

REVISED AND EXPANDED

THE
**Mismeasure
of Man**

BY STEPHEN JAY GOULD

If the misery of our poor be caused not by the laws of nature, but by our institutions, great is our sin. —CHARLES DARWIN, Voyage of the Beagle

W • W • NORTON & COMPANY • NEW YORK • LONDON

The Hereditarian Theory of IQ

An American Invention

Alfred Binet and the original purposes of the Binet scale

Binet flirts with craniometry

When Alfred Binet (1857–1911), director of the psychology laboratory at the Sorbonne, first decided to study the measurement of intelligence, he turned naturally to the favored method of a waning century and to the work of his great countryman Paul Broca. He set out, in short, to measure skulls, never doubting at first the basic conclusion of Broca's school:

The relationship between the intelligence of subjects and the volume of their head . . . is very real and has been confirmed by all methodical investigators, without exception. . . . As these works include observations on several hundred subjects, we conclude that the preceding proposition [of correlation between head size and intelligence] must be considered as incontestable (Binet, 1898, pp. 294–295).

During the next three years, Binet published nine papers on craniometry in *L'Année psychologique*, the journal he had founded in 1895. By the end of this effort, he was no longer so sure. Five studies on the heads of school children had destroyed his original faith.

Binet went to various schools, making Broca's recommended measurements on the heads of pupils designated by teachers as their smartest and stupidest. In several studies, he increased his sample from 62 to 230 subjects. "I began," he wrote, "with the idea,

impressed upon me by the studies of so many other scientists, that intellectual superiority is tied to superiority of cerebral volume" (1900, p. 427).

Binet found his differences, but they were much too small to matter and might only record the greater average height of better pupils (1.401 vs. 1.378 meters). Most measures did favor the better students, but the average difference between good and poor amounted to a mere millimeter—"extrêmement petite" as Binet wrote. Binet did not observe larger differences in the anterior region of the skull, where the seat of higher intelligence supposedly lay, and where Broca had always found greatest disparity between superior and less fortunate people. To make matters worse, some measures usually judged crucial in the assessment of mental worth favored the poorer pupils—for anteroposterior diameter of the skull, poorer students exceeded their smarter colleagues by 3.0 mm. Even if most results tended to run in the "right" direction, the method was surely useless for assessing individuals. The differences were too small, and Binet also found that poor students varied more than their smarter counterparts. Thus, although the smallest value usually belonged to a poor pupil, the highest often did as well.

Binet also fueled his own doubts with an extraordinary study of his own suggestibility, an experiment in the primary theme of this book—the tenacity of unconscious bias and the surprising malleability of "objective," quantitative data in the interest of a preconceived idea. "I feared," Binet wrote (1900, p. 323), "that in making measurements on heads with the intention of finding a difference in volume between an intelligent and a less intelligent head, I would be led to increase, unconsciously and in good faith, the cephalic volume of intelligent heads and to decrease that of unintelligent heads." He recognized the greater danger lurking when biases are submerged and a scientist believes in his own objectivity (1900, p. 324): "Suggestibility . . . works less on an act of which we have full consciousness, than on a half-conscious act—and this is precisely its danger."

How much better off we would be if all scientists submitted themselves to self-scrutiny in so forthright a fashion: "I want to state very explicitly," Binet wrote (1900, p. 324), "what I have observed about myself. The details that follow are those that the

majority of authors do not publish; one does not want to let them be known." Both Binet and his student Simon had measured the same heads of "idiots and imbeciles" at a hospital where Simon was in intern. Binet noted that, for one crucial measurement, Simon's values were consistently less than his. Binet therefore returned to measure the subjects a second time. The first time, Binet admits, "I took my measures mechanically, without any other preconception than to remain faithful to my methods." But the second time "I had a different preconception. . . . I was bothered by the difference" between Simon and myself. "I wanted to reduce it to its true value. . . . This is self-suggestion. Now, capital fact, the measures taken during the second experiment, under the expectation of a diminution, are indeed smaller than the measures taken [on the same heads] during the first experiment." In fact, all but one head had "shrunk" between the two experiments and the average diminution was 3 mm—a good deal more than the average difference between skulls of bright and poor students in his previous work. Binet spoke graphically of his discouragement:

I was persuaded that I had attacked an intractable problem. The measures had required traveling, and tiring procedures of all sorts; and they ended with the discouraging conclusion that there was often not a millimeter of difference between the cephalic measures of intelligent and less intelligent students. The idea of measuring intelligence by measuring heads seemed ridiculous. . . . I was on the point of abandoning this work, and I didn't want to publish a single line of it (1900, p. 403).

