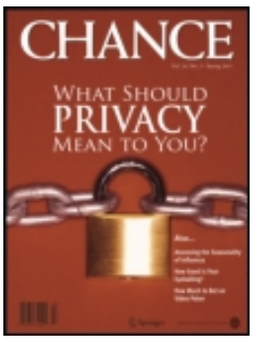


CHANCE

ISSN: 0933-2480 (Print) 1867-2280 (Online) Journal homepage: http://www.tandfonline.com/loi/ucha20

Misuses of Statistics in the Study of Intelligence: The Case of Arthur Jensen

Jack Kaplan , Conor V. Dolan & Arthur R. Jensen

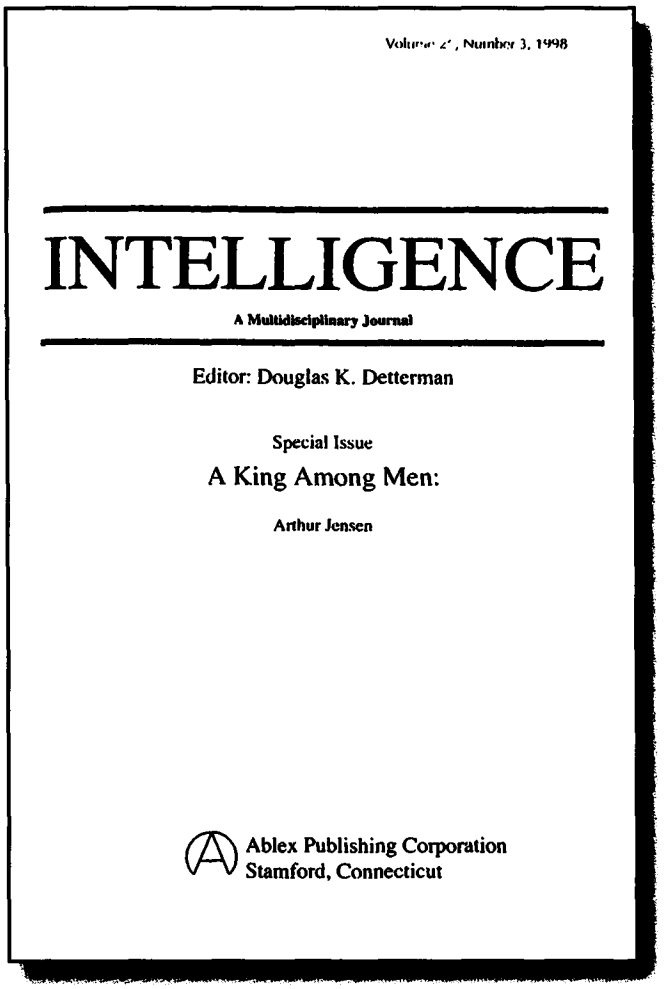
To cite this article: Jack Kaplan , Conor V. Dolan & Arthur R. Jensen (2001) Misuses of Statistics in the Study of Intelligence: The Case of Arthur Jensen, CHANCE, 14:4, 14-26, DOI: 10.1080/09332480.2001.10542294

To link to this article: http://dx.doi.org/10.1080/09332480.2001.10542294

Statistics plays a key role in the nature/nurture debate about intelligence.

Misuses of Statistics in the Study of Intelligence: The Case of Arthur Jensen

Jack Kaplan



Few scientific topics generate as much controversy as the study of intelligence. Practically from the moment the first intelligence test was published in France by Albert Binet in 1905, scholars and members of the general public have debated what intelligence is; whether it is, in fact, reliably measured by IQ tests; the degree to which it is genetically determined; whether some racial, ethnic, and social class groups are on average more intelligent than other groups; the relationship between intelligence and social problems such as crime, poverty, and immorality; and related questions. These debates are, to a great degree, arguments about the proper use and interpretation of statistical methods.

In the United States there have been three periods of particularly intense debate about intelligence. The first such period occurred in the 1920s, when the results of IQ tests given to Army recruits during World War I became ammunition for advocates of immigration restriction who argued that immigrants from Southern and Eastern Europe were lowering the genetic quality of the country's population. These arguments played a role — how much of a role is a matter of dispute — in the passage of the racist Immigration Restriction Act of 1924, which heavily favored immigrants from Western and Northern Europe (members of the so-called Nordic races) over the allegedly inferior Alpine and Mediterranean races from Southern and Eastern Europe who in recent decades had been coming to the country in large numbers.

A second period of intense debate took place in the late 1960s and the 1970s, centered largely on the theories advanced by William Shockley, a Nobel prize-winning physicist who turned to the study of intelligence later in life, and Arthur Jensen, an educational psychologist at the University of California at Berkeley.

The last period occurred just recently following the 1994 publication of *The Bell Curve* by Richard Herrnstein and Charles Murray (Simon and Schuster, New York).

The Role of Arthur Jensen

Although Shockley's writings are largely forgotten, Jensen has continued to be a central figure in the IQ debate up until the present day. His writings are still widely cited, and he continues (he is now in his late seventies) to be actively engaged in research. Anyone making even a cursory examination of recent books and articles on intelligence will inevitably run across many citations to his five books and literally hundreds of articles about intelligence. Herrnstein and Murray, for example, cite his 1980 book, *Bias in Mental Testing* (Free Press, New York), as the primary reference for their claim that certain assertions about intelligence — that intelligence is reliably measured by IQ tests, that intelligence is highly heritable, that IQ tests are not culturally biased, and so forth — have been scientifically proven (or, in their words, "are by now beyond significant technical dispute").

One indication of Jensen's continued influence is that a few years ago the journal *Intelligence* dedicated an issue in Jensen's honor. The issue was largely devoted to essays effusively praising Jensen, whom the journal extolled on its title page as "A King Among Men" (see illustration on previous page).

Jensen recently published a major book, *The g Factor* (Praeger, West Port, CT, 1998) and an article of his, "The Puzzle of Nongenetic Variance," was included in a recently published collection, *Intelligence, Heredity, and Environment* (Cambridge University Press, Cambridge, NY, 1997) edited by Robert Sternberg and Elena Grigorenko, that also included articles by Sternberg, Howard Gardner, H. J. Eysenck, Sandra Scarr, Robert Plomin, John Loehlin, Thomas Bouchard, and other leading intelligence researchers.

Jensen's prominence dates from the publication in 1969 of a long article about intelligence in the *Harvard Educational Review*, "How Much Can We Boost IQ and Scholastic Achievement?" The article begins with the provocative sentence, "Compensatory education has been tried and it apparently has failed."

Jensen went on to argue that the reason for its failure is that it assumes, incorrectly, that environmental factors are primarily responsible for the low achievement levels of most children from disadvantaged families in the United States. On the contrary, according to Jensen, a child's potential for educational achievement is largely determined by his or her intelligence — what some psychometricians call g — and that g is largely determined by a person's genes and therefore cannot be substantially raised by an improved environment.

Jensen estimated in his article that the heritability of intelligence is around 80%. He quoted approvingly a comment by the psychologist Edward Thorndike in 1905 that "In the actual race of life ... the chief determining factor is heredity." Jensen wrote that "the preponderance of evidence has proved [Thorndike] right, certainly as concerns those aspects of life in which intelligence plays an important part."

