

drawal after surgery ($P = .025$), while the responses of the sham-operated were not significantly changed. After the operation the scores of the deganglionated group were significantly lower than the scores of the sham-operated group ($P = .013$). These results suggest that the siphon withdrawal component of the defensive reflex is mediated primarily by the central nervous system. This does not mean that all siphon withdrawal is centrally mediated. In the absence of the abdominal ganglion, the peripheral plexus is still able to move the siphon when it is directly stimulated. However, deganglionated animals are unable to perform an integrated withdrawal reflex that removes the siphon from view.

Since the siphon can move in the deganglionated animal, we sought to determine whether the sensitization of the reflex might be due to the facilitation of the relatively small peripherally mediated siphon movements. Twenty-four animals were matched as in previous experiments and all animals were given sensitization training (four shocks per day for 4 days) beginning 6 days after the habituation session. On the day after the last shock, half of the animals were deganglionated and the other half were sham-operated. All animals were given a 1-day postoperative retention test using a blind procedure (Table 1b). The deganglionated animals failed to show any effect of the previous sensitization training. In fact, these animals showed significantly less siphon withdrawal than they did previously ($P = .025$) (Table 1b). By contrast, the sham-operated animals showed significant sensitization ($P = .005$) despite the intervening anesthetic and surgical procedure. Two days after the last shock, the median value of the sham-operated group was about 440 percent of the presentization value. These results suggest that long-term sensitization of siphon withdrawal is centrally mediated.

The demonstration of long-term sensitization of defensive siphon withdrawal in *Aplysia* is a further extension of the short- and long-term forms of behavioral modifications that this simple reflex can undergo. Whereas habituation is perhaps the most elementary behavioral modification, sensitization is more complex because it involves changes in one reflex pathway as a result of activity in another one. As Groves and Thompson (14) have pointed out, sensitization represents an aspect of arousal, or a "state

variable," characteristic of vertebrates. By developing a simple system for the analysis of sensitization, it may become possible to gain some understanding of the cellular mechanisms of arousal and elementary forms of learning.

HAROLD M. PINSKER*

WAYNE A. HENING

THOMAS J. CAREW

ERIC R. KANDEL

Departments of Physiology and Psychiatry, New York University Medical School, and Department of Neurobiology and Behavior, Public Health Research Institute of New York, New York 10016

References and Notes

1. I. Kupfermann and E. R. Kandel, *Science* **164**, 849 (1969).
2. H. Pinsker, I. Kupfermann, V. Castellucci, E. R. Kandel, *ibid.* **167**, 1740 (1970); I. Kupfermann, V. Castellucci, H. Pinsker, E. R. Kandel, *ibid.*, p. 1743; V. Castellucci, H. Pinsker, I. Kupfermann, E. R. Kandel, *ibid.*, p. 1745.
3. T. Carew, V. Castellucci, E. R. Kandel, *Int. J. Neurosci.* **2**, 79 (1971).
4. T. Carew, H. Pinsker, E. R. Kandel, *Science* **175**, 451 (1972).
5. For discussions of sensitization (or non-specific alpha conditioning) see G. A. Kimble, *Hilgard and Marquis' Conditioning and Learning* (Appleton-Century-Crofts, New York, 1961), pp. 64-65; J. W. Kling and J. G. Stevenson, in *Short-term Changes in Neural Activity and Behaviour*, G. Horn and R. Hinde, Eds. (Cambridge Univ. Press, London, 1970), pp. 42-46; E. R. Kandel and W. A. Spencer, *Physiol. Rev.* **48**, 65 (1968) (see pp. 97-99). For an example of the use of the term sensitization to describe the facilitation of a response to a single stimulus when that stimulus is repeated, see Groves and Thompson (14).
6. M. J. Wells, in *Symposium on Neurobiology of Invertebrates*, J. Salanki, Ed. (Plenum, New York, 1967), p. 391; M. J. Wells and J. Wells, *Anim. Behav.* **19**, 305 (1971).
7. Each shock (6 ma a-c between bipolar electrodes in seawater) lasted about 1 second and successive shocks within a day were separated by a minimum of 2 hours. The anterior mantle region is innervated by the branchial nerve, whereas the siphon region is innervated by the siphon nerve. It has been shown that reflex habituation produced by repetitive tactile stimulation of the siphon does not generalize to the anterior mantle region (3). This suggests that the electrical shock reached the central nervous system via a different pathway than the tactile stimulus.
8. Each animal contributed a single score, the sum of the duration of siphon withdrawal for all ten trials. Nonparametric statistics were used to compare scores either within a group (Wilcoxon matched-pairs signed-ranks test), or between groups (Mann-Whitney U test).
9. Using a restrained preparation, we attempted to relate sensitization of siphon withdrawal to sensitization of gill withdrawal (4). We found no significant difference in gill withdrawal between experimental and control animals. It may be that restraining the animals caused sufficient sensitization to obscure group differences.
10. E. R. Kandel, in *The Neurosciences*, F. O. Schmidt, Ed. (MIT Press, Cambridge, Mass., in press).
11. B. Peretz, *Science* **169**, 379 (1970); I. Kupfermann, H. Pinsker, V. Castellucci, E. R. Kandel, *ibid.* **174**, 1252 (1971).
12. I. Kupfermann and M. Kesselman, cited by E. Kandel and W. Spencer [*Physiol. Rev.* **48**, 65 (1968)]; K. Lukowiak and J. Jacklett, *Science* **178**, 1306 (1972).
13. Animals were anesthetized with injections of isotonic $MgCl_2$ (25 percent of body weight), and the ganglion was removed through a longitudinal incision in the foot. Data are presented only for animals surviving 5 days after surgery.
14. P. M. Groves and R. F. Thompson, *Psychol. Rev.* **77**, 419 (1970).
15. We thank A. Perlman and K. Hiltner for assistance and I. Kupfermann for helpful discussions. Supported by NIMH grants 19795-02 and NS 09361-03, a NIH career scientist award MH 18558-06 to E.R.K., and a special fellowship from NIMH MH 5355 1-01 to T.J.C.

