

## The Psychometric Paradigm: Some Problems

Hernnstein and Murray, the author's of *The Bell Curve* make several independent arguments. They argue that intelligence is largely “determined” by heredity, that it is single-scaled ability that can be measured (hence the term “psychometric”) and that it, rather than factors in the social environment (including the wealth of parents) is more critical in determining one's position in a host of outcomes, including income distribution and crime. The authors of *Inequality by Design* show that (1) AFQT (the test of intelligence used by Herrnstein and Murray) is not “construct valid”—it does not measure what it pretends to measure, (2) that on their own data, “Parental SES” has greater “predictive validity” than AFQT (i.e., the correlation between SES and income inequality is stronger than between AFQT and income inequality) and that (3) in any case, the “predictive validity” of IQ is very weak. Neither (1) nor (2) is discussed in what follows. In what follows, however, the effort is made to clarify and raise problems regarding causality (with special reference to human capacities and social outcomes) and then, more generally, to show the limits of “multiple regression,” the statistical technique which Herrnstein and Murray employ, and which is in wide use in the social sciences.

### Human Development

Multiple regression is used to establish which of several factors (“the independent variables”) is more (or less) important in explaining “variance,” or differences in “the dependent variable.” Presumably the “independent variables” are the factors which “determine” the outcome. In this case, there are two questions: Given that there are other factors involved, what is the relative importance of heredity as regards IQ (presumably the measure of intelligence) and what is the relative importance of IQ (or AFQT) as regards income inequality.

Begin with the question of heredity (or what is “determined” by DNA). As seems clear, the psychological and physical traits of persons are causal outcomes of a complex epigenetic process which begins with conception and ends with the death of the organism. To be human, one must have human DNA sequences and one must be reared in a human environment. A particular and *unique* genome, itself the product of the conjunction of haploid sex cells, is, in embryogenesis, the locus of continuous transactions both in itself and in relation to its “environment,” beginning in the womb and then later, in transaction with the natural and human environment. As the infant grows, she develops a particular physiognomy, she acquires a language, a set of habits and dispositions, various capacities and traits, a self and a personality. Through both of these conditions, in the womb and then after, as the biologist Paul Weiss remarks, ‘the latitude for epigenetic vagaries of the component elements on all levels is immense.’ To say that the process is epigenetic is to say that in this immensely complicated process—misleadingly called a chain—the causes are transactional and not additive.<sup>1</sup> Indeed, with the exception of mechanical causation where forces are interactive and can be added (as vectors), *this is true of all causality*. For example, a plant's growth depends upon the nature of seed, water, fertilizer and a host of other “factors.” All of these “cause” the eventual outcome. But As Lewontin remarks, it would “absurd to say what proportion of the plant's height is owed to the fertilizer it

received and what proportion to the water...:<sup>2</sup> More generally, all outcomes are the result of complex causality working conjunctively and while each identifiable “cause” contributes to the outcome, often as necessary conditions, we cannot in general simply add them up as sum of factors so that they equal 1.0.

As regards human development, there are many causes, physical, chemical, biological, and social, working at various levels. Biochemical processes and interventions at the biochemical level, e.g., nicotine or radiation, have effects on molecular processes which then have effects on higher-order complexes, e.g., the brain or heart, and finally to the intact functioning organism, a conscious person acting in the world (and being acted on by it!). And, as noted, from birth on, there are the continuous effects of the social environment: all the problems and opportunities presented by your social situation, what you see, hear, feel, read: You discover something that you didn’t know and you change your mind—sometimes with critical consequences, some intended, some not.

Since each moment of development establishes new conditions for what comes after, time is also critical. Some changes (biological and psychological) are small and unimportant relatively; others, including those which are irreversible, are not.<sup>3</sup> And while change continues until death, the earlier periods of life are much more important in establishing both the psychological and the physical traits of the person. Poor nutrition can leave enduring problems; so can poor “socialization.” We develop habits, this includes not only ways of acting, but ways of thinking as well.

The epigenetic “vagaries,” of course, are not unlimited: They are, if you will, restricted by our “biologically determined” human nature. It may be useful here to give a restricted meaning to a term used widely but vaguely (and usually wrongly). We can say that some trait, capacity, or difference is a feature of our (biologically determined) “human nature” only if the person realizing that trait would have realized it had she been in any other time or place. That is, time and place are irrelevant as regards features of our “human nature.”

