRA studies, (c) our TRA results, (d) the documentation that the criticism of the previous TRA studies were flawed/fallacious (reference 20); we concluded that 'general intelligence or IQ is strongly affected by genetic factors'. These results were replicated in Sweden two years later (Pedersen et al., 1992) using a design that included both MZ and DZ twins reared apart and together. For their IQ measure (the 'First Principal Component') the heritability was .81 with no shared environmental component. They reported no influence of age of separation, degree of separation or number of years separated, on twin similarity. More recently, similar results (.86) were reported using the large Vietnam Twin Study of Aging (Panizzon et al., 2014)

Questionable Research Practices and Assumptions

p-Hacking

p-hacking can be defined in several different ways and Joseph provides a few. He does not provide, in a full page of text devoted to the topic (588 words), any examples of *p*-hacking in MISTRA. The best he could do is make two misleading claims.

In the first claim he cited Segal (1999):

Bouchard cautioned that the Minnesota [IQ] data are preliminary and require further analysis (p. 136). Neither Segal nor Bouchard, however, provided a valid reason why the Minnesota data were preliminary and required further analysis. (p. 57)

What I wrote was, 'The MISTRA IQ correlations have not yet been fully analyzed. We are awaiting completion of the study before conducting a full analysis' (Bouchard, 1998, p. 262).

The second misleading claim implies that we engaged in 'questionable research practices': 'Questionable research practices of this type can occur when researchers are not required to adhere to a stated data collection stop point, which would be established and documented in a pre-registered study' (p. 57).

Note that *p*-hacking, establishing a stated date collection stop point, and pre-registering studies, are three different concepts (rules?). Joseph is implying that we violated the latter two rules knowing full well that they were not in place when MISTRA was conducted. Gelman and Loken (2014) provide a thoughtful discussion of both topics and make it clear that such rules do not apply to every type of study. Given that his criticism of TRA studies is largely based on studies (Kamin, Taylor and Farber) plagued with *p*-hacking, it is ironic that Joseph's claims of *p*-hacking and questionable research practices by MISTRA are both false and disingenuous.

Psychology's 'Replication Crisis'

The material presented in this section, nearly a page of text (457 words) is irrelevant to MISTRA. The IQ studies using reared-apart MZ twins constituted constructive replications, a term coined by a member of our research team (Lykken, 1968). The studies were conducted by different research teams, at different times (1937

to 1992), with different instruments, three different languages, different protocols, different recruitment methods and different samples. Regardless of this enormous variation across studies, the findings are consistent. Regarding TRA studies there is no replication crisis.

A Key MISTRA Assumption is Not Supported by the Evidence

In order to conclude that above-zero MZA IQ test score correlations are caused only by genetic influences, TRA researchers must control for the potential environmental confounds and cohort influences seen in Table 3, or they must assume that these environmental confounds and cohort influences do not exist. For the most part, the MISTRA researchers chose the latter course. (p. 58)

The first sentence would be true if the supposed confounds listed in Table 3 caused MZA twin similarity. As I show below, some work in the opposite direction from that claimed most do not cause similarity, some of the claims are specious and others are irrelevant. Since we also demonstrated within our sample that many of the claimed causes of MZA twin similarity do not cause similarity it follows that the last sentence is false. Contrary to Joseph's claim, we tested for many possible causes of TRA similarity and found them wanting (Bouchard et al., 1990, p. 225).

The Environment Is Mostly Genetic

Bouchard and colleagues based their conclusions about IQ heritability on the claim that the MZA correlation alone 'directly estimates heritability.' However, they reached their conclusions only because they decided to count most environmental influences as genetic influences. (p. 58)

The last sentence is false and refuted in the next section as the environmental influences referred to (his Table 3) have not been shown to make MZA twins similar.

There is a common confusion regarding the logic of population genetic research. All such researchers recognize that a supportive environment is necessary for an organism to grow and develop. A corn seed sitting in glass jar will not grow into a corn plant. It must be planted in an environment conducive to growth and development. This does not mean we cannot study the genetics of corn. Joseph has confused two levels of explanation, the population level and the individual level. MISTRA was a population-level study. We showed that genes influence the expression of individual differences in a wide variety of traits. We proposed, but did not prove, that this population outcome might be explained by processes at the individual level.

Specific mechanisms by which genetic differences in human behavior are expressed in phenotypic differences are largely unknown. It is a *plausible conjecture* [my emphasis] that a key mechanism by which the genes affect the mind is indirect, and that genetic differences have an important role in determining the effective psychological environment of the developing child. (p. 227)

There are multiple levels of behavioral causation and many different associated disciplines (Bouchard & Johnson, 2021, Figure 2). They will all be needed in order for us to fully understand the 'sources of genotype-phenotype association in humans' as some are direct and other indirect (Young et al., 2019).

Fifteen Nonfamilial Environmental Influences Experienced or Potentially Experienced By Monozygotic Twin Pairs Separated Near Birth and First Reunited When Studied

As mentioned earlier, Joseph claims we assumed 'none of the influences seen in Table 3 increased MZA IQ correlations for

nongenetic reasons'. The statement is false. As I show for the first item, the empirical evidence demonstrates that the 'effects' are difference-producing, not similarity-producing. The direction of effects for the various items is an empirical problem that needs to be determined and, in most cases, it would be reasonable to conclude that the evidence favors difference-producing.

