# 5 IQ similarity in twins reared apart: Findings and responses to critics

*Thomas J. Bouchard, Jr.*

Research on genetic influence on intelligence has a long and contentious history (Brand, 1993; Fancher, 1985; Kamin, 1974). Both the idea of a general factor of cognitive ability, Spearman's *g*, and the idea that genetic factors might be an important source of variance in cognitive ability have been continuously debated since they were first systematically expounded by Galton (1869, 1876). Reviews of Galton's books published in the *London Times* at the time of their appearance could, if slight changes were made, be published today. The debate on the nature of mental abilities and the influence of heredity on such abilities (as well as most other psychological traits) initiated by Galton continues unabated.

## The current status of *g*

There should be no doubt that the issues of the measurability of IQ and its usefulness are still controversial issues. Consider the following recommendation regarding the measurement of abilities and other psychological traits:

> Make explicit to everyone (pupils, parents, public and professionals of all kinds) that a person's abilities, activities, and attitudes cannot be measured. The public, especially, misperceive that hard data exist, and that test scores constitute these data. The public does not realize how quickly the point is reached where we do not know how to discriminate validly among people, but where data mislead us to think we do. This is what is meant by the myth of measurability [Tyler & White, 1979, p. 376].

This recommendation is from a report published by the National Institute of Education in the United States. I would be inclined to argue that such a claim (that abilities, activities, and attitudes cannot be measured) reflects an abysmal level of ignorance about psychometrics and the accomplishments of social and behavioral scientists over the last 100 years (Bollen, 1989; Ghiselli, Campbell, & Zedeck, 1981; Robinson, Shaver, & Wrightsman, 1991). However, R. W. Tyler, the lead author of the report, is a senior scholar with a distinguished career in education. He cannot be unfamiliar with the evidence. The reason for such a ludicrous claim, consequently, must lie outside the body of empirical evidence generated by social scientists. Indeed, if such a claim is true, then the physical sciences must also be in a dismal state, as there is good evidence that measurement in the social sciences is not nearly as poor, in comparison with that of the physical sciences, as many people believe (Hedges, 1987). The measurement of IQ is more precise and has been more fully explored for sources of artifact than any other construct in psychology (Barrett & Depinet, 1991; Gottfredson, 1986; Gottfredson & Sharf, 1988; Hartigan & Wigdor, 1989; Humphreys, 1992; Jensen, 1980).

Here is a second example:

> Because intelligence is not the objectively defined explanatory concept it is often assumed to be, it is more an obstacle than an aid to understanding abilities [Howe, 1990, p. 100].

This quote is by a distinguished professor of educational psychology at Exeter University in England. His claims, in my opinion, are also unsupportable by the evidence (Detterman, 1993; Matarazzo, 1992; Vernon, 1993).

A last example demonstrates that derisive, but unsupportable, comments are not the exclusive domain of educational psychologists:

> Spearman's *g* is not an ineluctable entity; it represents one mathematical solution among many equivalent alternatives. The chimerical nature of *g* is the rotten core of Jensen's edifice, and of the entire hereditarian school [Gould, 1981, p. 320].

These views can be contrasted with those of equally eminent scholars. John Carroll (1993), for example, has recently completed the most comprehensive survey of the factor-analytic literature ever published; he is confident that there is a factor of general intelligence and that it is influenced by genetic factors. More importantly he argues:

> In *The Abilities of Man*, Spearman (1927) developed what was probably the first formal theory of cognitive abilities, the so-called two-factor theory whereby any cognitive test was conceived to be "saturated" with a general factor *g* and a specific factor *s* unique to that test. . . . In the main, I accept Spearman's concept of *g*, at least to the extent of accepting for serious consideration his notions about the basic processes measured by *g* – the apprehension of experience (what might now be called metacognition) and the eduction of relations and correlates [pp. 636–7].

Nathan Brody (1992) has recently reviewed the substantive research on both theories and correlates of IQ measures and draws a similar conclusion:

> The first systematic theory of intelligence presented by Spearman in 1904 is alive and well. At the center of Spearman's paper of 1904 is a belief that links exist between abstract reasoning ability, basic information-processing abilities, and academic performance. Contemporary knowledge is congruent with this belief [p. 349].

Neither of these reviewers believes that the "problem of intelligence" is completely solved, but they do agree that considerable progress has been made. I agree with this evaluation. Stated in different terms I believe the descriptive problem – namely, the answer to the question, "At the level of phenotypic test scores derived from cognitive tests, what is the structure of human cognitive abilities?" – has been largely answered. Technical arguments about the proper method of rotation and so forth are simply irrelevant distractions. In assessing *g*, it simply makes little difference what method is used (Jensen, 1994). This is not to deny that more research will clear up many details. Second, I believe and will argue that genetic factors play a profound role in the determination of an individual's ultimate level of cognitive ability when that individual is reared under a normal range of circumstances (Scarr, 1992, 1993). The mechanisms or processes that control these outcomes remain largely a mystery, although we do have some clues and tentative theories (Bouchard, Lykken, Tellegen, & McGue, in press; Byrne & Whiten, 1988; Reed, 1984, 1990).

Why is there so much controversy over the construct of intelligence or *g*? Because it is one of the most important and powerful constructs in the armamentarium of psychology (Miller, 1984), and taking it seriously has immense repercussions. This argument has been brought to the fore with a vengeance by Herrnstein and Murray in their recent book *The Bell Curve* (1994). With very few exceptions, virtually any dimension of behavior scaled from the less valued end to the more valued end correlates positively with IQ (Jensen, 1980, chapter 8; Matarazzo, 1972, chapter 12). The correlations are modest, but they are seldom zero, and they are almost never negative. None of the correlations are high enough to allow one to conclude that *g* is an overall measure of goodness or human worth or anything else of the sort. The correlations are modest enough so that one can easily find "bad people" with high IQs – a common complaint against IQ tests. Such cases, no matter how often they are cited, do not constitute evidence sufficient to refute stable statistical trends. As I will show, one of the major blocks to advancing our understanding of these issues, both in psychology as a profession and in the public at large, is the abysmally low level of quantitative understanding in both populations. Verbal sophistry – bolstered by anecdotes, linked to emotional appeals, and buttressed by claims of evil intent – masquerade as explanations of embarrassing findings even though they cannot withstand the most elementary quantitative scrutiny. It does not seem to be very widely understood that virtually all these wordy arguments can, if they are sensible, be reformulated into quantitative arguments and evaluated. In order for them to be taken seriously, numbers must be attached! As I show in this chapter, when numbers are properly attached, the explanatory power of most of these arguments evaporates.

## Genetic influence on mental ability: Current status

Psychology no less than other human endeavors is subject to fads and fashions. The view that heredity is an important source of human individual differences has waxed and waned over the years (Degler, 1991; Richards, 1987), and these changes were often unrelated to the amount and quality of evidence available. It now seems clear that part of the problem was that psychology (and many other social sciences as well) (1) was wedded to the concept of testing the null hypothesis and testing for statistical significance, (2) lacked a systematic means of integrating data from multiple studies of different kin, and (3) failed to put the evidence and arguments into systematic quantitative form. These problems have now been largely solved. The null hypothesis is known, to put it lightly, to be "bunk" and the testing of statistical significance downright misleading in almost all instances where it is used (Cohen, 1994; Lykken, 1968; Meehl, 1990; Schmidt, 1994). Formal testing of substantive hypotheses via model fitting has now become the norm in behavior-genetic research (Neale & Cardon, 1992) and is likely to become widespread throughout psychology in the future (Schmidt, 1993). Model fitting can, like any other methodology, be abused. This approach to data analysis does, however, force researchers and critics alike to state their claims in testable forms. The failure to specify a model underlying a verbal claim reveals the claim for what it often is: an unsubstantiated assertion disguised as knowledge.

The most recent round in the long running debate about the importance and validity of IQ measures, as well as the debate about the influence of heredity on IQ, was launched by Arthur Jensen in a now famous article entitled, "How much can we boost IQ and scholastic achievement?" (1969). In this article, Jensen claimed on the basis of his review of the evidence that

> compensatory education has been tried and it apparently has failed. . . . Why has there been such uniform failure of compensatory programs wherever they have been tried? What has gone wrong? In other fields, when bridges do not stand, when aircraft do not fly, when machines do not work, when treatments do not cure despite all conscientious efforts on the part of many persons to make them do so, one begins to question the basic assumptions, principles, theories, and hypotheses that guide one's efforts. Is it time to follow suit in education [p. 2]?

Jensen then presented a systematic body of evidence to show that what had failed were two theories that continue to permeate American social science – namely, the *average-child concept* and the *social-deprivation hypothesis*. The average-child concept encompasses the belief that all children are basically equivalent in their capacity to learn and develop. Observed differences are due to their upbringing (socioeconomic status) and to other general social and/or idiosyncratic influences. The social-deprivation hypothesis is ancillary to the average-child hypothesis and asserts that children in minority groups and children of the poor are invariably less capable only because of the environmental deprivations that they experience as excluded groups.

After demonstrating that this theory had failed to explain the observed differences in IQ, Jensen argued that we should replace it with a Genetic-Diversity Theory of Individual Differences and its natural complement on the environmental side, a diversity of learning opportunities. Specifically, he asserted that

> if diversity of mental abilities, as of most other human characteristics, is a basic fact of nature, as the evidence indicates, and if the idea of universal education is to be successfully pursued, it seems a reasonable conclusion that schools and society must provide a range and diversity of educational methods, programs, and goals, and occupational opportunities, just as wide as the range of human abilities [p. 117].

The Genetic-Diversity Theory of Individual Differences proposed by Jensen is a continuation of the Galtonian model. It asserts that there are fundamental differences between human beings in their capacity to develop intellectual skills (IQ, special mental abilities) and most other characteristics. In Darwin's (1871) words:

> So in regard to mental qualities, their transmission is manifest in our dogs, horses and other domestic animals. Besides special tastes and habits, general intelligence, courage, bad and good tempers, etc., are certainly transmitted.

As with Galton, who had concerned himself with, among other things, individual differences, social class differences and race differences in ability, Jensen also addressed social-class differences and race differences. I forego discussion of race differences here as it is not germane to this chapter. Jensen argued that social-class differences were in part genetic in origin. Jensen's work set off a storm of protest (Hirsch, 1975), and since then a great deal of research has been carried out with the goal of refuting his claims. As I will show in this chapter, using the MZA (monozygotic, or identical, twins reared apart) data, and as I have shown elsewhere using the entire array of kinship data (Bouchard, 1993b; McGue, Bouchard, Iacono, & Lykken, 1993), the evidence for a large degree of genetic influence on individual differences in intelligence, as measured by IQ test scores, is now irrefutable. The evidence for genetic influences on SES (socioeconomic status) differences in IQ has also grown (Bouchard, 1976; Bouchard, Lykken, McGue, Segal, & Tellegen, 1990a). Jensen drew on the work of Barbara Burks (1938) among others. Scarr and Weinberg (1978) report, on the basis of their adoption study: "Burks estimated that genetic differences among the occupational classes account for about .67 to .75 of the average IQ differences among children born into those classes. Our studies support that conclusion" (p. 689). This interpretation of the evidence is reasonable even though there are studies that demonstrate SES effects in the context of extreme placement – namely, the French cross-fostering study (Capron & Duyme, 1989) and the French adoption study (Schiff & Lewontin, 1986). It is crucial to keep in mind that the average age of the children in the Capron and Duyme study is 14 years. As I will show shortly, common family environmental influences appear to attenuate to near zero as adulthood is reached (McGue et al., 1993). As McGue (1989) has pointed out, it will be interesting to see the results of follow-ups of the French adoption studies.