There are some obvious biologically determined traits: Our human anatomy and physiology is one. Others, related to this, include manifest physical traits which mark family resemblance, such as facial features, body type, and skin color. Race is *not* biologically determined since on all the evidence there are no *biological* grounds for grouping people according to race.<sup>4</sup> Race (like ethnicity) is a social construction: We simply employ a socially agreed upon difference as the criteria for the grouping. Of course, there are historical reasons which explain why the cluster of differences associated with skin color became the difference which became the criterion for distinguishing “races.”

We can also distinguish *realized capacities*, e.g., the ability to speak (say) Dutch, from *capacities as potentialities*, the ability to acquire language, and hold that the capacities as potentialities are biological determined although realized capacities are not. There are a host of distinctly human (species-specific) capacities as potentialities which presuppose our distinct evolutionary history. These are a critical part of “human nature” (as defined above). Humans cannot fly and (lacking gills), they cannot breath in water.

*Homo sapiens* everywhere and anywhen has the potential to be linguistically competent. But of course, depending upon the time and place, children acquire some very different languages. That is, the potential is concretely realized in societies which differ. There are probably as well biologically grounded propensities or tendencies of other sorts, e.g., toward cancer and schizophrenia. Like the human powers just mentioned, these need not be realized, and like these other capacities, they are still very poorly understood. We do know, however, that *all the distinctly human capacities require for their realization a human environment*. It follows, accordingly, that all these realized capacities are social in a very obvious sense.<sup>5</sup>

But if we all share in being *biologically* humans, there is no doubt also that differences in the DNA sequences of the unique genome (which was you on Day I of your life) can make a difference in terms of what you will grow up to be. Plainly, the “genes” play the critical role in whether you are male or female, dark or light skinned, potentially tall or potentially stocky. They probably also play a role in matters of “temperament,” whether you tend to be impulsive, willing to take risks, quick to anger (or the reverse). But almost certainly also there will be genetic differences in your potentiality to be musical, mathematical, athletic or artistic. Some people have a “tin ear,” and others have “perfect pitch,” some people cannot “draw a straight line,” some people have poor hand/eye coordination (and can’t hit the curve ball). But these are potentialities in just the sense that a good deal more has to be happen if they are to be realized. You will not be tall if you have been protein starved: you will not be able to play the violin if you never had one; you will not be able to hit the curve if you have not tried—and tried.

Four points are here critical. First, it is hardly clear that “intelligence” compares with (say) musical or even mathematical ability. *That is, there is good reason to believe that intelligence is not a single-scaled capacity*. Some people seem better than others with words and some with spaces; some with numbers and some with relations; some seem more “creative,” others more “logical.” Charles Spearman first suggested (in 1904) that there was some common “cause” of correlations between different tests. He designated it *g*, what came to be thought of as “intelligence”. But as Gould showed, *g* is a statistical artifact which can be made to disappear “by simply rotating the dimensions to different positions.” These days, psychologists are speaking about “multiple intelligences.”<sup>6</sup>

Second, many potentialities of persons are either not realized at all or are barely realized. And there are many reasons for this. One obvious reason, already noted, was that other conditions necessary to realize the capacity were absent: Insufficient protein, no violin, no teacher. Another obvious reason is that realizing capacities often requires work, often at a sacrifice of other goals and interests. In the spirit of the foregoing, we may well ask: How much of any realized capacity is due to work and how much to the “natural” ability?

Third, while there will be differences in individuals across a host of realized capacities, the capacities which are valued, and the importance of the differences between individuals—*like the importance of all differences between persons, e.g., skin color*---is determined socially. No society (at least as far as I know) has placed value on the ability

of person to belch or to snore. Sarah Chang plays the violin better than me, so people will pay to hear her play. How much better than me must she play in order to be paid to play? That depends. And it depends pretty much on the same social mechanisms, including for example, existing labor markets, which determine whether somebody who can belch louder than anybody will be paid to belch, or how much clerks or quarterbacks get paid.

Third, since such realized capacities are epigenetic and “causes” are not additive, we cannot say how much of any such trait was due to your DNA on Day I of your life and how much was due to the environmental factors, natural and social, or to effort. Just as it would be “absurd” to ask what proportion of the plant's height is owed to genes and what proportion to the fertilizer or water, it is equally absurd to say that what proportion of one’s cognitive or athletic (or belching) ability is due to DNA and what proportion to the social environment—including here, of course, the teachers and time and energy spent in realizing these capacities.