Prenatal Effects, Including Prenatal Exposure to Toxins and Other Influences

This item is indexed with three references. The first tells us that toxins in the environment have adverse effects on human beings. The second deals with the influence of poverty on infant health. Neither study involves nor mentions twins. I am unaware of any behavior geneticist who would dispute the influence of these factors on development. Joseph, however, fails to tell us why and how these factors should bias studies of monozygotic twins reared apart in the direction of similarity. The third reference explains why he did not do so. Line 7 of the introduction mentions the work of Bronson Price (1950, 1978). The title of this classic paper is 'Primary Biases in Twin Studies, A Review of Prenatal and Natal Difference-Producing Factors in Monozygotic Pairs'.³ The first sentence of the article reads as follows: 'In all probability the net effect of most twin studies has been underestimation of the significance of heredity in the medical and behavior sciences.' I am unaware of any research that would change this conclusion and Joseph does not cite any.

We discussed prenatal and perinatal environmental influences and cited Price (Bouchard et al., 1990, p. 225). I discussed this topic in more detail in Bouchard (1984). Price is not mentioned by Joseph.

Perinatal-Infancy Health Care, Nutrition, and Exposure to Environmental Toxins

This item is simply a repeat of the first item and was dealt with above. The extent of effect of any one of these factors would not necessarily be the same for twins in different families. For example, the dose could be different for each twin; consequently, it would be difference-producing, not similarity-producing. Joseph's assumption of it being similarity-producing (p. 227) is just that: an assumption.

Birth Cohort (Same Age), Which, in IQ Terms, Might Create More Similar Scores Based on the 'Flynn Effect' and Exposure to Similar Methods of Education

We age-corrected the data, as Joseph acknowledges. There is no explanation as to *how* MZA twin similarity might be influenced. Do all people in the same cohort and of the same age (children in a classroom) have the same IQ? In the MISTRA paper we discuss these issues on page 227.

Flynn did not believe that his findings (the Flynn effect) negated the causal influence of genetic factors on behavioral traits within a population. In his last book, in a section under the heading 'Psychological Research', he recognized the importance of 'heritability' and pointed out a fundamental flaw in much developmental research.

I often read the advice eminent psychologists give to the discipline about methodology. Concerning individual differences within groups, no journal

³The 1978 paper provides all the references many of which were not provided in the 1950 paper due to page restrictions but, were available from the author at the time.

should accept a correlation study as evidence that parenting is causal without performing the necessary controls for genetic relatedness and hence heritability. To do so is actively deceptive. (Flynn, 2020, p. 195)

Selective Placement Status (Adoptive)

The citations here are Kamin (1974) and Richardson and Norgate (2006). I and others have addressed Kamin's analysis (discussed above) and shown it to be flawed beyond redemption and deceptive. Richardson and Norgate do not deal with twins reared apart. Table 3, page 225 in MISTRA deals directly and quantitatively with nine selective placement factors and is not mentioned by Joseph. The appropriate control for placement is the similarity in IQ found for unrelated individuals reared together when measured in adulthood. It is extremely small (McGue et al., 1993, Figure 4). The extremely low correlations between adoptive parents and their offspring constitute a constructive replication of such findings (Willoughby et al., 2022, Figure 3). It is incumbent on Joseph to report such findings. We reported such effects in footnote 21 of the MISTRA paper.

Gender Cohort (Sex)

There are three citations here. None of them discuss twins reared apart. As Joseph acknowledges, we deal with the issue of sex differences, so that is not an issue. Sex effects on kin correlations are discussed in detail in the 1981 meta-analysis (Bouchard & McGue, 1981, Tables 2 and 3). There is nothing relevant here.

Developmental Stage, Maturational Change

There are two citations here. Neither source deals with twins or twins reared apart, nor explains why the topics they discuss would increase the similarity between twins reared apart. As noted earlier, they may well be difference-producing factors.

Striking Physical Resemblance, Including Facial Appearance and Height

There are three references here (Cropanzano & James, 1990; Hu, 2018; Zebrowitz & Montepare, 2008). Cropanzano and James is a critique of Arvey et al. (1989), a paper that established a new domain of research in behavior genetics, namely the genetics of work behavior (Colarelli & Arvey, 2015). The criticisms are largely the same ones that we have already dealt with, and we provided a rejoinder (Bouchard et al., 1992). The rejoinder was not cited by Joseph.

The second paper, 'First Impression of Personality from Body Shapes' (Hu, 2018), provides no useful information regarding the similarity of MZA twins. I infer that this study of body stereotypes is included because it might lead one to infer that MZ twins who typically have a similar body shape, elicit a common response from the people they interact with and this influences their personality — thus, the similarity between MZ TRAs. This is a common argument made by psychologists. Lois Hoffman (1991), for example, discusses the role of attractiveness as an elicitor of treatment by caregivers and others in the shaping of personality. She cites numerous studies showing that attractive people are treated differently from unattractive people.