The relevance of Jensen's work to this chapter is that Jensen relied heavily on the studies of Sir Cyril Burt of identical twins reared apart. Jensen's work evoked a scathing review of the IQ literature by Leon Kamin (1974). Kamin was especially critical of Burt and is credited with exposing Burt as a fraud. The case against Burt has, however, weakened considerably in recent years (Aldhous, 1992; Fletcher, 1991, 1993; Joynson, 1990). In collaboration with like-mined colleagues, Kamin eventually went on to criticize the entire enterprise of behavior genetics (Lewontin, Rose, & Kamin, 1984). I will address here only the issue of IQ. In his 1974 book, Kamin concluded that "there exists no data which should lead a prudent man to accept the hypothesis that IQ test scores are in any degree heritable" (p. 1). This conclusion and his criticisms, particularly the criticism of the MZA studies, over the last 20 years have been widely repeated in introductory psychology texts and elsewhere. The most recent publication in which he is repeatedly quoted, demonstrating that he has not changed his mind, is the attack on behavior genetics published by *Scientific American* (Horgan, 1993).

It should be noted that, Kamin's criticisms not withstanding, there is a strong consensus among experts regarding the findings in this domain. Snyderman and Rothman (1987, 1988) carried out a survey of expert opinion about IQ tests, their meaning, and the nature–nurture controversy. The results show that experts agree with the conclusions drawn by Carroll and by Brody, cited earlier, regarding *g* and the findings regarding the heritability of IQ discussed in this chapter.

## A brief historical background on the study of MZAs

The origin of the study of twins adopted early in life and reared apart – the twin-reared-apart (TRA) method – is unknown. Although Francis Galton can be credited with introducing the twin method and the adoption method (Bouchard, 1993c), he never mentioned the TRA method, even though after the publication of his now famous paper on twins (Galton, 1876), one of his correspondents mentioned the existence of one such pair (Townsend, 1874–5).

The first systematic collection of quantitative data from a pair of monozygotic twins reared apart (Bessie and Jessie) was carried out by H. J. Muller who followed up on a more discursive treatment of the same pair by Popenoe (1922). It is of considerable interest that while Muller published his paper on this pair of MZA twins in the biologically oriented *Journal of Heredity* (1925), the study dealt primarily with psychological, not physical or medical, variables and was subtitled, "The extent to which mental traits are independent of heredity, as tested in a case of identical twins reared apart." Muller understood the value of such cases and articulated the logic of this experiment of nature even before the necessary statistical methods were developed to analyze properly the data collected from a series of such cases.

> Cases are required in which the identical twins are reared apart, under environments differing as much as those *commonly met with do, in order that we may gain an idea of the amount of effect of such environmental differences as distinguish separate families in a community* [italics added]. Not one such case has heretofore been systematically investigated by modern methods, as such cases are very rare. Each such case is extremely valuable, however, since in any one such case, if a mental trait is found which shows marked similarity in the two members of the pair, and wide diversity in other individuals, in spite of the fact that the environments of the twin members differed considerably in such features as would be most likely to influence the trait, it may be pretty safely concluded that the trait in question, when measured by the method used, is genetically narrowly determined, and is reliable as a genetic indicator; where on the other hand, great differences appear, it is highly probable that the latitude of genetic indetermination is great, that the character differences so indicated are largely non-genetic, and that some other method of observation or testing must be used for estimating the genes which may be concerned with such characters. The results, then, may indicate not only the amount of variation caused by environment in the trait measured, but also the reliability of the method of measurement used, for indicating genetic facts [p. 434].

Muller made a number of important methodological points that are apparently not widely understood 70 years later and still warrant discussion. I will return to them after I discuss the analytic methods currently applied to data gathered on such twins and summarize the previous findings.

## The quantitative analysis of twins-reared-apart data

The simplest way to conceptualize the quantitative evaluation of TRA data is via path analysis. Figure 5.1 shows three path diagrams, one for identical twins reared together, one for identical twins reared apart, and one for unrelated individuals reared together. For simplicity I have left out sources of variance that do not contribute to similarity. Path diagrams allow us to quantify our intuitive notions of influence and correlation. The notations are as follows: items in circles indicate underlying (latent) variables; items in boxes indicate measured phenotypes (scores) for the kinship indicated

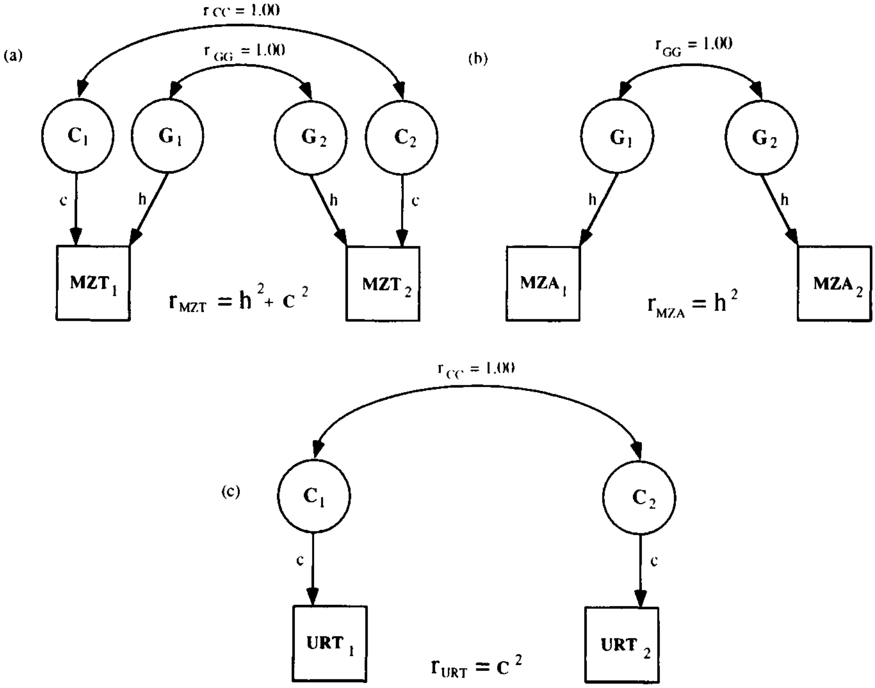

Figure 5.1. Path diagrams for (a) monozygotic twins reared together, (b) monozygotic twins reared apart, and (c) unrelated individuals reared together.

(e.g., $MZT_1$ is the score, on the trait under consideration, for the first member of a twin pair raised together); G = genotype; C = common (shared) environment; and h and c equal, respectively, genetic and shared-environmental path coefficients. Single-headed arrows denote causal influences, with the lower-case letters representing the degree to which the phenotypic standard deviation is a function of the variability in the latent causal entities. Double-headed arrows indicate correlations. Figure 5.1a diagrams the model for monozygotic twins reared together. We see that $G_1$ and $G_2$ are correlated 1.00, reflecting the identical genotypes of monozygotic twins – for DZ (dizygotic, or fraternal) twins, the correlation would be .5, thereby showing what we know from genetic theory, that on average they share 50% of their segregating genes. Both twins are influenced by G (genes cause similarity), and the magnitude of influence is indexed by the path coefficient h. Because we are discussing correlations, not covariances, the path coefficients in these models are standardized. We see that common environment, which by definition is correlated 1.00, also influences (causes similarity between) the twins, and its influence is indexed by the term $c$. The rules of path analysis allow us to estimate the influence of an underlying latent trait by multiplying the terms of the path. There are two paths in the case of MZ twins reared together (MZT). The first path is through the genes shown by the term $(h * 1.00 * h)$, or more simply $h^2$. The second path is through the common environment $(c* 1.00*c)$, or $c^2$. The correlation between MZT twins is the sum of these influences, or:

$$r_{\mathrm{MZT}} = h^2 + c^2$$

This equation formalizes our intuitive notion that MZ twins are alike because we recognize that two different factors can be the cause of their similarity – heredity and common environmental influences. For DZ twins, we can state on the basis of genetic theory that under certain assumptions (no nonadditive genetic variance and no assortative mating), the genetic influence is half that of MZ twins. These equations show that genetic and environmental factors are confounded when relatives are reared together (Bouchard & Segal, 1985; Scarr, chapter 1, this volume). The most common criticism of the TRA design is the argument that the design assumes that MZ and DZ twins experience similar, common family environmental influences – the so-called *equal environment assumption.* Consider the following quote from Lewontin et al. (1984):

> There are also some obvious environmental reasons to expect higher correlations among MZ than among DZ twins, especially when one realizes the degree to which an MZ pair creates or attracts a far more similar environment than that experienced by other people. Because of their striking physical similarity, parents, teachers, and friends tend to treat them much alike and often even confuse them for one another.... *There is no great imagination required to see how such a difference between MZs and DZs might produce the reported difference in IQ correlations* [italics added]. It is entirely clear that the environmental experiences of MZs are much more similar than those of DZs [pp. 115–16].

It is certainly true that MZ twins experience more similar environments than do DZ twins, but it is also true, if perhaps surprising, that no one has been able to show that such imposed similarities in treatment are trait-relevant. The critical assumption being made, when this argument is brought forward, is the trait relevance of the treatment. Loehlin and Nichols (1976) studied this problem using very large samples of twins. They related differences within pairs of twins to differences in treatments as reported by the twins' mothers. Consider the dressing-alike argument; it is often claimed that because MZ twins dress alike much more than do DZ twins, they are made more similar. Measured differences in dress were related to differences on the 18 California Psychological Inventory (CPI) Scales for 451 MZ twins. The average correlation was .004. The corresponding correlation with a composite measure of differential experiences (as reported by the mothers) was .056. These effects are obviously trivial. Statistically sophisticated readers may note that difference scores are notoriously unreliable and discount these findings on that basis. However, similar difference scores have been shown to be sufficiently reliable to capture artifacts in our own MZA analyses and yield correlations in the .60 range. In addition, the quantitative findings are replicable using alternate methods (Bouchard & McGue, 1990). A large number of studies have now been carried out on this problem (DeFries & Plomin, 1978; Kendler, Neale, Kessler, Heath, & Eaves, 1993; McCartney, Harris, & Bernieri, 1990; Rose, 1981; Rose, Kaprio, Williams, Viken, & Obremski, 1990; Rowe & Clapp, 1977; Rowe, Clapp, & Wallis, 1987; Scarr, 1968; Scarr & Carter-Saltzman, 1979; Scarr, Scarf, & Weinberg, 1980). Most of this evidence was available and had been brought to their attention prior to the time Lewontin et al. wrote their book. Nevertheless, the only citation on the subject they provide demonstrates that there are treatment differences, not that such differences are trait-relevant.