Finally, even among those factors which we think are largely due to the social environment, there are very good reasons to believe that we must be careful in distinguish the developmental effects of the home environment, the influence of parents and siblings and the influence of peers, “groupness” in the neighborhood and schools.<sup>7</sup> That is, the dichotomy “genes” versus “environment” is both crude and misleading.

### **The Limits of Multiple Regression**

It is too often thought that multiple regression overcomes this problem, that through its use can establish such proportions.

Even careful writers often confuse the following ideas:

1. A (some “variable,” e.g., IQ) correlates with B (some other variable, e.g., income)
2. A “predicts” B
3. A “explains the variance” in B
4. A “explains” B
5. A “causes” B

We can handle 4. and 5. together. We can say 4., “A explains B” only if we can say, “A causes B.” But first, correlations do not establish causes. Causes are “mechanisms” which produce outcomes. We can have a correlation where there is no conceivable mechanism, e.g., the price of eggs in a Beijing market and the price of Microsoft on the New York Stock Exchange. Second, as already noted, there are always many causes of any outcome. In order to make a fire, we need in addition to some combustible material, a source of heat and oxygen. Absent any of these, no fire. So which is more important? We get a fire only if the *right combination* is present. (It takes a good deal more heat to ignite a vinyl fabric than it does to ignite cotton.) If we pick out a source of heat as “the cause,” that is because we assume the presence of oxygen and the combustible material. We forget about the oxygen and say, the spark “caused” the fire. This is both convenient and unsurprising. But the fact remains: *all* the factors are important: you will not get a fire if *any* are absent. Consider then Sarah’s ability to score big on the SAT. What is “the cause”? Which the “factors” (causes) will be more important? Sarah may be “bright,” but she also was well-motivated, got some terrific education—and she felt good on the day of the test.

1. and 2. also can be treated together. A correlation between A and B can be between 0.0 and 1.0. 0.0 is no correlation; 1.0 is a perfect correlation: every change in A has commensurate change in B. Perfect correlations are rare indeed. This raises the question: when is a correlation (or in multiple regression, a “coefficient of correlation”) “significant.” In part this will depend upon the problem and what we need to know. Correlations down near 0.0, e.g., .02 or .1, are plainly not “significant.” (As Gould noted, most of the coefficients of correlation in the *Bell Curve* are less than 0.1.)<sup>8</sup>

Where we have a significant correlation, we can predict. If there is, e.g., a positive correlation of .6 between the price of eggs in China and the market price of Microsoft on the New York Stock Exchange, then if I know that there is drop in the price of eggs in China, I had best sell my Microsoft quickly. There is no suggestion that there is a causal connection here. Were it to happen, it would be a statistical fluke. But for purposes of prediction, the relation is all that I need. Nor because there is no causal mechanism involved here, can I *explain* the drop in the price Microsoft by appeal to the fall in the price of eggs in China.

Smoking and cancer is good example. There is *some* causal mechanism at work in cancer production and smoking is related to this in ways that we do not yet understand. Some people surely do smoke all their lives and never get cancer. And some people who never smoke do. But we know that the probability of getting cancer significantly increases if you smoke: A “predicts” B.

Sophisticated scientists are very often careless when they speak about 3., “explaining the variance.” What they intend by this can be briefly summarized. Assume that there are a number of “factors” which taken together presumably “determine” some outcome. The idea then is find out how significant each factor is “producing” this outcome. The language of “producing an outcome” or “determining an outcome” is causal language. But indeed, such language is entirely inappropriate. We need to go a little deeper to see what is at issue here.

Assume first a standard regression equation, a set of dependable, meaningful independent variables (a, b...) with a linear relation to the dependent variable (Y).

$$Y = a + b_1 + b_2 + b_1b_2 + e \text{ (Equation 1)}$$

“Y,” the “dependent variable,” presumably is “determined” by the independent variables, “a + b<sub>1</sub>...” The problem is then one of variable selection. The goal of the analysis is a “good fit.” If we do our work well, what we end up is “a useful statistical description defensible against plausible alternative interpretations.”<sup>9</sup> It is critical to emphasize that *the very best result is a statistical description*, a point nearly always missed. At best, the result is a highly simplified picture, a statistical snapshot, of a fantastically complicated concrete social situation. For example, as an abstract ratio, the crime rate represents a picture of crime in the real world. It leaves much out—obviously. On the other hand, as Achen remarks: “A picture of a friend is useless if it covers a football field and exhibits

every pore. What one looks for instead is an interpretable amount of information, with the detailed workings omitted (p. 13). As regards the crime rate, the 'detailed workings' include, of course, the specific structured actions of *everyone* in society: both criminals and non-criminals. While it would be agreed that a crime rate is such snapshot taken from a very long distance, the same is true of all other statistical results, including the results of regressions.

