Explaining Race Differences in IQ
The Logic, the Methodology, and the Evidence

Brian Mackenzie University of Tasmania, Australia

ABSTRACT: The claim that there are genetic racial differences in IQ has been based mainly on two grounds: the evidence for high within-race heritability for IQ and the failure of various environmental explanations to account completely for mean racial differences. However, neither of these has direct relevance to the question of whether race differences in IQ have mainly a genetic or mainly an environmental origin. The assumption that these factors are relevant and that they support a genetic account is criticized as the "hereditarian fallacy." The choice between a genetic and an environmental account of race differences is most properly based on "jointly genetic/environmental designs," which control for both genetic and environmental differences in a behavior genetic framework. The available evidence from such designs tends to support an environmental over a genetic account, although the evidence is not presently sufficient and gives little insight into what the relevant environmental differences might be. Identification of those environmental differences may require turning away from the traditional "heredity versus environment" question and focusing on the detailed causal influences on the development of intellectual level generally.

In a number of articles in the past few years, I have criticized the logic, methodology, and assumptions involved in some of the attempts to demonstrate the likelihood of genetic racial differences in IQ (Mackenzie, 1979, 1980a, 1980b). Along with several other authors (e.g., Feldman & Lewontin, 1975; Flynn, 1980; Lewontin, 1970, 1975; Vetta, 1975, 1980), I have shown that in several cases the evidence and reasoning are not merely inadequate but irrelevant to the issues involved. Like Vetta (1975, 1980), but unlike the other authors cited, I have not tried to answer the substantive question of whether or not there are genetic factors involved in the measured race differences in mean IQ, particularly between the black and white populations in the United States. Instead, I have focused solely on the faults in some of the reasoning and evidence that have been presented on the question. This "agnostic" approach may seem restrictive and narrowly methodological, but it does have some advantages. The scientific issues involved in the debate over IQ and race are complex and technical, although some of the positions that have been advanced are simplistic and ideological. Even technical arguments and criticisms often seem to be developed and put forward on an ad hoc basis, because they can be used to justify an already favored position. To the extent that this impression is justified, it is very hard to assess the relevance and importance of particular arguments that are offered, even if we can accept their validity. An analysis of the logic and methodology in the area can throw more light on how the substantive questions may possibly be resolved, if it is not already tied to one particular resolution of them.

In the present article I will extend these methodological analyses and discuss their significance for the problem of explaining black–white IQ differences. In particular, I will try to show that (a) much of the debate on the issue has suffered from confusion over what relevance most of the evidence has for genetic versus environmental explanatory hypotheses; (b) contrary to the claims of some writers, the inadequacy of the specific environmental hypotheses so far put forward does not provide support for a general genetic hypothesis; (c) the particular genetic hypotheses so far put forward have little or no scientific weight; (d) there is only a small body of existing evidence suitable for testing a general environmental hypothesis against a general genetic one; (e) although this body of evidence tends to support an environmental hypothesis, it is inadequate both in size and in quality to resolve the issue; (f) evidence potentially sufficient to resolve the issue can be gathered with currently available methods; (g) the general question of genetics versus environment is an extremely limited form of the question of how to explain race differences in IQ; and (h) progress in dealing with less limited forms of the question will be aided by closer relationships between theories of cognition, behavior genetic analysis, and research designs that focus on environmental manipulation and assessment.

In criticizing the main arguments and evidence for genetic racial differences in IQ, I will focus...
largely on the writings of Arthur Jensen. It is appropriate to focus on Jensen's statements because they contain the clearest, fullest, and most influential presentation of the hypothesis of genetic racial differences in IQ. For this same reason, a thorough criticism specifically of Jensen's position may have more general applicability. However, I will not be trying to show that Jensen in particular holds an untenable position on questions of race and IQ. Most of the arguments and evidence that Jensen puts forward have been offered by others as well, as I will show. Furthermore, Jensen's own position on the question is a more modest one than many critics have maintained. In a recent book he explains that one of his "chief aims" has been merely to show that the causes of race differences in IQ constitute "an unsettled question for which our present evidence cannot justify the pretense of a definitive answer"; he has proposed and argued for a hypothesis of genetic racial differences in IQ, he says, mainly in order to provide "goad and grist for scientific action" (Jensen, 1981a, p. 214).

Three Kinds of Evidence

Throughout, I will assume the reliability of the general finding that blacks in the United States score, on average, about 15 points lower than whites on a variety of IQ tests. Equally, I will not dispute the conclusions emphasized by Jensen (1980) that the difference is greater on nonverbal than on verbal tests, that it is greater on culture-reduced than on culture-loaded ones, and that item analyses reveal few, if any, significant race-by-item interactions. The issue is whether this 15-point mean difference is due almost entirely to genetic differences between blacks and whites, due almost entirely to environmental differences, or due to a significant amount of both. It may be noted that few, if any, serious researchers support the first of these positions, that the IQ difference is all or almost all genetic. Instead, those who maintain that there is a genetic difference in IQ usually support the third position, that both genetic and environmental factors are involved. Those who support an environmentalist position, however, typically take a stronger stand, insisting that there are no significant genetic differences in IQ between races. Three broad classes of evidence can be used to choose between the competing hypotheses.

The first is genetic evidence: Genetic mechanisms or evolutionary processes that would produce or index IQ differences can be derived from genetic theory or evolutionary biology, and we can look for evidence that these processes occur differentially in the two racial groups. Alternatively, we can hypothesize that the difference is genetic (in whole or part), propose candidates for genetic variables that might produce or index the IQ differences, measure the black–white difference in these, and test for the relationship of these variables to IQ.

The second is environmental evidence: Environmental processes that would produce IQ differences can be derived from psychological or sociological theory, and we can look for evidence that these processes occur differentially in the two racial groups. Alternatively, we can hypothesize that the difference is environmental, propose candidates for the relevant environmental variables, measure the black–white differences in these areas, and test for the relationship of these variables to IQ.

In each of these two cases, the second strategy is theoretically weaker than the first. In the first, strong form, we actually derive the mechanisms or processes from the genetic or environmental theory. In the second, weaker form, we try merely to ensure that the proposed processes do not contradict the theory; we actually propose them on the basis of our intuitions, hunches, or wide clinical or experimental experience. The distinction is worth making here, because it is sometimes suggested that genetics is in a stronger position than psychology or sociology to account for race differences in IQ. The reason is that genetics has a formal, fully elaborated, and general theory, whereas the other two have only a bewildering variety of often incompatible microtheories. I do not dispute the stronger theoretical status of genetics here, not because it is indisputable but because it is irrelevant. Genetic hypotheses to account for the observed IQ differences are, at least as much as environmental hypotheses, based on the second, weaker strategy of proposing possibly relevant variables and effects that at best do not contradict the theories of their respective disciplines. No one has ever both derived and tested a mechanism that would produce or index the mean black–white IQ difference from genetic theory or evolutionary biology as such. Those that have been derived from the theories have not been testable. Those that can be tested have not been derived from the theories. I will return to this point in reviewing the genetic evidence.

The third class is jointly genetic/environmental evidence. We acquire this kind of evidence by focusing on those situations in which genetic and environmental hypotheses lead to clearly opposite predictions. In doing so we empirically oppose the two hypotheses to each other. This is the research strategy that Platt (1964) has called the method of strong inference, in which we do not merely test two rival theories on their own and see which makes

Requests for reprints should be sent to Brian Mackenzie, Department of Psychology, University of Tasmania, Hobart, Tasmania, Australia 7001.
more successful predictions, but rather test them against each other, seeking events that simultaneously confirm the one and disconfirm the other. This method would be fairly easy to apply if we had opposing predictions clearly derived from the relevant theories. Because we usually do not, we must look for cases where as many as possible of the genetic variables that have been proposed as relevant, and as many as possible of the environmental variables that have been proposed as relevant, are largely controlled for in the same or comparable samples of individuals.

This third class of evidence is clearly the strongest of the three. If any particular genetic hypothesis is refuted, that fact does not prove that the difference is environmental. At most (assuming adequate testing procedures, provision for interaction effects, etc.), it proves either that the difference is nongenetic or that the genetic influence or effect being tested is not the appropriate one. The same holds for environmental predictions and influences. Only designs that yield jointly genetic/environmental evidence permit a relatively clear choice between genetic and environmental hypotheses overall. Even with such designs, of course, absolute certainty cannot be achieved. The strength of the conclusions that can be reached by pitting genetic and environmental hypotheses against each other is relative to our level of understanding of genetic and environmental influences and mechanisms. Later discoveries may make us conclude that we did not pit the right hypotheses against each other, so that the results of the confrontation have to be revised. This limitation, however, applies to any scientific conclusions.

In the field of individual differences, jointly genetic/environmental designs are most familiar in studies of separated identical twins that were undertaken to estimate environmental and genetic variance in IQ within racial groups. In the twin design, genetic variables are controlled by being identical within each twin pair, and environmental variables are controlled by random allocation of twins to environments. The defects in studies that use the twin design have nothing to do with their logic, which is sufficiently compelling that twin studies often have more credibility than they probably deserve. The defects arise instead from failure to meet the logical requirements of the design in that, with the small sample of separated identical twins available for study, it is very hard to find enough pairs in which allocation to environments can be considered random. It remains to be seen whether, in the study of race differences in IQ, jointly genetic/environmental designs are available that are not so plagued with sampling difficulties.