* Present address: Marine Biomedical Institute, University of Texas Medical Branch, Galveston 77550.

5 May 1973; revised 16 July 1973

Heritability of IQ by Social Class: Evidence Inconclusive

In her provocative article on race and intelligence (1), Scarr-Salapatek may give the mistaken impression that "two major, competing hypotheses," or some combination of them, are the only plausible explanations of the relation among social class, race, and IQ (intelligence quotient). Either (i) racial differences in intelligence result from environmental disadvantage that simultaneously retards mental development and prevents full expression of genetic differences or (ii) racial differences reflect genetic differences that contribute a similar proportion of variance in all social classes. Scarr-Salapatek attempts to exclude the second hypothesis and thereby, perhaps, to strengthen the environmental explanation of race differences.

It is sometimes supposed that an optimum environment will result in maximum expression of genetic factors,

but the fallacy of this view becomes apparent when one asks, "Optimum for what?" or "Expression of which genetic factors?" Different environments elicit the expression of different sets of genes. Scarr-Salapatek's restriction of explanations to two models tends, albeit unintentionally, to affirm the above fallacious view and to perpetuate the widespread idea that genetic factors set limits on an individual's potential, while the environment determines how closely he will approach these limits. Neither heredity nor environment sets absolute limits on quantitative traits.

If we discard simplistic formulations, many more than two models have to be considered in any attempt to understand racial and class differences in intelligence. A complete and testable model should predict at least three things: the effect of socioeconomic environment on intelligence test scores,

the relative magnitude of the phenotypic (total) variance in different classes, and class differences in the proportion of that variance which is genetic (heritability in the broad sense). Scarr-Salapatek's two models make very simple predictions: either favorable environments increase the mean, the variance, and the heritability of intelligence or environments do not significantly affect intelligence at all.

Another hypothesis that might be as easy to test is that environmental advantages increase the mean and variance of intelligence, while reducing its heritability. Different favorable conditions might provide people with different mental skills almost independently of their genetic endowment, and the genetic endowment would be expressed most distinctly in basic or deprived cultures. However, if disadvantaged monozygotic twins are no more similar in intelligence than Scarr-Salapatek has estimated, we must agree with her that "genetic factors cannot be seen as strong determinants of aptitude scores in the disadvantaged groups" (1, p. 1292). One might then modify this hypothesis or look at a few others.

Several models could be proposed in which lower-class environments, more than upper-class environments, contain diverse stimuli that produce deviations from an individual's "most probable" IQ. The diversity and magnitude of stresses in some economically deprived groups are formidable, and, unlike chronic deprivation, stresses may have positive behavioral consequences (2). If some stresses in a lower-class environment produce positive, and others, negative, deviations in intelligence, this could account for its low heritability in low socioeconomic classes. Particular models would further specify whether a low mean IQ in these social classes reflected cultural impoverishment or economic selection, and what effect either phenomenon might have on variance. One such particular model would invoke the effects of stress in a lower-class environment to modify the hypothesis, proposed in the preceding paragraph, that environmental advantages tend to lower the heritability of intelligence.

These hypotheses are all more complicated than the two discussed by Scarr-Salapatek, but some of them might be closer to reality.