This line of reasoning is more fully developed in the third reference provided (Zebrowitz & Montepare, 2008). These authors provide empirical data in support of their argument: 'Moreover, these trait impressions are accompanied by preferential treatment of attractive people in a variety of domains, including interpersonal

relations, occupational settings, and the judicial system (Langlois, 2000; Zebrowitz, 1997)'. By referring to this paper, I believe Joseph wants us to infer that attractiveness causes attractive people to have higher IQs than less attractive people due to the way they are treated by caregivers and others. For this mechanism to work, among other things, it would require accurate impressions of the intelligence of people across the entire spectrum of intelligence. The authors of the paper make it clear that this is not the case.

First, accurate impressions of health and intelligence are limited to perceptions of people in the bottom half of the attractiveness continuum. Although greater attractiveness is associated with impressions of greater intelligence and health across the whole continuum, attractiveness and actual intelligence or health are related only among people who range from unattractive to average and not among those who range from average to attractive. (Zebrowitz & Rhodes, 2004, p. 1501)

The problem with this line of reasoning is the long chain of *causation* necessary to make it plausible (Bouchard, 1993a, p. 30). No one has ever presented a quantitative model and data that support this idea.

There are some individuals who are unrelated and have shown extraordinary physical resemblance. A small but fascinating study shows that such individuals even have some genes in common and the genes may prove to influence traits beyond facial features (Joshi et al., 2022). Nevertheless, studies of unrelated look-alikes demonstrate that 'appearance is not meaningfully related to personality similarity and social relatedness. The criticism that MZ twins are alike in personality because their matched looks invite similar treatment by others is refuted' (Segal et al., 2018).

Condition of Being an Adopted Child with Accompanying Abandonment, Attachment, and Mental Health Issues

Three studies are cited here. Because each twin is reared in a different family, how this issue manifests itself will differ from family to family. Consequently, it is likely to be difference producing. The assumption of it being similarity producing is just that, an assumption. It is an assumption with no evidence.

National, Regional, Ethnic, Religious, and Political Culture

There are two references here. Like several previous items, the references have nothing to do with twins and Joseph does not provide us with any meaningful causal mechanisms or evidence. In cases where the TRAs were reared in different national, regional, ethnic, religious and political conditions, they likely would create differences, not similarities.

Socioeconomic Status

There are three citations here. Since Joseph does not present any theory regarding the causal chain from SES to TRA similarity in IQ, we are left to speculate. If he means that placement in homes similar in SES is the cause of TRA similarity in MISTRA, one must wonder if he even read the paper. Table 3 from the MISTRA reports the relevant findings for SES and was discussed above. There are placement effects, but they are shown quantitatively *not* to influence the MZA correlation. We discussed this issue previously under selective placement.

If Joseph is arguing that the TRA similarity is due to simply being born in a family of a particular SES, then he needs to provide a mechanism and evidence that is independent of genetic influences.

Oppression, Racism, Discrimination, or Privilege on the Basis of Common Racial or National Background, Gender, Socioeconomic Status, Disabilities and So Forth

There are three references here and like several previous items, they have nothing to do with twins and Joseph does not provide us with any causal mechanisms or evidence. As noted earlier they may well be difference-producing factors.

Restricted Range of Adoptive Family Environments

There are two references here. As noted earlier we have responded to Cropanzano and James. This issue has also been dealt with by Loehlin and Horn (2000) and McGue et al. (2007). Both studies address restriction of range, and both demonstrate that it is not a serious problem.

In Johnson et al. (2007, Tables 8 and 9), we discuss the influence of 21 family measures generally considered by psychologists to be importance sources of influence on psychological traits. The range of environmental factors was generally larger for adoptees than for non-adoptees (mostly spouses of TRAs), not smaller. We demonstrated quantitatively their degree of influence on IQ and concluded that 'The similarity in placement data produced no indications of substantive influence' (p. 558). The hypothesis was refuted. This information was ignored by Joseph.

Shifting Gender Roles and Increased Career Opportunities for Women; Access to Birth Control

There are two references here and neither addresses the question of how these factors influence twin similarity in IQ. If a model of some sort were presented than it could be addressed.

Exposure to the Mass Media, Internet, Social Media

There are two references here and neither addresses the question of how these factors influence twin similarity in IQ. If a model of some sort were presented, then it could be addressed. Any given cohort is exposed to these effects. Does this mean that they are all similar in IQ?

Diet/Nutrition

It is unclear to me, and others whom I have consulted with, how the two articles cited are relevant (causal) of TRA similarity? A mechanism and evidence would need to be provided.

Replication and Multiple Corroboration (Converging Evidence)

According to Joseph:

Some people might defend the MISTRA IQ study's conclusions on the grounds that, as mentioned in the *Science* article abstract, researchers performing other types of behavioral genetic studies arrived at similar conclusions. There are at least two ways to counter such a point. (a) A psychological study, test, or method must stand or fall on its own logic and soundness and cannot be validated by supposed 'converging evidence' from other methods (Lilienfeld, Lynn, & Lohr, 2003). (b) Several authors have challenged the findings of the other TRA studies . . . (p. 62).