Given the greater cogency and relevance of jointly genetic/environmental evidence, one might expect that it would receive more attention than the other two kinds, both in data collection and in interpretation. In fact, it has received much less. Only a small amount of this type of evidence has ever been collected, and what little exists is only occasionally emphasized. Instead, attention has mostly, and unprofitably, focused on the second approach, that of proposing and testing hypotheses about specific environmental influences that might account for the mean IQ difference between races. The reasons for this misplaced emphasis are probably to be found in the recent history of the race–IQ controversy. As both Jensen (1980) and Vernon (1979) have pointed out, there has been a widespread assumption in the social sciences since the middle 1950s that the black–white IQ difference is clearly and obviously environmental in origin, with poverty, prejudice, and malnutrition as the most easily specified candidates for deleterious environmental influences. The critical examination and substantial rejection of these candidates, brought to public notice by Jensen (1969) and extended in later writings by Jensen and others, stimulated a sometimes feverish search for more credible alternative candidates. Thus, the focus in the debate on race and IQ has continued to be on the criticism and defense of environmentalist assumptions, rather than on the development of the most effective and unbiased research strategies for resolving the issues.

Environmental Models and Evidence

A large number of hypotheses have been advanced to account for the black–white IQ difference in terms of the different social, economic, educational, cultural, motivational, nutritional, and medical situations of blacks and whites in the United States. Some of these have been purely speculative, but many of them have been stated in precise enough terms that they could be empirically tested. I will not attempt to review these here, because thorough reviews are already available (e.g., Loehlin, Lindzey & Spuhler, 1975; Minton & Schneider, 1980; Vernon, 1979), because there is far more material available than could be discussed in one article, and because it is the systematic import of such research that is of concern here, rather than the details of individual studies.

It is clear that some of the natural environmental variables investigated by researchers in the field are significantly related to IQ level and to the mean racial difference. It is likewise clear that some of the programs of environmental enrichment and compensatory education have been successful in raising the IQ and other indices of intellectual performance in the target populations, largely of black children. The extent of this success, however, is disputable. Many authors, such as Vernon (1979) and Minton...
and Schneider (1980), maintain that the intellectual gains are usually short-lived, almost disappearing in 1- or 2-year follow-up studies. Palmer and Andersen (1979), by contrast, find evidence in a large number of studies of long-term gains following early intervention, although the gains are greater on measures such as age-grade placement and school performance than on IQ. Such studies therefore provide stronger evidence on the overall social effectiveness of such programs than on the specific question of the source of race differences in IQ. What is finally clear from such research, therefore, is that environmental factors have not been identified that are sufficient to account for all or even most of the 15-point mean difference in IQ between blacks and whites in the United States. Jensen’s (1973a) conclusion that half to two thirds of the gap remains unaccounted for by any proposed combination of environmental influences is still unfurled.

Criticism of Environmental Models: The Sociologist’s Fallacy

This lack of firm evidence for an environmental explanation has not stopped some writers from vigorously insisting that an environmental explanation must be the correct one. In doing so, they sometimes resort to special pleading and ad hoc arguments. The most obvious kind of special pleading is to make the description of hypothesized environmental factors more and more vague, subtle, and (as a consequence) untestable, as more precise versions are shown to be inadequate. As Jensen complained, “Some environmental factors are formulated clearly enough to be put to the test of evidence; as each of the hypothesized factors is rejected on the basis of evidence, other increasingly subtle environmental deficits are postulated to explain the differences” (Jensen, 1973a, p. 232; Jensen gives specific examples on pp. 125–129 and 231–236).

The most important kind of special pleading is what Jensen calls the “sociologist’s fallacy.” This is the fallacy of assuming that any environmental difference between individuals or groups reveals a purely environmental influence responsible for producing IQ differences. This fallacy is typically displayed in discussions of socioeconomic status (SES) and related sociocultural factors and their relationship to IQ. The idea is that because there are substantial black–white differences in SES, and substantial differences in mean IQ between different SES groups, the black–white IQ differences should be corrected for the black–white SES differences. What makes this position fallacious is the assumption that any SES differences in IQ must be purely environmental rather than partly genetic. If there are any genetic social class differences in IQ, then equating black and white samples for SES would equate them on some genetic as well as some environmental factors. Furthermore, a hypothesis of genetic social class differences in IQ, although by no means established, is a genuine and meaningful one; it is not merely an abstract speculation. A mechanism that could produce such differences, through IQ-related intergenerational social mobility, has been clearly spelled out and shown to operate in some cases (Waller, 1971). Some writers have offered analytical support for it (e.g., Herrnstein, 1973). None of this amounts to strong evidence in favor of genetic social class differences in IQ, of course, but it shows that it is a mistake to ignore the possibility of them or to dismiss that possibility out of hand.

Sociologists, and psychologists too, easily slip into the sociologist’s fallacy when they attempt to interpret survey data on the different social situations of blacks and whites in the United States. Mercer and Brown (1973), for instance, found that the mean IQ differences between black, “Anglo” (white English speaking), and Hispanic children (180 of each in a California school sample) were almost completely eliminated following statistical control of nine social variables: the mother’s participation in formal organizations (scouts, church groups, etc.); racial composition of the neighborhood (more versus less than 40% Anglo); mother’s fluency in English and familiarity with child’s school; SES; parents’ geographical origins (northern or western United States versus southern United States or Mexico); home ownership; mother’s individualistic achievement values; family structure (intact nuclear versus other); and the child’s general anxiety level. They interpreted their results as showing that the nine social factors, as purely environmental influences, were responsible for producing the mean IQ differences between the three racial–ethnic groups. Indeed, as influences, most of the nine would be environmental (although child’s anxiety level is a dubious candidate for an environmental influence on the child). The fallacy of their conclusion lies in the assumption that there were, within each group, no IQ-relevant genetic differences associated with position on the nine social variables. It is not clear that there were any such associated genetic differences, of course, but Mercer and Brown’s data do not provide any evidence that there were not.  

1 To be fair, Mercer and Brown did try to rule out a genetic interpretation, but they did not do it successfully. They reasoned that if the correlation between the nine social variables and IQ was due to between-group genetic factors, then controlling for genotypic intelligence should eliminate any correlation between the nine social variables and racial-ethnic group membership. Using the child’s IQ as the estimate of genotypic intelligence, they partialed IQ from the correlation between group membership and each of the social variables. Most of the correlations remained significant, reinforcing their conclusion, they claimed, that the social variables should be considered purely as environmental.
A more ambitious attempt along similar lines was made by Blau (1981), in a Chicago school sample of over 500 white and 500 black children and their families. Blau regressed scholastic achievement and IQ on a wide variety of social structure and socialization variables. The social structure variables included four components of SES, two of religious practice, and three of family structure. Among her many findings was that, when the structural variables were controlled, the mean racial IQ difference was reduced from 12 points to 2.7 points for girls, and from 9 points to 2.9 points for boys. Blau went on to propose many ingenious and reasonable social influence pathways through which these structural and socialization variables might have led to racial differences in IQ and cognitive skills. She concluded that her data provided strong evidence that these differences “are not inherent but result from identifiable environmental deficits of a cumulative and complex nature that are associated with poverty, but which constitute independent impediments to the intellectual development of minority children” (Blau, 1981, p. 222). At no time, however, did she address the question of whether there might be genetic differences associated with the structural and socialization variables within each racial group. She, therefore, could not address the question of the extent to which racial IQ differences were caused by the social differences and the extent to which they were caused by genetic differences that could underlie some of the social ones as well. Her conclusion, therefore, is another clear example of the sociologist’s fallacy.

Social survey studies such as those of Mercer and Brown (1973) and Blau (1981) are among the most extensive and detailed research we have on race differences in IQ. Are we to dismiss such research as worthless, with no relevance to explaining race differences in IQ? Clearly not; research of this sort can be critically important when used and interpreted properly. Such research, when considered alone, cannot provide answers to the question of whether race differences in IQ have a genetic component. But if the question involved identifying relevant environmental variables (rather than deciding whether any genetic ones exist), then survey research of this sort would come into its own. When carried out on particular additional populations, it could be a powerful tool for identifying what the between-race environmental differences were that were responsible (whether in whole or in part) for between-race IQ differences. (I will return to this point in the concluding section.) As long as such research is directed to the general “heredity versus environment” question, however, it cannot provide any valid answers, and the answers it seems to provide are fallacious.

Environmental Models and Evidence: Conclusion

The import of the findings of environmental research for the general hypothesis that the mean black-white IQ difference is entirely or almost entirely due to environmental differences can, therefore, be summed up very simply: They fail to confirm it. Of course, they do nothing to refute it either, but they certainly set the stage for acceptance of any competing hypothesis that is more able to generate confirmatory evidence. Assessing the status of the environmental hypothesis in this way is relatively simple because, despite some conceptual confusion and special pleading, it has generated a large number of specific, testable hypotheses. Assessing a general genetic hypothesis is not so simple because it has not generated specific, testable hypotheses in the same way. Instead, the support for a genetic explanation, especially as presented by Jensen (1969, 1973a, 1973b, 1981a), comes from a complicated mixture of evidence and argument. In the next section I will disentangle the threads that are used to support the genetic hypothesis, so as to permit a comparative assessment of the models and evidence for both hypotheses.