Although dozens (if not hundreds) of articles and quite a few books have been written criticizing Jensen, his use of statistics in particular has not been subjected to the sort of scrutiny given to Herrnstein and Murray...

Jensen's article provoked a huge controversy, among both scholars and the general public, not unlike the one that occurred more recently following the 1994 publication of *The Bell Curve*. The two subsequent issues of the *Harvard Educational Review* were largely devoted to responses, mostly critical, to Jensen's article, and many other articles appeared in other scholarly and general interest publications. Some of these articles were later collected and published as a book, *The IQ Controversy: Critical Readings* (Pantheon Books, New York, 1976) edited by N. J. Block and Gerald Dworkin.

Over the next decade or so a number of books were written, at least in part as a critical response to Jensen, including *The Mismeasure of Man* (Norton, New York, 1981) by Stephen Gould; *The Science and Politics of IQ* (L. Erlbaum Assoc, Potomac, MD, 1974) by Leon Kamin; *Not in Our Genes* (Pantheon Books, New York, 1984) by R. Lewontin, S. Rose, and L. Kamin; *Race, IQ, and Jensen* (Routledge and Kegan Paul, Boston, 1980) by J. R. Flynn; *The IQ Game* (Rutgers University Press, New Brunswick, NJ, 1980) by H. F. Taylor; and *Arthur Jensen, Consensus and Controversy* (Farmer Press, Philadelphia, 1987) edited by Sogan and Cecil Mogdil. Jensen himself elaborated his arguments in three books — *Educability and Group Differences* (Methuen, London, 1973) *Bias in Mental Testing*, and *Straight Talk About Mental Tests* (Free Press, New York, 1981) — and dozens of articles. (A fourth book, *Genetics and Education* [Methuen, London, 1972] reprinted some of Jensen's journal articles, including the one from the 1969 *Harvard Educational Review*.) Jensen became such a controversial figure personally that his classes and out-of-town lectures were sometimes disrupted by demonstrators. For a while, two plainclothes policemen were assigned to accompany him whenever he walked to and from his classes.

One indication of Jensen's prominence is that the hereditarian school of thought about intelligence is often referred to as "Jensenism." The term is even included in some dictionaries.

Jensen's Use of Statistics

Herrnstein and Murray's use of statistics has been severely criticized by a number of scholars. In a book review of *The Bell Curve* that was published in 1995 in the *Journal of the American Statistical Association*, two statisticians (Stephen Fienberg and Kathryn Roeder), a historian (Daniel Resnick), and a psychiatrist (Bernie Devlin) at Carnegie Mellon University wrote the following:

> *The Bell Curve* is superficially a powerful book, with a clear mes-

sage for all societies. It offers intellectual reinforcement to many beliefs and prejudices that continue to be voiced both publicly and privately. We believe that it should be read, but with a skeptical eye. *Where we have delved with care into its arguments and analyses, we have found these to be deeply flawed and misleading* ... The premises of [the authors'] main theme need to be strongly qualified at the very least. What's worse, the conclusions that they draw from their premises are not supported by the empirical evidence marshaled in their support. [Italics added] (pp. 1497–1498)

The same authors also wrote an article, *Wringing The Bell Curve*, in the Summer 1995 issue of *Chance* and edited a book, *Intelligence, Genes, and Success: Scientists Respond to The Bell Curve* (Springer, New York, 1997) that included a number of articles discussing the use and (mostly) misuse of statistics by Herrnstein and Murray. *Chance* also published another article about Herrnstein and Murray's use of statistics, *A Statistical Error in the Bell Curve,* by Jack Kaplan, in the Winter 1997 issue.

Although dozens (if not hundreds) of articles and quite a few books have been written criticizing Jensen, his use of statistics in particular has not been subjected to the sort of scrutiny given to Herrnstein and Murray by Fienberg, Roeder, and other statisticians, and as far as I am aware no articles have been published about Jensen intended primarily for an audience of statisticians. This is unfortunate because Jensen's use of statistics is, in my opinion, at least as flawed as Herrnstein and Murray's, and the fact that (unlike Herrnstein and Murray) he has published numerous journal articles and several books aimed at specialists tends to give his opinions some prima facie credibility.

In reading through Jensen's books and some of his articles, I have numerous times run across statistical statements that are, at a minimum, highly questionable. Described here are four such statements. Not all of them concern issues that are important in and of themselves, but they all serve to illustrate what I would argue is Jensen's tendency to draw clear-cut, strongly stated conclusions from woefully inadequate evidence. The examples concern the following questions about intelligence:

- Is the black/white difference in average IQ scores explained, at least in part, by genetic differences between the populations?
- Is the black/white difference in average IQ scores explained, at least in part, by blacks' being at greater risk of having neurological "accidents" during pregnancy and childbirth?
- Do blacks have less variability in intelligence than whites?
- Is intelligence (as distinct from scores on intelligence tests) normally distributed?

This is a standard illustration that differences between populations can be entirely environmental even when differences within populations is largely genetic.

Jensen on Genetics

American blacks on average score about one standard deviation lower than American whites on most IQ tests. Whether this difference is at least partly genetic is probably the most controversial question raised by the study of intelligence.

Jensen's Statement

Jensen wrote in *Educability and Group Differences* that genetics most likely accounts for between 50% and 75% of the average difference between the two populations. As evidence he pointed to studies of sibling regression to the population mean. Such studies, according to Jensen, show that siblings of whites have IQ scores that regress to the mean IQ of the white population, whereas siblings of blacks have IQ scores that regress to the mean of the black population. For example, if one takes a sample of blacks with IQs of 120, the average IQ of their siblings will be around 100; whereas for a sample of whites with IQs of 120, their siblings will have an average IQ of around 110. Jensen concludes the following:

> [This result] is entirely expected if one assumes a genetic model of intragroup and intergroup differences .... [but] seems difficult to reconcile with any strictly environmental theory ... that has yet been proposed. (pp. 117-118)

Discussion

Actually the result is entirely expected either way. Regression to the mean applies regardless of whether the populations differ for genetic or environmental reasons. (This point was made by J. R. Flynn in *Race, IQ, and Jensen.*)

Consider, for example, planting corn on two plots of land, one with fertile soil and the other with poor soil. Assume you use seed from the same source for both plots. Naturally the yield will be higher for corn planted in the fertile soil. This will be entirely due to the environment (the soil) rather than genetics (the seed), since both plots use the same seed. This is a standard illustration (used by Herrnstein and Murray, among others) that differences *between* populations can be entirely environmental even when differences *within* populations is largely genetic.

Now consider what happens with regression to the mean. For purposes of discussion, let us assume the average plant from the fertile plot yields 40 ounces of corn compared to 32 ounces for the average plant from the unfertile plot. Now take a sample of plants from the fertile plot that all yielded exactly 40 ounces of corn, and re-plant the same field with seed from those plants. On average we would expect the new plants to again yield 40 ounces of corn.