Psychologist Leon Kamin was a pioneering analyst of TRA research (Joseph, 2018; Kamin, 1974), and psychologist Susan Farber (1981) published an exhaustive critical analysis of the three TRA studies conducted prior to the MISTRA. (p. 49)

Statement (a) is false. No study is perfect, including MISTRA, and that is why research must rely on constructive replication and multiple corroboration (triangulation). Gelman and Loken (2014) have pointed out that 'Criticism is easy, research is hard. Flaws can be found in any research design if you look hard enough' (p. 464). Sandra Scarr (1981) has described the logic underlying constructive replication succinctly:

the most important fact is that the flaws of one study are not the same as those of another; there are nonoverlapping cracks in the evidence. Even though one adoption study confounds age of placement with preadoptive experience, the next does not; the second study compares samples of biological and adoptive families with different parents, whereas the first study sampled only adoptive parents-most of whom had their own biological children. Each study can be criticized for its lack of perfection, but laid on top of one another, the holes do not go clear through.

Joseph has utterly failed to undermine either the logic or the soundness of MISTRA. Citing Lilienfeld is ironic. As a graduate student he gathered psychophysiological measurements from TRAs. If he were still alive, he would refute Joseph (Lilienfeld, 2010, p. 284 on radical environmentalism, p. 286 on pseudoscience). Studies of both genetic and environmental influences require a combination of research strategies (Rutter et al., 2001).

At no point does Joseph refute the converging evidence (e.g., the animal work cited at the beginning of this manuscript) in favor of the hypothesis that genetic factors influence human traits. Virtually all traits in all species that have been studied demonstrate genetic influence assessed by estimates of heritability; 'the interesting questions remaining are, How does the magnitude of h^2 differ among characters and species and why?' (Lynch & Walsh, 1998, p. 175). How the human brain and mind evolved continues to be controversial and puzzling (Gangestad & Simpson, 2007; Mitchell, 2018), but there is little doubt that they evolved (Bouchard, 2014). Belief that things are otherwise is 'cognitive creationism' (Shermer, 2017).

The 'so-called' challenges by Kamin, Farber (both cited above), and Taylor to the previous TRA studies, have been refuted (failed to replicate), a fact well known to Joseph, but omitted from his report. The only replication crisis in this domain of research is the failure of the proposed environmental explanations to replicate.

Policy Implications

Joseph attacks MISTRA for publishing results but not reporting the case studies and not making the raw data publicly available (lack of transparency discussed below). These actions supposedly result in support of 'far-right white-nationalist political groups' and 'racist research and eugenics'.

His argument is that we are somehow responsible for what others do with our findings. This reasoning is reminiscent of John Horgan's tying the MISTRA IQ research to Nazism and eugenics (Horgan, 1993).

Vincent Sarich (1993) has pointed out the fallacy in this argument:

Horgan includes the obligatory connection of eugenics (incorporating the dubious link between genes and behavior) and the Nazis, but why do we hear nothing of what might be termed 'eumemics'?⁴ After all, Stalin and Mao, in the name of 'eumemics', each systematically murdered far more

⁴Eumemics is the belief and practice of improving the mememic quality of the human population. This can be done either voluntarily, by the individuals making an effort to acquire and share beneficial ideas and reject harmful ones, or involuntarily, by statist aggression that engages in censorship and other means of forcible memocide.

human beings than Hitler did. Human beings can do nasty things to one another appealing to most any ideology, but surely the entirety of human evolutionary history tells us that knowledge is preferable to ignorance. We may get it wrong sometimes; and we may even get it right, yet misapply our knowledge; but it is going to be very difficult to sustain the argument which says that we are better off not knowing. How can we be better off not knowing? How can you do something with nothing? And ignorance, by definition, is nothing.”

The similarity of Joseph’s paper to Horgan’s is striking. Even though I had supplied Horgan with reprints (most cited in this manuscript) refuting Kamin’s ‘causal claims’ regarding the similarity of TRAs. Horgan cited Kamin as though the assertions were true causes, stating that ‘in his investigation of other twin studies, Kamin has shown that identical twins supposedly raised apart are often raised by members of their families or by unrelated families in the same neighborhood; some twins had extensive contact with each other while growing up’.⁵

Joseph’s and Horgan’s comments are examples of virtue signaling. In the last paragraph of his book on the suppression of speech and writing, James Flynn (2020) asserted the following:

Looming over this whole debate is a terrible temptation: the assumption that since you know that virtue is on your side, truth must be on your side — and that an honest effort to perceive the truth is immoral. That is the surest road to hell for an otherwise honorable human being. (p. 300)

Gladys and Helen

According to Joseph, in 1976, Bouchard ‘recognized that being raised in a “less than-favorable” environment — as opposed to a favorable one could lower a person’s expected IQ score by 24 points . . . For the pre-MISTRA Bouchard, environmental influences were just that — environmental influences — and at least in this example, for him they had a powerful effect on determining a person’s IQ score’ (p. 60). The implication appears to be that somehow I had ignored the 1976 insight in the 1990 paper. The truth is the opposite: ‘It is not that we have found no evidence of environmental influence; in individual cases environmental factors have been highly significant (for example, the 29 IQ point difference in Fig. 1)’ (Bouchard et al., 1990, p. 225).

Data Kept Secret

Because we have not published case studies and the raw data, Joseph accuses us of ‘data hoarding’. Anyone who has conducted research of the type that characterizes MISTRA (which was both a psychological and medical study with some of the mental ability data being gathered by hospital staff) would recognize that MISTRA was required by the University of Minnesota Institutional Review Board to gather informed consent from all participants and guarantee confidentiality (discussed in detail by Segal, 2012, p. 113). Many of the participants were highly visible during the study and would not have participated without the assurance of confidentiality regarding both their data and participation. Some of them are easily identifiable even today. Informed consent became more and more rigorous regarding confidentiality as the study progressed. Given Joseph’s meticulous examination of the TRA literature, he cannot have been unaware as far back as 1937 when one of the separated twin pairs in the Newman et al.