Genetic Models and Evidence

Historically, many writers have thought it was obvious that there were genetic racial differences between groups, just as many now think it is obvious that there are not. As late as 1952, R. A. Fisher maintained that “available scientific knowledge provides a firm basis for believing that the groups of mankind differ in their innate capacity for intellectual and emotional development, seeing that such groups do differ undoubtedly in a very large number of their genes” (United Nations Educational, Scientific, and Cultural Organization, 1952, p. 56). Loehlin et al. (1975) made an effective reply and set the problem in a modern genetic context:

Lewontin (1972), on the basis of a survey of published frequency data for 17 polymorphic genes, estimated that only about 6 percent of the genetic diversity in the total human population of the earth is accounted for by racial membership. (Genetic differences between local populations within races accounted for about 8 percent more, but by far the largest portion of the genetic variation, more than 85 percent, represented differences between individuals within populations.) . . . And yet at some blood-group
loci, and presumably at loci affecting skin color, physical appearance, and a number of other biological characteristics, these populations have strikingly different distributions of allelic frequencies. Do the genes that influence intelligence fall into the major category—in which there are only small differences in frequencies between races—or the minor category—in which differences in frequencies may be quite large? This is one way of putting the central question of this book. (pp. 37-39)

**Theoretical Genetic Models**

It is not hard to derive a mechanism from theories in genetics and evolutionary biology that would produce or index between-group genetic differences in IQ. Loehlin et al. (1975), for instance, have shown that if two noninterbreeding groups differ very slightly in their reproductive selection for IQ, the mean IQ of the two groups will gradually diverge. In their example, which assumed a moderately high narrow heritability for IQ, 250 generations, less than 5,000 years, are enough to produce a genetic between-group difference of 10 points. Because the major racial groupings have almost certainly been differentiated for much longer than that, there has been more than enough time for such evolutionary divergence in IQ to occur. On the other hand, there is of course no evidence at all of any such differential reproductive selection over such a time span. The example is a good one of a mechanism that can be derived from genetic theory, but not tested.

For another example, Thoday (1969) neatly derived a genetic effect that would index differences in the genetic value of a trait in contemporary populations. His measure involved comparing the genetic variance on the trait in hybrids with the genetic variance in the parent populations; if the parent populations have different genetic values, hybrids (results of intergroup mating) should show increased genetic variance. However, Thoday's measure could be confounded by effects of heterosis and by between-group differences in dominance variance and would at most reveal that the groups had different genetic values on the trait, not which one was higher.

Finally, DeFries (1972) developed an elegant measure of the heritability of between-group genetic differences. Unfortunately, his measure of between-group heritability required measuring the between-group genetic differences first, and therefore it does not help in deciding whether these differences occur.

Genetic models such as these are not applicable to answering the question of whether between-group genetic differences in IQ occur. They are developed for quite another reason, to show how genetic theory could accommodate, make sense of, or explain such differences. Demonstrating that they exist is another matter altogether.

**Jensen's Genetic Models**

Arthur Jensen has repeatedly and consistently argued that a genetic explanation for the mean black–white IQ difference is supported by testable genetic effects. In some instances, other psychologists have presented similar arguments. However, all three of the major claims that Jensen has made in this area have been decisively refuted. I made possibly the most systematic of the refutations (Mackenzie, 1979, 1980a, 1980b), but my analyses did little more than put the icing on the cake by giving a more systematic expression to some insights that were already distributed through the literature.

In his first attempt, Jensen made the bald claim that high trait heritability within groups should lead us to expect genetic control of trait differences between groups, as long as the groups had been “geographically or socially isolated from one another for many generations” (Jensen, 1969, p. 80). In a later version of the same article he removed the need for geographical isolation, stating simply that “characteristics that vary genetically among individuals within a population also vary genetically between different breeding populations of the same species” (Jensen, 1972, p. 162). Lewontin (1970) countered with a now famous pair of thought experiments in plant genetics. In one, within-group variation was entirely genetic whereas between-group differences were entirely environmental. In the other, within-group variation was entirely environmental, whereas between-group differences were entirely genetic. Lewontin's point was that high (or low) within-group heritability should not lead us to expect anything at all about the control of between-group differences, because the two kinds of variation can be controlled independently. Jensen (1970) replied that the expectation was still a reasonable one, even if it was sometimes incorrect:

The real question is not whether a heritability estimate, by its mathematical logic, can prove the existence of a genetic difference between two groups, but whether there is any probabilistic connection between the magnitude of the heritability and the magnitude of the group differences. Given two populations (A and B) whose means on a particular characteristic differ by x amount, and given the heritability (h_A^2 and h_B^2) of the characteristic in each of the two populations, the probability that the two populations differ from each other genotypically as well as phenotypically is some monotonically increasing function of the magnitudes of h_A^2 and h_B^2 (pp. 21-22)

I showed (Mackenzie, 1980a) that, despite the technical sounding language, Jensen's statement was not based on genetic theory at all. It was merely a more general statement of what he was trying to establish in the particular case, that within-group genetic differences index between-group ones. There
is no genetic mechanism that would require such a relationship, and Jensen did not offer any empirical support for it.

In his second attempt, Jensen appealed to the model for between-group heritability developed by DeFries (1972), mentioned above. He maintained (Jensen, 1973a, 1973b, 1978) that because between-group heritability increases as a function of within-group heritability, as well as of a number of other factors, high within-group heritability is evidence for nonzero between-group heritability and hence, for genetic control of between-group differences. He noted that the relationship between the two kinds of heritability did not yield any quantitative estimates of the degree of genetic control of between-group differences but made a point of its theoretical importance. I pointed out (Mackenzie, 1980a; cf. Feldman & Lewontin, 1975; Lewontin, 1975) that because one of those “other factors” that between-group heritability depends on, in fact, genetic between-group differences, which have to be specified before DeFries' model can be applied, Jensen's claim was conceptually as well as empirically empty.

In his third attempt, Jensen (1969, 1973a, 1973b; cf. Scarr-Salapatek, 1971; Vernon, 1979) focused on differential regression effects in the black and white populations. The IQ scores of the offspring or siblings of blacks regress to the black population mean of 85, whereas the IQ scores of the offspring or siblings of whites regressed to the white population mean of 100. Jensen claimed that a genetic hypothesis for between-race IQ differences predicts these differential regression effects and that the confirmation of the prediction therefore provides strong evidence for the hypothesis. I showed (Mackenzie, 1979, 1980b; cf. Vetta, 1975, 1980) that regression is a purely statistical phenomenon with no separate biological or genetic significance, that the differential regression effect is due solely to the phenotypic difference regardless of whether there are genotypic differences, and that the IQ scores of the two racial groups would regress to the combined mean if the sibling or parent–offspring correlations were calculated in the combined sample rather than in each racial sample separately.

As pointed out above, my demonstration of these points added only a little to what was already available in the literature in a less systematic form. It is clear that the attempt to demonstrate genetic control of race differences by models such as these is now a dead issue. Although Jensen has not acknowledged his errors in proposing any of these models, he has tacitly dropped them from his more recent discussions of race differences (e.g., Jensen, 1981a). I have briefly reviewed them here because, in evaluating the significance that Jensen attached to the failure of environmental models, it will be necessary to have the failure of his genetic models available for comparison.

The Strategy of Jensen's Arguments for Genetic Racial Differences

Jensen's main argument for a genetic hypothesis to explain race differences in IQ is not based on these fallacious genetic models. It is based instead on the failures of environmental explanations, and he leads up to it gradually. After criticizing “environmentalists” for their ad hoc arguments and closed-mindedness he wrote:

Therefore, in studying subpopulation differences in mental abilities, does it not seem a more scientific approach to consider all factors which are known to cause individual differences within groups? And is it not reasonable, if for purposes of research strategy we must assign some priority to the hypothesized causes we wish to consider, that the evidence derived from studies within groups should serve as a guide to the kinds of hypotheses most worth entertaining about the causes of differences between groups? And does this not lead us directly to the hypothesis of genetic factors as being among the undoubtedly multiple causes of racial subpopulation differences in mental abilities? Furthermore, it is practically axiomatic in biology that any characteristics showing individual variation within subgroups of a species will also show variation between subgroups of the species. (Jensen, 1973a, p. 129)

Jensen's argument in this passage is ambiguous. At one level he may be taken as arguing for the likelihood of the existence of between-race genetic differences in IQ. This interpretation is strengthened by the similarity of these comments to the ones Jensen made (1970, pp. 21–22, quoted above) in reply to Lewontin, in which he explicitly made such an argument on similar grounds. The fallacy of that argument has already been pointed out.