The TRA design largely overcomes the objection of a highly similar common rearing environment. If the twins are not subject to placement bias, a testable proposition I will discuss later, then they no longer share a common environmental source of similarity so that (see Figure 5.1b):

$$r_{\mathrm{MZA}} = h^2$$

Those with a psychometric background will recognize that this model is of the same form as the true-score model for test–retest and (more pertinently) parallel form reliability (Hayes, 1973). The unsquared correlation between the two forms represents the proportion of variance explained by the true scores. The MZA correlation similarly represents the variance explained by genetic influences (Bouchard, Lykken, McGue, Segal, & Tellegen, 1990b; Jensen, 1971; Miller & Levine, 1973).

The correlation between MZA twins estimates the *broad heritability* of a trait as opposed to the *narrow heritability*. The broad heritability includes all genetic factors that make MZA twins alike. These include nonadditive genetic factors (dominance, epistasis), which while genetic in origin are nontransmissible from parents to offspring (Lykken, McGue, Tellegen, & Bouchard, 1992). Methods that estimate the narrow heritability of IQ find a somewhat lower figure than methods that estimate the broad heritability, suggesting that nonadditive variance may be important for this trait (Pedersen et al., 1992). This distinction gives rise to the common practice of claiming that the heritability of IQ is between .4 and .8 (Herrnstein & Murray, 1994). For the kinds of samples ordinarily studied, the narrow heritability is probably between .4 and .6, and the broad heritability is, as I show in this chapter, around .75.

In other sciences, investigators go through a great deal of trouble to create efficient model systems for investigating a phenomenon. The goal is to create a system that gives the most direct and clearest answer to a question. In animal behavior genetics, the most obvious example is the widespread use of inbred strains of animals. The comparison of random samples of strains allows the investigator to hold heredity constant in order to allow investigation of the influence of various environmental manipulations. Conversely, different strains exposed to identical environments are compared to detect genetic influences. Of course, if animals respond to selective breeding for a behavioral trait, the evidence is even more conclusive (DeFries, Gervais, & Thomas, 1978). In human behavior genetics, monozygotic twins are the closest we can come to inbred strains of animals. Monozygotic twins reared apart combine an experiment of nature (twins) and an experiment of nurture (adoption). The intraclass correlation between MZA twin members is the most powerful and most direct way to estimate the broad heritability of a trait (Plomin, DeFries, & McClearn, 1990). The statistical power of this design is remarkable. For a trait with a heritability of about .50 (an estimate close to that found for many psychological characteristics), 50 pairs of MZA twins have roughly the same statistical power as 1,000 pairs (500 MZ and 500 DZ) of twins reared together – the heritability estimates have the same 95% confidence interval (Lykken, Geisser, & Tellegen, 1978).

An efficient design to detect both the broad heritability and the influence of common family environment is one that contains equal proportions of MZA and MZT twins (Eaves, 1970). A simple, powerful, and direct design for estimating common family environmental influences is the study of unrelated individuals reared together as siblings (URT); the path diagram for this design is shown in Figure 5.1c. As with the MZA correlation, the URT correlation is also a direct estimate of a parameter – in this instance common (shared) environmental influence. While it is understandable that, because of their rarity, MZA twins have been infrequently studied, it is a mystery why URTs have been studied so seldom (Scarr & Weinberg, 1994). Compared to twins reared apart, URTs are relatively common. It almost as though psychologists did not wish to collect data using a sample that would refute their favorite hypotheses.

## IQ findings from twins reared apart

Table 5.1 summarizes the entire world literature on the IQ correlations between twins reared apart, including recent data from the Minnesota Study of Twins Reared Apart (MISTRA) (Bouchard et al., 1990a) and the Swedish Adoption Study of Aging (SATSA) (Pedersen et al., 1992).

As shown by the path model for MZA twins, introduced in Figure 5.1b, the MZA intraclass correlation gives us what Muller called "an idea of the

Table 5.1. *Intraclass correlations, confidence intervals, sample sizes, and tests utilized for IQ in five studies of MZA twins*

| Study and Test used (Primary/Secondary/ Tertiary) | *N* for each Test | Primary Test | Secondary Test | Tertiary Test | Mean of Multiple Tests |
|---|---|---|---|---|---|
| Newman, Freeman, & Holzinger (1937) (Stanford-Binet/Otis) | 19/19 | .68 ± .12 | .74 ± .10 | | .71 |
| Juel-Nielsen (1980) (Wechsler-Bellevue/ Raven) | 12/12 | .64 ± .17 | .73 ± .13 | | .69 |
| Shields (1962) (Mill-Hill/Dominoes) | 38/37 | .74 ± .07 | .76 ± .07 | | .75 |
| Bouchard, Lykken, McGue, Segal, & Tellegen (1990a) WAIS/Raven-Mill-Hill First principal component | 48/42/43 | .69 ± .07 | .78 ± .07 | .78 ± .07 | .75 |
| Pedersen, Plomin, Nesselroade, & McClearn (1992) First principal component | 45 | .78 ± .06 | | | .78 |
| Weighted Average | | | | | .75 |

amount of effect of such environmental differences as distinguish separate families in a community." Kinship studies, like any kind of scientific study, allow us to generalize only to populations similar to the one sampled, in this case to the range of environments found in the community from which the MZA twins had been sampled. This well-known restriction has a long history, and Galton similarly restricted the range of his generalizations when he asserted:

> There is no escape from the conclusion that nature prevails enormously over nurture *when the differences of nurture do not exceed what is commonly to be found among persons of the same rank of society and in the same country* [italics added] [Galton, 1876, p. 576].

The goal of determining the magnitude of genetic and environmental influence on IQ should be clearly distinguished from the goal of determining the full reaction range of a trait (Turkheimer, 1991). I discuss the concept of reaction range in the next section.

The Minnesota Study of Twins Reared Apart has been very explicit regarding the range of environments to which its conclusions can be generalized.

> The IQs of the adult MZA twins assessed with various instruments in four independent studies correlate about 0.70, indicating that about 70% of the observed variance in IQ in this population can be attributed to genetic variation. Since only a few of these MZA twins were reared in real poverty or by illiterate parents and none were retarded, this heritability estimate should not be extrapolated to the extremes of environmental disadvantages still encountered in society. Moreover, these findings do not imply that traits like IQ cannot be enhanced [Bouchard et al., 1990a, p. 227].

It is often argued that the heritability statistic is uninformative because like any statistic it may vary from population to population and from one set of circumstances to another. I find this argument nonsensical because it directly implies that we should do away with all descriptive statistics. Furthermore, the implied claim that heritability varies greatly is an empirical one and can only be answered by obtaining estimates of this statistic in a variety of settings. We have reason to believe that genetic estimates are more generalizable than often claimed (Rushton, 1989), and the data in Table 5.1 confirm this conjecture for IQ. A second argument against its use is that heritability is misunderstood and that its use furthers that misunderstanding. I would simply argue that what is needed is more education, not less, and certainly not the suppression of statistics. If there is one important point being made in this chapter, it is that the implications of a wide variety of verbal arguments and claims regarding IQ are quantifiable and testable. The only way that we can resolve these disagreements is via statistical evidence, as the phenomena are inherently probabilistic (Bouchard, 1993c).

As Table 5.1 indicates, the weighted average of the MZA IQ correlations is .75. This figure should be compared to the reliability of the types of tests used in these studies, a figure that is unlikely to be above .90 (Parker, Hanson, & Hunsley, 1988). Clearly, the MZA method tells us that a very significant portion of the reliable variance in measured IQ is genetic in origin. How dependable and meaningful are these findings? I turn to these questions next.

### Reaction range – Genotype × environment interaction

It always bears repeating that IQ is a phenotype and the genotype is a biochemical code. Measurement of a phenotype presumes a previous developmental process. If that developmental process is nonlinearly related to the genotype, there may be complex interactions. The possibility of such interactions has been repeatedly put forward as a reason for not computing heritabilities (Feldman & Lewontin, 1975; Layzer, 1974; Lewontin, 1974; Wahlsten, 1990; and numerous commentators).

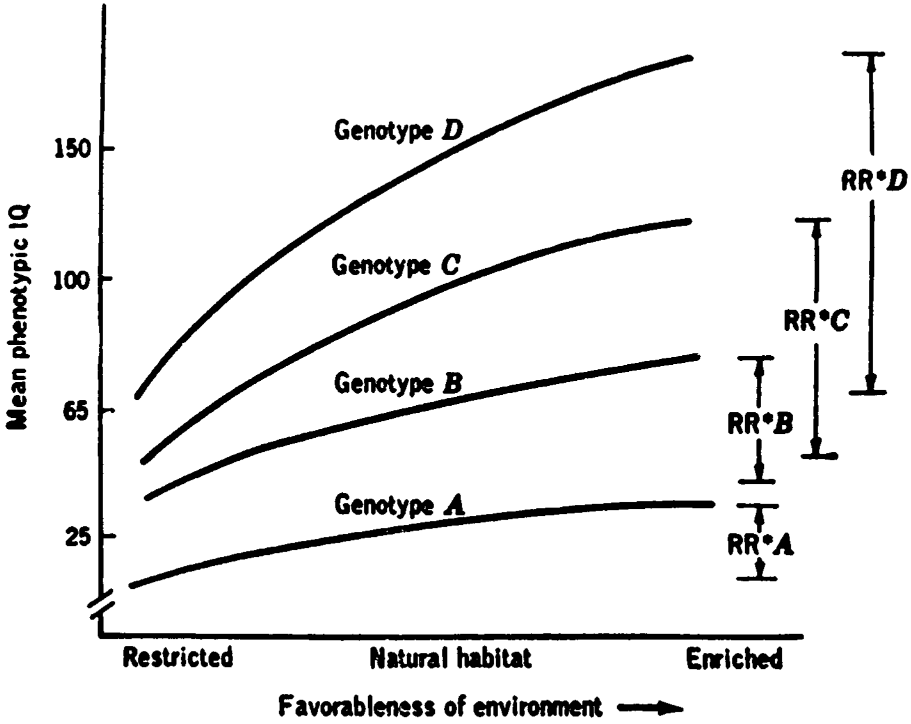

Figure 5.2. Scheme of the reaction-range concept for four hypothetical genotypes. Note: Marked deviation from the natural habitat has a low probability of occurrence. RR signifies reaction range in phenotypic IQ. (From Gottesman, 1963, p. 255.)

The concept of reaction range was initially introduced into the discussion of genetic influence on IQ by I. I. Gottesman (1963, 1968; Turkheimer & Gottesman, 1991). Gottesman's classic hypothetical reaction-range curves for IQ are shown in Figure 5.2.

Unlike the heritability statistic, which summarizes the bottom-line outcome for a population of individuals exposed to a range of environments, the reaction-range curve attempts to illustrate the degree to which genotypes have variable expression in different environments. The hypothetical curves in Figure 5.2 convey a number of important ideas. First, there is a strong genetic main effect and no disordinal interaction (i.e., a genotype keeps the same rank order under all environments, and none of the curves cross). There is an ordinal interaction (a fan-shaped spread): Genotype D responds much better to the enriched environment than Genotype A (the reaction range for each genotype is given at the far right). *One of the features of a reaction-range curve is that it specifically attempts to character-*

*ize the degree of expression of a trait for different genotypes under varying environments.* Gottesman's curves have the added advantage of pointing out that there is a range of environments which we tend to characterize as the *natural habitat* for a particular organism. Gottesman's range of environments could, of course, be extended to include what might be called, using the plant analogy, *hothouse environments* – environments in which every conceivable effort is made to enhance the trait of interest. It is possible, for example, under very special conditions to make a tomato plant grow into the size of a tree and produce enormous tomatoes. There is a lower tail to the environmental dimension as well. Without the right conditions, no organism will survive.