At the end, Binet snatched a weak and dubious victory from the jaws of defeat. He looked at his entire sample again, separated out the five top and bottom pupils from each group, and eliminated all those in the middle. The differences between extremes were greater and more consistent—3 to 4 mm on average. But even this difference did not exceed the average potential bias due to suggestibility. Craniometry, the jewel of nineteenth-century objectivity, was not destined for continued celebration.

Binet's scale and the birth of IQ

When Binet returned to the measurement of intelligence in 1904, he remembered his previous frustration and switched to other techniques. He abandoned what he called the "medical"

approaches of craniometry and the search for Lombroso's anatomical stigmata, and decided instead on "psychological" methods. The literature on mental testing, at the time, was relatively small and decidedly inconclusive. Galton, without notable success, had experimented with a series of measurements, mostly records of physiology and reaction time, rather than tests of reasoning. Binet decided to construct a set of tasks that might assess various aspects of reasoning more directly.

In 1904 Binet was commissioned by the minister of public education to perform a study for a specific, practical purpose: to develop techniques for identifying those children whose lack of success in normal classrooms suggested the need for some form of special education. Binet chose a purely pragmatic course. He decided to bring together a large series of short tasks, related to everyday problems of life (counting coins, or assessing which face is "prettier," for example), but supposedly involving such basic processes of reasoning as "direction (ordering), comprehension, invention and censure (correction)" (Binet, 1909). Learned skills like reading would not be treated explicitly. The tests were administered individually by trained examiners who led subjects through the series of tasks, graded in their order of difficulty. Unlike previous tests designed to measure specific and independent "faculties" of mind, Binet's scale was a hodgepodge of diverse activities. He hoped that by mixing together enough tests of different abilities he would be able to abstract a child's general potential with a single score. Binet emphasized the empirical nature of his work with a famous dictum (1911, p. 329): "One might almost say, 'It matters very little what the tests are so long as they are numerous.'"

Binet published three versions of the scale before his death in 1911. The original 1905 edition simply arranged the tasks in an ascending order of difficulty. The 1908 version established the criterion used in measuring the so-called IQ ever since. Binet decided to assign an age level to each task, defined as the youngest age at which a child of normal intelligence should be able to complete the task successfully. A child began the Binet test with tasks for the youngest age and proceeded in sequence until he could no longer complete the tasks. The age associated with the last tasks he could perform became his "mental age," and his general intellectual level

was calculated by subtracting this mental age from his true chronological age. Children whose mental ages were sufficiently behind their chronological ages could then be identified for special educational programs, thus fulfilling Binet's charge from the ministry. In 1912 the German psychologist W. Stern argued that mental age should be divided by chronological age, not subtracted from it,* and the intelligence *quotient*, or IQ, was born.

IQ testing has had momentous consequences in our century. In this light, we should investigate Binet's motives, if only to appreciate how the tragedies of misuse might have been avoided if its founder had lived and his concerns been heeded.

In contrast with Binet's general intellectual approach, the most curious aspect of his scale is its practical, empirical focus. Many scientists work this way by deep conviction or explicit inclination. They believe that theoretical speculation is vain and that true science progresses by induction from simple experiments pursued to gather basic facts, not to test elaborate theories. But Binet was primarily a theoretician. He asked big questions and participated with enthusiasm in the major philosophical debates of his profession. He had a long-standing interest in theories of intelligence. He published his first book on the "Psychology of Reasoning" in 1886, and followed in 1903 with his famous "Experimental Study of Intelligence," in which he abjured previous commitments and developed a new structure for analyzing human thinking. Yet Binet explicitly declined to award any theoretical interpretation to his scale of intelligence, the most extensive and important work he had done in his favorite subject. Why should a great theoretician have acted in such a curious and apparently contradictory way?