At another level, however, Jensen may be taken as arguing, not for the truth of a genetic hypothesis, but only for the reasonableness of investigating it. On this interpretation, the existence of genetic differences within groups is not evidence for their existence between groups, but it directs attention to the possibility that they exist; it serves as a source of research hypotheses that must then be tested on their own. The relevance of within-group genetic factors thus lies in the context of discovery, in which hypotheses are invented, rather than in the context of justification, in which they are tested and either rejected or provisionally accepted.

Jensen is perfectly right in this interpretation. Within-group genetic differences can quite properly suggest a search for between-group ones. It is, therefore, quite appropriate for us to entertain a genetic hypothesis, as Jensen puts it, as long as we do not show it more hospitality than it can prove it deserves. Before we trust it enough to conclude that it is a
It is essential to be quite clear on just what Jensen's argument consists of. The structure, as well as the influence, of the argument was set out most clearly and systematically by Flynn (1980), who said that Jensen has fashioned a powerful argument in two steps: first, most of the IQ variance within white America is due to genetic factors and most of IQ variance within black America may be, although here the evidence is just beginning to come in, second, we can falsify every current hypothesis which attempts to explain the gap between black and white in environmental terms. Therefore, we have good reason to suspect that most of the gap between black and white is genetic. (p. 23)

Flynn (1980) went on to outline the way Jensen used this two-step argument and to testify to its almost overwhelming power:

It is hard to tell whether it is more potent when the $h^2$ estimates come first and the attack on candidates for the role of blindfold second, or when the steps come in reverse order. He usually begins with his first step. He concedes that high heritability estimates do not in themselves constitute a case for a genetic hypothesis about group differences, rather such estimates merely set the stage for evidence which renders such a hypothesis highly probable. They set the stage of course by driving the environmentalist to posit what I have called a blindfold and what Jensen calls a factor X: a factor "which is present in one population and not in the other and which affects all individuals in one population and none in the other"; and which must have "an equal or constant effect on all members of the population in which it is present." And once we have falsified every specific candidate for the role of factor X, a genetic hypothesis is highly probable. At other times he reverses the order. ... Using an $h^2$ for both black and white of .85 and a between-families environmental component of .05, he calculates that a strictly environmental hypothesis about the 15-point gap between black and white must assume ... that most of black America suffers from an environment so bad that it falls virtually off the bottom end of the scale of environmental effects in the white population. Jensen challenges environmentalists to find any significant environmental factor on which the gap between black and white is even close to [the necessary] 4.5 standard deviations. He provides a series of estimates: the gap for SES is about 1.24 SDs, for income .80 SDs, for unemployment rates .33 SDs, and so forth. (pp. 61-63)

In short, if there is relatively little environmental variance (i.e., high heritability) in IQ within racial groups, then the environmental factors needed to account for substantial IQ differences between racial groups must be ones that, although having a strong effect on IQ, are relatively uniform within each group. It is difficult even to imagine any credible environmental factors that meet these requirements, and no satisfactory ones have ever been proposed. Therefore, it can reasonably be concluded that the sources of race differences in IQ are most likely to be genetic. The genetic hypothesis can win the day by default, on the basis of the weakness of its
opponent, without having to survive any independent tests of its own. Flynn (who accepted Jensen's arguments but opposed his conclusions) therefore maintained that if an environmental explanation of the mean black–white IQ difference is to retain any credibility, or even to be interpretable if it is independently supported by some other evidence, it is necessary either to discover what he called a "semi-blindfold," or to show that Jensen's estimates of IQ heritability within racial groups are inflated. He attempted to do both, but he was more hopeful about the prospects for showing that IQ heritability is lower than Jensen estimated. If he could do this, he suggested, then more room would be left for the influence of environmental factors overall. That is, environmental factors that could be proposed to account for IQ differences between races would have a large enough field of application (as they must, to be credible) to account for IQ differences within races as well.

Criticism of Jensen's Argument: The Hereditarian Fallacy

As it happens, there is good evidence that Jensen's estimates of IQ heritability are indeed inflated. Plomin and DeFries (1980) reviewed a large body of modern data that jointly indicated that the broad heritability of IQ in contemporary Western populations is around .50, rather than the .75–.80 that Jensen estimated. Their conclusion was based on reviews of heritability analyses alone and did not require the identification of any major environmental sources of variation; instead, they confessed almost complete ignorance of what the relevant environmental variables might be.

However, such a demonstration was not necessary for the purposes that Flynn claimed it was. Considering the enormous influence that Jensen's writings have had on his supporters, his critics, and those who have tried to remain uncommitted in the controversy, it comes as something of a shock to realize that both Jensen's reasoning and Flynn's faithful reconstruction of it are totally invalid. Jensen's two-step argument does not lead to the conclusions that Flynn said it does and that Jensen in fact reached, even if each step is correct. Instead, the two-step argument is a clear example of what may be called the "hereditarian fallacy." The hereditarian fallacy is the assumption that unless an IQ difference can be given a detailed and specific environmental explanation, it can most reasonably be concluded to have a genetic one. This assumption is clearly expressed in Flynn's (1980) reconstruction of Jensen's argument, "Once we have falsified every specific candidate for the role of factor X, a genetic hypothesis is highly probable" (p. 62). Jensen did not express this assumption and its place in his argument as clearly as Flynn did or as baldly as I have. Nevertheless, it is the linchpin of his case: It is only this assumption that can provide a basis for Jensen's emphasis on refuting environmental hypotheses as the major way of gathering support for a genetic one. The assumption, however, is a fallacy. In fact, it is a double fallacy, a fallacy in two parts.

First, it assumes that our inability to identify the relevant environmental variables provides good grounds for judging that there are none to be identified and therefore for rejecting an environmental hypothesis generally. It is for this reason that Jensen spent the greater part of his writings on race and IQ in attempting to "falsify every current hypothesis which attempts to explain the gap between black and white in environmental terms" (Flynn, 1980, p. 23). Once these current hypotheses are all falsified, Jensen could then assume that "environmentalism" in general has been refuted, even if some obdurate environmentalists refuse to accept the fact. What makes this assumption a fallacy is not the trivial logical truth that more successful environmental hypotheses may one day be proposed. It is more basic than that. The success or failure of an environmental hypothesis of the source of race differences may have nothing to do with our ability to identify environmental influences. In the domains of both personality (Loehlin & Nichols, 1976) and intelligence (Plomin & DeFries, 1980) we can, through behavior genetic techniques, make estimates of the considerable variance attributable to environmental sources, while confessing almost complete ignorance of what the specific environmental influences might be. Knowing whether there are important environmental influences and knowing what the relevant environmental factors are, are thus entirely separate pieces of knowledge, gained by asking very different research questions. Given our present research methods, failure to answer the second question (the "what" question) has no implications for the answers we give to the first (the "whether" question).

The constraints imposed on explanations of race differences by high within-race heritability estimates are therefore purely logical ones. If within-race heritability is high, then an environmental explanation of race differences must involve environmental factors with greater between-race than within-race variance. Conversely, if race differences are environmental, then either the environmental factors must be ones with greater between-race than within-race variance, or within-race heritability must be lower than Jensen estimated. These logical considerations do not, however, amount to evidence against an environmental explanation, any more than they amount to evidence against Jensen's heritability estimates. They can, once again, raise suspicions about where the explanation of race differ-
ences is to be found, and they can just as legitimately raise suspicions about Jensen's heritability estimates. In each case, however, the suspicions can do no more than stimulate a search for evidence; they cannot substitute for it.

The first part of the hereditarian fallacy, then, is to assume that the failure to identify specific, relevant environmental influences is sufficient to refute an environmental hypothesis generally. The second part is to assume that the refutation of an environmental hypothesis is sufficient to confirm a genetic one. It is for this reason that refuting environmentalism is so important to Jensen's strategy. His genetic hypothesis is waiting in the wings and need only await the collapse of its environmental rival before taking its place at center stage. The assumption, however, is fallacious, and the strategy invalid. The competing general hypotheses, genetic and environmental, must start off on an equal footing. They both need to be elaborated, articulated, and defended through the development and testing of specific, empirical research hypotheses. In each case, that is, a specific research hypothesis must be offered and corroborated in order to give the general hypothesis any credibility. In the absence of such specifics, it is no more an explanation of race differences to say, "We can't find the environmental cause, therefore it must be due to the genes," than it is to say, "We can't find the genetic cause, therefore it must be due to the environment."

The failure of the specific tests of one hypothesis can indeed be used to the advantage of the other, but only if that other hypothesis already has clear empirical content of its own—only, that is, if it has given rise to its own research hypotheses that have been tested and confirmed. Successful tests of the competing hypotheses are essential because, except in the special cases where the method of strong inference can be used, the failure of one hypothesis has no direct implications for the validity of the other. All it can affect is the relative weight of the evidence on the two sides. There must be evidence on the other side, therefore, for the failure of the first hypothesis to make any difference. For this reason, the idea that one hypothesis can win by default turns out to be illusory. There is no substitute for confirmatory evidence.