An examination of the extremes of the favorableness dimension is actually very informative. It quickly becomes clear that a single dimension of favorableness is misleading. At the low end, a tomato plant can die from drought, heat, excessive dampness, or frost. At the high end, a giant tomato plant needs support for its branches lest they break off; it also needs protection from the wind or else it will be blown over and uprooted; and any of these events can quickly lead to infection and death. This brings us to the issue of natural selection for organismic characteristics. One of the most important limiting selective forces that shape a species are the extremes of environment that it faces. Thus, a species of trees lives in a natural habitat where the most extreme winds experienced over long periods of time do not destroy all the trees, or the ground does not flood sufficiently frequently to drown the roots of all members of the species. Organisms in the natural world colonize ranges or niches that are close to the ones in which they evolved. There are interesting examples of organisms changing the environment to fit their needs. Eucalyptus trees in California pull enough moisture out of the fog with their leaves to make up for the lack of rain. The occasional freezes are, however, a continual threat to their existence and limit their spread.

Can we think of comparable examples for the trait of IQ? Clearly, there are environments so bad that they are incompatible with life. There are also environments so intellectually impoverished that mental growth is stunted, and there is a cumulative deficit in IQ (Jensen, 1977). If there are many individuals living in such environments, and they are sampled properly in an MZA study, the computed heritability will reflect this fact ($h^2$ will be higher if they are not sampled). It is widely agreed that such environments are undersampled by MZA studies (Bouchard et al., 1990a, p. 227), and in that sense, the figures overestimate the degree of genetic influence in the entire population. As I have pointed out elsewhere, however, in the Minnesota study, 90% of the population have IQs in the range we studied (Bouchard, Lykken, McGue, Segal, & Tellegen, 1991). What does not seem to be adequately appreciated, and is a direct implication of this rendition of the reaction-range curve, is that as environments get better, genetic differences generally become a more important source of individual differences than environment ($h^2$ becomes larger). It is also important not to underestimate what can be done at the highly favorable end of the environmental continuum. Unlike the dangers that exist for a tomato plant grown in a hothouse, and despite widespread beliefs to the contrary, there is no reason to believe that accelerating the educational progress of gifted children is detrimental to their well-being (Stanley, 1973). Nor is there any doubt that the opportunity for extensive practice is important in the development of high-level skills (Ericsson & Charness, 1994; Ericsson, Krampe, & Heizman, 1993).

If the reaction-range curves in Figure 5.2 were largely disordinal – that is, if different genotypes performed very differently as they moved along the environmental continuum and repeatedly crossed each other – this would complicate the computation of heritabilities. There are those who argue that this is likely to be the case, although they seldom refer to human IQ data. One particular example is often presented by Lewontin (Lewontin, 1982; Lewontin, 1975; Schiff & Lewontin, 1986, p. 172) and others (Byne, 1994). It involves demonstrating that it is possible to find phenotypic features of some organisms, in this case the Achillea plant, that interact in a disordinal manner with the environments in which they are raised. In their example, seven different genotypes, rank-ordered according to their height when grown at a low elevation (the Stanford University Botanical Garden), are then shown to differ in their rank order when grown at medium (California Foothills) and high elevations (Mountains of the High Sierra). After detailed scrutiny of the original source (Clausen, Keck, & Hiesey, 1940), I have been unable to locate the precise figure or set of data used by Lewontin. I am not implying bias here; he could easily and legitimately have combined data from a variety of tables and figures presented in the book. The problem is one of incomplete reporting. The figure most similar to Lewontin's figure is figure 122 (p. 310). That figure and the legend are reproduced here as Figure 5.3.

I would argue that this figure also shows, as Clausen et al. indicate in the legend, that there are ecotypes and ecospecies of the California Achilleas, and that the type chosen from a particular environment generally does best in its natural environment or one close to it. In addition, extremes kill. The Maritime plant (*A. borealis arenicola*) picked near sea level does best at sea level but quite well at medium elevation and does not even flower at Timberline. The Mid-Sierran race (1315–1) actually does better at sea level than Mather (where it was picked) but does not flower at Timberline. The High Sierran form (2459–1) appears to do well at all altitudes, but the

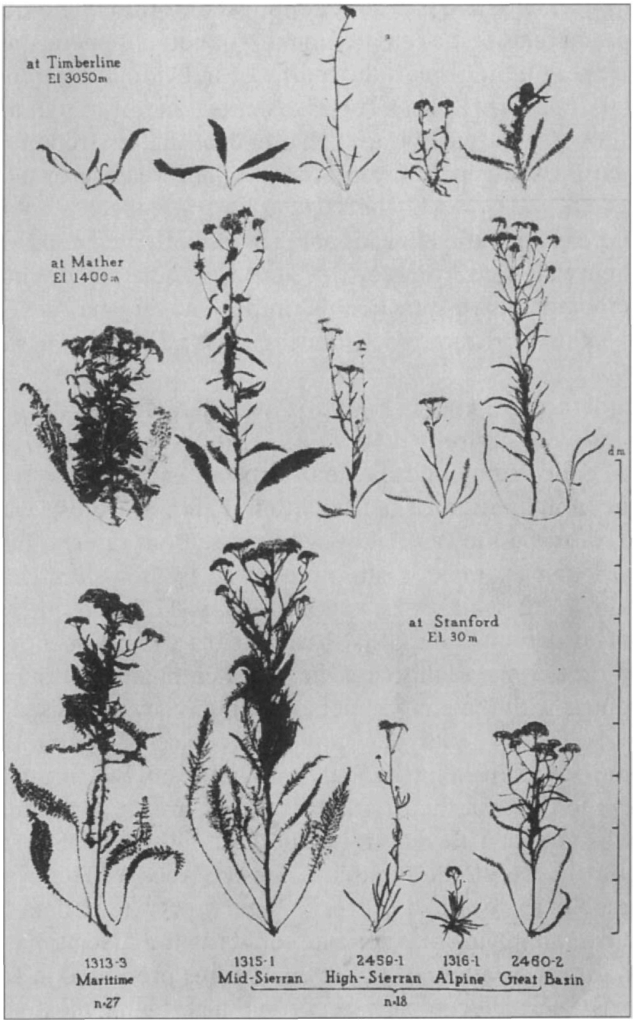

Figure 5.3. Modifications at three transplant stations in five clones representing altitudinal ecotypes and ecospecies of the California Achilleas. The lowest row consists of specimens grown at Stanford, the middle row at Mather, and the top row at Timberline. (From Clausen, Keck, & Hiesey, 1940, p. 310.)

Alpine form (1316–1) appears to do poorly at Stanford. The Great Basin Form appears to do best at its own altitude. We might be able to claim that forms selected for extremes do well in milder environments, but, of course, we have no idea if they would survive against competitors over time. In this study, natural selection has been removed. We can, of course, say that two of the types do not survive in the Timberline environment as they do not flower. Clausen et al. also show that plants that do flower in one season sometimes simply do not survive because in a subsequent season the frost occurs on an earlier date (cf. their figure 123). Interestingly, Clausen et al. point out that Achilleas has "pronounced individual differences in earliness (time of flowering) within the same ecotype" (p. 314). The fact is, however, that the kinds of disordinal effects emphasized by Lewontin in this plant are exhibited at the extremes (many do not survive more than a few years at Timberline), and we do not know what would happen under natural circumstances. The meaningful interactions may indeed be slightly ordinal. This is not to argue that this approach to the matter is uninformative. As Turkheimer and Gottesman (1991) point out, "In some contexts, it is perfectly reasonable to ask how individuals in the natural environment come to vary as they do; in others, it is reasonable to ask how they might vary if the environments were to be altered radically" (p. 19). An excellent discussion of this issue can also be found in Haldane (1946).

Another widely cited complex interaction should also be dealt with at this point. Cooper and Zubeck (1958) demonstrated that if two strains of rats (Bright and Dull) were raised in three different environments (enriched, normal for laboratory rats, and restricted) their performance (error rate) would yield a strong interaction. The primary finding is that both strains do poorly in deprived environments; they differ in the environment under which they were selected (as they should) but do not differ in enriched environments. Unfortunately, this article is extremely misleading, as the authors themselves admit that the results may be due to an artifact – namely, "the ceiling of the test may have been too low to differentiate the animals, that the problems may not have been sufficiently difficult to tax the ability of the brighter rats" (p. 162). It must also be mentioned that studies of environmental influence using inbred strains of animals, while of great interest theoretically, create serious problems with regard to generalizability to hybrid organisms, which almost all species of animals, including human beings, are. Inbred strains of animals are hardly representative of their own species (most originating lines do not survive the process of inbreeding) and are unduly sensitive to most environmental variation relative to hybrids. Hybrids are probably buffered from environmental influences (Hyde, 1973). These issues have been discussed in detail in the technical literature (Bouchard, 1993b, pp. 72–3; Crow, 1990; Falconer, 1990; Henderson, 1990; Hyde, 1974).

While the possible existence of complex interactions may make the analysis of the main effects of genes and environment a futile exercise (Feldman & Lewontin, 1975), it must be kept in mind that there is very little evidence for such effects on IQ even though a great deal of work has been carried out on the problem (Eaves, Last, Martin, & Jinks, 1977a; Eaves, Last, Young, & Martin, 1977b; Jinks & Fulker, 1970). The fact is that "everything in the world can be explained by factors about which we know nothing" (Urbach, 1974, p. 253). More to the scientific point, as Rao, Morton, and Yee (1974) have argued, "since armchair examples of significant interactions in the absence of an additive effect are pathological and have never been demonstrated in real populations, we need not be unduly concerned about interaction effects. The investigator with a different view should publish any worthwhile results he may obtain" (p. 357).

### The misuse of environment as an explanation of MZA similarity in IQ: Trait relevance and the partialling fallacy

Another common argument against the MZA method is that few of the twins are reared in extremely different environments. This argument is not relevant if one is attempting to describe the source of variation found in a specific community. An additional flaw that often accompanies this argument is that the environment can be characterized along one dimension (e.g., good ↔ bad). In point of fact, it is virtually certain that different environments are relevant to different traits (Muller's "features as would be most likely to influence the trait"). The term *trait-relevant environments* has been introduced to deal with this problem. Cases from the extremes of one trait-relevant environment will not necessarily be at the extreme of another trait-relevant environment. One can easily imagine a pair of twins who, while reared apart, both live in affluent homes with unlimited access to books, good education, and so on, but where one twin is loved and showered with affection while the other twin is abused and treated with scorn. We can only hope to capture these types of differences in a sample of MZA twins.