Binet did seek "to separate natural intelligence and instruction" (1905, p. 42) in his scale: "It is the intelligence alone that we seek to measure, by disregarding in so far as possible, the degree of instruction which the child possesses. . . . We give him nothing to read, nothing to write, and submit him to no test in which he might

* Division is more appropriate because it is the relative, not the absolute, magnitude of disparity between mental and chronological age that matters. A two-year disparity between mental age two and chronological age four may denote a far severer degree of deficiency than a two-year disparity between mental age fourteen and chronological age sixteen. Binet's method of subtraction would give the same result in both cases, while Stern's IQ measures 50 for the first case and 88 for the second. (Stern multiplied the actual quotient by 100 to eliminate the decimal point.)

succeed by means of rote learning" (1905, p. 42). "It is a specially interesting feature of these tests that they permit us, when necessary, to free a beautiful native intelligence from the trammels of the school" (1908, p. 259).

Yet, beyond this obvious desire to remove the superficial effects of clearly acquired knowledge, Binet declined to define and speculate upon the meaning of the score he assigned to each child. Intelligence, Binet proclaimed, is too complex to capture with a single number. This number, later called IQ, is only a rough, empirical guide constructed for a limited, practical purpose:

The scale, properly speaking, does not permit the measure of the intelligence, because intellectual qualities are not superposable, and therefore cannot be measured as linear surfaces are measured (1905, p. 40).

Moreover, the number is only an average of many performances, not an entity unto itself. Intelligence, Binet reminds us, is not a single, scalable thing like height. "We feel it necessary to insist on this fact," Binet (1911) cautions, "because later, for the sake of simplicity of statement, we will speak of a child of 8 years having the intelligence of a child of 7 or 9 years; these expressions, if accepted arbitrarily, may give place to illusions." Binet was too good a theoretician to fall into the logical error that John Stuart Mill had identified—"to believe that whatever received a name must be an entity or being, having an independent existence of its own."

Binet also had a social motive for his reticence. He greatly feared that his practical device, if reified as an entity, could be perverted and used as an indelible label, rather than as a guide for identifying children who needed help. He worried that schoolmasters with "exaggerated zeal" might use IQ as a convenient excuse: "They seem to reason in the following way: 'Here is an excellent opportunity for getting rid of all the children who trouble us,' and without the true critical spirit, they designate all who are unruly, or disinterested in the school" (1905, p. 169). But he feared even more what has since been called the "self-fulfilling prophecy." A rigid label may set a teacher's attitude and eventually divert a child's behavior into a predicted path:

It is really too easy to discover signs of backwardness in an individual when one is forewarned. This would be to operate as the graphologists did

who, when Dreyfus was believed to be guilty, discovered in his handwriting signs of a traitor or a spy" (1905, p. 170).

Not only did Binet decline to label IQ as inborn intelligence; he also refused to regard it as a general device for ranking all pupils according to mental worth. He devised his scale only for the limited purpose of his commission by the ministry of education: as a practical guide for identifying children whose poor performance indicated a need for special education—those who we would today call learning disabled or mildly retarded. Binet wrote (1908, p. 263): "We are of the opinion that the most valuable use of our scale will not be its application to the normal pupils, but rather to those of inferior grades of intelligence." As to the causes of poor performance, Binet refused to speculate. His tests, in any case, could not decide (1905, p. 37):

Our purpose is to be able to measure the intellectual capacity of a child who is brought to us in order to know whether he is normal or retarded. We should therefore study his condition at the time and that only. We have nothing to do either with his past history or with his future; consequently, we shall neglect his etiology, and we shall make no attempt to distinguish between acquired and congenital idiocy. . . . As to that which concerns his future, we shall exercise the same abstinence; we do not attempt to establish or prepare a prognosis, and we leave unanswered the question of whether this retardation is curable, or even improvable. We shall limit ourselves to ascertaining the truth in regard to his present mental state.

But of one thing Binet was sure: whatever the cause of poor performance in school, the aim of his scale was to identify in order to help and improve, not to label in order to limit. Some children might be innately incapable of normal achievement, but all could improve with special help.