Now take a sample of plants from the unfertile plot that also yielded exactly 40 ounces of corn and replant that field with seed from those plants. On average we would expect the new plants to yield something less than 40 ounces of corn, how much less depending on the heritability of yield.

Thus a plant with yield 40 ounces from the fertile plot will tend to have higher-yielding offspring than a plant

with yield 40 ounces from the unfertile plot, despite the fact that the two populations are genetically identical. Intuitively, the reason for this is that plants from the unfertile plot that yield 40 ounces of corn will mostly be genetically superior plants whose superior genes are offset by a poor environment. Offspring of such plants will tend to be less genetically superior and thus have an average yield below 40 ounces. Plants from the fertile plot that yield 40 ounces of corn, on the other hand, will mostly be genetically average plants. Offspring will also tend to be average and will therefore have the same 40-ounce average yield.

Regression to different populations means is not in any way inconsistent with an environmental explanation for population differences.

Jensen on Neurological Accidents

One possible explanation for black/white IQ differences is that blacks may be at higher risk for suffering neurological damage from adverse events occurring prior to or during birth. Such events could result in a degree of intellectual impairment without being severe enough to cause a detectable abnormality.

There are a number of plausible reasons why blacks may be more vulnerable than whites to such incidents — poorer maternal health care, a higher rate of teen pregnancy, a higher prevalence of substance abuse, greater exposure to lead and other environmental pollutants, and so on. Moreover, blacks are known to experience a much higher rate than whites of measurable pregnancy complications such as prematurity, low birthweight, and infant mortality. It would seem reasonable to expect that whatever factors cause these measurable complications may, in less severe form, cause undetected harm that can impair intellectual functioning later in life.

Jensen's Statement

To evaluate this possibility, Jensen hypothesized in *Educability and Group Differences* (p. 346) that fetal damage results from stresses that can be measured by a variable that might be called "organismic viability" or "freedom from impairment." He then assumes that, above a certain threshold, values of this variable can lead to observable harm in an individual fetus, whereas values that are somewhat high but below the threshold can lead to nonobservable consequences, including impaired mental capacity.

Jensen then compared nationwide fetal death rates for blacks and whites — 25.8 deaths per 1,000 live births for blacks, 13.3 per 1,000 for whites. He assumes that these deaths correspond to values on the "organismic viability" variable that are above a certain threshold. Assuming that this variable is normally distributed with equal variances in the two populations, one can calculate that the average black value is .46 standard deviations below the average white value.

But blacks and whites differ on IQ scores by a full standard deviation. Jensen concludes that reproductive casualty can at most explain a small fraction of the black/white IQ difference.

Discussion

Although the basic premise of this argument is plausible — factors that in severe form cause observable harm may, in less severe form, cause unobservable mental impairment — to view these factors as values of a single underlying "organismic viability" quantitative variable is artificial and probably unhelpful. To assume furthermore that this variable is normally distributed with equal variances in both the black and white populations seems to me clearly unjustified.

Jensen on Variability

In most studies comparing the IQs of blacks and whites, IQs of the black sample have a smaller standard deviation than IQs of the white sample, as well as a smaller mean.

Jensen's Statement

Jensen, in *Educability and Group Differences*, spent six pages discussing in considerable detail possible reasons for this (pp. 211–216). He suggested and evaluated several possibilities — that the black population has less genetic variability than the white population, that blacks engage in assortative mating to a lesser extent than the white population, and that blacks experience less variability in their environment than the white population.

Discussion

The smaller standard deviation of IQ scores for blacks compared to whites is most likely nothing more than a statistical artifact — a consequence of the fact that IQ scores have been normalized on a predominantly white population. Since most blacks score lower than the average white score, their scores are mostly clustered in the lower half of the IQ distribution. If IQ scores were normalized on a predominantly black population, scores for whites would be mostly clustered in the upper half of the distribution, in which case the standard deviation for blacks would probably be *greater* than the standard deviation for whites — although it's not possible to know this for certain without knowing the underlying distributions of each population.

As a test of how this would work out with real data, I analyzed scores on a vocabulary test given to 2,281 black and 4,333 white high school students as part of a study conducted by the federal government. The scores of blacks on the original scale had a mean of 46.1 and a standard deviation of 9.04. The scores of whites had a mean of 55.0 and a standard deviation of 9.46.

Transforming the data so that the scores of whites were normally distributed with mean 100 and standard deviation 15, scores of blacks had a standard deviation of 13.4. But transforming the data so that the scores of blacks were normally distributed with mean 100 and standard deviation 15, the scores of whites had a standard deviation of 14.7.

Which population had more variability thus depended, as expected, on which population was used for normalizing the scores.

Jensen on Normality

Even assuming that IQ tests do measure intelligence, the question of whether *intelligence* is normally distributed is different from the question of whether *intelligence test scores* are normally distributed. Scores are forced to be nor-

mally distributed by a normalizing transformation. This can be done for any continuous variable and says nothing about the underlying trait being measured.

Jensen's Statement

Jensen argued in his *Harvard Educational Review* article that intelligence itself must be approximately normally distributed because the normalized test scores behave like an interval variable:

> If we *assume* that intelligence is "really" normally distributed in the population, and then measure it in such a way that we obtain a normal distribution of scores, our measurements (IQs) can be regarded as constituting an interval scale. If, then, the scale in fact behaves like an interval scale, there is some justification for saying that intelligence itself (not just IQ) is normally distributed. What evidence is there of the IQs behaving like an interval scale? The most compelling evidence, I believe, comes from studies of the inheritance of intelligence, in which we examine the pattern of intercorrelations among relatives of varying degrees of kinship. [*Genetics and Education,* p. 91]

Jensen went on to compare the statistical behavior of IQ scores of kinship pairs (e.g. parents and children, siblings, cousins, etc.) with the statistical behavior of heights of kinship pairs. He concluded that they both show regression to the mean, with the degree of regression to the mean varying with the closeness of the biological relationship. From this he concluded that IQ, like height, must be an interval variable and that the normal distribution of IQ scores therefore implies the normal distribution of intelligence.

Discussion

What Jensen appears to be saying here is that regression to the mean applies only to variables that are measured on an interval scale. Since IQ scores of related individuals show regression to the mean, according to this view, they must be measured on an interval scale — implying, for example, that the difference in intelligence between a person with an IQ of 100 and a person with an IQ of 105 is the same as the difference between a person with an IQ of 150 and a person with an IQ of 155. Furthermore, if IQ is an interval variable *and* IQ scores are normally distributed *and* IQ is a measure of intelligence, then intelligence must be normally distributed.

After going through a complex, lengthy, and superficially impressive statistical analysis he reaches a clear conclusion that, on closer consideration, has little or no justification.

Apparently Jensen does not understand that regression to the mean is a property that holds true for any variables which have a bivariate normal distribution, regardless of the type of variable. It makes no difference whether they are interval variables, transformed interval variables, or transformed ordinal variables.

The Puzzle of Nongenetic Variance

The four statistical errors described previously are particularly blatant and obvious ones, but Jensen makes many others as well. They are part of a larger pattern of careless thinking and sloppy analysis.