Failure to list other alternatives would not detract from an effective exclusion of one hypothesis. Scarr-Salapatek's main purpose. When one

examines her calculations, one is forced to doubt whether she did, in fact, demonstrate lower heritability in disadvantaged groups, and this doubt can be made more explicit than was done by Dawes (3). Estimates obtained from differences between statistics may have relatively large errors because they combine the two sampling errors of the statistics from which they were calculated. Scarr-Salapatek has compounded her sampling errors by taking differences between differences. First, to estimate the intraclass correlation coefficient for the monozygotic twins, Scarr-Salapatek subtracted the coefficient of the opposite-sex pairs from that of the same-sex pairs, after converting to Fisher z scores and weighting them according to the estimated proportions of monozygotic and dizygotic same-sex twins (1, p. 1287). The same-sex pairs were, by her estimates, approximately equally divided between monozygotic and dizygotic pairs, and the error of the transformed monozygotic coefficient in her formula is therefore at least twice the error of the transformed same-sex coefficient. The formula for heritability again subtracts the correlation coefficient of opposite-sex twins, this time from the indirectly obtained coefficient of monozygotic twins (1, p. 1290). When at last she compares heritabilities, the observed differences may be explained by chance variation.

To appreciate the degree of uncertainty surrounding Scarr-Salapatek's estimates, consider the 95 percent confidence interval for her estimates. The limits of this interval can be calculated for her intraclass correlation coefficients by adding ± 1.96 times the square root of the sampling variances of the corresponding Fisher z scores. The conventional large-sample variances may be used for the coefficients of same-sex and opposite-sex twins. The estimated coefficient for monozygotic twins requires a different calculation, its sampling variance (var) being a weighted sum of the variances of the two coefficients from which it was calculated:

$$var(z_{r_{mz}}) = \left(\frac{1}{SS_{mz}} \right)^2 var(z_{r_{iss}}) + \left(\frac{SS_{dz}}{SS_{mz}} \right)^2 var(z_{r_{ios}})$$

where $z_{r_{mz}}$ is the transformed correlation coefficient for monozygotic twins, $z_{r_{iss}}$ that for same-sex twins, and $z_{r_{ios}}$ that for opposite-sex twins; SS_{mz} is the proportion of monozygotic twins among

same-sex pairs, and SS_{dz} is the proportion of dizygotic twins among same-sex pairs. This assumes the validity of her method of estimating the monozygotic intrapair correlation.

The coefficient for the middle and above median group of dizygotic black twins with respect to verbal aptitude scores, calculated by Scarr-Salapatek as .460, has 95 percent confidence limits at .241 and .635. For monozygotic twins in the same group, on the same tests, with a correlation coefficient estimated by her as .753, the possible range is .492 to .890. The wide overlap with the range for dizygotic twins would be even wider if one took into account the negative correlation between the coefficient of dizygotic twins and the derived coefficient of monozygotic twins. Yet this comparison was one of the most reliable (4). It is therefore not surprising that several of the groups in her study appeared to have negative heritabilities. Eaves and Jinks have presented a detailed mathematical criticism of this point (5).

Finally, Scarr-Salapatek's attempt to estimate the intraclass correlation coefficient of monozygotic twins by an extension of Weinberg's difference method, attributed to Burt (6), is of considerable methodological interest. Before other workers make the same attempt, the pitfalls should be noted, even though they do not affect Scarr-Salapatek's conclusions.

Burt's approach assumed that partitioning the z -transformation of the same-sex intraclass correlation coefficient was equivalent to partitioning the components of variance represented in that coefficient. This is only approximately correct, and it seems more appropriate to partition the mean squares, also available. The formula used by Scarr-Salapatek (1, p. 1287) can be applied separately to between-pair and within-pair mean squares instead of to converted correlation coefficients. The adjusted mean squares are then used in the usual formula for the intraclass correlation coefficient. This procedure yields corrections ranging from $-.028$ to $+.050$ in the coefficients estimated for monozygotic twins, but these corrections are smaller than the presumed sampling errors. Sampling variances of the improved estimates can be obtained only by approximation (7), but are probably rather similar to those we calculated for Scarr-Salapatek's estimates.

Both methods of estimating intraclass correlation coefficients of monozygotic

twins require three assumptions: (i) the usual Weinberg assumption, that same-sex dizygotic twins occur in the same number as opposite-sex twins or in a proportion that can be estimated from the sex ratio; (ii) that monozygotic and dizygotic twins have the same mean value (intelligence in this instance); and (iii) that variance within same-sex dizygotic pairs is equal to that within opposite-sex dizygotic pairs in all social classes. We are most interested in the third assumption. Actually, the variance among same-sex dizygotic pairs is almost always smaller than that among opposite-sex pairs, and subtracting the variance of opposite-sex twins from that of all same-sex pairs will remove too much of the variance. The remaining variance attributed to monozygotic twins will be an underestimate, and the intraclass correlation coefficient estimated by either method will be an overestimate. This exaggerates the heritability of the trait in question. It is an error in the conservative direction for Scarr-Salapatek's purpose of demonstrating low heritability of intelligence in the disadvantaged class.