⁵The article contains numerous additional elisions but there is not space to discuss them here. When asked about omissions and distortions in his article, Horgan’s response was ‘What’s wrong with writing a polemical article?’ (Wright, 1998, p. 225). My response is: when writing for a scientific magazine in which readers expect a high degree of objectivity, it is called dishonesty.

(1937) sued the authors because they revealed IQ scores. This event is discussed by Kamin (1974, p. 54). If the regulations we worked under had been in effect at the time of their work none of the previous TRA projects would have been able to publish case studies.

Genetic Confirmation Bias

Citing a footnote in one of our papers Joseph argues, ‘It appears that genetic confirmation bias was built into the MISTRA computer software program’ (p. 61). The program used is not MISTRA software. The program, as the footnote indicates, is Mx, one of the most widely used publicly available behavior genetics software programs.

Conclusion

In an astute critique of both physical and social science research, the distinguished physicist Richard Feynman (1974) put forth the following axiom.

The first principle is that you must not fool yourself — and you are the easiest person to fool. So you have to be very careful about that. After you’ve not fooled yourself, it’s easy not to fool other scientists . . . In summary, the idea is to try to give *all* of the information to help others to judge the value of your contribution; not just the information that leads to judgment in one particular direction or another.

Joseph points out that ‘Some parts of the submitted work have been adapted from the author’s non-peer-reviewed online articles. Areas adapted from the author’s 2015 book on twin research are cited (Joseph, 2015)’ (p. 62). Joseph cites the 2015 book 13 times.

In his book review, Eric Turkheimer (2015) characterized Joseph’s 2015 book as follows.

It is not a given that both sides of every argument are being reasonable. In the final analysis, this book is not reasoning forward from a known set of facts, seeking their explanation; it is confabulating backwards from a fixed conclusion, eliding any segments of the evidence that don’t lead to the preordained destination. *The Trouble With Twin Studies* is science denial.

Joseph has violated the ‘total evidence rule’ (Lubinski, 2000, p. 443). He has fooled himself. His work is not science — it is, to use Feynman’s term, pseudoscience.

Acknowledgments. Several colleagues have made helpful suggestions on how to improve the manuscript. The views expressed are my own as are any errors or omissions.

Statement of ethics. No human or animal participants were used in this study. There was no study protocol to approve.

Financial support. This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Competing interest. None.

Ethical standards. The author asserts that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

References

- Arvey, R. D., Bouchard, T. J., Jr., Segal, N. L., & Abraham, L. M. (1989). Job satisfaction: Environmental and genetic components. *Journal of Applied Psychology*, 74, 187–192. doi: 10.1037/0021-9010.74.2.187
- Border, R., Johnson, E. C., Evans, L. M., Smolen, A., Berley, N., Sullivan, P. F., & Keller, M. C. (2019). No support for historical candidate gene or