The authors of studies of MZA twins have repeatedly been accused of not adequately studying environments. They are accused, for example, of only examining crude indicators such as education of parent, socioeconomic status of family, family size, physical features of the environment, and fallible self-reports of child-rearing practices by parents. There is no question that this accusation is in part correct. Our reply, however, is that these features are measurable, and because they have often been put forward as explanations of individual differences in ordinary families, the validity of these claims must be tested in the context of an adoption design. Consider the recent discussions of SES and health (Adler, Boyce, Chesney, Cohen, Folkman, Kahn, & Syme, 1994), SES and achievement (White, 1982), and family size and IQ (Blake, 1989). All these authors fail to realize that the correlations they are discussing are confounded, or they hand-wave the possibility of genetic effects away. Adler et al. dismiss the genetic argument and fail to cite the most relevant competing paper that asserts a genetic explanation (West, 1991). It is our contention that in spite of years of concerted effort by psychologists, there is very little knowledge about the trait-relevant environments that influence IQ (Bouchard, 1993b; Jensen, chapter 2, this volume; Locurto, 1988, 1990, 1991) and ordinary personality traits (Bouchard, 1993a). This is not to assert that there are no findings in the environmental domain; rather, the findings are so inconsistent that it is necessary to appeal constantly to higher-order interactions (Wachs, 1992), which are notoriously difficult to replicate. Brand (1993) provides a trenchant critique of this position.

Finally, it is worth mentioning that many so-called environmental variables are not entirely environmental at all. Many of them often have a genetic component. As pointed out previously, SES differences in IQ are now known to have a significant genetic component (Scarr & Weinberg, 1978). Partialling out parental SES from a relationship that involves IQ and some other variable (e.g., occupational success) results in the removal of more genetic variance than environmental variance. In MISTRA, the following correlations between adoptive parental measures and participant's IQ were found: Father's education .10, Mother's education –.001, Father's SES .174. These correlations must be squared in order to estimate variance accounted for. When fit to an appropriate path model, these correlations, taken in conjunction with the degree of placement bias, accounted for only a trivial portion of the MZA similarity in IQ (Bouchard et al., 1990a, Table 3). The best estimate of the correlation between biological parent's SES and offspring IQ, based on a meta-analysis, is .33 (White, 1982). Clearly, a large part of this correlation is genetic in origin. The partialling out of genetic variance in the guise of equating for environmental differences is called the *partialling fallacy*, and it permeates the social science literature (Jensen, 1973; Meehl, 1970, 1971, 1978). It is embarrassing to point it out, but this problem has been well known for over 50 years (Burks, 1938). Recent embarrassing examples can be found in Hoffman (1991) and Tomlinson-Keasy (1990), and detailed criticisms can be found in Bouchard (1993a). Plomin (1994) provides an excellent review of the relevant literature showing that environmental variables are not always what they seem to be.

## Constructive replication

Muller was also alert to the problems of method specificity and reliability. He found that Bessie and Jessie were very much alike on two intelligence tests: the Army Alpha Test (Form 8, July 1918) and the Otis Advanced Intelligence Test (Form A, 1922). Their scores were 156 and 153 on the

Alpha and 64 and 62 on the Otis. Since both scores were high (very superior intelligence), it was less likely that the findings were due to chance than if their scores were in the middle range. The replication across tests (a version of constructive replication) also suggested that the findings were reliable. Nevertheless, the twins did not differ very much across their social backgrounds, so Muller felt that he could not draw sweeping conclusions from the IQ findings.

The findings for personality were quite different. The twins differed considerably on the measures used, the Pressey X-O tests and the Downey Individual Will-Temperament tests. The differences were in fact larger than the expected differences between randomly chosen individuals in the norm group for each test. None of these early personality tests has survived in the face of scientific advance in measurement. They simply did not prove to be reliable or valid enough in ordinary usage to become a part of the psychological armamentarium. Very different and much more reliable and valid methods of measuring personality have taken their place (Goldberg, 1971).

One of the striking findings from the IQ data in the MZA studies is the replicability of findings across studies, measures, countries, and cohorts. The studies span over 50 years, involve many different measures of IQ, took place in five different countries, and were conducted in three languages. All the settings were, however, modern industrialized societies. The findings from all kinship studies show similar robustness but are also limited to modern industrialized societies (Bouchard & McGue, 1981).

## A closer look at some previous criticisms

### The Farber Analysis

Susan Farber in the introductory chapter of her book *Identical Twins Reared Apart: A Reanalysis* (1981) argues, regarding the previous analyses of the MZA data, that

> my own evaluation, particularly of the allegedly scientific analysis made of the IQ data, is more caustic. Suffice it to say that it seems that there has been a great deal of action with numbers but not much progress – or sometimes not even much common sense [p. 22].

Farber is echoing the complaints of Leon Kamin (1974), but she and Howard Taylor, who is discussed in the next subsection, go about their debunking in a much more systematic manner, so I will deal with their analyses while recognizing that the results also apply to Kamin. The interested reader may also wish to examine the following reviews of Kamin's work (Bouchard, 1982b; Fulker, 1975; Jackson, 1975). Farber, following

Kamin, spends much of her time trying to demonstrate that various forms of contact and degree of separation account for the similarities. Bouchard reviewed her book in detail elsewhere (1982a), and only a few summary points from that review can be presented here. Regarding the statistical analyses, Bouchard concluded that

> the results seriously abuse statistical theory and reinforce the widespread belief that scientists can prove anything with statistics. In sum, the treatment of the IQ data is an exercise in obfuscation. Perhaps this new approach needs a name. I suggest the term "pseudoanalysis" [p. 190].

The most important reanalysis of Farber's data involves the cases that *she herself classifies as highly separated.* Bouchard's comments on this analysis, part of which, it should be noted, Farber did not carry out, are as follows:

> By this point I was persuaded that separation had little or no effect on similarity between twins. I decided to calculate intraclass *r*'s for the Highly Separated group for whom I had expected to find an analysis but had not. The results were surprising! For the entire group: $n = 39$, $r_i = .76$, mean = 97.42, SD = 14.28. For the females: $n = 26$, $r_i = .76$, mean = 97.96, SD = 14.29. For the males: $n = 13$, $r_i = .76$, mean = 96.35, SD = 14.20. The three arrays show the slight depression in IQ characteristic of most twin samples, a standard deviation comparable to the normative population, identical intraclass *r*'s that are indistinguishable from the full sample where separation is ignored [p. 191].

These twins admittedly constitute a modest subgroup, but it is large enough to address Taylor's argument regarding the myth of separated twins discussed in the next subsection. As a historical note, it is interesting to find that a full analysis of all the cases reported in Farber's book yield a correlation of .771 (Bouchard, 1982a), precisely the correlation that Kamin accused Burt of fabricating (Kamin, 1974). Farber's data set did not, of course, include Burt's twins. The correlations of .78 and .78 reported in the two recent replications (see Table 5.1) are surprisingly close to this figure as well.

To make it clear that Bouchard's very negative review of this book is not idiosyncratic, I cite a review by Loehlin (1981). After pointing out numerous errors in the reporting of birthweights, he states:

> A second aspect of the book is an elaborate statistical treatment of the IQ data from the separated MZ twin studies. Some interesting analyses are provided, but readers are hereby cautioned to watch out for the graphs and summaries in Chapter 7. These suggest that the amount of contact between separated MZ twins accounts for some 20–30% of the IQ variance. Perhaps, but only if one assumes that the mechanisms involved work in the opposite directions in males and females (see Appendix E, p. 350). For the sexes combined, the amount of contact between the twins does not predict their resemblance [p. 297].

Locurto (1991) and Brody (1992) have provided similar critical examinations of the Farber analysis.

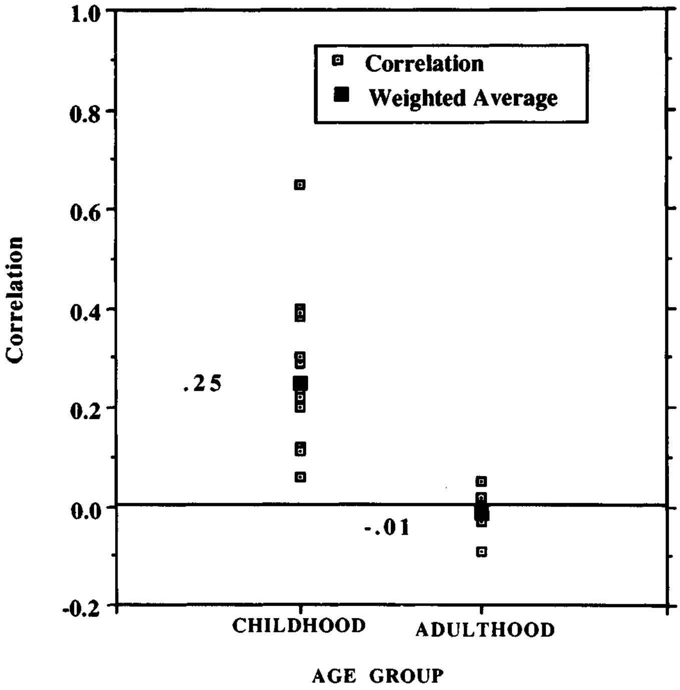

Figure 5.4. IQ correlations among nonbiologically related, but reared-together relatives (both adopted-adopted and adopted-biological pairs). Weighted average correlations were derived using the Fisher Z transformation method. (From McGue, Bouchard, Iacono, & Lykken, 1993, p. 67).

### *The Taylor Analysis*

Howard Taylor, again following Kamin, has also carried out a detailed analysis of the MZA data in his book *The IQ Game* (1980). The chapter dealing with this topic is entitled "The myth of separated identical twins." According to Taylor:

> The similarity in educational, socioeconomic, and interpersonal environments, referred to here as social environment, is a central reason why monozygotic twins regarded in the professional literature as separately raised reveal similar IQ scores. MZ twin pairs who have had similar social environment (such as similar schooling) have similar IQs, and twin pairs who have relatively different social environments (especially different schooling) have different IQs [p. 92].

None of these claims is true. Using Taylor's own classification of the twins, Bouchard showed that his findings simply did not replicate when they were tested with a different IQ measure independently obtained from the same sample (Bouchard, 1983). As indicated above, the analysis of Farber's highly separated sample also refutes this claim. Bouchard concluded:

> Taylor's conclusion that "it seems reasonable to suggest that the IQ correlations characterizing pairs of individuals with absolutely identical genes and absolutely uncorrelated environments would be extremely low" cannot be substantiated from the evidence at hand [p. 175].

These findings were provided to Taylor prior to submission of the paper for publication, and he did not comment on them. They have yet to be refuted.

The failure to find any influence from the kinds of biasing variables explored by Farber and Taylor extends to both SATSA (Pedersen et al., 1992) and MISTRA (Bouchard et al., 1990a). Neither study has been able to find significant effects on IQ due to placement on these types of variables. Interestingly enough, had Farber's and Taylor's findings been replicable, they would have been an anomaly. We now know that unrelated individuals reared in the same home show no similarity whatsoever if their IQs are measured in adulthood. The results of those studies are shown in Figure 5.4.

Data gathered on unrelated individuals reared together (URTs) and measured as children does show an effect. This influence apparently fades with time and disappears in adulthood. Only one study in Figure 5.4 is longitudinal. We clearly need more studies of this sort. In any event, it is clear that if URT individuals show no similarity in IQ in adulthood after having lived together for years, it is not a surprise that the various measures of environmental similarity used in the MZA studies have failed to explain the similarity in MZA twins.