Perhaps Jensen's most bizarre use of statistics is to be found in "The Puzzle of Nongenetic Variance," the article that was published a few years ago in *Intelligence, Heredity, and Environment.*

The purpose of the article was to study the sources of IQ variance that cannot be explained genetically. Such variance is usually attributed to a person's social and cultural environment — such things as how a person was raised by his or her parents, the quality and length of a person's education, the type of neighborhood lived in, and so on. But IQ can also be affected by the biological environment — such things as nutrition, exposure to lead, prenatal exposure to alcohol or drugs, trauma before or during birth, and so on.

In an effort to understand the nature of this nongenetic variance, Jensen goes through a lengthy (18-page) and confusing statistical analysis of the IQ scores of 180 pairs of identical twins. The analysis focuses on the nonnormality of the distribution, on comparing the variance of the lower-scoring members of the twin pairs (the IQL's) to the variance of the higher-scoring members (the IQH's) in different subsets of the data, and on comparing the correlation between D, the difference in IQ scores between members of a twin pair, and IQL, the IQ of the lower-scoring member, to the correlation between D and IQH, the IQ of the higher-scoring member. (See sidebar for details of Jensen's analysis.)

Following this odd and, in my opinion, not very meaningful analysis, Jensen somehow manages to arrive at the following conclusions:

> The neural basis of mental development is affected in each individual by a limited number of physical events beginning shortly after conception, each with a biologic effect usually too small to be detected individually ... These small biologic events are a random selection from among all such micro-environmental events that may affect development ... Because the net deviations have resulted from many small, independent events, they are normally distributed in the population.
>
> Superimposed on this normal distribution of random environmental effects on IQ is a distribution resulting from a small number of comparatively large environmental effects, more often negative than positive, that "hit" only a fraction of the population. They are attributable to (1) a nonrandom, stochastic snowball effect on a few unlucky individuals ... and (2) the occurrence of rare events with large effects that "hit" only a small fraction of the population. The composite of these two distribu-

Jensen on Nongenetic Variance

In an effort to understand the nature of the nongenetic portion of the variability in IQ scores, Jensen went through the following series of calculations using the IQ scores of 180 pairs of identical twins:

For each pair of twins, Jensen calculated the absolute difference (D) between the two IQ scores. He then graphically compared the distribution of these differences to what would be expected under normality and found that it deviates substantially from the normal.

Next Jensen calculated, for each pair of twins, the deviations of the two individual IQ scores from the mean of the two scores. (For example, if the IQ's for a pair of twins are 106 and 112, D is 6, the mean is 109, and the deviations are -3 and +3.) He then calculated the kurtosis of the resulting 360 deviations. The kurtosis works out to 3.93, substantially greater than the value of 3 for a normal distribution.

Jensen next looked at the variance of IQ scores for only the lower-scoring members of the 180 twin pairs (IQLs) and compared it to the variance of scores for the higher-scoring members (IQHs). He calculated the ratio of these variances, which comes to .99.

For each value of D, Jensen then calculated the variance of the IQLs for all pairs of twins with that value of D or larger. He called these the cumulative variances of the IQLs, which he labeled cVIQL. Similarly, he calculated the cumulative variances (cVIQH) for the IQHs. For each value of D, he then calculated the ratio of cVIQL to cVIQH. He found that this ratio is greater than 1 for values of D that are 6 or larger but less than 1 for most values of D that are 5 or smaller.

Jensen then calculated the cumulative means of the IQLs and IQHs, analogous to the cumulative variances. These he labeled cIQL and cIQH. He then calculated the correlation between D and cIQL (.293) and between D and cIQH (-.075). He then recalculated these correlations separately for twin pairs with values of D that are 10 or larger and twin pairs with values of D that are 9 or small. In the first group, the correlation between D and cIQL is -.76 and between D and cIQH is -.45. In the second group, the corresponding correlations are .035 and .678.

The calculations of D, cIQL, cIQH, cVIQL and cVIQH for a hypothetical set of data are illustrated below. Jensen uses the results to argue that the different behavior in the large D pairs indicates large environmental effects (see text for more details). Does this follow from these calculations?

IQL	IQH	D	cIQL	cIQH	cVIQL	cVIQH
93	113	20	93.0	113.0	---	---
108	126	18	100.5	119.5	112.5	84.5
116	130	14	105.7	123.0	136.3	79.0
109	122	13	106.5	122.8	93.7	52.9
82	89	7	101.6	116.0	190.3	267.5
120	122	2	104.7	117.0	208.7	220.0
109	110	1	105.3	116.0	176.6	190.3
106	106	0	105.4	114.8	151.4	175.6

Jensen's calculations from "The Puzzle of Non-Genetic Variance," illustrated using artificial data. The first two columns represent the raw data - IQ scores for eight pairs of identical twins. IQL is the IQ of the lower-scoring twin. IQH is the IQ of the higher-scoring twin. D, cIQL, cIQH, cVIQL, and cVIQH are described above.

> tions of net environmental effects forms a population distribution that is leptokurtic, with excess frequencies in the two tails, especially in the tail on the negative side ...
>
> In the picture we see emerging from behavior-genetic analyses of mental abilities, the psychometric construct called *g* appears to be a biological phenomenon.

This is vintage Jensen. After going through a complex, lengthy, and superficially impressive statistical analysis he reaches a clear conclusion that, on closer consideration, has little or no justification. Statistics becomes little more than window dressing for Jensen's speculative theorizing.

Intensive, Detailed, Exhaustive

That, at any rate, is how I see it. Others see it differently.

One person who sees it differently is Thomas Bouchard of the University of Minnesota, one of the country's most prominent scholars in the area of intelligence testing. In an article entitled "Intensive, Detailed, Exhaustive" that appeared in the 1998 issue of *Intelligence* that was dedicated in Jensen's honor, Bouchard had this to say:

> These three terms [intensive, detailed, exhaustive] capture much of the flavor of Jensen's writings. I should also add fair-minded, temperate, and courageous ... Jensen's writings are virtual tutorials on how to write science and how to deal with controversy — stick to the available evidence in it's [sic] full context, carefully explain the methods, their rationale and the assumptions, acknowledge the lack of evidence when it does not exist and avoid ad hominem arguments.

Douglas Detterman of Case Western Reserve University, the journal's editor, had this to say in the same issue:

> I have never known anybody with fewer prejudices [than Jensen] ... Jensen has no loyalty whatsoever

to any theory or hypothesis even if they come from his own ideas. He would gladly know the truth even if it proved him wrong. In fact, he would be excited to know the truth.

Sandra Scarr, another prominent scholar in the field, had this to say:

> Art Jensen's contributions to psychological science are enormous, and they continue to mount. His work includes the impeccable tome on test bias, the most thoughtful research on learning and intelligence, and some critical studies on race and environment. The massive body of work will persist for generations.

Scarr dismissed Jensen critics Marcus Feldman, Stephen Jay Gould, and Leon Kamin as being "thugs with pens ... politically driven liars ... [and] despicable." Strong language indeed for a scholarly journal, and an indication of just how emotionally charged the whole issue of intelligence testing has become.

Should We Care?

Should statisticians care about all this?