What tests has Jensen's genetic hypothesis passed, therefore, that put it in a strong position in its own right, able to capitalize on the failure of its competitors? I have already reviewed Jensen's attempts to provide testable articulations of his genetic hypothesis. These attempts proved wholly inadequate, because on examination they turned out not to be significant articulations of his hypothesis at all, but misconceptualizations that were tested and refuted on simple logical grounds. His "alternative genetic hypothesis" with "high credibility" (Jensen, 1973a, p. 129) is therefore not one for which he has provided any separate empirical or theoretical justification. The only source of its high credibility is the success and prestige of genetic theory in general. Even there, however, Jensen's position is without support. Jensen did not, as his writings frequently seemed to suggest, have the great weight of genetic theory on his side in arguing for genetic between-race differences in IQ. Genetic theory is entirely neutral on the issue. All that Jensen could take from genetics was some of the technical language of genetics. He could phrase his position in genetic terminology because genetic theory could accommodate genetic racial differences in IQ. It could make sense of them and provide a technical description of them. But it could just as easily accommodate and make sense of the nonexistence of any such differences (for more on Jensen's use of technical language, see Mackenzie, 1980a).

In summary, the first part of the hereditarian fallacy is to assume that the failure to identify specific environmental causes of race differences in IQ is sufficient to refute an environmental hypothesis generally. It is a fallacy because our research methods do not require us to pinpoint the environmental influences, any more than the genetic ones, in order to estimate the relative contribution of each. The second part of the hereditarian fallacy is to assume that the refutation of an environmental hypothesis is sufficient to confirm a genetic one. It is a fallacy because the failure of one hypothesis can strengthen the case for a competing one only if that competing one has some independent empirical or theoretical backing of its own. In the absence of such backing the failure of an environmental hypothesis could not in itself provide support for any alternative hypothesis. Indeed, how could it possibly do so? To conclude that "it must be due to the genes," on the grounds that we cannot find anything in particular that is due to, would be to reach a conclusion that was not merely made in ignorance but that was explicitly based on ignorance. Therefore, it would not be a scientific conclusion at all and would be more like concluding that "it must be due to destiny" or "it must be due to the will of God." In each case the conclusion would have the effect mainly of declaring that the resolution of the issue lay beyond the reach of scientific inquiry.

Genetic Models and Evidence: Conclusion

With regard to producing even tentative conclusions about the source of race differences in IQ, the genetic models and arguments have been found to be, at best, no more helpful than the environmental ones. On reflection, this result should not be surprising. Neither quantitative nor molecular genetics
have yet arrived at the point where they can begin to trace the expression of a polygenic trait in natural human populations. When they do, their ability to detail the mechanism of gene expression may nullify current debates on the relative weights of genetic and environmental influences on a trait. Until they do, however, questions about the distribution of influence on traits that may have substantial variance from both genetic and environmental sources are best answered by research designs that consider genetic and environmental variables together. It is to such designs that we now turn.

**Jointly Genetic/Environmental Models and Evidence**

Jointly genetic/environmental designs are commonplace in the behavior genetic study of natural populations. The study of separated identical twins, mentioned earlier, is only the most obvious example; the strategy is a general one in all kinship studies. Genetic variables are controlled through specifying the degree of genetic similarity of the subjects being studied (from identical twins to unrelated persons), and environmental variables are controlled through specifying the degree of their environmental similarity. This latter variable does not have to be quantified as long as it can be considered equal for two or more grades of kinship. The strategy is not foolproof; serious questions can be raised, especially on the environmental side (Taylor, 1980). But it is to some extent self-correcting; inappropriate assumptions about such issues as the degree of environmental similarity will lead to inconsistent and meaningless results.

In the study of race differences, jointly genetic/environmental designs are less common but not unknown. They involve the study of individuals with different racial backgrounds (e.g., blacks and whites, or persons with different proportions of African and European ancestry), who live in equivalent cultural or home environments, or alternatively, of individuals with a single racial background who live in different cultural or home environments (e.g., associated with the white or the black community). Generally, therefore, such designs require the independent variation of racial composition and cultural environment. Within this general framework many more designs are possible than have been investigated, and many are impractical. Collection of data on whites living in a black cultural environment, for instance, has never been seriously considered except on an anecdotal level (e.g., Griffin, 1977). Studies of blacks living in a white cultural environment are only slightly less scarce: They include studies of interracial adoption (Scarr & Weinberg, 1976) and several older studies on “passing” (e.g., Wirth & Goldhamer, 1944). Such studies can provide valuable insights on the problems and accomplishments that accompany assimilation into a racially atypical environment, but can yield only very limited information on the source of race differences in IQ. The reason is that the black and white cultural environments in the United States are relatively distinct. Studies of “passing” typically focused on adults or adolescents who had been socialized in the black community. Studies of interracial adoption focus on black children who can be expected, outside of the specific home environment, to be assimilated to a greater or lesser extent to the black cultural environment. In neither case can it be maintained that the blacks involved are living in an environment equivalent to that of any group of whites.2

What is left? Only two sources of data, in my view, escape these limitations (although they have problems of their own, which will be discussed). They are data from racially mixed persons with different proportions of African and European ancestry: persons who live in the black community in the U.S., on one hand, and in communities (outside the U.S.) without distinct black and white cultural contexts, on the other. In this section I will briefly discuss the major studies of these sorts, point out their strengths and limitations, and suggest how some of them could be made the basis for a strong test of environmental versus genetic hypotheses on the source of race differences in IQ.

**Racial Admixture Designs**

The most direct test of genetic versus environmental hypotheses comes from studies of racial admixture and its relationship to IQ. Social classifications of race have traditionally outweighed biological ones in the United States; individuals with any visible trace of African ancestry have typically been labeled black. As a result, the black population in the United States is racially heterogeneous, with an estimated 20% to 30% of European (“white”) ancestry (Reed, 1969). The percentage of white ancestry of individual blacks would range from nil to over 90%. This interindividual variation in ancestry could be expected to have an effect on the distribution of IQ. Disregarding heterosis effects and any race differences in dominance variance, a simple genetic hypothesis would predict that an increase in white ancestry beyond the black mean of about 22% would produce a proportional reduction in the genetic component (8 to 11 points in the estimate by Jensen, 1973a, p. 363) of the mean IQ difference. An environmental

---

2 It is worth noting that the only substantial interracial adoption study of recent times (Scarr & Weinberg, 1976) produced equivocal results with regard to IQ. White adoptees did not differ significantly from adoptees with one white and one black parent, but both scored significantly higher than adoptees with two black parents.
The hypothesis would predict that so long as blacks with different proportions of white ancestry do not live in systematically different environments, racial admixture will have no relationship to IQ.

Racial admixture has been assessed in three ways: biochemical assays (of blood and protein markers), genealogical records, and visible racial indicators such as skin color. There are problems associated with all three. Biochemical techniques for estimating proportions of individual ancestry derived from different ancestral populations are in their infancy. Reed (1973) has shown that to produce even moderately reliable point estimates of ancestry (i.e., of the ancestral racial mix of individuals), much more information would be needed on population differences in blood group frequencies than is currently available. Studies that use biochemical estimates of racial admixture have therefore had to use measures that are weaker than actual point estimates. Genealogical records, on the other hand, are notoriously unreliable; it is never clear how much weight can be accorded to any results based on them. Visible racial markers such as skin color would seem to be the easiest and most reliable to use, but are actually the least suitable, for both genetic and environmental reasons. The genetic reason is that most of the variance in skin darkness in U.S. blacks is due, not to proportion of white ancestry, but to factors within the African ancestry (Harrison & Owen, 1964). The environmental reason is that skin color has always been socially relevant within the black community, although the social desirability of lighter versus darker skin appears to have lessened and in some cases reversed in recent years (Udry, Bauman, & Chase, 1971). For both reasons the many older studies (reviewed by Shuey, 1966) that found a low positive correlation between IQ and skin lightness are useless as sources of evidence for genetic racial differences in IQ. They provide, perhaps, stronger evidence on the social consequences of having lighter skin and thereby point to the desirability of statistical control of visible race markers in conjunction with biochemical or genealogical assessment of racial admixture.

Fairly complete reviews of racial admixture studies are given by Loehlin et al. (1975). Most of them yield no usable results, mainly because of the unavailability of reliable measures of racial admixture. I will focus on two studies that, although not completely satisfactory in themselves, go much of the way toward providing a design that would yield strong evidence.

The Scarr, Pakstis, Katz, and Barker study.

The most ambitious racial admixture study is one carried out by Scarr, Pakstis, Katz, and Barker (1977), on samples of black and white Philadelphia schoolchildren. As an approximation to point estimates of ancestry, they used the distribution of 43 different blood markers to calculate two kinds of odds coefficients for their subjects. These were measures of the similarity of the individual's blood group markers to those in the black and white samples overall (sample odds) or to those in the black and white ancestral populations (ancestral odds). Then, in the black sample, they correlated both kinds of odds coefficients with scores on four standard cognitive tests, the first principal component derived from correlations among the four tests (used as an estimate of g, or general intelligence) and a test of paired associate learning. The results showed a marginally significant tendency (p < .10, two tailed) for the "more black" (on the sample odds) to score higher on the paired associates test. However, there were no significant correlations of either odds coefficient with any of the cognitive tests or with the principal component.