- candidate gene-by-interaction hypotheses for major depression across multiple large samples. *American Journal of Psychiatry*, 176, 376–387. doi: [10.1176/appi.ajp.2018.18070881](https://doi.org/10.1176/appi.ajp.2018.18070881)
- Bouchard, T. J., Jr. (1982a). Identical twins reared apart: Reanalysis or pseudo-analysis? [Review of Farber, S. L., (1981). Identical twins reared apart: A reanalysis]. *Contemporary Psychology*, 27, 190–191. doi: [10.1037/021001](https://doi.org/10.1037/021001)
- Bouchard, T. J., Jr. (1982b). Review of 'The Intelligence Controversy' by H. J. Eysenck and L. Kamin. *American Journal of Psychology*, 95, 346–349.
- Bouchard, T. J., Jr. (1984). Twins reared together and apart: What they tell us about human diversity. In S. Fox (Ed.), *Individuality and determinism: Chemical and biological bases*. New York: Plenum.
- Bouchard, T. J., Jr. (1993a). Genetic and environmental influences on adult personality: Evaluating the evidence. In J. Hettema & I. Deary (Eds.), *Foundations of personality* (pp 15–44). Kluwer Academic Publishers.
- Bouchard, T. J., Jr. (1993b). The genetic architecture of human intelligence. In P. A. Vernon (Ed.), *Biological approaches to the study of human intelligence* (pp. 33–93). Ablex.
- Bouchard, T. J., Jr. (1998). Genetic and environmental influences on adult intelligence and special mental abilities. *Human Biology*, 70, 257–279.
- Bouchard, T. J., Jr. (2014). Genes, evolution and intelligence. *Behavior Genetics*, 44, 549–577. doi: [10.1007/s10519-014-9646-x](https://doi.org/10.1007/s10519-014-9646-x)
- Bouchard, T. J., Jr., Arvey, R. D., Keller, L. M., & Segal, N. L. (1992). Genetic influences on job satisfaction: A reply to Cropanzano and James. *Journal of Applied Psychology*, 77, 89–93. doi: [10.1037/0021-9010.77.1.89](https://doi.org/10.1037/0021-9010.77.1.89)
- Bouchard, T. J., Jr., & Johnson, W. (2021). Keith Hayes's experience-producing drives: An appreciation and extension. *Personality and Individual Differences*, 169, 110082. doi: [10.1016/j.paid.2020.110082](https://doi.org/10.1016/j.paid.2020.110082)
- Bouchard, T. J., Jr., Lykken, D. T., McGue, M., Segal, N. L., & Tellegen, A. (1990). Sources of human psychological differences: The Minnesota Study of Twins Reared Apart. *Science*, 250, 223–228. doi: [10.1126/science.2218526](https://doi.org/10.1126/science.2218526)
- Bouchard, T. J., Jr., & McGue, M. (1981). Familial studies of intelligence: a review. *Science*, 212, 1055–1059. doi: [10.1126/science.7195071](https://doi.org/10.1126/science.7195071)
- Colarelli, S. M., & Arvey, R. D. (Eds.). (2015). *The biological foundations of organizational behavior*. University of Chicago Press.
- Cropanzano, R., & James, K. (1990). Some methodological considerations for the behavior genetic analysis of work attitudes. *Journal of Applied Psychology*, 75, 433–439. doi: [10.1037/0021-9010.75.4.433](https://doi.org/10.1037/0021-9010.75.4.433)
- Darlington, C. D. (1954). Heredity and environment. In *Proceedings of the 9th International Congress of Genetics* (pp. 370–381).
- Darwin, C. (1859). *On the origin of the species by means of natural selection, or the preservation of favored races in the struggle for life: A facsimile of the first edition with an introduction by Ernst Mayr* (1964). Harvard University Press.
- Darwin, C. (1871). *The descent of man, and selection in relation to sex*. John Murray.
- Eysenck, H. J., & Kamin, L. J. (1981). *The intelligence controversy*. Wiley.
- Fancher, R. E. (1985). *The intelligence men: Makers of the IQ controversy*. W. W. Norton & Co.
- Farber, S. L. (1981). *Identical twins reared apart: A reanalysis*. Basic Books.
- Feynman, R. P. (1974). *Cargo cult science: Some remarks on science, pseudoscience, and learning how to not fool yourself*. Caltech's 1974 commencement address. <https://calteches.library.caltech.edu/51/2/CargoCult.htm>. Reprinted in Feynman, R. P. (1985). 'Surely you're joking Mr. Feynman'. W. W. Norton.
- Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society Edinburgh*, 52, 399–433.
- Flynn, J. R. (2020). *A book too risky to publish: Free speech and universities*. Academic Press.
- Fulker, D. W. (1975). Review of: Kamin, L. The Science and Politics of IQ. *American Journal of Psychology*, 88, 505–537.
- Galton, F. (1892/1962). *Hereditary genius: An inquiry into its laws and consequences (with an introduction by C. D. Darlington)*. World Publishing Company.