I think we should. It's bad enough when statistics is misused in a book like *The Bell Curve*, written by nonspecialists for the general public. But when statistics gets extensively misused by a prominent, widely cited scholar who is admired practically to the point of hero worship by other prominent scholars, then clearly something is seriously amiss.

References and Further Reading

Gottfredson, L. S. (1998), "Jensen, Jensenism, and the Sociology of Intelligence," *Intelligence*, 26, 291–299.

Hirsch, J. (1975), "Jensenism: The Bankruptcy of 'Science' Without Scholarship," *Educational Theory*, 25, 3–27.

Jacoby, R., and Glauberman, N. (eds.) (1995), *The Bell Curve Debate: History, Documents, Opinions*, New York: Random House.

Jensen, A. R. (1998), "Jensen on 'Jensenism'," *Intelligence*, 26, 181–208.

Kempthorne, O. (1978), "Logical, Epistemological, and Statistical Aspects of Nature-Nurture Data Interpretation," *Biometrics*, 34, 1–23.

Comment:
Unhelpful Criticism in the Study of Intelligence

Conor V. Dolan

According to Jack Kaplan, Jensen misuses statistics in his studies of intelligence. To demonstrate this, Kaplan discusses a number of Jensen's results. Although I welcome critical examination of results presented in the study of intelligence, especially if they have a bearing on the issue of black–white differences in IQ, I find Kaplan's criticism largely unhelpful. First, the serious accusation of misuse is hard to judge, because Kaplan fails to state what he means by misuse of statistics. Second, Kaplan criticizes Jensen but at times does so without cogent argumentation. Finally, although this is somewhat removed from the criticism per se, I find Kaplan's juxtaposition of his critical comments with comments made in a special issue of *Intelligence* in honor of Jensen facile.

What Is a "Misuse of Statistics"?

As the theme of Kaplan's article is Jensen's *misuse of statistics*, it is unfortunate that Kaplan does not provide the reader with his definition of this term. Consider the following. Imagine a person whose job it is to sell cigarettes. I would consider it a misuse of statistics if this person presents the positive correlation between number of cigarettes smoked a day and lung capacity obtained in a sample of 13- to 18-year-olds as evidence in favor of smoking (conditioning on age will quickly reveal the actual effect of smoking on lung capacity). Assuming that this person is cognizant of the source of the positive correlation, he is misusing a statistic in pursuit of his hidden agenda. It should be clear to anyone who reads Jensen carefully that Jensen does not engage in this kind of misuse.

An examination of Kaplan's actual criticism does not help to pinpoint his meaning of this term. Consider the section "Jensen on Variability." The standard deviations of IQ test scores in black samples are consistently smaller that those in white samples. Jensen suggested a number of hypotheses to account for these differences, including differences in genetic and environmental variance and differences in the degree of assortative mating. Kaplan offers the explanation that the differences are due to the normalization of scores in the white population. Kaplan presents a result in support of his hypothesis, but the absence of any details con-

cerning his calculations makes his results hard to judge. However, with respect to the problem at hand — namely the definition of misuse — we have this: Jensen offers three hypotheses to account for black/white differences in variance and Kaplan offered one of his own. I fail to see the misuse of statistics that Jensen has perpetrated by positing his hypotheses. Even if Kaplan is completely right about the causes of the difference in variance, what misuse of statistics has Jensen actually committed ? And why does Jensen's formulation of hypotheses constitute a "particularly blatant and obvious" error ?

Next, consider the section "Jensen on Neurological Accidents." Here Jensen related a dichotomy (survival/death among infants) to a hypothetical latent variable, which he called "organismic viability" (OV). He did this by employing the so-called liability threshold model, a model that is used in genetic epidemiology (e.g., see Faraone, Tsuang, and Tsuang 1999). OV is modeled as a normal variable. An infant with a score beyond a certain value on the variable, the so-called threshold, dies. Jensen posited that the threshold is fixed and that the variance of OV is identical in both populations. He then modeled black/white differences in infant mortality by positing a difference in mean on the latent OV variable. Kaplan finds the assumption of equal variance artificial. This may be so, but in a model like the liability-threshold model one has to introduce some constraints. One cannot allow a difference in both means and variances because this gives rise to problems of identification. Jensen chose to consider the implications of a difference in means. The resulting model may be artificial, but why exactly is Jensen's application of it a misuse of statistics?

Unhelpful Criticism

Kaplan, reiterating the point made by Jensen critics Flynn and Brody, explains why sib regression is not convincing support of the plausibility of the genetic hypothesis. I consider this to be constructive criticism because Kaplan explains why such results cannot be used in this fashion (see also the section "Jensen on Normality"). It is, however, unfortunate that Kaplan fails to maintain this degree of constructiveness throughout his criticism. Consider again the section "Jensen on Neurological Accidents." Kaplan views the conceptualization of the latent variable OV as a single quantitative variable as "artificial and probably unhelpful." As criticism, "probably unhelpful" is *definitely unhelpful*. Moreover, Kaplan considers the conceptualization of the OV variable as a normal variable with equal variances in the black and white populations to be "clearly unjustified." As mentioned, equal variances should be seen as a modeling constraint, which is often employed in the liability-threshold model. Kaplan does not explain why aspects of this model are "probably unhelpful" or "clearly unjustified." We see the same in Kaplan's criticism of Jensen analysis of nongenetic variance. Kaplan finds this analysis to be "confusing," "odd," and not "very meaningful." Again Kaplan may well be right, but he makes little effort to explain why this is so. As such, I regard these criticisms to be gratuitous and therefore unhelpful.

Is Something Seriously Amiss?

Kaplan juxtaposes his criticism of Jensen's work with the words of praise that Bouchard, Detterman, and Scarr expressed in their contributions to a special issue of *Intelligence*, which was dedicated in Jensen's honor. The message seems to be that, although Kaplan is highly critical, prominent scholars like Bouchard, Detterman, and Scarr admire Jensen to the point of hero worship. Something, therefore, is "seriously amiss," as Kaplan puts it. I find this facile. To explain this, I shall speak for myself. Jensen proposed in his book *The g Factor* that *g*, general intelligence, is the dominant dimension along which black and whites differ. Differences on the latent variable *g* are hypothesized to account for most (but not all) of the differences in psychometric IQ test scores. To investigate this hypothesis, Jensen devised the method of correlated vectors, which in this context involves correlating factor loading of the observed subtest scores on *g* and the standardized black/white differences on the subtest. A high correlation is supposed to be indicative of the importance of *g* in black/white differences. Jensen called this method "efficient, practical and statistically rigorous" (Jensen 2000, p. 42). I believe that this method is neither efficient, nor statistically rigorous (Lubke, Dolan, and Kelderman 2001). I also believe that there are better ways to investigate this issue (Dolan 2000; Dolan and Hamaker 2001). So I like to think that I am critical. Still, should I agree to contribute to a Jensen Festschrift, I would have no trouble in expressing my appreciation for Jensen's efforts in trying to advance the understanding of the causes of black/white differences in IQ test scores. My point is that one can be appreciative of Jensen's contributions to the study of intelligence but still disagree with him on a variety of issues. I imagine that this is the case with Bouchard, Detterman, and Scarr (e.g., see Scarr 1981, p. 515 ff.). I see no problem here and doubt that in this respect something is "seriously amiss."