In summary, Scarr-Salapatek has presented a plausible model and a helpful approach to a difficult problem, but her data are insufficient. The approach might permit the exclusion of not one, but several significant hypotheses if the blood types of such a series of twins were determined or, given a much larger series, even if they were not.

GORDON ALLEN

*Mental Health Intramural Research
Program, National Institute of Mental
Health, Bethesda, Maryland 20014*

KAREN D. PETTIGREW

*Biometry Branch,
National Institute of Mental Health*

References

1. S. Scarr-Salapatek, *Science* **174**, 1285 (1971).
 2. V. H. Denenberg and A. E. Whimbey, *Develop. Psychobiol.* **1**, 55 (1968).
 3. R. M. Dawes, *Science* **178**, 229 (1972).
 4. S. Scarr-Salapatek, *ibid.*, p. 235.
 5. L. J. Eaves and J. L. Jinks, *Nature* **240**, 84 (1972).
 6. C. Burt, *Brit. J. Psychol.* **57**, 137 (1966).
 7. O. Kempthorne, *An Introduction to Genetic Statistics* (Wiley, New York, 1957), p. 246.
- 29 November 1972

Considerable heat, 1 or 2 million words of discussion, and several pounds of printed paper have been generated during the past few years in controversy over genetic versus environmental interpretations of racial and social class differences in mean IQ scores. No satisfactory resolution has been possible be-

cause of the inadequacies of available data. The latest major article, Scarr-Salapatek (1), furnishes a fresh set of data collected in a study of school-aged black and white twins grouped by social class. Studies of twins are frequently used to derive estimates of a trait's heritability (that is, the ratio of the genetic variance to the phenotypic variance) within a given population. That approach has yielded relatively consistent estimates of the heritability of IQ within white populations in the course of a number of investigations, but such estimates have been lacking for blacks and members of lower social classes. Scarr-Salapatek proposes that by filling the gap and by comparing the heritabilities estimated for each race and for the different social classes, competing predictions of simple nature and nurture hypotheses about the origins of between-group IQ differences can be put to the test. Thus, at first glance, this new study seems to promise the kinds of data that are needed to settle the issue at last.

Indeed, some readers will be tempted to believe that Scarr-Salapatek's report contains the definitive answer, especially because the sentiments expressed in the concluding paragraphs are so clearly fair-minded. Scarr-Salapatek states (and who would disagree with her?) that "Group differences in IQ scores and phenotypic variability that exist because of environmental deprivation can and should be ameliorated" (1, p. 1294). We wish we could as readily agree that her data convincingly establish that the between-group differences in IQ observed in her study do exist largely because of environmental deprivation. Nevertheless, we are compelled to question whether such a conclusion—or, in fact, any conclusion—can be drawn from these data, just as we seriously doubt that conclusions can be based upon the lines of evidence that other authors (2) have assembled in attempting to demonstrate the existence of group differences because of genetic factors.

Several technical difficulties in Scarr-Salapatek's material will be obvious to most readers. They include: the loss of one-third of her starting sample, with the reasons for the losses apparently being differently distributed in the two racial groups (3); the need to estimate social class from census tract data rather than from known characteristics of the individual twins' families; and the extreme nonnormality of the test

score distributions mentioned by the author. All combine to introduce into the analyses an unknown, but possibly substantial, amount of "noise." Confusion is added, too, by a number of discrepancies in the tables (4). Yet, we are troubled chiefly by another problem, one that is less likely to be recognized by many readers but that is more fundamental than the above shortcomings: All of Scarr-Salapatek's main analyses are based on the twin method, which, in turn, depends upon comparisons between monozygotic and dizygotic pairs and, hence, upon accurate zygosity determinations. But no tests of zygosity were made on this sample (5); not a single same-sex pair can be classified as to zygosity.

The author has sought to cope with this important omission by calling upon Weinberg's differential rule (6), which postulates that same-sex and opposite-sex pairs occur in about equal frequency among dizygotic twins. Presumably, therefore, one has only to subtract twice the number of opposite-sex pairs from the total sample size to find the number of monozygotic pairs in the sample. This is the procedure that Scarr-Salapatek follows. Reliance on the Weinberg rule, however, has been called into question by several authors (7). A recent review (8) of eight studies of twins shows that the proportion of same-sex dizygotic pairs predicted by the differential rule may be considerably less than the proportion actually found when blood-grouping is done. If the proportion of dizygotic pairs is underestimated, then, of course, monozygotic pairs are proportionately overestimated. In that case, analyses like Scarr-Salapatek's will almost certainly undervalue the genetic contribution to phenotypic variance.

We see three specific reasons to believe that the Weinberg rule fits Scarr-Salapatek's sample poorly.

1) The correlations reported for the test scores of opposite-sex twins are frequently—in three out of nine comparisons within the black group and four out of nine comparisons within the white group—higher than the correlations for same-sex pairs as a whole and than the estimated correlations for monozygotic twins (1, tables 5 to 8, pp. 1290–1291). In a letter on Scarr-Salapatek's article, Dawes (9) points out that such a finding is not to be expected on genetic grounds. We would add that it is not to be expected on environmental grounds either.