- Gangestad, S., & Simpson, J. A. (Eds.). (2007). *The evolution of mind: fundamental questions and controversies*. Guilford Press.
- Gärtner, K. (1990). A third component causing random variability beside environment and genotype. A reason for the limited success of a 30 year long effort to standardize laboratory animals? *Laboratory Animals*, 24, 71–77. doi: [10.1258/002367790780890347](https://doi.org/10.1258/002367790780890347)
- Gauch, H. G., Jr. (2006). Winning the accuracy game. *American Scientist*, 94, 133–141.
- Gelman, A., & Loken, E. (2014). The statistical crisis in science. *American Scientist*, 102, 460–465.
- Grandin, T. (2022). *Genetics and the behavior of domestic animals* (3rd ed.). Academic Press.
- Heath, A. C. (1982). Review of: Taylor, H. The IQ Game. *Psychological Medicine*, 12, 213–215.
- Hoffman, L. W. (1991). The influence of the family environment on personality: Evidence for sibling differences. *Psychological Bulletin*, 110, 187–203.
- Horgan, J. (1993). Eugenics revisited. *Scientific American*, 268, 122–132. doi: [10.1038/scientificamerican0693-122](https://doi.org/10.1038/scientificamerican0693-122)
- Hu, Y. (2018). First impression of personality from body shapes. *Psychological Science*, 29, 1969–1983.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine*, 2, 696–701. doi: [10.1371/journal.pmed.0020124](https://doi.org/10.1371/journal.pmed.0020124)
- Jackson, D. N. (1975). Intelligence and ideology. *Science*, 189, 1078–1080.
- Johnson, W., Bouchard, T. J., Jr., McGue, M., Segal, N. L., Tellegen, A., Keyes, M., & Gottesman, I. I. (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. doi: [10.1016/j.intell.2006.10.003](https://doi.org/10.1016/j.intell.2006.10.003)
- Joseph, J. (2015). *The trouble with twin studies: A reassessment of twin research in the social and behavioral sciences*. Routledge/Taylor & Francis Group.
- Joseph, J. (2018, April 4) Leon J. Kamin (1927–2017): A nemesis of genetic determinism and scientific racism. Mad in America. <https://www.madinamerica.com/2018/04/leon-j-kamin-nemesis-genetic-determinism/>.
- Joseph, J. (2022). A reevaluation of the 1990 'Minnesota Study of Twins Reared Apart' IQ Study. *Human Development*, 66, 48–65. doi: [10.1159/000521922](https://doi.org/10.1159/000521922)
- Joshi, R. S., Rigau, M., García-Prieto, C. A., Castro de Moura, M., Piñeyro, D., Moran, S., Davalos, V., Carrión, P., Ferrando-Bernal, M., Olalde, I., Lalueza-Fox, C., Navarro, A., Fernández-Tena, C., Aspandi, D., Sukno, F. M., Binefa, X., Valencia, A., & Esteller, M. (2022). Look-alike humans identified by facial recognition algorithms show genetic similarities. *Cell Reports*, 40, 111257. doi: [10.1016/j.celrep.2022.111257](https://doi.org/10.1016/j.celrep.2022.111257)
- Kamin, L. J. (1974). *The science and politics of IQ*. Erlbaum.
- Koshland, D. E., Jr. (1984). Individuality in bacteria and its relationship to higher species. In S. W. Fox (Ed.), *Individuality and determinism: Chemical and biological bases* (pp. 1–31). Plenum.
- Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, S. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, 126, 390–423. doi: [10.1037/0033-2909.126.3.39](https://doi.org/10.1037/0033-2909.126.3.39)
- Lewontin, R. C., Rose, S., & Kamin, L. J. (1984). *Not in our genes: Biology ideology, and human nature*. Pantheon.
- Lilienfeld, S. O. (2010). Can psychology become a science. *Personality and Individual Differences*, 49, 281–288.
- Lilienfeld, S. O., Lynn, S. J., & Lohr, J. M. (Eds.). (2003). *Science and pseudoscience in clinical psychology*. Guilford Press.
- Loehlin, J. C. (1981). Review of: Farber, S. L. Identical twins reared apart: A reanalysis. *Acta Geneticae Medicae et Gemellologiae*, 30, 297–298.
- Loehlin, J. C., & Horn, J. M. (2000). Stoolmiller on restriction of range in adoption studies: A comment. *Behavior Genetics*, 30, 245–247. doi: [10.1023/a:1001922509893](https://doi.org/10.1023/a:1001922509893)
- Lubinski, D. (2000). Scientific and social significance of assessing individual differences: 'Sinking shafts at a few critical points'. *Annual Review of Psychology*, 51, 405–444. doi: [10.1146/annurev.psych.51.1.405](https://doi.org/10.1146/annurev.psych.51.1.405)
- Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin*, 70, 151–159. doi: [10.1037/h0026141](https://doi.org/10.1037/h0026141)
- Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sinauer.
- Mackintosh, N. J. (1998). *IQ and human intelligence*. Oxford University Press.