Conclusion

If Jack Kaplan's article results in his fellow statisticians addressing the problem of the black/white test score gap, I should consider his article a great success. Few researchers are willing to consider this problem (Jensen is certainly an exception). In the introduction to their recent book on the black–white IQ test score gap, Jencks and Phillips (1998) wrote that their book tries to update the reader's knowledge about many aspects of the problem. However, they stated "... almost every chapter raises as many questions as it answers. This is partly because psychologists, sociologists and educational researchers have devoted far less attention to the test score gap ... than its political and social consequences warranted. Most social scientists have chosen safer topics and hoped the problem would go away. It didn't. We can do better" (p. 47). I agree: We can do better.

Note: References from this comment appear following Jensen's comment.

[I thank Sofie van der Sluis and Ellen Hamaker for their critical but constructive remarks on an earlier version of this comment. The research of CVD has been made possible by a fellowship of the Royal Dutch Academy of Arts and Sciences.]

Comment:
Misleading Caricatures of Jensen's Statistics

Arthur R. Jensen

Ignoring the old-hat song-and-dance routine that accompanies Kaplan's specific complaints about my use of statistics, I will comment only on each of his four main points of contention. Because he frequently omits page references to his quotations and paraphrases from my writings, I will provide these for readers so they can see for themselves what I have actually said on these topics. Kaplan's critique aims to create the impression that I have all along been somewhere between naive and bizarre in what he purports to be my misuse of statistics.

Sibling regression

In my discussion of this topic (Jensen, 1973, pp. 117–119; 1998, pp. 467–472) I pointed out that a simple polygenic model of IQ differences that predicts the correlations between individuals of various degrees of genetic kinship shows the same sibling regression effect for both whites and blacks, and the regressions are linear throughout the full range of IQs in both racial groups. There is no purely environmental/cultural theory that makes any specific prediction of the degree of regression that would be found for any particular degree of kinship. A specific model that makes quantitative predictions of a phenomenon (in this case sibling regression) is more valuable scientifically than one that can only come up with ad hoc explanations after the fact. Of course regression could reflect either genetic, environmental, or error effects; the point I was making is that genetic theory can make empirically testable quantitative predictions, which purely environmental theories of IQ variance cannot do. Moreover, a polygenic model that shows essentially the same regression effects in representative white and black samples is consistent with my "default hypothesis" — namely, that the difference between the white and black distributions on psychometric *g* (i.e.,the common factor in all cognitive tests) has the same genetic and environmental causes, in about the same degree, as individual differences within each population [Jensen 1998, chap. 12; and note my discussion (p. 457) of the overworked "corn analogy" used by Kaplan]. There is no need to posit unique environmental factors for either group. Hence there is no scientific basis for treating members of various racial populations differently than one would treat comparable individuals within each population. Group differences in *g*, according to this theory, are just aggregated individual differences.

The point I was making is that genetic theory can make empirically testable quantitative predictions, which purely environmental theories of IQ variance cannot do.

I refer readers to my most comprehensive explication of Galton's so-called "law of filial regression," in which I state: "The phenomenon of regression is a valid argument to support a hypothesis of genetic inheritance of a given trait only if the amount of regression is closely consistent with an explicit genetic model that predicts the degree of regression that should be theoretically expected for any given degree of kinship. . . . There is nothing in the phenomenon of regression *per se* that proves either genetic or environmental causes or some combination of these" (Jensen 1984, p. 312). Clearly, Kaplan's complaint is vacuous and misleading in view of what I have actually written about kinship regressions toward their population mean.

Reproductive casualty

My literature review on this topic (Jensen 1973, pp. 341–348) shows that the risk of reproductive casualty (RC) is higher in the black population than in whites and Asians. The observed effects of RC are not an all-or-none disadvantage but appear as a continuous variable. These disadvantageous prenatal and perinatal effects are associated with the mother's age, lack of prenatal care, drug abuse, immunogenic incompatibilities between mother and fetus, prematurity, low birth weight, poor obstetrics, nutritional deficiencies, and the like. I have suggested that these conditions are among the various environmental factors that may adversely affect later mental development. However, empirical evidence, which I cited, also indicates that the frequency of neurologically detectable RC, although showing significant racial differences, was not great enough to account for more than a relatively small part of the average IQ differences between the major racial

populations in the United States. Since I first wrote on this subject, new evidence has appeared that I believe strengthens support for the hypothesis that a variety of prenatal and perinatal conditions are not a negligible casual factor in *g* variance, including subpopulation differences in *g* (Jensen 1998, pp. 500–509). This hypothesis, for which there is considerable empirical evidence (cited by Jensen 1997, 1998), is germane to my theory of microenvironmental effects on *g* (discussed in the last section).

Variance of IQ in the black and white populations

In 65% of 200 samples of white and black samples that took various IQ tests, whites had the larger variance (Jensen 1973, pp. 211–216). In the largest samples, which show this variance difference most clearly, both the black and white distributions are spread across the full range of test scores; there appears to be no scale artifact that constrains the variance of the score distributions in either sample. This phenotypic difference in variances (or standard deviations) is of interest psychometrically because it enters into the calculation of the "effect size" (ES) of race (e.g., black/white) on IQ scores, ES = the difference between the group means divided by the square root of the within-groups variance. In the disputed work (Jensen 1973) I pointed out that mental test scores, including IQ, are certainly not a ratio scale, and may not even be an interval scale throughout the full range of scores in the population. Without assuming approximate normality of the population distribution of intelligence, IQ scores cannot be treated as other than an ordinal, or rank-order, scale. After pointing out the effects of various transformations on the IQ distributions, I concluded, "Thus, the smaller IQ variance of Negroes than of whites could be merely an artifact of our scale for measuring intelligence" (p. 213). Then I go on to explain why this question itself (assuming an interval or ratio scale) is theoretically important for understanding the nature of the black–white IQ difference in relation to the heritability of IQ within each population, where a heritability analysis (based on various kinship correlations) estimates the proportion of the total variance of a given metric trait into its genetic and nongenetic components. It is suggested that such methods might help to discover the answer to this question by testing whether IQ behaves in kinship regression analyses as theoretically would be expected if the IQ measurements were truly an interval scale. Kaplan's claim that any transformation of scale would result in the same degree of regression toward the mean as predicted by the genetic model is either unclear or incorrect. The degree of regression predicted by the additive genetic model originally proposed by R. A. Fisher (1918), for example, would not predict the same absolute values found with an interval scale if it were subjected to a nonlinear transformation. Moreover, the Pearson *r* between relatives predicted by Fisher's genetic model, would necessarily be affected by any nonlinear transformations of the correlated variables measured on an interval or ratio scale, such as height and weight.)