## A more detailed analysis of a commonly hypothesized mediating mechanism: Physical attractiveness

Ford (1993) has argued that the "reported striking physical similarity of MZAs for facial features, height, weight, gait, posture and voice . . ." might explain the personality concordance in MZAs. Ford goes on to assert that

> there is a plethora of research indicating that physical features and attractiveness strongly influence how others react to us (Alley, 1988b; Bull & Rumsey, 1988; Feingold, 1992). The literature that examines the related question of the association of personality traits with physical features and attractiveness is less extensive; however, there is evidence of a connection (Borkenau, 1991; Melamed, 1992) [p. 1294].

This criticism has been put forward numerous times at public presentations of the MISTRA findings. It tends to be applied indiscriminately to our findings for abilities, personality, vocational interests, and social attitudes. It is worth pointing out that I have no a priori objection to this or any other explanation that involves mediating mechanisms. Such a process may well be involved in the development of various traits, and it is the goal of psychological research to explicate them. They cannot, however, simply be proposed with a wave of the hand without a close examination of their quantitative implications. Indeed, the adherents of this alternative explanation (physical attractiveness) and others like it do not appear to appreciate its complexity and implausibility. In words, a simple version of the causal model is as follows:

> Parent makes judgments about their child's physical attractiveness → they consistently treat the child in some manner → this treatment influences the child's traits (regardless of other attributes which we know are only weakly correlated with physical attractiveness) → physical attractiveness comes to be correlated with a whole host of traits.

I have used a parent–child model because we know that IQ stabilizes early in life and that spouses have a trivial influence over each other's personality over the course of a marriage (Caspi & Herbener, 1993). In order for MZA twins to become correlated, a second set of parents must carry out the same process in the same manner. We also must assume that some of these parents have the capacity to push a trait like IQ as low as 79 and as high as 130 (the range of IQs in MISTRA). I leave it to the reader to decide if such an argument is plausible. If this particular model is unacceptable, simply state another one; nevertheless, the same constraints will apply. Note that we need not assume that the model explains all of the similarity between MZA twins. Even a modest amount of explained variance would make physical attractiveness an interesting variable. In virtually any model of this process, however, physical attractiveness, or whatever trait is considered important, is expected to correlate eventually with the dependent trait (IQ, personality, etc.).

What is the actual evidence regarding physical attractiveness? Let us begin with the meta-analysis of the correlations between physical attractiveness and various traits cited by Ford. Feingold (1992) reports that the experimental literature demonstrates that physically attractive people, as opposed to physically unattractive people, are *perceived* as more dominant, intelligent, mentally healthy, sexually warm, and sociable. In addition, *self-ratings* of physical attractiveness are positively correlated with a wider range of attributes than objectively rated physical attractiveness. The correlational literature, where physical attractiveness is determined *objectively* (not self-ratings), indicates generally trivial relationships. The findings (sample size, median and mean correlations for each trait) taken from Feingold's (1992) table 6, are as follows: Sociability ($N$ = 1,710, .00, .04), Dominance ($N$ = 2,858, .04, .07), General mental health ($N$ = 2,597, .02, .05), Self-esteem ($N$ = 4,942, .04, .06), Internal locus of control ($N$ = 3,683, .00, .02), Freedom from loneliness ($N$ = 430, .04, .15), Freedom from general social anxiety ($N$ = 1,155, .06, .09), Freedom from heterosexual anxiety ($N$ = 1,539, .19, .22), Freedom from public self-consciousness ($N$ = 578, –.20, –.18), Freedom from self-absorption ($N$ = 746, .00, –.08), Freedom from manipulativeness ($N$ = 252, .03, –.01), Social skills ($N$ = 1,050, .25, .23), Popularity ($N$ = 982, .04, .08), Intelligence ($N$ = 3,497, .00, –.04), and Grades ($N$ = 3,445, .07, .02). Feingold concludes, as one might expect, that "good-looking people are not what we think." I would add, given these results, that "the influence of attractiveness on stable personality traits is unlikely to be what some people thought." It is well worth repeating that there is a powerful physical-attractiveness stereotype. People believe that attractive people are at the high end of most of these traits. These data are, of course, not dispositive regarding the influence of differential treatment due to physical attractiveness in childhood and adolescence, the periods during which most traits coalesce. They do, however, throw considerable doubt on the idea that physical attractiveness is an important determinant of any of the traits studied. Note that I am not asserting that physical attractiveness does not influence any social behavior and attitudes. The correlations for Social skills, Freedom from heterosexual anxiety, and Freedom from public self-consciousness are higher than for other traits. In addition, the comparable figures for Noncoital sexual experience are ($N$ = 1,167, .16, .13) and for Global sexual experience are ($N$ = 1,896, .18, .18). These findings suggest that physical attractiveness has a very modest and a very narrow and specific influence on sexual behavior and attitudes. Other evidence supports this conjecture (Mazur, Halpern, & Udry, 1994).

How about the other references cited by Ford? The Bull and Rumsey (1988) book is in our opinion a thorough, critical, and scholarly work. Nevertheless, it does not deal with the topic under discussion in any straightforward way; moreover, it fails to support the view that Ford would like to foster – namely, that facial appearance is an important determinant of personality and thus significantly influences twin similarity. Each chapter in the Bull and Rumsey book has an excellent summary paragraph, and the skeptical reader is urged to look at each of them. The conclusions reported in the book were such a surprise that I report selected results in the following numbered paragraphs, where I simply cite enough to convey the flavor of the findings.

1. Regarding facial appearance in liking and dating (chapter 2), the authors argue that in the 1960s and 1970s, attractiveness was seen to play a major role. After a review of that literature, they argue that the experimental conditions were very artificial, and while subsequent studies often found a statistically significant effect, "the power or strength of effect was rarely mentioned" (p. 39). They go on to argue that while facial appearance does play a role in marital selection, "no consistent relationship between facial appearance and marital adjustment has been found, although there is the suggestion that facial appearance may play a contributory role" (p. 39).

2. Regarding facial appearance in Persuasion, Politics, Employment,

and Advertising (chapter 3), we find that vis-à-vis advertising, "there is little evidence that actual behavior is affected" (p. 79).

3. Regarding facial appearance in the criminal justice system (chapter 4), the authors conclude that with respect to attributions of responsibility for crimes, the literature is "replete with contradictory findings" (p. 120).

4. Regarding facial appearance and education (chapter 5), the authors conclude, like Feingold, that it is easy to find effects, in this case expectation effects, by using studies that manipulate photographs. "However, the evidence that such expectations have any meaningful resulting effects is much weaker, several studies having found rather limited or no such effects" (p. 150).

The Alley (1988b) book contains little not covered by Feingold (1992). The most relevant chapter is chapter 8, written by the editor (Alley, 1988a). He can speak for himself:

> Is facial structure related to personality? Folklore, literary characterizations, and requests for photographs from job applicants, not to mention the practice and promulgation of physiognomy for centuries, indicate that such relationships exist. Scientific research, however, has generally found little or no validity in physiognomy. For instance, in a relatively recent and thorough study, Cohen (1973) "Found it impossible to discover any meaningful relations – even through use of multiple correlations – between physiognomic and psychological characteristics, which could maintain their statistical significance in cross validation on other data: (p. 107)" [p. 172].

Now let's look at the two empirical papers that Ford cites, Borkenau (1991) and Melamed (1992). I quote Borkenau's abstract:

> Self-report personality correlates of wearing glasses were investigated. To control for possible effects of social stereotypes on self-reports of personality, judgments by strangers were also collected. The trait that perceivers inferred from spectacles differed from the self-reported traits that actually co-occurred with the presence of glasses. Thus a substantial influence of social stereotypes on self-reports of personality was not reasonable [p. 1125].

The abstract from Melamed (1992) reads as follows:

> Physical height was correlated with the 16 PF. Height was significantly related to suspiciousness for both sexes, and to dominance and independence for males [p. 1349].

The correlations with suspiciousness were .29 and .27. As the authors point out, "The correlation found with assertiveness and independence followed expectation. Yet, it is hard to explain the correlation with suspicion and the lack of significant correlation with self-assurance (factor O)" (p. 1349). With results like these, I believe that it might be wise to wait for a few replications.

Rowe (1994, p. 47) presents a path model to represent how weak the influence of physical attractiveness is on twin similarity for personality under unlikely assumptions (correlations between attractiveness and personality are much higher than reported by Feingold). More conclusive, however, is the empirical evidence. Rowe et al. (1987) showed that controlling for attractiveness does not remove the similarity of MZ twins reared together. In addition, two other studies (Matheny, Wilson, & Dolan, 1976; Plomin, Willerman, & Loehlin, 1976) showed that twins who were more alike in personality were not rated more alike in appearance. Burks and Tolman (1932) long ago failed to find such an effect with siblings.

I have devoted a great deal of space to this topic because it is prototypic of the kinds of explanations that are repeatedly bandied about in attempts to explain the similarity of MZA twins. Evidence is needed to support these explanations, and it seldom supports the claims that are made. On numerous occasions, I have found that these explanations are assented to (by well-trained psychologists) within seconds of their being proposed, and the parties wander away with a self-satisfied look that indicates they believe that psychology has again provided a simple and powerful explanation of striking phenomena when it has done no such thing. I note that this is a very general problem in psychology (Dawes, 1994). The striking similarity of IQ and other traits in MZA twins may be mediated by environmental processes of this sort, and I have suggested what such processes might look like (Bouchard et al., 1990a; Bouchard et al., in press). Such theories might, however, also be wrong. Potential IQ, personality, and other psychological traits may simply reside in the brain with a wide range of nonspecific environmental influences being sufficient to mediate their development. Such theories must be tested and shown to be in accord with the best evidence available, not ordinary intuition.

## Conclusions

As far as I am aware, no plausible alternative to genetic influence exists to explain the IQ similarity in monozygotic twins reared apart. Since these findings are highly consistent with heritability estimates from other adult kinships, also collected in similar settings (Bouchard et al., in press; McGue et al., 1993), I conclude that genetic factors are the predominant source of variation in adult measured intelligence in modern Western societies.

## Acknowledgments

The Minnesota Study of Twins Reared Apart has been supported by grants from the Pioneer Fund, the David H. Koch Charitable Trust, the Seaver Institute, the Spencer Foundation, the National Science Foundation (BNS-7926654), the University of Minnesota Graduate School, and the Harcourt Brace Jovanovich Publishing Co. I would like to thank my colleagues

Margaret Keyes, Auke Tellegen, and David Lykken for numerous suggestions that greatly improved the manuscript. The errors that remain are mine.

## References

Adler, N. E., Boyce, T., Chesney, M. A., Cohen, S., Folkman, S., Kahn, R. L., & Syme, S. L. (1994). Socioeconomic status and health: The challenge of the gradient. *American Psychologist, 49,* 15–24.

Aldhous, P. (1992). Psychologists rethink Burt. *Nature, 356,* 5.

Alley, T. R. (1988a). Physiognomy and social perception. In T. R. Alley (Ed.), *Social and applied aspects of perceiving faces.* Hillsdale: Erlbaum.

Alley, T. R. (1988b). *Social and applied aspects of perceiving faces.* Hillsdale, NJ: Erlbaum.

Barrett, G. V., & Depinet, R. L. (1991). A reconsideration of testing for competence rather than for intelligence. *American Psychologist, 46,* 1012–24.