- McGue, M., Bouchard, T. J., Jr., Iacono, W. G., & Lykken, D. T. (1993). Behavior genetics of cognitive ability: A life-span perspective. In R. Plomin & G. E. McClearn (Eds.), *Nature, nurture and psychology* (pp. 59–76). American Psychological Association.
- McGue, M., Keyes, M., Sharma, A., Elkins, I., Legrand, L., Johnson, W., & Iacono, W. G. (2007). The environments of adopted and non-adopted youth: Evidence on range restriction from the Sibling Interaction and Behavior Study (SIBS). *Behavior Genetics*, 37, 449–462. doi: [10.1007/s10519-007-9142-7](https://doi.org/10.1007/s10519-007-9142-7)
- Mitchell, K. J. (2018). *Innate: How the wiring of our brains shapes who we are*. Princeton University Press.
- Molenaar, P. C. M., Boomsma, D. I., & Dolan, C. V. (1993). A third source of developmental differences. *Behavior Genetics*, 6, 519–524.
- Newman, H. H., Freeman, F. N., & Holzinger, K. J. (1937). *Twins: A study of heredity and environment*. University of Chicago Press.
- Ostrander, E. A., & Wayne, R. K. (2005). The canine genome. *Genome Research*, 15, 1706–1716. doi: [10.1101/gr.3736605](https://doi.org/10.1101/gr.3736605)
- Panizzon, M. S., Vuoksima, E., Spoon, K. M., Jacobson, K. C., Lyons, M. J., Franz, C. E., Xian, H., Vasilopoulos, T., & Kremen, W. S. (2014). Genetic and environmental influences on general cognitive ability: Is g a valid latent construct? *Intelligence*, 43, 65–76. doi: [10.1016/j.intell.2014.01.008](https://doi.org/10.1016/j.intell.2014.01.008)
- Pearson, K. (1924). *The life, letters, and labours of Francis Galton* (vol. 2). Cambridge University Press.
- Pedersen, N. L., Plomin, R., Nesselroade, J. R., & McClearn, G. E. (1992). A quantitative genetic analysis of cognitive abilities during the second half of the life span. *Psychological Science*, 3, 346–353. doi: [10.1111/j.1467-9280.1992.tb00045.x](https://doi.org/10.1111/j.1467-9280.1992.tb00045.x)
- Price, B. (1950). Primary biases in twin studies, a review of prenatal and natal difference-producing factors in monozygotic pairs. *American Journal of Human Genetics*, 2, 293–352.
- Price, B. (1978). Bibliography on prenatal and natal influences in twins. *Acta Geneticae Medicae et Gemellologiae*, 27, 97–113.
- Richardson, K., & Norgate, S. (2006). A critical analysis of IQ studies of adopted children. *Human Development*, 49, 319–335. doi: [10.1159/00009653](https://doi.org/10.1159/00009653)
- Ritchie, S. J. (2020). *Science fiction: How fraud, bias, negligence, and hype undermine the search for truth*. Metropolitan Books.
- Rutter, M., Pickels, A., Murray, R., & Eaves, L. J. (2001). Testing hypotheses on specific environmental causal effects on behavior. *Psychological Bulletin*, 127, 291–324.
- Sarich, V. (1993). Unpublished letter to the editors of *Scientific American*—26 May 1993.
- Scarr, S. (1976). Review of: The science and politics of IQ. *Contemporary Psychology*, 21, 98–99.
- Scarr, S. (Ed.). (1981). Having the last word. In *Race, social class, and individual differences in IQ* (pp. 523–532). Erlbaum.
- Schmidt, F. L., & Oh, I.-S. (2016). The crisis of confidence in research findings in psychology: Is lack of replication the real problem? Or is it something else? *Archives of Scientific Psychology*, 4, 32–37. doi: [10.1037/arc0000029](https://doi.org/10.1037/arc0000029)
- Scott, J. P. (1990). Foreword. In M. E. Hahn, J. K. Hewitt, N. D. Henderson, & R. H. Benno (Eds.), *Developmental behavior genetics: Neural, biometrical, and evolutionary approaches*. Oxford.
- Segal, N. L. (1999). *Entwined lives: Twins and what they tell us about human behavior*. Dutton.
- Segal, N. L. (2012). *Born together-reared apart: The landmark Minnesota Twin Study*. Harvard University Press.
- Segal, N. L., Hernandez, B. A., Graham, J. L., & Ettinger, U. (2018). Pairs of genetically unrelated look-alikes: Further tests of personality Similarity and social affiliation. *Human Nature*, 29, 402–417. doi: [10.1007/s12110-018-9326-2](https://doi.org/10.1007/s12110-018-9326-2)
- Shermer, M. (2017). For the love of science. *Scientific American*, 318, 77. doi: [10.1038/scientificamerican011877](https://doi.org/10.1038/scientificamerican011877)
- Shields, J. (1978). MZA twins: Their use and abuse. In W. Nance (Ed.), *Twin research psychology and methodology*. Allan R. Liss.
- Szucs, D., & Ioannidis, J. P. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS Biology*, 15, e2000797. doi: [10.1371/journal.pbio.2000797](https://doi.org/10.1371/journal.pbio.2000797)
- Szucs, D., & Ioannidis, J. P. A. (2021). Correction: Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS Biology*, 19, e3001151. doi: [10.1371/journal.pbio.3001151](https://doi.org/10.1371/journal.pbio.3001151)
- Taylor, H. F. (1980). *The IQ game: A methodological inquiry into the heredity environment controversy*. Rutgers University Press.
- Turkheimer, E. (2015). Arsonists at the Cathedral—A review of: ‘The trouble with twin studies: A reassessment of twin research in the social and behavioral sciences’. *PsyCRITIQUE*, 60. doi: [10.1037/a0039763](https://doi.org/10.1037/a0039763)
- Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, 76, 105–110.
- Urbach, P. (1974a). Progress and degeneration in the ‘IQ Debate’ (I). *British Journal of the Philosophy of Science*, 25, 99–135.
- Urbach, P. (1974b). Progress and degeneration in the ‘IQ debate’ (II). *British Journal of the Philosophy of Science*, 25, 235–259.
- Wainer, H. (2007). The most dangerous equation: Ignorance of how sample size affects statistical variation has created havoc for nearly a millennium. *American Scientist*, May–June, 249–256.
- Willoughby, E. A., Giannoulis, A., Lee, J. J., Iacono, W. G., McGue, M., & Vrieze, S. (2022). A century of behavior genetics at the University of Minnesota. *Twin Research and Human Genetics*, 25, 211–225. doi: [10.1017/thg.2023.1](https://doi.org/10.1017/thg.2023.1)
- Wright, W. (1998). *Born that way: Genes, behavior, personality*. Knopf.
- York, R. A. (2018). Assessing the genetic landscape of animal behavior. *Genetics*, 209, 223–232. doi: [10.1534/genetics.118.300712](https://doi.org/10.1534/genetics.118.300712)
- Young, A. I., Benonisdottir, S., Przeworski, M., & Kong, A. (2019). Deconstructing the sources of genotype-phenotype associations in humans. *Science*, 365, 1396–1400. doi: [10.1126/science.aax3710](https://doi.org/10.1126/science.aax3710)
- Zebrowitz, L. A. (1997). *Reading faces: Window to the soul?* Westview Press.
- Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why it matters. *Social and Personality Compass*, 2, 1497–1517.
- Zebrowitz, L. A., & Rhodes, G. (2004). Sensitivity to ‘bad genes’ and the anomalous face overgeneralization effect: Cue validity, cue utilization, and accuracy in judging intelligence and health. *Journal of Nonverbal Behavior*, 28, 167–185.
- Zwaan, R. A., Alexander, E., Lucas, M. B., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, 41, e120. doi: [10.1017/S0140525X17001972](https://doi.org/10.1017/S0140525X17001972), e120