Structural Equation Modeling. Probably the most rigorous methodology presently available to test the default hypothesis is the application of structural equation modeling to what is termed the biometric decomposition of a phenotypic mean difference into its genetic and environmental components. This methodology is an extraordinarily complex set of mathematical and statistical procedures, an adequate explanation of which is beyond the scope of this book, but for which detailed explanations are available.[40] It is essentially a multiple regression technique that can be used to statistically test the differences in "goodness-of-fit" between alternative models, such as whether (1) a phenotypic mean difference between groups consists of a linear combination of the same genetic (G) and environmental (E) factors that contribute to individual differences within the groups, *or* (2) the group difference is attributable to some additional factor (an unknown Factor X) that contributes to variance *between* groups but not to variance *within* groups.

Biometric decomposition by this method requires quite modern and specialized computer programs (LISREL VII) and exacting conditions of the data to which it is applied—above all, large and representative samples of the groups whose phenotypic means are to be decomposed into their genetic and environmental components. All subjects in each group must be measured with at least three or more different tests that are highly loaded on a common factor, such as *g*, and this factor must have high congruence between the two groups. Also, of course, each group must comprise at least two different degrees of kinship (e.g., MZ and DZ twins, or full-siblings and half-siblings) to permit reliable estimates of WGH for each of the tests. Further, in order to meet the assumption that WGH is the same in both groups, the estimates of WGH obtained for each of the tests should not differ significantly between the groups.

Given these stringent conditions, one can test whether the mean group difference in the general factor common to the various tests is consistent with the default model, which posits that the between-groups mean difference comprises the same genetic and environmental factors as do individual differences within each group. The goodness-of-fit of the data to the default model (i.e., group phenotypic difference = G + E) is then compared against the three alternative models, which posit *only* genetic (G) factors, or *only* environment (E), or *neither* G nor E, respectively, as the cause of the group difference. The method has been applied to estimate the genetic and environmental contributions to the observed sex difference in average blood pressure.[40]

This methodology was applied to a data set[41] that included scores on thirteen mental tests (average *g* loading = .67) given to samples of black and white adolescent MZ and DZ twins totaling 190 pairs. Age and a measure of socioeconomic status were regressed out of the test scores. The data showed by far the best fit to the default model, which therefore could not be rejected, while the fit of the data to the alternative models, by comparison with the default model, could be rejected at high levels of confidence ($p < .005$ to $p < .001$). That is, the observed W-B group difference is probably best explained in terms of both G and E factors, while either G or E alone is inadequate, given the assumption that G and E are the same within both groups. This result, however, does not warrant as much confidence as the above *p* values would indicate, as these particular data are less than ideal for one of the conditions of the model. The data set shows rather large and unsystematic (though nonsignificant) differences in the WGHs of blacks and whites on the various tests. Therefore, the estimate of BGH, though similar to the overall WGH of the thirteen tests (about .60), is questionable. Even though the WGHs of the general factor do not differ significantly between the races, the difference is large enough to leave doubt as to whether it is merely due to sampling error or is in fact real but cannot be detected given the sample size. If the latter is true, then the model used in this particular method of analysis (termed the psychometric factor model) cannot rigorously be applied to these particular data.

A highly similar methodology (using a less restrictive model termed the biometric factor model) was applied to a much larger data set by behavioral geneticists David Rowe and co-workers.[42] But Rowe's large-scale preliminary studies should first be described. He began[42a,b] by studying the correlations between objective tests of scholastic achievement (which are substantially loaded on *g* as well as on specific achievement factors) and assessment of the quality