2) As Scarr-Salapatek correctly notes, "The heritability of intelligence in white, middle-class populations . . . has been repeatedly estimated to account for 60 to 80 percent of the total variance in general intelligence scores . . ." (1, p. 1285). For her own group of white, middle-class children, however, the heritabilities of the test scores (1, tables 9 to 12, pp. 1292-1293) range between only 4 and 44 percent. Problems with the Weinberg differential rule mentioned above could account for the failure to obtain figures in line with most other studies. In the absence of previous data on black and disadvantaged subjects, it is of course not possible to judge whether Scarr-Salapatek's data minimize the genetic contribution to differences in IQ as drastically within those groups as they appear to do for middle-class white subjects.

3) Finally, the sex distributions presented in table 13 (1, p. 1293) make it evident that, for the black group at least, the Weinberg rule is inappropriate. The rule rests on the assumption that the distribution of sexes is nearly equal among twins in any population. That assumption is certainly not met in the sample of black twins, which contains 194 female-female and 139 male-male pairs. (Carrying through on the Weinberg formula for this sample with 169 opposite-sex pairs, one would obtain estimates of 109 female and only 55 male monozygotic pairs. Or, if the basic assumption of the Weinberg rule is waived, the deviation in sex ratio of 41 percent males to 59 percent females found among the same-sex black pairs may be assigned equally to monozygotic and dizygotic pairs, yielding 67 male to 97 female monozygotic pairs and 69 male to 100 female dizygotic pairs—a biased sample at best. And, then, how much further distortion occurs when this sample is subdivided by estimated ratings of social class?)

For the foregoing reasons, we believe that the zygoty estimates in Scarr-Salapatek's study cannot be accepted with any degree of certainty. It is difficult to see how the analyses, which hinge upon such estimates, can be considered meaningful.

Perhaps it is just as well that the data are not to be taken too seriously, for, otherwise, a true puzzle might confront us all. According to her own formulation, Scarr-Salapatek would have to demonstrate that heritabilities of the

test scores are higher in whites than in blacks, and higher in middle than in lower social classes, in order to support the theoretical model, which attributes group differences to the depressing effects of environmental disadvantages rather than to genetic differences. The author holds that her data on social class are consonant with the environmental disadvantage hypothesis. Actually, as noted by Dawes' letter (9) and in the author's reply (10), heritability estimates are missing for lower-class whites on both the verbal and nonverbal aptitude tests and for lower-class blacks on the nonverbal tests owing to the methodological problems detailed above. Hence, we contend that hypotheses about social class differences in IQ are untestable with Scarr-Salapatek's data.

For blacks and whites within social classes, however, some comparisons are possible, and there is where the puzzle would come in. Of the four possible comparisons [using either h_r^2 or h_n^2 in table 9 (1, p. 1292)], three show the estimated heritability ratios for blacks to exceed, by at least 50 percent, those for the white group. The environmental disadvantage explanation of black and white differences in IQ would predict the reverse. Fortunately, the methodological difficulties that we have noted make it unnecessary to worry over the seeming contradiction between the reported results and expectations of the environmental hypothesis.

Emotionally and intellectually, we concur in the belief that the environmental hypothesis is the correct explanation for observed differences in IQ between groups, at least between blacks and whites. Our point, however, is that Scarr-Salapatek's data do not provide the longed-for evidence in support of that hypothesis.

L. ERLNMEYER-KIMLING

Department of Medical Genetics,
New York State Psychiatric Institute,
New York 10032

SAMUEL E. STERN

Department of Sociology, Georgia
State University, Atlanta 30303

References and Notes

1. S. Scarr-Salapatek, *Science* 174, 1285 (1971).
2. A. R. Jensen, *Harv. Educ. Rev.* 39, 1 (1969); H. J. Eysenck, *The IQ Argument: Race, Intelligence and Education* (Library Press, New York, 1971).
3. According to Scarr-Salapatek's note 26 (1, p. 1295), exclusion of twins enrolled in special classes for the "retarded" accounted for 23 percent of the lost black pairs and 10 percent of the lost white pairs.
4. For example, there are unexplained small fluctuations in N 's. From table 1 (1, p. 1288),

it appears that 506 black pairs and 282 white pairs took the aptitude tests on which the main statistics are based. In tables 5 through 8 (1, pp. 1290-1291), N 's shown for black pairs range between 501 and 504, and N 's shown for white pairs range from 274 to 282. No explanation is given for these differences. It is especially difficult to understand why the number of pairs listed for the total aptitude scores consistently exceeds the number shown for either the verbal or the nonverbal sections of the tests alone.

