- Title: "Identical Twins Reared Apart: Reanalysis or Pseudo-analysis?"
- Author: Thomas J. Bouchard Jr.
- Journal: *PsycCRITIQUES* v. 27, n. 3 (Mar 1982) : 190-191
- DOI: 10.1037/021001
- Portico Content Set: ISSN_00107549, (American Psychological Association)
- ISSN: 15540138, 00107549
- Portico Item ID: ark:/27927/phzkfk9w1
- Type: book review of *Identical Twins Reared Apart: A Reanalysis*, by Susan L. Farber, 1981.

Identical Twins Reared Apart: Reanalysis or Pseudo-analysis?

by Thomas J. Bouchard Jr.

This book attempts to pull together in a single place all of the data available on identical twins separated early in life and reared apart (MZA twins). It is a Herculean effort that few would have the courage to attempt. The author has arranged the data in a logical fashion. There is an introductory chapter that deals with methodological issues (biases, design problems, etc.), a chapter that reviews prior studies, a chapter that describes the entire population of such twins and assigns them case numbers, and chapters on physical traits, physical symptoms and disorders, psychosis, IQ, personality, and a summing up.

Twenty-five percent of the book deals with IQ. This is the most controversial topic covered and the one least satisfactorily dealt with. Consequently, I will devote most of the review to it, with the expectation that other reviewers will focus on different topics.

Farber's approach to the analysis of the IQ portion of this data set is akin to the approach of Leon Kamin (1974). The data are subgrouped using a variety of criteria that, although plausible on their face, yield the smallest genetic estimates that can be squeezed out. Statistical significance tests are liberally applied and those favorable to the investigator's prior position are emphasized. Lack of statistical significance is overlooked when it is convenient to do so, and multiple measurements of the same construct (constructive replications within a study) are ignored. There is repeated use of significance tests on data chosen post hoc. The sample sizes are often very small, and the problem of sampling error is entirely ignored. It cannot be argued strongly enough that significance tests simply do not protect against drawing false conclusions when post hoc procedures are used. The results seriously abuse statistical theory and reinforce the widespread belief that scientists can prove (or disprove) anything with statistics. In sum, the treatment of the IQ data is an exercise in obfuscation. Perhaps this new approach needs a name. I suggest the term "pseudo-analysis."

This approach should be contrasted with the rapidly developing methodologies of meta-analysis introduced by Glass and colleagues in educational psychology (1981), and Hunter, Schmidt, and colleagues in industrial/organizational psychology (1981). For years reviewers in both domains bemoaned the inconsistency of findings and postulated innumerable moderator effects. It is now known that the conflicting results are most often artifactual. That is, they depend entirely on such artifacts as sampling error, differing quality of measurement from study to study, differing reliabilities, and so forth. The meta-analyst depends heavily on pooled data and the idea of constructive replication.

The literature on the IQs of monozygotic twins reared apart (MZAs) is unique in that almost all the actual IQs are available (often on more than one test). Farber has prepared a table with MZA twins ordered by degree of separation. Group I is made up of *Highly Separated* cases ($n = 45$). Group II is composed of *Mixed Separation* cases ($n = 23$). Group III consists of *Little Separation* cases ($n = 27$), and Group IV includes twins separated after four years ($n = 11$). The table lists (when available) sex, handedness, birth weight, birth order, age of separation, knowledge of twinship prior to reunion, age learned of twinship, age met, age seen by investigator, a measure of contact over four developmental periods, years apart and degree of contact, rearing status (parent, relative, other), and reference for case. It is an extremely useful table. Later in the book the same ordering is used to report twin IQs. I was somewhat suprised and distressed to find that the author chose not to include Otis IQs for the Newman, Freeman, and Holzinger sample, nor did she include the Raven raw scores for the Juel-Nielsen sample. The point is not minor. Both Farber and Kamin have criticized the Stanford-Binet used by Newman, Freeman, and Holzinger. The Otis gives almost the same intraclass correlations and means as the Stanford-Binet, however (r_i = .74, mean = 97.16, SD = 13.58 for Otis; r_i = .68, mean = 95.68, SD = 13 for the Stanford-Binet). Multiple measurement thus reveals that the criticisms of the Stanford-Binet data are simply spurious. The same story applies to the Juel-Nielsen data. Farber cites Kamin regarding problems of standardization of the Wechsler-Bellevue test (W-B) in Denmark. The Raven scores, however, when transformed to IQ equivalents using Raven's percentile table "where the raw score values have been noted with the corresponding probit-values (with regard to age grouping)" yield a twin correlation of .73, the Raven IQ and W-B IQs correlate .82. Again, two different tests, based on different norms, give very similar results. The criticism of the single test is thus shown to be spurious. Farber, like Kamin, makes a variety of criticisms of the Dominoes and Mill-Hill used by Shields. Shields (1978) in his last publication before he died defended himself well against a number of unjustified accusations, and I will not repeat his arguments here. I will show below that Shields's results are so comparable to that of the others that these criticisms must also be considered spurious.