A useful description—a good fit—is not so easy to come by. One test of this is the “coefficient of correlation,”  $R^2$ . It is usually said that  $R^2$  gives “the percentage of variance explained” in the dependent variable by the regression. But as Achen comments: This is an expression that, “or most social scientists, is of doubtful meaning but great rhetorical value” (58f.) The rhetorical value lies in the supposition that first, a large  $R^2$  guarantees “good fit” and second, in the more radical confusion, that the number represents the causal importance of the factor in the regression.

Neither supposition can be sustained. As Achen says,  $R^2$  “is best regarded as characterizing the geometric shape of the regression points and nothing more” (p. 59). It is easy to see why it is nothing more than this. Achen says: “The central difficulty with the  $R^2$  for social scientists is that the independent variables are not subject to experimental manipulation.” In the natural sciences, one tests theories about causality with an experiment. The experiment seeks to “control” the conditions to see if the hypothesized cause actually produces the outcome which the theory predicted. This is not possible in the social sciences. “Regression,” which presumes to “control” variables, *mathematically*, is often thought to be an adequate substitute for experiment.

There are several lines of argument that it is not. One regards the problem that “variances are a function of the *sample*, not the underlying relationship.” That is, the linear model (eq. 1) is a *local analysis* whose result depends upon the actual distributions of the variables in the population sampled. Thus, “in some samples, they vary widely, producing large variance; in other cases, the observations are more tightly grouped and there is little dispersion.” (One needs some further understanding of statistical analysis to fully grasp this criticism.) For this reason, then, “they cannot have any real connection to the ‘strength’ of the relationship as social scientists ordinarily use the term, i.e., as a measure of how much effect a given change in the independent variable has on the dependent variable...”

Second, there is the problem of assuming that the measured variables “add up” to 1.0, the problem of “additivity” and independence which we noted earlier. Achen offers an example:

If the regression describes, say, domestic violence in countries as a function of violence in prior years plus economic conditions, can one say which variable is more important in causing violence? For most purposes the answer is no. The units of one variable are violence per amount of prior violence; the units of the other are violence per amount of economic dislocation. One can say only that apples differ from oranges. *As theoretical forces abstracted from any historical circumstances, they have no common measure* (p. 70).

Equation 1 makes us believe that the variables are both additive and independent (with  $b_1b_2$  taking into account the interaction effects of the variables.) But this is never the case. The best sort of example to illustrate the general principle is to see the confusion in the mostly meaningless discussions of the relative effects of heredity and the environment. Consider a parallel (idealized) biological study, a study which requires a controlled experiment.

Take a genotype replicated by inbreeding or cloning. This minimizes genotypic individuality. Place them in a various carefully controlled environments. It is then possible to establish rough tables of correspondence between phenotype on the one hand and genotype-environment combinations on the other. The results, called the “norm of reaction,” are *never* predictable in advance. They are not predictable since genetic and environmental factors are not additive (and hence cannot be represented by linear equations.) They are causes in transaction in exactly the sense that genes cause different outcomes in different transactional environments.

If such norms could be experimentally established for persons in their development, then across the range of controlled environments and (cloned?) genotypes, one could relate the variances in outcomes with the changes in the independent variables. This would still not provide the proportion of causation since causation does not suddenly become additive. But one could talk sensibly about their relative “importance.” One could “explain the variance” sensibly. More dramatically, as Achen says, put some children in middle-class homes and the others in closets. There surely will be differences in cognitive ability, personality, etc. Almost certainly, most of the differences in these realized capacities will be “explained” by environment. Conversely, put them all (*per impossible*) in *the same* environment, most of the variation surely will be “explained” by heredity.