5. We fully appreciate the tactical difficulties that Scarr-Salapatek mentions in connection with differentiating monozygotic and dizygotic twins in large samples. Nevertheless, we believe that this must be done. Many large twin studies were carried out before blood-grouping was generally available. Data from those studies are often criticized for that reason, but at least some effort was made by the investigators to classify individual pairs by one of the phenotypic similarity methods.
6. W. Weinberg, *Arch. Gesamte Physiol. Menschen Tiere (Pfluegers)* 88, 346 (1902).
7. I. I. Gottesman and J. Shields, in *Progress in Experimental Personality Research* (Academic Press, New York, 1966), vol. 3, pp. 1-84; P. P. S. Nylander and G. Corney, *Ann. Hum. Genet.* 33, 31 (1969); K. O. Renkonen, *ibid.* 30, 277 (1967).
8. W. H. James, *Nature* 232, 277 (1971).
9. R. M. Dawes, *Science* 178, 229 (1972).
10. S. Scarr-Salapatek, *ibid.*, p. 235.
11. We thank Arthur Falek, Leonard L. Heston, and Jacques Rutschmann for their helpful comments and suggestions.

27 November 1972

My first attempt to explore possible differences in the sources of variance in aptitude scores among children from several populations (1) has been roundly, and often correctly, criticized because it failed to settle all of the methodological, statistical, ethical, and social issues arising from the observation of individual and group differences in intelligence.

Further, the discussion section seems to have enraged some hereditarians by its emphasis on environmental differences, even though the sentiments expressed have been labeled as "fair-minded." First, let me discuss briefly the difficulties of model testing in human populations and, second, deal with specific criticisms raised by the two technical comments.

Model Testing

The posing and testing of competing models to explain the human data on intellectual variation is an extremely difficult task, made nearly impossible by the requirement that each study meet all possible criticisms. Many potential investigators, especially the biometricians (2, 3), can specify ideal designs for genetic research on behavior. Their specifications for ideal studies are so extraordinary, however, that no research is likely to meet their criteria of sample size, composition, minimum standard errors of estimate, and so forth, unless a giant, collaborative effort were launched. To predict

from past performance, the critical research will certainly not be done by those who demand such rigor from others.

There is also an irony in their demands: as Barker (4) has pointed out, the higher the estimated heritability, the fewer the pairs of related persons needed to detect statistically significant genetic variance, because the power of the test increases as heritability estimates increase. Thus, if heritabilities are low, as predicted for disadvantaged populations, their detection is nearly impossible by biometrical standards. There is an overwhelming bias in favor of accepting the results of studies with high heritability estimates.

There has never been a study of the effects of genetics on human behavior that could withstand all of the criticisms leveled at mine. Does this mean that we know nothing about the effects of genetic and environmental differences on behavior? Nonsense. I believe we do know that genetic differences play an important role in the distribution of individual differences for many characteristics in some populations. Our knowledge is based not on one critical study, but on the accumulated weight of evidence from many partially flawed investigations. Strong inferences can often be made on the basis of such data (5).

I agree that we do not yet have a sufficiently sound basis for making strong inferences about possible differences in the expression of genetic variants within and between many populations and subgroups. The pattern of results I obtained suggested one set of interpretations regarding environmental differences, but more definitive studies are obviously needed.

I agree with Allen and Pettigrew that more models than the two simple ones proposed can and should be tested. In fact I said so (1, p. 1287), but not as eloquently or explicitly as they have. My choice of the two simple and opposing models was not random, however, but was based on prevailing views in the controversy over the relative importance of genetic and environmental differences in intellectual differences.

The environmental disadvantage model is supported by Tanner's (6) analysis of variation in physical growth. He concluded (6, pp. 40-41):

The rate of growth at any age is clearly the outcome of the interaction of genetic and environmental factors. The child inherits possible patterns of growth from

his parents. The environment, however, dictates which (if any) of the patterns will become actual. In an environment where nutrition is always adequate, where the parents are caring, and where social factors are adequate, it is the genes that largely determine differences between members of the population in growth and adult physique. In an environment that is suboptimal and perhaps changes from time to time, as in periodic famines characteristic of much of the world, differences between members of the population reflect the social history of the individuals as much as their genetic endowment.

Tanner went on to discuss the fact that the growth of some individuals is affected more severely by deprivation than the growth of others. In other words, environmental deprivation—in this case nutritional, social, and emotional disadvantages—has a generally depressing effect on average physical growth in a total population and both a depressing and variable effect on the expression of genetic differences among individuals. A principal effect is lowered heritability of differences in physical growth in disadvantaged populations.

To the extent that intellectual growth is similar to physical growth (by being cumulative and subject to the effects of continuous or periodic deprivation), the same simple environmental disadvantage model may well apply. I hope that more studies of intellectual differences within and between populations will further test the appropriateness of this model.

