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.