Blake, J. (1989). Number of siblings and educational attainment. *Science, 245,* 32–6.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Borkenau, P. (1991). Evidence of a correlation between wearing glasses and personality. *Personality and Individual Differences, 12,* 1125–8.

Bouchard, T. J., Jr. (1976). Genetic factors in intelligence. In A. R. Kaplan (Ed.), *Human behavior genetics.* Springfield, IL: Charles C. Thomas.

Bouchard, T. J., Jr. (1982a). [Review of "Identical twins reared apart: A reanalysis."] *Contemporary Psychology, 27,* 190–1.

Bouchard, T. J., Jr. (1982b). [Review of "The intelligence controversy."] *American Journal of Psychology, 95,* 346–9.

Bouchard, T. J., Jr. (1983). Do environmental similarities explain the similarity in intelligence of identical twins reared apart? *Intelligence, 7,* 175–84.

Bouchard, T. J., Jr. (1993a). Genetic and environmental influences on adult personality: Evaluating the evidence. In I. Deary & J. Hettema (Eds.), *Basic issues in personality.* Dordrecht: Kluwer Academic Publishers.

Bouchard, T. J., Jr. (1993b). The genetic architecture of human intelligence. In P. A. Vernon (Ed.), *Biological approaches to the study of human intelligence.* Norwood, NJ: Ablex.

Bouchard, T. J., Jr. (1993c). Twins: Nature's twins told tale. In T. J. Bouchard, Jr., & P. Propping (Eds.), *Twins as a tool of behavior genetics.* Chichester, England: Wiley & Sons Ltd.

Bouchard, T. J., Jr., Lykken, D. T., McGue, M., Segal, N. L., & Tellegen, A. (1990a). Sources of human psychological differences: The Minnesota study of twins reared apart. *Science, 250,* 223–8.

Bouchard, T. J., Jr., Lykken, D. T., McGue, M., Segal, N. L., & Tellegen, A. (1990b). When kin correlations are not squared. *Science, 250,* 1498.

Bouchard, T. J., Jr., Lykken, D. T., McGue, M., Segal, N. L., & Tellegen, A. (1991). IQ and heredity. *Science, 252,* 191–2.

Bouchard, T. J., Jr., Lykken, D. T., Tellegen, A. T., & McGue, M. (in press). Genes, drives, environment and experience: EPD theory – revised. In C. P. Benbow & D. Lubinski (Eds.), *Psychometrics and social issues concerning intellectual talent.* Baltimore: Johns Hopkins University Press.

Bouchard, T. J., Jr., & McGue, M. (1981). Familial studies of intelligence: A review. *Science, 212,* 1055–9.

Bouchard, T. J., Jr., & McGue, M. (1990). Genetic and environmental influences on adult personality: An analysis of adopted twins reared apart. *Journal of Personality, 58,* 263–92.

Bouchard, T. J., Jr., & Segal, N. L. (1985). Environment and IQ. In B. J. Wolman (Ed.), *Handbook of intelligence: Theories, measurements, and applications.* New York: Wiley.

Brand, C. R. (1993). Cognitive abilities: Current theoretical issues. In T. J. Bouchard, Jr., & P. Propping (Eds.), *Twins as a tool of behavioral genetics.* Chichester, England: Wiley.

Brody, N. (1992). *Intelligence* (2nd ed.). San Diego: Academic Press.

Bull, R., & Rumsey, N. (1988). *The social psychology of facial appearance.* New York: Springer-Verlag.

Burks, B., & Tolman, R. (1932). Is mental resemblance related to physical resemblance in sibling pairs? *Journal of Genetic Psychology, 40,* 3–15.

Burks, B. S. (1938). On the relative contributions of nature and nurture to average group differences in intelligence. *Proceedings of the National Academy of Sciences, 24,* 276–82.

Byne, W. (1994). The biological evidence challenged. *Scientific American, May,* 50–5.

Byrne, R. W., & Whiten, A. (Eds.). (1988). *Machiavellian intelligence.* Oxford: Clarendon.

Capron, C., & Duyme, M. (1989). Assessment of effects of socio-economic status on IQ in a full cross-fostering study. *Nature, 340,* 552–4.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Caspi, A., & Herbener, E. S. (1993). Marital assortment and phenotypic convergence: Longitudinal evidence. *Social Biology, 40,* 48–60.

Clausen, J., Keck, D. D., & Hiesey, W. M. (1940). Experimental studies on the nature of species. I. Effects of varied environments on western North American plants. *Carnegie Institution Washington Publication No. 520,* 1–452.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49,* 997–1003.

Cohen, R. (1973). *Patterns of personality judgment.* New York: Academic Press.

Cooper, R., & Zubeck, J. (1958). Effects of enriched and restricted early environments on the learning ability of bright and dull rats. *Canadian Journal of Psychology, 12,* 159–64.

Crow, J. F. (1990). How important is detecting interactions? *Behavior and Brain Sciences, 13,* 126–7.

Darwin, C. (1871). *The descent of man and selection in relation to sex.* London: John Murray (New York: Modern Library, 1967).

Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth.* New York: Free Press.

DeFries, J. C., Gervais, M. C., & Thomas, E. A. (1978). Response to 30 generations of selection for open-field activity in laboratory mice. *Behavior Genetics, 8,* 3–13.

DeFries, J. C., & Plomin, R. (1978). Behavior genetics. *Annual Review of Psychology, 29,* 473–515.

Degler, C. N. (1991). *In search of human nature: The decline and revival of Darwinism in American social thought.* New York: Oxford University Press.

Detterman, D. K. (1993). Giftedness and intelligence: One and the same? In G. R. Bock & K. Ackrill (Eds.), *The origins and development of high ability.* Chichester, England: Wiley.

Eaves, L. J. (1970). The genetic analysis of continuous variation: A comparison of experimental designs applicable to human data. II. Estimation of heritability and comparison of environmental components. *British Journal of Mathematical and Statistical Psychology, 23,* 189–98.

Eaves, L. J., Last, K., Martin, N. G., & Jinks, J. L. (1977a). A progressive approach to nonadditivity and genotype–environment covariance in the analysis of human differences. *British Journal of Mathematical and Statistical Psychology, 30,* 1–42.

Eaves, L. J., Last, K. A., Young, P. A., & Martin, N. G. (1977b). Model-fitting approaches to the analysis of human behavior. *Heredity, 41,* 249–320.

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologists, 49,* 725–47.

Ericsson, K. A., Krampe, T. T., & Heizman, S. (1993). Can we create gifted people? In G. R. Bock & K. Ackrill (Eds.), *The origins and development of high ability.* Chichester, England: Wiley.

Falconer, D. S. (1990). *Introduction to quantitative genetics* (3rd ed.). New York: Longman Group Ltd.

Fancher, R. E. (1985). *The intelligence men: Makers of the IQ controversy.* New York: W. W. Norton & Co.

Farber, S. L. (1981). *Identical twins reared apart: A reanalysis.* New York: Basic books.

Feingold, A. (1992). Good-looking people are not what we think. *Psychological Bulletin, 111,* 304–41.

Feldman, M. W., & Lewontin, R. C. (1975). The heritability hand-up. *Science, 190,* 1163–8.

Fletcher, R. (1991). *Science, ideology, and the media: The Cyril Burt scandal.* New Brunswick, NJ: Transaction Books.

Fletcher, R. (1993). The "Miss Conway" story. *The Psychologist: Bulletin of the British psychological Society, 6,* 214–15.

Ford, B. D. (1993). Emergenesis: An alternative and a confound. *American Psychologist, 48,* 1294.

Fulker, D. W. (1975). [Review of "The science and politics of IQ."] *American Journal of Psychology, 88,* 505–37.

Galton, F. (1869). *Hereditary genius: An inquiry into its laws and consequences.* London: Macmillan.

Galton, F. (1876). The history of twins, as a criterion of the relative powers of nature and nurture. *Journal of the Anthropological Institute of Great Britain and Ireland, V,* 391–406.

Ghiselli, E. E., Campbell, J. P., & Zedeck, S. (1981). *Measurement theory for the behavioral sciences.* San Francisco: Freeman.

Goldberg, L. R. (1971). A historical survey of personality scales and inventories. In P. McReynolds (Ed.), *Advances in psychological assessment.* Palo Alto: Science and Behavior Books.

Gottesman, I. I. (1963). Genetic aspects of intelligent behavior. In N. R. Ellis (Ed.), *Handbook of mental deficiency* (pp. 253–96). New York: McGraw-Hill.

Gottesman, I. I. (1968). Biogenetics of race and class. In M. Deutsch, I. Katz, & A. Jensen (Eds.), *Social class, race and psychological development* (pp. 11–51). New York: Holt, Rinehart & Winston.

Gottfredson, L. S. (1986). The *g* factor in employment: A special issue of the *Journal of Vocational Behavior. Journal of Vocational Behavior, 29,* 3293–450.

Gottfredson, L. S., & Sharf, J. C. (1988). Fairness in employment testing: A special issue of the *Journal of Vocational Behavior. Journal of Vocational Behavior, 33,* 225–463.

Gould, S. J. (1981). *The mismeasure of man.* New York: W. W. Norton.

Haldane, J. B. S. (1946). The interaction of nature and nurture. *Annals of Eugenics, 13,* 197–205.

Hartigan, J. A., & Wigdor, A. K. (Eds.). (1989). *Fairness in employment testing: Validity generalization, minority issues, and the General Aptitude Test Battery.* Washington, DC: National Academy Press.

Hayes, W. L. (1973). *Statistics for the social sciences* (2nd ed.). New York: Holt, Rinehart & Winston.

Hedges, L. V. (1987). How hard is hard science, how soft is soft science? The empirical cumulativeness of research. *American Psychologist, 42,* 443–55.

Henderson, N. (1990). Why do gene–environment interactions appear more often in laboratory animal studies than in human behavioral genetics? *Behavior and Brain Sciences, 13,* 136–7.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York: Free Press.

Hirsch, J. (1975). Jensenism: The bankruptcy of "Science" without scholarship. *Educational Theory, 25,* 3–28.

Hoffman, L. W. (1991). The influence of the family environment on personality: Evidence for sibling differences. *Psychological Bulletin, 110,* 187–203.

Horgan, J. (1993). Eugenics revisited. *Scientific American* (June), 122–32.

Howe, M. J. A. (1990). *Sense and nonsense about hothouse children.* Leicester: British Psychological Society.

Humphreys, L. G. (1992). Commentary: What both critics and users of ability tests need to know. *Psychological Science, 3,* 271–4.

Hyde, J. (1973). Genetic homeostasis and behavior: Analysis, data and theory. *Behavior Genetics, 3,* 233–45.

Hyde, J. (1974). Inheritance of learning ability in mice: A diallel-environment analysis. *Journal of Comparative and Physiological Psychology, 86,* 116–23.

Jackson, D. N. (1975). Intelligence and ideology. *Science, 189,* 1078–80.

Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review, 39,* 1–123.

Jensen, A. R. (1971). Note on why genetic correlations are not squared. *Psychological Bulletin, 75,* 223–4.

Jensen, A. R. (1973). Equating for socioeconomic status. In A. R. Jensen (Ed.), *Educability and group differences.* New York: Harper and Row.

Jensen, A. R. (1977). Cumulative deficit in IQ of blacks in the rural south. *Developmental Psychology, 13,* 184–91.