Specific Criticisms

Both technical comments question the appropriateness of the Weinberg rule, which was used to estimate the monozygotic twin correlations and, subsequently, the heritabilities. Interestingly, Allen and Pettigrew conclude that limitations on the appropriateness of the Weinberg rule probably led to an *overestimate* of genetic variance in my study, while Erlenmeyer-Kimling and Stern conclude that the Weinberg rule probably led to an *underestimate* of the genetic variance in the same data. The reasoning behind their criticisms is sufficiently different to lead to conflicting opinions on the effects of the Weinberg rule.

The technical comments agree, however, in questioning the statistical significance of the pattern of results I reported and interpreted to support primarily the environmental disadvantage hypothesis. In response to the same criticism from Dawes (7), I professed ignorance of any known statistical

technique to calculate the reliability of an estimated correlation coefficient. By ignoring the unreliability introduced by estimation, I calculated the usual Fisher formula to show that the advantaged groups of both races had (statistically) "significantly" higher monozygotic than dizygotic correlations, while the disadvantaged groups did not. Since then, several statisticians have contributed error terms that preclude any statistical significance without samples consisting of many thousands of pairs. I stand corrected on the parametric front. The only other comment I would make is that the distribution of monozygotic : dizygotic correlations is still quite interesting: the monozygotic coefficients exceeded the dizygotic in all six comparisons in advantaged groups, but in only one comparison in the disadvantaged groups. This is the pattern of results that I discussed.

Several "technical difficulties" are cited by Erlenmeyer-Kimling and Stern as criticisms of the study. Some of these I acknowledged in the article: (i) individual zygoty could not be determined for each pair because the twins were not seen; (ii) social class ratings depended upon census tract data and thus described neighborhood, not individual, characteristics (which may have been an asset, not a liability, if one goal is to describe the school-aged child's environment); (iii) the raw test data were skewed and had to be normalized; and (iv) small fluctuations in sample size (of less than .02 percent) occurred in the tables. This "bias" occurred because a few children failed to correctly answer a sufficient number of items on a particular subtest to obtain a scaled score; total scores were extrapolated from other subtests by the school testing service (a very trivial point).

Less obvious "technical difficulties" cited by Erlenmeyer-Kimling and Stern pertain to sample losses and to the appropriateness of the Weinberg method.

1) Sample losses, they say, may be differently distributed in the two racial groups. In fact, the total public school twin population, as reported (1), was 64 percent black and 36 percent white; the final sample with aptitude scores was 64.7 percent black and 35.3 percent white. There was no differential loss by racial group. It is true that more black children than white were lost to special classes where standard tests were not given. A larger portion of the lower tail of the black tested-ability distribution was probably lost. As noted (1,

note 26), the results can only be applied to children in normal, public school classrooms.

2) They state that one-third of the starting population was lost. This is not true. As explained in note 27 (1), the aptitude tests were given in every other grade from 2 through 12. Thus, 282 pairs were too young to take the tests, and five grades were not tested in the year we collected data. We actually tried to go back to the previous year's records to obtain aptitude scores on those not currently tested, but this was only possible if a child had not changed schools (because test records were kept only by school building at that time). Of the 1115 pairs in regular classrooms of grades 2 through 12, the sample tested should have included six-elevenths of the total (660) plus some others who remained in the same school building. Since we had aptitude test scores on both members of 778 pairs, I cannot concede that one-third of the sample was lost for biased reasons.

3) Erlenmeyer-Kimling and Stern suggested that the Weinberg differential rule, based on equal numbers of same-sex and opposite-sex dizygotic twins, may be inappropriate. If James (8) is correct in saying that the ratio of same-sex to opposite-sex dizygotic twins is 7 : 6, then the proportion of monozygotic twins was lower than calculated. Therefore, the estimated monozygotic correlations should have been slightly higher than calculated in all groups. The pattern of results would remain exactly the same, however.

4) They assert that higher opposite-sex than same-sex correlations were sometimes obtained, a finding not to be expected on genetic or environmental grounds. I certainly agree, except that Erlenmeyer-Kimling and Stern must recognize that these slight differences fall well within the range of the sampling errors they apply so rigorously to other aspects of the study. Furthermore, I replied to this point previously (7).

5) They correctly note that the sex ratio in the black sample was not the ideal 1 : 1, and they claim that the unusual sex ratio makes the Weinberg rule inapplicable. Let me examine the consequences of this bias.