This section in my 1973 book also discussed in theoretical terms the diverse possible causes — genetic, environmental, and psychometric — of a difference between populations in the variance of a trait. There is nothing at all in this discussion that is in the least contradicted by anything Kaplan has to say about it, and the generality of his one possible explanation for the difference in population variances is unsupported by large studies that fail to show any scale artifact that would restrict the IQ variance within either group, assuming an equal-interval scale throughout (see also Jensen 1980, pp. 98–100). The matter could be settled definitively, of course, if there existed an undisputed interval or ratio scale of mental measurement. Chronometric measures, such as choice reaction time and inspection time, are the only behavioral ratio scales that are significantly correlated with IQ (Jensen 1998, chap. 8).

Normality of the IQ distribution

My most comprehensive and detailed discussion of the normality of the IQ distribution begins with the following sentence: "Nothing of fundamental empirical or theoretical importance is revealed by the frequency distribution per se of the scores on any psychometric test composed of items. This is true regardless of whether we are dealing with raw scores, z scores, or any otherwise transformed scores" (Jensen 1998, pp. 100–103). I further explain (1) how test constructors can manipulate the moments (i.e., mean, *SD*, skewness, and kurtosis) of any distribution of test scores by item selection based on item difficulty and inter-item and item-total score correlations, or simply by normalizing the z distribution of test scores via their percentile ranks in the normal curve, (2) the purely statistical advantages of approximating a normal (Gaussian) distribution as closely as possible for population "norms," and (3) the several theoretical and empirical arguments for the plausibility of assuming that the latent trait (*g*) that IQ tests attempt to measure would approximate a normal distribution. I believe that it soon will be possible empirically to test the normality of *g* in the general population by means of mental chronometry, based on physical, or ratio, scales, such as reaction time in elementary cognitive tasks, and various physiological measurements that are known to be related to *g* (Jensen 1998, chap. 8).

The puzzle of nongenetic variance

Kaplan refers to my book chapter (Jensen 1997) with this title as my "most bizarre use of statistics" and claims that it merely serves as irrelevant "window dressing" for the conclusions that follow. He is wrong. The analysis I performed on the IQs of monozygotic (MZ) twins is neither " bizarre" nor "window dressing" but is the mainstay of the article. The background for the analysis, spelled out in the article, happens to be one of the most surprising and well-established findings of behavior-genetic research in recent years — namely, that by late adolescence the between-families (BF) component of the nongenetic (or environmental) variance virtually disappears, leaving only the within-families (WF) component of nongenetic variance. The BF variance is called the *shared* envi-

ronment because it is attributable to variables on which families differ (social class, ethnicity, culture, income, diet, etc.) but is shared by individuals reared together in the same family. It is the shared environment that increases similarity between siblings (or any other children) who are reared together. The WF source of variance is *unshared* by children reared together; it makes them less similar to one another. Because the BF variance dwindles to near 0 by late adolescence, the heritability of IQ is around 70%, and measurement error is around 5% of the total variance, we are left with at least 25% of the IQ variance consisting of unshared (WF) variance. The as-yet-unsolved puzzle is the causal nature of this substantial source of nongenetic WF variance. I decided that a good beginning point for hypothesizing on this puzzle was to suppose that there is an indefinitely large number of advantageous (+) and disadvantageous (–) microenvironmental events, each of which has some small effect on mental growth. The hypothesis holds that these random microenvironmental events and their effects on mental development are effectively random with respect to most individuals reared in the same family; the random effects are additive and normally distributed, and some individuals have good luck (+) and others bad luck (-) in the cumulative effect of these microenvironmental events, just as there are winners and losers leaving a casino.

Metric data on MZ twins reared together afford the most direct measure of the WF environmental variance. Because MZ twins have identical genotypes, any difference between a pair of MZ twins reared together that is not attributable to measurement error (which can be separately taken into account) is by definition a WF environmental difference. Therefore, I tested the microenvironmental model by analyzing the distribution of MZ twin differences on the Stanford–Binet IQ. If the model were correct, these IQ differences should closely approximate the chi distribution for 1 *df*, which is the distribution of absolute differences between all possible pairs of values in a normal distribution.

This analysis revealed at least two important findings that could not have been discovered just by eyeballing a large collection of MZ twin data: (1) The chi distribution model perfectly fits only the 80% of twin pairs who show the smaller IQ differences (<9 points), but 20% of the twin pairs departed from the chi distribution — that is, they showed larger IQ differences than could be accounted for by the summation of entirely random microenvironmental effects — and (2) the WF environmental factors in MZ twins, on average, have larger negative effects than positive effects on IQ; that is, environmental effects are not symmetrically + or - and, at least for MZ twins, the WF environment is more often harmful than beneficial. After presenting the evidence for these quantitative effects in considerable detail, I cited evidence in the literature on factors in the biological environment that seemed most likely to account for these results and suggested them as hypotheses worth testing as a likely explanation for my findings. Isn't this how science works? I would urge readers to consult this article itself if they care to know what I have done and why, rather than relying on Kaplan's confusing caricature of it, in which his sidebar is meaningless absent my article's whole theoretical context.

References and Further Reading from the Comments

Brody, N. (1991), *Intelligence* (2nd ed.), San Diego: Academic Press.

Dolan, C. V. (2000), "Investigating Spearman's Hypothesis by Means of Multi-group Confirmatory Factor Analysis," *Multivariate Behavioral Research*, 35, 21–50.

Dolan, C. V., and Hamaker, E. L. (2001), "Investigating Black–White Differences in Psychometric IQ: Multi-group Confirmatory Factor Analyses of the WISC-R and K-ABC and a Critique of the Method of Correlated Vectors," in *Advances in Psychological Research*, (vol. VI), ed. F. Columbus, Commack, NY: Nova Science Publishers.

Faraone, S. V., Tsuang, M. T., and Tsuang, D. W. (1999), *Genetics of Mental Disorders. A Guide for Students, Clinicians, and Researchers*, New York: Guildford Press.

Fisher, R. A. (1918), "The Correlation Between Relatives on the Supposition of Mendelian Inheritance," *Transactions of the Royal Society of Edinburgh*, 52, 399–433.

Flynn, J. R. (1980), *Race, IQ and Jensen*, London: Routledge & Kegan Paul.

Jencks, C., and Phillips, M. (1998), "The Black–White Test Score Gap: An Introduction," in *The Black–White Test Score Gap*, eds. C. Jencks and M. Phillips, Washington DC: Brookings Institute Press, pp. 1–54.

Jensen, A. R. (1973), *Educability and Group Differences*, London: Methuen.

———. (1980), *Bias in Mental Testing*, New York: The Free Press.

———. (1984), "Law of Filial Regression," in *Encyclopedia of Psychology* (Vol. 2), ed. R. J. Corsini, New York: Wiley, pp. 280–281.

———. (1997), "The Puzzle of Nongenetic Variance," in *Intelligence, Heredity, and Environment*, eds. R. J. Sternberg and E. L. Grigorenko, Cambridge: Cambridge University Press.

———. (1998), *The g Factor. The Science of Mental Ability*, Westport: Praeger.

———. (2000), *The g Factor: Psychometrics and Biology*, In *The Nature of Intelligence*, eds. G. R. Bock, J. A. Goode, and K. Webb, Chichester, U.K.: Wiley, pp. 37–47.