Jensen, A. R. (1980). *Bias in mental testing.* New York: Free Press.

Jensen, A. R. (1994). What is a good *g*? *Intelligence, 18,* 231–58.

Jinks, J. L., & Fulker, D. W. (1970). Comparison of the biometrical genetical, MAVA, and classical approaches to the analysis of human behavior. *Psychological Bulletin, 73,* 311–49.

Joynson, R. B. (1990). *The Burt affair.* New York: Academic Press.

Juel-Nielsen, N. (1980). *Individual and environment: Monozygotic twins reared apart (revised edition of 1965 monograph).* New York: International Universities Press.

Kamin, L. J. (1974). *The science and politics of IQ.* Potomac, MD: Erlbaum.

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., & Eaves, L. J. (1993). A test of the equal-environment assumption in twin studies of psychiatric illness. *Behavior Genetics, 23,* 21–7.

Layzer, D. (1974). Heritability analysis of IQ scores: Science or numerology. *Science, 183,* 1259–66.

Lewontin, R. C. (1982). *Human diversity.* New York: Scientific American Books.

Lewontin, R. C. (1974). The analysis of variance and the analysis of causes. *American Journal of Human Genetics, 26,* 400–11.

Lewontin, R. C. (1975). Genetic aspects of intelligence. *Annual Review of Genetics, 9,* 387–405.

Lewontin, R. C., Rose, S., & Kamin, L. J. (1984). *Not in our genes: Biology, ideology, and human nature.* Pantheon: New York.

Locurto, C. (1988). On the malleability of IQ. *The Psychologist: Bulletin of the British Psychological Society, 11,* 431–5.

Locurto, C. (1990). The malleability of IQ as judged from adoption studies. *Intelligence, 14,* 275–92.

Locurto, C. (1991). *Sense and nonsense about IQ: The case for uniqueness.* New York: Praeger.

Loehlin, J. C. (1981). [Review of Farber, S. L., "Identical twins reared apart: A reanalysis."] *Acta Geneticae Medicae et Gemellogiae, 30,* 297–8.

Loehlin, J. C., & Nichols, R. C. (1976). *Heredity, environment, & personality: A study of 850 sets of twins.* Austin: University of Texas Press.

Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin, 70,* 151–9.

Lykken, D. T., Geisser, S., & Tellegen, A. (1978). *Heritability estimates from twins studies: The efficiency of the MZA design.* Unpublished manuscript, Department of Psychology, University of Minnesota.

Lykken, D. T., McGue, M., Tellegen, A., & Bouchard, T. J., Jr. (1992). Emergenesis: Genetic traits that may not run in families. *American Psychologist, 47,* 1565–77.

Matarazzo, J. D. (1972). *Wechsler's measurement and appraisal of adult intelligence* (5th ed.). Baltimore: Williams & Wilkins.

Matarazzo, J. D. (1992). Biological and physiological correlates of intelligence. *Intelligence, 16,* 257–8.

Matheny, A. P., Wilson, R. S., & Dolan, A. B. (1976). Relations between twins' similarity of appearance and behavioral similarity: Testing an assumption. *Behavior Genetics, 6,* 343–51.

Mazur, A., Halpern, C., & Udry, J. R. (1994). Dominant-looking male teenagers copulate earlier. *Ethology and Sociobiology, 15,* 87–94.

McCartney, K., Harris, M. J., & Bernieri, F. (1990). Growing up and growing apart: A developmental meta-analysis of twin studies. *Psychological Bulletin, 107,* 226–37.

McGue, M. (1989). Nature–nurture and intelligence. *Nature, 340,* 507–8.

McGue, M., Bouchard, T. J., Jr., Iacono, W. G., & Lykken, D. T. (1993). Behavior genetics of cognitive ability: A life-span perspective. In R. Plomin & G. E. McClearn (Eds.), *Nature, nurture and psychology* (pp. 59–76). Washington, DC: American Psychological Association.

Meehl, P. E. (1970). Nuisance variables and the ex post facto design. In M. Radner & S. Winokur (Eds.), *Minnesota studies in the philosophy of science IV.* Minneapolis: University of Minnesota Press.

Meehl, P. E. (1971). High school yearbooks: A reply to Schwarz. *Journal of Abnormal Psychology, 77,* 143–8.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46,* 806–34.

Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports, 80,* 195–244.

Melamed, T. (1992). Personality correlates of physical height. *Personality and Individual Differences, 13,* 1349–50.

Miller, G. A. (1984). The test. *Science, 84, 5,* 55–7.

Miller, J. K., & Levine, D. (1973). Correlation between genetically, matched groups versus reliability theory: A reply to Jensen. *Psychological Bulletin, 79,* 142–4.

Muller, H. J. (1925). Mental traits and heredity. *Journal of Heredity, 16,* 433–48.

Neale, M. C., & Cardon, L. R. (Eds.) (1992). *Methodology for genetic studies of twins and families.* Dordrecht: Kluwer Academic Publishers.

Newman, H. H., Freeman, F. N., & Holzinger, K. J. (1937). *Twins: A study of heredity and environment.* Chicago: University of Chicago Press.

Parker, K. C., Hanson, R. K., & Hunsley, J. (1988). MMPI, Rorschach, and WAIS: A meta-analytic comparison of reliability, stability and validity. *Psychological Bulletin, 103,* 367–73.

Pedersen, N. L., Plomin, R., Nesselroade, J. R., & McClearn, G. E. (1992). A quantitative genetic analysis of cognitive abilities during the second half of the life span. *Psychological Science, 3,* 346–53.

Plomin, R. (1994). The nature of nurture: Family environment. In R. Plomin (Ed.), *Genetics and experience: The interplay between nature and nurture* (pp. 104–48). Beverly Hills: Sage.

Plomin, R., DeFries, J. C., & McClearn, G. E. (1990). *Behavioral genetics: A primer* (3rd ed.). New York: W. H. Freeman.

Plomin, R., Willerman, L., & Loehlin, J. C. (1976). Resemblance in appearance and the equal environments assumption in twin studies of personality traits. *Behavior Genetics, 6,* 43–52.

Popenoe, P. (1922). Twins reared apart. *Journal of Heredity, 5,* 142–4.

Rao, D. C., Morton, N. E., & Yee, S. (1974). Analysis of family resemblance. II. Linear model for familial correlation. *American Journal of Human Genetics, 26,* 331–59.

Reed, T. E. (1984). Mechanisms for heritability of intelligence. *Nature, 311,* 417.

Reed, T. E. (1990). Evolutionary and neurophysiological arguments for the heritability of intelligence. *European Bulletin of Cognitive Psychology, 10,* 659–67.

Richards, R. J. (1987). *Darwin and the emergence of evolutionary theories of mind and behavior.* Chicago: University of Chicago Press.

Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (Eds.). (1991). *Measures of personality and social psychological attitudes.* San Diego: Academic Press.

Rose, R. J. (1981). [Review of Farber, S. L., "Identical twins reared apart: A reanalysis."] *Science, 215,* 959–60.

Rose, R. J., Kaprio, J., Williams, C. J., Viken, R., & Obremski, K. (1990). Social contact and sibling similarity: Facts, issues, and red herrings. *Behavior Genetics, 20,* 763–78.

Rowe, D. C. (1994). *The limits of family influence: Genes, experience, and behavior.* New York: Guilford Press.

Rowe, D. C., Clapp, M., & Wallis, J. (1987). Physical attractiveness and the personality resemblance of identical twins. *Behavior Genetics, 17,* 191–201.

Rushton, J. P. (1989). The generalizability of genetic estimates. *Personality and Individual Differences, 10,* 985–89.

Scarr, S. (1968). Environmental bias in twin studies. *Eugenics Quarterly, 15,* 34–40.

Scarr, S. (1992). Developmental theories for the 1990's: Development and individual differences. *Child Development, 63,* 1–19.

Scarr, S. (1993). Biological and cultural diversity: The legacy of Darwin for development. *Child Development, 64,* 1333–53.

Scarr, S., & Carter-Saltzman, L. (1979). Twin method: Defense of critical assumption. *Behavior Genetics, 9,* 527–42.

Scarr, S., Scarf, E., & Weinberg, R. A. (1980). Perceived and actual similarities in biological and adoptive families: Does perceived similarity bias genetic inferences? *Behavior Genetics, 10,* 445–58.

Scarr, S., & Weinberg, R. A. (1978). The influence of family background on intellectual attainment. *American Sociological Review, 43,* 674–92.

Scarr, S., & Weinberg, R. A. (1994). Educational and occupational achievement of brothers and sisters in adoptive and biological related families. *Behavior Genetics, 24,* 301–25.

Schiff, M., & Lewontin, R. (1986). *Education and class: The irrelevance of IQ genetic studies.* Oxford: Clarendon Press.

Schmidt, F. L. (1993). Data, theory, and meta-analysis: Response to Hoyle. *American Psychologist, 48,* 1096.

Schmidt, F. L. (1994). Quantitative methods and cumulative knowledge in psychology: Implications for the training of researchers. *Presidential Address to APA Division 5, August 13, 1994.*

Shields, J. (1962). *Monozygotic twins: Brought up apart and brought up together.* London: Oxford University Press.

Snyderman, M., & Rothman, S. (1987). Survey of expert opinion on intelligence and aptitude testing. *American Psychologist, 42,* 137–44.

Snyderman, M., & Rothman, S. (1988). *The IQ controversy: The media and public policy.* New Brunswick, NJ: Transaction Books.

Spearman, C. (1927). *The abilities of man: Their nature and measurement.* New York: Macmillan.

Stanley, J. C. (1973). Accelerating the educational progress of intellectually gifted youths. *Educational Psychology, 10,* 133–46.

Taylor, H. F. (1980). *The IQ game: A methodological inquiry into the heredity environment controversy.* New Brunswick, NJ: Rutgers University Press.

Tomlinson-Keasey, C., & Little, T. D. (1990). Predicting educational attainment, occupational achievement, intellectual skill and personal adjustment among gifted men and women. *Journal of Educational Psychology, 82,* 442–55.

Townsend, M. (1874–75). Letter to Francis Galton. In Galton Archives in the Library of University College, London, List # 112/3.

Turkheimer, E. (1991). Individual and group differences in adoption studies of IQ. *Psychological Bulletin, 110,* 392–405.

Turkheimer, E., & Gottesman, I. I. (1991). Individual differences and the canalization of human behavior. *Developmental Psychology, 27,* 18–22.

Tyler, R. W., & White, S. H. (Chairmen) (1979). *Testing, teaching, and learning: Report of a conference on research on testing.* August 17-26, 1978. Washington, DC: National Institute of Education.

Urbach, P. (1974). Progress and degeneration in the "IQ debate" (II). *British Journal of the Philosophy of Science, 25,* 235–59.

Vernon, P. A. (Ed.). (1993). *Biological approaches to the study of human intelligence.* Norwood, NJ: Ablex.

Wachs, T. H. (1992). *The nature of nurture.* Newbury Park, CA: Sage.

Wahlsten, D. (1990). Insensitivity of the analysis of variance to heredity–environment interaction. *Behavior and Brain Sciences, 13,* 109–61.

West, P. (1991). Rethinking the health selection explanation for health inequalities. *Social Science and Medicine, 32,* 373–84.

White, R. K. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin, 91,* 461–81.