As noted earlier, the ratio of black to white pairs was the same in the total twin population and in the final sample. The ratio of same-sex to opposite-sex pairs (the central requirement of the Weinberg rule) was also the same in

the twin population and in the final sample. Black opposite-sex pairs were 34 percent of the original population and 33.6 percent of the tested sample; white opposite-sex pairs were 30 percent of both groups. Upon further examination, we discovered that proportionally fewer black males and more black females had actually been tested. For unknown reasons, the larger number of black same-sex females tested had compensated for the loss of black same-sex males, thereby maintaining the racial balance and the same-sex to opposite-sex ratio. One could speculate about the reasons for the unequal sex ratio of black pairs in the public schools and in the tested sample, but the main concern here is how the overrepresentation of female pairs could affect the Weinberg rule. Since the same-sex to opposite-sex ratio was constant, and since there were no sex differences in test scores (1), I do not believe that the final sample was biased in any important way.

6) Erlenmeyer-Kimling and Stern criticize the study's failure to replicate the high heritabilities often reported for general IQ scores in studies of white, middle-class samples. Upon closer inspection of the reported twin studies, one finds the claimed unanimity of results to be highly misleading, based primarily on the questionable reports of Burt's studies (3, 9) and on the use of median data (10). Erlenmeyer-Kimling has unfortunately perpetrated the view that the heritability of IQ can be calculated for any population. Others have long ago shown that multifactorial approaches to intellectual skills yield not only different heritabilities for different measures at different ages in the same population, but also that various components of intelligence may have different sources of genetic variance (11).

7) They ridicule the suggestion that disadvantaged and black children have lower heritabilities for aptitude scores than advantaged and white children. I agree that statistically the pattern of results I obtained was not strictly defensible, but a new study, with improved methodology, is forthcoming. Four hundred pairs of adolescent twins, stratified by race and social class, were studied in Philadelphia (12). Five cognitive skills and many other variables of personality, self-esteem, physical growth, and medical-dental status were assessed. All twins were given extensive blood tests. Several models of genetic and environ-

mental differences will be tested. The study will surely not settle all of the issues raised by Allen and Pettigrew and Erlenmeyer-Kimling and Stern, but our preliminary results do add weight to the environmental disadvantage hypothesis.

Let me emphasize that other partially flawed studies can increase our knowledge of the roles of genetic and environmental differences in relatively unexplored populations and environments. Studies of separated siblings, half-siblings, and adopted children will be particularly valuable contributions to our knowledge, even if no one study can include 10,000 pairs. Over the next several years my colleagues and I plan to collect data on the similarities in intellectual skills among adopted and natural children and separated siblings to add to our twin data. No one study will settle all of the issues, but I hope that others will join us in seeking new knowledge about diverse human groups.

SANDRA SCARR-SALAPATEK
Institute of Child Development,
University of Minnesota,
Minneapolis 55455

References and Notes

1. S. Scarr-Salapatek, *Science* **174**, 1285 (1971).
2. L. J. Eaves and J. L. Jinks, *Nature* **240**, 84 (1972); J. L. Jinks and D. W. Fulker, *Psychol. Bull.* **73**, 311 (1970); L. J. Eaves, *ibid.* **77**, 144 (1972). For very different reasons, a perfectionist stance has also been adopted by Kamin (3).
3. L. Kamin, "Heredity, intelligence, politics, and psychology," paper presented at the Eastern Psychological Association, May 1973.
4. W. F. Barker, "The relationship between sample size, power, and heritability estimates," paper presented at Behavior Genetics Association, Chapel Hill, N.C., April 1973.
5. For example, A. R. Jensen, *Educability and Group Differences* (Harper & Row, New York, 1973).
6. J. M. Tanner, *Sci. Amer.* **229**, 34 (1973).
7. R. M. Dawes, *Science* **178**, 229 (1972); S. Scarr-Salapatek, *ibid.*, p. 235.
8. W. H. James, *Nature* **232**, 277 (1971). James claims that the ratio of same-sex to opposite-sex DZ twin pairs at birth is 7 : 6 rather than 1 : 1. If true, the bias introduced into the estimated proportion of MZ twins in the same-sex group was about 9 percent. Whether the bias persists into middle childhood is unknown because twin studies never begin by ascertaining whether all pairs in the population are living later.
9. A. R. Jensen, *Behav. Genet.*, in press.
10. L. Erlenmeyer-Kimling and L. F. Jarvik, *Science* **142**, 1477 (1963).
11. For example, see J. B. Block, in *Progress in Human Behavior Genetics*, S. G. Vandenberg, Ed. (Johns Hopkins, Baltimore, 1968), pp. 221-228; R. D. Bock and S. G. Vandenberg, *ibid.*, pp. 233-260; J. C. Loehlin and S. G. Vandenberg, *ibid.*, pp. 261-278; R. D. Bock and D. Kolakowski, *Amer. J. Hum. Genet.* **25**, 1 (1973); R. S. Wilson, *Science* **175**, 914 (1972).
12. In collaboration with Solomon H. Katz, William B. Barker, and others. Reports of the entire study are in preparation in two volumes: *Black and White Children: A Study of Physical, Intellectual and Personality Development at Adolescence and Black and White Children: Biomedical Aspects of Development at Puberty*.

2 October 1973