Lubke, G. H., Dolan, C. V., and Kelderman, H. (2001), "Investigating Group Differences on Cognitive Tests Using Spearman's Hypothesis: An Evaluation of Jensen's Method," *Multivariate Behavioral Research*, 36, 299–324.

Scarr, S. (1981), *Race, Social Class and Individual Differences in IQ*, Hillsdale, NJ: Lawrence Erlbaum.

Reply

Jack Kaplan

Dolan

First of all, I want to make clear that all I mean by the term *misuses* of statistics is *incorrect uses* of statistics. I am not suggesting (as Sandra Scarr does concerning Stephen Gould and Leon Kamin) that Jensen is being intentionally deceptive, only that he is mistaken.

Conor Dolan doesn't find my criticism helpful, but it isn't clear that he actually *disagrees* with any of it.

Concerning the point I made about whether blacks have less variability in intelligence than whites, he writes, "Jensen offers three hypotheses to account for black/white differences in variance and Kaplan offers one of his own ... Even if Kaplan is completely right ... what misuse of statistics has Jensen actually committed?"

Concerning the point about the degree to which neurological accidents prior to or during birth affect intelligence, he writes, "Kaplan finds the assumption of equal variance artificial. This may be so, but in a model like the liability-threshold model one has to introduce some constraints ... The resulting model may be artificial, but why exactly is Jensen's application of it a misuse of statistics?"

Concerning the point about nongenetic variance, he writes, "Again Kaplan may well be right, but he makes little effort to explain why this is so. As such, I regard these criticisms to be gratuitous and unhelpful."

None of these comments implies that I'm actually wrong about anything.

Dolan writes about the example I use to illustrate the point about black versus white variability in intelligence, "Kaplan presents a result in support of his hypothesis, but the absence of any details concerning his calculations makes his results hard to judge." If Dolan (or anyone else) wants further details, I will be happy to supply them.

Dolan writes, "If Jack Kaplan's article results in his fellow statisticians addressing the problem of the black/white test score gap, I should consider his article a great success."

I agree — sort of. Both Dolan and I would like to see a dialogue develop between statisticians and hereditarians. But we envision different outcomes from that dialogue.

I envision statisticians telling hereditarians that their methods are deeply flawed and that many of the conclusions they've drawn are simply not true.

Dolan presumably envisions statisticians assisting hereditarians in refining their methods and extending their results. I envision statisticians telling hereditarians that their methods are deeply flawed and that many of the conclusions they've drawn are simply not true. Among their wrong conclusions, in my opinion, are their two most basic ones — that IQ scores are a measure of general intelligence (using any reasonable definition of general intelligence) and that the heritability of IQ scores is greater than 40%.

Jensen

Most of Jensen's comments do not address the specific points I made in my article. Instead he repeatedly digresses into irrelevant discussions of other issues or refers to articles or books he's written other than the one I was commenting on.

Sibling Regression

My article addresses statements Jensen made in his 1973 book, *Educability and Group Differences*. But his response relies mainly on a quote from an article he wrote in 1984 and a summary of arguments he made in a book in 1998.

My article accurately describes (and gives page references for) what Jensen wrote in 1973. What he wrote then — that regression to the mean provides evidence that the black/white difference in average IQ is at least partly genetic — is clearly a misinterpretation of statistics. Regression to the mean applies just as well to populations that differ for environmental reasons as it does to populations that differ for genetic reasons.

Reproductive Casualty

Jensen's comments are not responsive to the point I made in my article.

I don't disagree with the basic concept that "the observed effects of [reproductive casualty] are not an all-or-none disadvantage" and that the higher fetal death rate for blacks as compared to whites suggests that blacks may also suffer more often from less severe problems that don't cause detectable harm but may impair later mental development. What I disagree with is Jensen's jump from population differences in fetal death rates to average population dif-

ferences on a continuous "organismic viability" variable. That jump requires assuming that fetal death corresponds to a value above some threshold on a variable that is not only continuous but is normally distributed with equal variances in both populations. That is quite an assumption!

It's true that I paraphrased Jensen on this point rather than quoting him verbatim. The verbatim quote (from page 346 of *Educability and Group Differences*) is

> In 1965, fetal deaths (for gestation periods of twenty weeks or more) nationwide had almost twice as high a rate among Negroes as among whites (25.8 *v.* 13.3 per 1000 live births) ... Assuming fetal death to be a threshold effect on a normally distributed variable, the Negro and white populations can be said to show a mean difference of 0.46σ on this variable. This is a large difference by any standard. But even if this variable (organismic viability, freedom from impairment, or whatever it is) were *perfectly* correlated with intelligence, it could account for less than half of the Negro-white difference [italics in original].

Readers can judge for themselves whether my paraphrase of Jensen's comments is in any way misleading.

Variability

Jensen writes, "The generality of [Kaplan's] one possible explanation for the difference in population variances is unsupported by large studies that fail to show any scale artifact that would restrict the IQ variance within either group, *assuming an equal-interval scale throughout*" [italics added].

If one assumes an interval scale, then my argument makes no sense whatsoever. But there is, in my opinion, no reasonable statistical argument that IQ scores constitute an interval scale.

Weight is an example of an interval scale. The difference in weight between two people who weigh 170 and 175 pounds is the same as the difference between two people who weigh 200 and 205 pounds. But is the difference in "intelligence" (or whatever it is that IQ measures) between two people with IQ scores of 100 and 105 the same as the difference in "intelligence" between two people with IQ scores of 130 and 135? I think not.

Normality

My article addresses statements Jensen made about normality in 1969 in his famous *Harvard Educational Review* article (as reprinted in *Genetics and Education*). But Jensen's response refers entirely to statements he made on the subject in his 1998 book, *The g Factor*.

What Jensen says about normality in the *Harvard Educational Review* article is nonsense. I'm not familiar with what he says about normality in *The g Factor*.

Nongenetic Variance

I don't think bizarre is too strong a word to use to describe Jensen's use of statistics in "The Puzzle of Nongenetic Variance." Readers who doubt this should do three things:

1. Read Jensen's description of his complex and confusing statistical analysis of the IQ's of 180 pairs of identical twins — all 14 pages of it (pp. 61–74).
2. Read Jensen's conclusions (pp. 80–82).
3. Ask themselves, What is the relationship — if any — between the analysis and the conclusions?

The underlying question Jensen is trying to answer is whether nongenetic variation of IQ scores is explained by social environmental factors or by biological environmental factors. His statistical analysis is basically a clumsy attempt to characterize the way in which his dataset departs from a normal distribution.

But how can departures from normality possibly distinguish between social and biological factors? What type of departure from normality would one expect from biological factors? What type of departure would one expect from social factors?

My suspicion is that Jensen, who has spent more than three decades arguing (unpersuasively) that variation in intelligence is explained primarily by genetics, is pre-disposed to favor a biological rather than a social interpretation for the remaining nongenetic variation. But the distribution of his data can be explained just as readily by either one. His assertion that biological factors are the probable source is nothing more than speculative theorizing. The statistical analysis is just window dressing.

I join with Jensen in urging readers to consult the full article. But I doubt that the article's "whole theoretical context" will make Jensen's analysis seem any more reasonable.