Simple significance tests can be used to assess the fit of these results to genetic and environmental hypotheses, if we know what the hypotheses are. The environmental hypothesis predicts a zero correlation between the odds coefficients and the cognitive measures. What about the genetic hypothesis? Jensen (1981b) stated that the genetic hypothesis predicts a correlation of only .039, on the basis that there was relatively little variation in African ancestry in the black sample. Scarr (1981) replied that Jensen's prediction is invalid, as it is based on a misreading of her statements on the amount of variation in African ancestry. Using Jensen's estimates of correlations between the odds coefficients and African ancestry and between African ancestry and IQ, she calculated that the genetic hypothesis predicts a correlation of .245 between the odds coefficients and the cognitive measures. The results certainly could not differentiate between predicted correlations of .039 and 0, but they can easily differentiate between predicted correlations of .245 and 0.

The simplest way to combine the results from the four cognitive tests is to use their first principal component as the measure of their common cognitive content. The first principal component correlated -.03 and -.05 with the ancestral and sample odds, respectively. These correlations are not significantly different from an environmental prediction of zero (p > .5 in each case) but are significantly different from a genetic prediction of .245 (p < .02 in each case). Thus, the correlation of the principal component with the odds coefficients appears to be consistent with an environmental hypothesis and inconsistent with a genetic one.

Alternatively, one can combine the probabilities associated with each correlation of odds coefficients and cognitive tests, with a technique devised by R. A. Fisher and discussed by Rosenthal (1978).
This technique is usually used to combine the results of independent studies but can be used here to combine the results of the four independent measures of cognitive ability. It tests the hypothesis that the overall trend of the correlations is significantly different from an environmental prediction of zero, or from a genetic prediction of .245. The outcome is similar to that for the principal component measure. The correlation of the four cognitive tests with each odds coefficient do not significantly diverge from zero ($p > .3$ in each case) but do significantly diverge from .245 ($p < .001$ in each case). Again, therefore, the correlational measures overall appear to be consistent with an environmental hypothesis and inconsistent with a genetic one.

At least they are inconsistent with a genetic hypothesis that predicts a correlation of .245 between the odds coefficients and the cognitive tests. The status of that prediction must be dubious, however, because of the uncertain status of the odds coefficients. The correlation of the odds coefficients with the variable of interest, proportion of white or of African ancestry, is unknown (although Scarr accepted Jensen's estimate of .49). More important, the mean and variance of the proportion of white ancestry in the black sample are also unknown, and the odds coefficients do not yield estimates of them. Jensen's variance estimate might be unreasonably low, as Scarr maintained, but she was unable to provide a definite estimate of her own. Without such information, the genetic prediction is indeterminate, and the results of studies such as that of Scarr et al. must remain open to conflicting interpretations.

The Witty and Jenkins study. In their reviews of studies on race differences in IQ, Loehlin et al. (1975) and Flynn (1980) both placed considerable emphasis on a study of 63 black Chicago schoolchildren by Witty and Jenkins (1936). In fact, the study cannot support a heavy weight of interpretation, but it does have some uniquely relevant features in its design. Witty and Jenkins did not compare the IQs of blacks with different proportions of white ancestry. Instead, they focused on the distribution of ancestry in a selected high IQ sample of black students. Their rationale was that "the hypothesis that Negroes are inferior to whites in mental ability" should generate the prediction that "Negroes who make the very highest scores on mental tests should be those who come from admixtures predominantly white" (Witty & Jenkins, 1936, p. 180). They therefore estimated their subjects' racial admixture, on the basis of parent interviews, as N (all Negro), NNW (more Negro than white), NW (equally Negro and white), or NWW (more white than Negro) and compared the distribution of reported ancestry to that found in a national sample of adult blacks by Herskovits (1930). There were no significant differences between the two distributions, and the small differences that existed were not in a consistent direction. They concluded that their findings provided fairly strong evidence against the "hypothesis of Negro inferiority" (p. 191).

In fact, however, these findings do not provide such evidence, even if the genealogical estimates are taken at face value. For Witty and Jenkins's findings to have any value as evidence, it is essential that their comparison sample (Herskovits's national sample) be an appropriate one. An appropriate sample would be one that represented the population from which the children were drawn, that is, a comparison sample of black Chicago schoolchildren. This point is not mere nit-picking; in Reed's (1969) figures, blacks in Chicago have less white ancestry than the average for the black population in the U.S. If the same was true in 1936, then a Chicago sample matched to a national sample would have more white ancestry than the local average. More important, Herskovits's national sample cannot be considered representative. About 32% of his sample was composed of Howard University undergraduates, and another 16% was taken from the "well-to-do and professional portion of the population of the Harlem district of New York City" (Herskovits, 1930, p. 5). Thus, almost half of Herskovits's national sample was highly selected for scholastic achievement or SES. A genetic hypothesis for race differences in IQ is not embarrassed by the finding that such a sample had a similar distribution of ancestry to a smaller sample that was highly selected for IQ. A genetic hypothesis is not actually supported by this finding, because it remains an open question whether the national sample itself had more white ancestry than the average for the black population. It is clear, however, that it stands, the Witty and Jenkins study yields no interpretable results.

Racial admixture studies: Conclusion. The study of racial admixture in United States blacks provides a potentially powerful tool for assessing genetic and environmental hypotheses about the source of race differences in IQ. Because of the complexities involved in measuring racial admixture, however, none of the studies so far carried out can be considered strong enough to lead to an unequivocal conclusion. Nevertheless, some of the elements of the designs of racial admixture studies can be given a central place in the design of stronger and more unambiguous studies, as I will try to show.
Interpretation of any racial admixture studies, however, rests on the assumption that the white ancestors of modern blacks were not selected (positively or negatively) for IQ. If the white ancestors were genetically superior in IQ to both the white and the black average, then a positive correlation between IQ and proportion of white ancestry would not be evidence that the mean racial difference was genetic. Conversely, if the white ancestors were genetically equal in IQ to the black average, then, as Loehlin et al. (1975) pointed out, there are two possibilities:

Either (1) the U.S. slave and white populations did not differ appreciably in their average genetic intellectual potential, or else (2) the populations did differ as a whole, but there was selection of whites who mated with blacks such that the white partners were of about the black population average in genetic intellectual potential. (p. 131)

In the second case, the absence of a correlation between IQ and proportion of white ancestry would not be evidence that the mean racial difference was environmental. Loehlin et al. did not consider the second case to be very likely. They commented that existing demographic evidence suggests that selection on both sides has characterized black-white marriages in this century; that is, the white and black partners tend on the average to be below the white and above the black population mean on educational and occupational criteria. . . . Projection of this tendency back to black-white matings in the period of U.S. slavery would probably be unjustified, however. (p. 131)

Nevertheless, Centerwall (1978) has argued that because the second possibility cannot be definitively ruled out, racial admixture studies can throw no light on the source of race differences in IQ. He wrote that a lack of correlation only demonstrates that there was no significant difference in genetic IQ between ancestral blacks and ancestral miscegenous whites. To complete the syllogism, it is necessary to demonstrate that there was no significant difference in IQ between ancestral miscegenous whites and other whites, or, if there was, to demonstrate that the difference was due to environmental rather than genetic causes. Since most of the principals are dead and the historical data almost non-existent, neither demonstration is possible. (Centerwall, 1978, p. 238).

As a result, he concluded, "Any inferences (from admixture studies) are scientifically invalid" (p. 237).

Centerwall did not offer the "negative white selection" hypothesis as a serious hypothesis with respect to IQ but rather as a logical or epistemological dilemma. Nevertheless, because of the potential importance of racial admixture studies, it is likewise important to evaluate the strength of Centerwall's argument. In fact, his conclusion is overly restrictive.

It is certainly true that the possibility of negative white selection cannot be definitively ruled out. But as philosophers of science from Duhem (1914/1954) to Lakatos (1970) have stressed, for any set of scientific findings there are always alternative explanations that cannot be definitively ruled out (see Mackenzie, 1977, ch. 5, for discussion). It is for this reason that scientific knowledge is always to some extent provisional and that scientific revolutions are possible. The alternative explanations can be assessed, however, on the grounds of their independent cognitive content, their ability to generate empirically verifiable predictions, and their conformance with known mechanisms.

On these criteria, a hypothesis of negative white selection fares badly. Clearly, it does not have much independent cognitive content. There is no independent way of estimating how much lower genetic potential the ancestral whites might have had; that estimate can only be tied to whatever current estimates are offered as to the lower genetic potential of blacks. If the selection was more than was needed for that purpose, there would be a negative correlation between IQ and proportion of white ancestry. If it was less than was needed, there would be a (reduced) positive correlation. It is only if the negative selection of whites was exactly enough to offset the genetic inferiority of blacks, within the limits of modern error of measurement of racial admixture, that the negative selection hypothesis could come into play. The ad hoc character of the hypothesis is apparent. Relatedly, the ability of the hypothesis to generate empirically verifiable predictions is negligible or nonexistent. The hypothesis could serve only to explain away certain outcomes of admixture studies and would not have any separate empirical implications. Again, its ad hoc nature is evident.