I hope, however, it is obvious that except, for identical twins, not only are no two genotypes the same, but that in the concrete real world, there is not any way *in principle* to specify all the relevant environmental “variables,” exactly because *these are not independent*. The social world is real enough, but the mere fact that *necessarily* it is mediated by the consciousness of agents makes it impossible to say how a condition will be experienced and understood by the agent, and thus what affect it will have on him and his behavior. Accordingly, not only will multiple regression not give the proportion of causality involved in some outcome, but in general, it will not even allow us “to explain the variance.”

Indeed, there are differences between individuals which are rooted in our genes, but if we want to explain inequalities in the real world, we had best find some other way. That, of course, was the task of *Inequality by Design*.

For class use

October 1999

## ENDNOTES

---

<sup>1</sup> Reporting on very recent research in evolutionary theory which shows that a common genetic key triggers the development of eyes of vastly different constructions, Peter Monaghan referred to the process as a 'cascade of biochemical events that take place in eye development' (*Chronicle of Higher Education*, 26 May 1995).

<sup>2</sup> Richard Lewontin, 'Analysis of Variance and the the Anaysis of Causes,' reprinted in N.J. Bloch and G. Dworkin (eds.), *The IQ Controversy* (New York : Pantheon, 1976.) This should be required reading for all social scientists.

<sup>3</sup> There is a body of evidence that suggests that there are "windows" in early development such that after they close, e.g., by an injury to the brain or by insufficient nutrition or stimulation, further development is either impossible or hextremely difficult. See for a very recent overview, see George Johnson, " How much can the Brain Take," *New York Times*, 24 October 1999.

<sup>4</sup> For an excellent review and summary of the evidence, see Emanuel Drechsel, 'The Invalidity of the Concept "Race,"' in M. Tehranian, *Restructuring for Ethnic Peace* (Honolulu, Hawai'i: Matsunaga Institute for Peace, 1991). Also see Stephen Jay Gould, *Mismeasure of Man* (New York: W.W. Norton, 1981). He concludes with a wonderful illustration from Lewontin: "If the holocaust comes and a small tribe in the New Guinea forests are the only survivors, almost all the genetic variation now expressed among the innumerable groups of our four billion [1980] people will be preserved' (p. 323). It was also a part of the *Bell Curve* argument that there were differences in the IQ scores of "races" and that these had to be explained by heredity.

<sup>5</sup> This is true, probably, of language, the emotions, perception and cognition. For language, see Derek Bickerton, *Language and Species* (Chicago: University of Chicago Press, 1990). For the emotions, see Rom Harre, 'Social Sources of Mental Content and Order,' in Margolis, Manicas, Harre and Secord, *Psychology: Designing the Discipline* (Oxford: Basil Blackwell, 1986). For perception, see J. Van Brakel, 'The Plasticity of Categories: The Case of Colour,' *British Journal for the Philosophy of Science*, 44 (1993). For cognition, see most recently, Marshall Sahlins, *How 'Natives' Think About Captain Cook, For Example* (Chicago: University of Chicago Press, 1995).

<sup>6</sup> For a discussion of Spearman's *g* and problems with "factor analysis" (which are analogous to those to be discussed below), see Gould, *Mismeasure of Man*. For the most recent defense of the idea of "multiple intelligences," See Howard Gardner, *The Theory of Multiple Intelligences* (New York: Basic Books).

<sup>7</sup> See Judith Rich Harris, *The Nurture Assumption* (New York: Free Press, 1998). Harris brilliantly rejects the idea that the main determinant in human development is the way that parents bring up their children. She offers instead "group socialization" and the critical variable. From the present point of view, while she uses the evidence of behavioral genetics to refute standard psychological misconceptions, especially in undermining assumptions of correlations, she seems a bit insensitive to the transactional or

---

epigenetic character of all development.

<sup>8</sup> Stephen Jay Gould, "Curveball," *New Yorker*, 28 April 1994, p. 147. For a wonderful critique of the methodological fallacies in the *Bell Curve*, see also, Ian Hacking, "Pull the Other One," *London Review of Books*, 26 January 1995. E.g., there is simply no reason to believe that "intelligence" (whatever it is) has to be distributed in the shape of bell!

<sup>9</sup> Christopher H. Achen, *Interpretating and Using Regression* (Beverly Hills, Ca.: Sage, 1982). I am indebted to Achen's clear account. In *Inequality by Design*, compare the *Bell Curve* analysis of the relative importance of AFQT as against Parental SES with the alternative provided by the authors. Both provide a "good fit". They are different because the "indexes" which measured Parental SES differ.