The overall analysis of the IQ data yields the following results: $r_i = .75$, mean 96.8, SD = 12.8. This set of data is, however, considered by Farber to be biased in a variety of ways—principally by contamination due to contact. This problem is treated in two ways. The first involves analysis of all of the cases. An analysis

of variance (A-Sex; B-Degree of Contact, 9 levels; A X B; C-Pairs; D-Individuals) was conducted on the Verbal, Performance, and Full Scale IQ scores. A second analysis with a second measure of separation was also used. The results of both analyses are shown in a table where intraclass correlations are given with and without "separation" taken into account. This table is a masterful example of the ignoring of significance tests when they are inconveniently insignificant. In 13 out of 18 significance tests, separation has no effect. For the combined sexes separation has no statistically significant effect for Verbal, Performance, or Full Scale IQ. A similar misleading analysis is carried out on a "PURE" sample (one significant effect out of six statistical tests).

By this point I was persuaded that separation probably had little or no effect on similarity between twins. I decided to calculate intraclass r's for the Highly Separated group for whom I had expected to find an analysis but had not. 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: $n = 13$, $r_i = .76$, mean = 96.35, SD = 14.20. The three arrays show the slight depression in IQ characteristic of most twin samples, a standard deviation comparable to the normative population, identical intraclass r's that are indistinguishable from the full sample where separation is ignored (full sample males, $n = 32$, $r_i = .74$; females, $n = 50$, $r_i = .76$; combined sex, $n = 82$, $r_i = .75$). If we drop the Shields data over which the author frets so much, the results are the following: $n = 28$, $r_i = .78$, mean = 99.36, SD = 14.94. Notice that the correlation goes up, not down. The Highly Separated group mimics the full sample perfectly. The inclusion or exclusion of the Shields data makes no difference whatsoever.

Farber, in her introductory chapter, quotes Fuller and Thompson regarding the limited contributions of statistical methodology to our understanding of the genetics of behavioral traits. She then concludes, "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 even much common sense."

Unfortunately analyses of the IQ scores in this book consist of the following: a great deal of action with numbers (there are 70 tables, graphs, and figures dealing with IQ); retrogression as opposed to advance in our understanding; inferences either flatly wrong or nonsensical; conclusions widely at variance with what we know about intelligence and IQ tests, irrespective of the MZA data.

The IQ correlation between MZA twins is between .75 and .80 and is probably not influenced by most of the so-called artifacts bandied about by critics of this literature. McNemar (1938) predicted precisely these values in his review of the very first MZA study when he corrected the Newman, Freeman, and Holzinger data for age and range—Binet .767, Otis .796.

I have passed a harsh but, I believe, a fair judgment on the IQ analysis. What

about the remaining topics? My general impression is that the compilations are dependable and a good starting place for someone working in a specialized area. Most of the conclusions are consistent with my understanding of the areas (an understanding that is sometimes limited, e.g., in the areas cancer, circulatory system disorders, etc.). In the personality area Farber suspects that "twins who had the least opportunity to influence each other were the most similar." Although reported elsewhere by others, this is a daring conclusion, and there is a suggestion in our own data that this may be confirmed.

The case index, subject index, and name index are excellent. The bibliography is good, but not excellent. The best single reference on primary bias in twin studies (Price, 1950) is not cited.

This is not the definitive work on identical twins reared apart, and because of the flaws outlined above I cannot recommend it as an introduction to the topic area. It will, however, serve as a useful resource for specialists in the area of twin studies.