With regard to its conformance with known mechanisms, the negative selection hypothesis fares worst of all. If we assume a narrow heritability for IQ in whites of .65 (which is a high estimate for narrow heritability) and a genetic racial difference of around 8 IQ points (the lower end of Jensen's 1973a estimate), then the ancestral whites would have needed a mean IQ of around 85. Lower heritabilities or greater genetic racial differences would, of course, require an even lower IQ for the ancestral whites. On average, therefore, the ancestral whites would have had a mean IQ lower than that of 85% of the white population. However, sexual selection

4 The genetic variance for a trait is equal to the heritability times the phenotypic variance. Thus, if the phenotypic variance for IQ is 225 (15²), the genetic variance will equal .65 \times 225 or 146. The genetic standard deviation is thus \sqrt{146} = 12 IQ points. If the between-race genetic difference is 8 IQ points, then the
is not typically made on the basis of IQ as such, but on the basis of associated characteristics such as occupational status, income, geographical location, and so on. Because these characteristics generally have only a modest correlation with IQ, any given degree of selection for IQ would have to be based on a much more stringent selection for the associated characteristics.

One of the highest reported correlations between IQ and any associated characteristic is given by Ball (1938): a correlation of .71 between IQ and occupational status. If we generously take this figure as an estimate of the correlation between IQ and all the sexual selection variables that operated in the white population, we find (from a normal curve) that the ancestral whites must have been, on average, selected from the lowest 18% of the white population on the selection variables. The types of variables that come to mind as candidates are lack of a stable family structure; low education, income, and occupational status; unemployment (providing plenty of free time); and so on. One can imagine fairly unpleasant social scenarios in which these would be sexual selection variables. However, because the greater portion of the white contribution to the gene pool of United States blacks occurred during the time of slavery (Wirth & Goldhamer, 1944), the actual selection variables would have included ownership of or access to (mainly female) slaves, sufficient affluence to support an informal liaison, and so on. Wirth and Goldhamer review historical records on the extent and distribution of miscegenation and its frequency in those who would have scored high on these selection variables. The idea that the white ancestry of blacks would have been drawn from the bottom 5th of the white population on these variables can be dismissed as derisory, as can the idea that these socioeconomic measures would have had a strong negative correlation with IQ.

Even if we disregard the associated variables, we are left with the task of accounting for how the ancestral whites could have had an average IQ lower than that of 85% of the population. Because the slave-owning class in the South had a relatively high proportion of university graduates and systematically engaged in the breeding of light-skinned slaves (who were more valuable than dark-skinned ones; again, Wirth & Goldhamer, 1944, have the details and the references), the hypothesis that the ancestral whites were genetically disadvantaged in IQ seems clearly untenable.

I conclude, therefore, that the hypothesis of negative white selection can be dismissed on the grounds that it is ad hoc, unsupported by evidence, and historically implausible. The evidence from racial admixture studies, present and future, can therefore be taken seriously.

Racial Crossing Designs: The Eyferth Study

A simple alternative to making biochemical or genealogical estimates of racial admixture is to compare the IQs of mixed-race children (those with one black and one white parent) with the IQs of black and of white children. A genetic hypothesis for race differences in IQ predicts that the IQ scores of mixed-race children will fall midway between the scores of their parent populations. An environmental hypothesis predicts that, to the extent that the environments of the mixed-race children are similar to those of one parent population or the other, the IQs of the offspring will tend toward the mean of the appropriate parent population. For this design to be workable it is necessary to focus on mixed-race children raised outside the relatively distinct racial environments of the United States. There, the tendency to assimilate mixed-race children into the black cultural environment would minimize the value of any comparisons. Black–white environmental differences outside the United States would not, however, be controlled for in this design.

The only substantial study along these lines is one by Eyferth (1961) that compared mixed-race and white children in a "white" environment. Eyferth obtained the IQ scores of (usually extramarital) children fathered by black and by white American (and in some cases French) occupation troops in Germany after World War II. From the total pool he selected comparison samples of white and mixed-race children (i.e., children with white and with black fathers), matched for age, sex, SES, family characteristics, racial mix of community, and numerous other social variables. The matching of environmental conditions could not be considered perfect because racial prejudice was widespread in Germany. Nevertheless, the children did not experience the differentiating effects of assimilation to putatively different black and white cultures. Furthermore, the differentiating effects of specifically racial prejudice were mitigated by the widespread prejudice against "American lovers" and illegitimate children in general. Eyferth found that the mean IQ (measured with the German version of the WISC) of the mixed-race children was 96.5 (girls 96, n = 89; boys 97, n = 81). The mean IQ of the white children was 97.2 (girls 93, n = 33; boys 101, n = 36).

The IQ difference between the children with white fathers and those with black is small, but the ancestral whites must be \( \frac{2}{3} \) of a standard deviation below the white genetic mean. This figure divided by the heritability gives their distance from the white phenotypic mean in standard deviation units, \( SD = 1.02 \), corresponding to 15 IQ points or a score lower than that of 85% of the white population.
results can be analyzed more precisely to contrast the genetic and environmental hypotheses. Let us assume that the standard deviation of IQ in each Race × Sex sample was 15, and that the black fathers had on average 80% black ancestry. The mixed-race children would thus have had, on average, 40% black ancestry. Taking the hypothesized genetic component of the mean racial difference to be 8 points (again, the bottom of Jensen's estimated range), and assuming no environmental addition to it, a simple genetic hypothesis would predict the mixed-race children to have a mean IQ 3.2 points lower than the white children. This predicted difference is small, but assessable. The mean IQ of the mixed-race children was significantly higher than this predicted mean \( t(169) = 2.17, p < .02 \) but was not even close to significantly different from the mean IQ of the white children \( t = 0.33, \) ns.\(^5\)

Eyferth's results thus appear consistent with an environmental hypothesis and inconsistent with a genetic one. There are, however, three possible objections to this conclusion. First, the black soldiers might have been a select group for IQ: American soldiers were routinely screened for minimum intellectual competence before being posted overseas. However, Flynn (1980), on the basis of very detailed examination of army records, concluded that the selection effect could have reduced the mean black-white IQ difference only to a minor degree. Loehlin et al. (1975), on the basis of a smaller body of published data, agreed that the IQ difference between the black and the white occupation troops remained around 1 standard deviation. Second, sexual selection effects could have occurred, in that either the black soldiers or their white German partners could have had IQs higher than average for their respective groups. Because no data are available on the IQs of any of the parents, this possibility cannot be definitively ruled out. However, Flynn (1980) reviewed army records showing that sexual liaisons with German women were extremely widespread among the occupation troops, with frequent sexual contact amounting to promiscuity slightly more common among soldiers with the lowest IQs. On the other side, for sexual selection of the mothers to offset genetic differences in the fathers, there would have to have been a strong and systematic pattern of higher IQ women mating with the black soldiers and lower IQ women mating with the whites. No such pattern was noted or has been proposed. Third, there was one difference between the white and mixed-race children; in both groups the males scored higher than the females, but the difference was significant only in the white children \( t(67) = 2.21, p < .04 \). Loehlin et al. (1975) considered this difference similar to the frequent United States finding, for blacks only, of a significant sex difference favoring females:

This has often been attributed to cultural factors, such as the greater prevalence of matriarchal families in the black community, but the presence of such differences in the present study suggests that a broader range of hypotheses, including genetic ones, may need to be considered. (p. 128)

This claim, that sex differences in one direction in whites should be considered equivalent to sex differences in the opposite direction in blacks, seems an odd one. On the German standardization of the WISC, there are no systematic sex differences. It was thus the white children who showed an anomalous sex difference, not the mixed-race ones. The distribution of sex differences, therefore, has no clear relevance to the results concerning race differences.

The Eyferth study provides, therefore, jointly genetic/environmental evidence that again tends to support an environmental over a genetic explanation for race differences in IQ. Because of its unique social circumstances, it cannot stand alone but joins the Scarr et al. (1977) study of racial admixture in providing evidence that tends to favor an environmental rather than a genetic hypothesis.

**Extensions to Genetic/Environmental Designs**

None of the studies described in this section is strong enough to warrant any firm conclusions on the source of race differences in IQ, even though their results are in marked agreement. The Scarr et al. (1977) study is handicapped by the uncertain status of the odds coefficients and by the unknown variance in racial admixture in the subject sample. The Witty and Jenkins (1936) study is vitiated by the use of an inappropriate comparison sample and by the unknown validity of the genealogical estimates of admixture. The Eyferth (1961) study is limited by its unique social circumstances and, perhaps, by the anomalous sex differences in the white sample. Nevertheless, all three studies can readily be extended, through the use of existing methods and technology, to the point where they would provide evidence that could be strong enough to settle the question.

The study that could most easily support the greatest degree of improvement is the one by Witty and Jenkins (1936). Although the methodology is naturally outdated, the idea of selecting a high IQ black sample and comparing its racial admixture with that of a control sample is a good one. The resolving power of such a design is higher than that of one that selects samples with different racial admixtures and compares them on IQ. A modern

---

\(^5\) The value assumed for the standard deviations is not critical. Any value between 10 and 20 would yield similar results.
replication of Witty and Jenkins's study would ignore Herskovits's national sample and compare the racial admixture of a high IQ sample of schoolchildren with that of a matched sample of average IQ children from the same school districts. Such a comparison could readily be made now with biochemical rather than genealogical estimates of admixture. Biochemical techniques can be applied here without undue difficulty because they are much more powerful in estimating the average admixture of groups than of individuals. In one of Reed's (1969) samples, for instance, his admixture estimate had an acceptably low standard error of .034 with a sample size of 179. These figures, furthermore, are based on data from only one gene locus, although the one showing the single greatest frequency differential between blacks and whites (for discussion see Cavalli-Sforza & Bodmer, 1971, ch. 8). Using as full a range of differentiating loci as Scarr et al. (1977) did, one should be able to reduce the required sample size to less than 100.

Group estimates of admixture could also be used to resolve the problems of the Scarr et al. (1977) admixture study. Within a large sample, iterative procedures could be used to assemble several groups differing in average level of admixture. The between-group differences could then be used to provide an estimate of admixture variance in the total sample, and the odds coefficients for individuals could be validated by assessing their ability to assign individuals to the appropriate group. This strategy, in fact, could be applied to a reanalysis of the Scarr et al. (1977) data, without testing additional subjects.

The limitations of the Eyferth (1961) study could be resolved simply through replication. As Loehlin et al. (1975) have noted, there are several other populations comparable to the one sampled by Eyferth. Comparisons of black-fathered and white-fathered children of American soldi ers could be carried out in Korea, Japan, and several other locations. Although the social circumstances of each sample would be unique, similar results across samples would lessen the need for caution in interpreting the results from each sample singly.

Further jointly genetic/environmental research along lines such as these should permit a fairly clear answer to the question of whether race differences in IQ between the black and white populations in the United States have mainly an environmental or mainly a genetic origin.

**Conclusion: The Limitations of the Question**

Suppose, then, that we do the research. Suppose that the results confirm those that we already have, so that we can confidently conclude that the differences are predominantly environmental. The question must then be put: How much more will we know? We will still not have an explanation for race differences in IQ. All we will have is an assurance that the explanation lies in one large domain rather than (or more than) in another. Is that rather limited conclusion worth all the time and effort it will have taken? What, if anything, will follow from the conclusion? If little or nothing follows, then is the question of the genetic or environmental origin of race differences even a legitimate scientific question? This concluding section will try to answer these more general questions.

Whether race differences in IQ have mainly a genetic or mainly an environmental origin is certainly a legitimate scientific question; any question that can be answered through scientific means is a legitimate scientific question. It is, nevertheless, an extremely limited one. It is a good example of what Lewontin (1974) called an “analysis of variance” question, a question about the apportioning of variance (into genetic and environmental sources) on a particular trait in a particular population in a particular set of environments at a particular time. It is a limited question because the answer to it will provide no insights into the causal pathways involved, will be of unknown generality (i.e., with regard to the range of environmental conditions in which it is valid), and will have no implications for how or even whether existing race differences can be modified or overcome. Kempthorne (1978) addressed these limitations in more detail. These questions are presumably the important ones, both scientifically and socially, but are left untouched by the question—regardless of the answer it receives—of “how much genetics versus how much environment.” That question has become a prominent one because it has become spuriously linked with the struggle of United States blacks for social equality, and with a number of ideological positions about the implications and prerequisites of social equality. Apart from these external relationships, the origin of race differences in IQ has few, if any, direct scientific implications and no immediate ones for social practice.

The question of the detailed causal explanation for race differences in IQ is of course a much more important question, both scientifically and socially. Knowledge of the mechanisms and influences involved would do much to further our understanding of the determinants of intellectual level in general and would at last provide an informed basis for social policy decisions. However, if this question can be freed from the intellectually polarized and ideologically loaded issue of genetic racial differences in intellectual capacity, it can be seen in a much more benign light. It becomes not simply a question about race differences but a part of the broader question...
of the determinants of IQ or intellectual level in general. Genetic as well as environmental factors will certainly enter the explanation here, whether or not they play a part in the explanation of mean racial differences as such.

Identifying the determinants of intellectual level involves much more than apportioning variance into genetic and environmental components. It involves an experimental and natural–environmental search for causal influences. It is part, therefore, of what Lewontin (1974) called the “analysis of causes” as distinguished from the “analysis of variance.” It is here that the detailed environmental hypotheses investigated by many researchers on race differences come into their own. The factors identified by these researchers—nutritional, economic, cultural, motivational, and so on—were often found to be related to race differences in IQ. Many of these factors may be among the causes influencing intellectual level, and identifying their contribution can help answer questions that focus on specific causes. The study of these factors was not sufficient to answer the analysis of variance question of the origin of race differences, because they did not account for all the between-race variance or, as with the correlational studies of Mercer and Brown (1973) and Blau (1981), did not adequately rule out the influence of genetic factors that might have been linked with the sociocultural ones. Residual unexplained variance therefore acted as an embarrassment to environmental hypotheses and an invitation to genetic ones. The analysis of causes question, however, does not demand a complete answer from the outset. Studies that identify some of the causal influences are very worthwhile on their own. Controlled intervention studies on the effects of nutritional supplements, cognitive enhancement programs, and so on are thus valuable, not to the extent that they completely account for racial differences, but to the extent that they identify some of the environmental variables implicated in the mean racial difference and important in cognitive development. Similarly, the Mercer and Brown (1973) and Blau (1981) studies are important, not because they nearly eliminated the mean racial differences in their original samples (where the results could have been confounded by associated genetic factors), but because they identified clear-cut candidates for major environmental variables. The influence of these variables, as purely environmental factors, can subsequently be assessed in samples in which a genetic confound can more easily be ruled out, such as samples of black and white adoptive children. It is conceivable that, in such samples, these candidate variables will account for the mean racial difference in IQ as convincingly as they did in the original samples. But the value of such a study will not depend on achieving exactly that result. Again, that is, it will depend not on how convincingly it eliminates a mean racial difference but on how much insight it provides into causal factors influencing cognitive development and IQ level of both black and white children.

To some extent, therefore, studies of these social, cognitive, and other factors had the wrong focus when they were directed toward answering the analysis of variance question. Instead, their natural connection is with empirical and theoretical models of the development and expression of intelligence. These models, whether they have a psychometric, Piagetian, or information-processing base, are attempts at specifying the processes involved in intelligent performance. Studies of how these processes depend on specific environmental influences, and on specific genetic ones as well, can thus help to explain race differences in intellectual level as a part of the task of accounting for the development of intellect generally.

A shift in emphasis from the analysis of variance question to the analysis of causes questions does not imply any downgrading of behavior genetic or other methods that emphasize the apportioning of variance. The idea that such a downgrading would follow seems to stem from a common but erroneous interpretation of a classic article by Anastasi (1958). Anastasi maintained that relatively unproductive questions about the proportions of genetic and environmental variance in a trait were due to be replaced by scientifically more interesting questions about the actual mechanisms of genetic and environmental influences. Anastasi underestimated the hold that analysis of variance questions would have on psychologists for another quarter of a century. Otherwise, her analysis was cogent and her examples productive. She did not suggest that proportions of variance attributable to specific genetic and environmental sources were scientifically uninteresting or unobtainable, although an uncritical reading of her statements on “interactionism” might have suggested such a view. As Plomin, DeFries, and Loehlin (1977; cf. Plomin, DeFries, & McClearn, 1980) have shown, Anastasi’s interactionist perspective and the aim of measuring the influence of specific environmental factors are often better served through using behavior genetic models than any others. Although the particular models and techniques differ in different cases, what is common to them is specifying an area of environmental or genetic influence precisely enough that the variance attributable to that influence, rather than to the whole domain, can be estimated. Pinning down those areas of influence requires input from researchers who have been most concerned with detailed causal hypotheses, on both the environmental and the genetic side, as well as
from those who have been mainly concerned with behavior genetics as such. It is clear that this shift in emphasis would be scientifically desirable. What has prevented it so far has not been any lack of willingness or ability on the part of psychological researchers to propose and test detailed causal hypotheses. It is rather that, in part for the social and ideological reasons I have already touched on, these hypotheses have been directed toward answering the wrong questions. They have been directed toward the analysis of variance question that they cannot answer until a complete explanation is possible, rather than toward the analysis of causes questions that they can answer bit by bit. Resolution of the analysis of variance question, through the powerful means that are appropriate to it, may thus facilitate a shift of emphasis to the detailed and scientifically more fundamental causal questions.

The resolution of the analysis of variance question of the overall source of race differences in IQ, limited though that question is, will thus be valuable and productive in a number of ways. It will be a scientific achievement of major importance in its own right, regardless of what specific implications it may have, simply because of the amount of effort and controversy it has attracted. It will enable detailed psychological research on the determinants of intellectual level to be applied more productively to the causal questions such research is better able to answer. Finally, it will do its part in reducing the ideological divisions and acrimonious debates that arise, within the psychological community as well as outside it, over all the questions dealing with race and IQ.

REFERENCES


Mackenzie, B. (1979). Regression to different population means as evidence for genetic racial differences in IQ. Australian Psychologist, 14, 216.