The difference between strict hereditarians and their opponents is not, as some caricatures suggest, the belief that a child's performance is all inborn or all a function of environment and learning. I doubt that the most committed antihereditarians have ever denied the existence of innate variation among children. The differences are more a matter of social policy and educational practice. Hereditarians view their measures of intelligence as markers of permanent, inborn limits. Children, so labeled, should be sorted,

trained according to their inheritance and channeled into professions appropriate for their biology. Mental testing becomes a theory of limits. Antihereditarians, like Binet, test in order to identify and help. Without denying the evident fact that not all children, whatever their training, will enter the company of Newton and Einstein, they emphasize the power of creative education to increase the achievements of all children, often in extensive and unanticipated ways. Mental testing becomes a theory for enhancing potential through proper education.

Binet spoke eloquently of well-meaning teachers, caught in the unwarranted pessimism of their invalid hereditarian assumptions (1909, pp. 16-17):

As I know from experience, . . . they seem to admit implicitly that in a class where we find the best, we must also find the worst, and that this is a natural and inevitable phenomenon, with which a teacher must not become preoccupied, and that it is like the existence of rich and poor within a society. What a profound error.

How can we help a child if we label him as unable to achieve by biological proclamation?

If we do nothing, if we don't intervene actively and usefully, he will continue to lose time . . . and will finally become discouraged. The situation is very serious for him, and since his is not an exceptional case (since children with defective comprehension are legion), we might say that it is a serious question for all of us and for all of society. The child who loses the taste for work in class strongly risks being unable to acquire it after he leaves school (1909, p. 100).

Binet railed against the motto "stupidity is for a long time" ("*quand on est bête, c'est pour longtemps*"), and upbraided teachers who "are not interested in students who lack intelligence. They have neither sympathy nor respect for them, and their intemperate language leads them to say such things in their presence as 'This is a child who will never amount to anything . . . he is poorly endowed . . . he is not intelligent at all.' How often have I heard these imprudent words" (1909, p. 100). Binet then cites an episode in his own baccalaureate when one examiner told him that he would never have a "true" philosophical spirit: "Never! What a momentous word. Some recent thinkers seem to have given their moral support to these deplorable verdicts by affirming that an individual's intel-

The dismantling of Binet's intentions in America

In summary, Binet insisted upon three cardinal principles for using his tests. All his caveats were later disregarded, and his intentions overturned, by the American hereditarians who translated his scale into written form as a routine device for testing all children.

1. The scores are a practical device; they do not buttress any theory of intellect. They do not define anything innate or permanent. We may not designate what they measure as "intelligence" or any other reified entity.
2. The scale is a rough, empirical guide for identifying mildly retarded and learning-disabled children who need special help. It is not a device for ranking normal children.
3. Whatever the cause of difficulty in children identified for help, emphasis shall be placed upon improvement through special training. Low scores shall not be used to mark children as innately incapable.

If Binet's principles had been followed, and his tests consistently used as he intended, we would have been spared a major misuse of science in our century. Ironically, many American school boards have come full circle, and now use IQ tests only as Binet originally recommended: as instruments for assessing children with specific learning problems. Speaking personally, I feel that tests of the IQ type were helpful in the proper diagnosis of my own learning-disabled son. His average score, the IQ itself, meant nothing, for it was only an amalgam of some very high and very low scores; but the pattern of low values indicated his areas of deficit.

The misuse of mental tests is not inherent in the idea of testing itself. It arises primarily from two fallacies, eagerly (so it seems) embraced by those who wish to use tests for the maintenance of social ranks and distinctions: reification and hereditarianism. The next chapter shall treat reification—the assumption that test scores represent a single, scalable thing in the head called general intelligence.

The hereditarian fallacy is not the simple claim that IQ is to some degree "heritable." I have no doubt that it is, though the degree has clearly been exaggerated by the most avid hereditarians. It is hard to find any broad aspect of human performance or anatomy that has no heritable component at all. The hereditarian fallacy resides in two false implications drawn from this basic fact:

intelligence is a fixed quantity, a quantity that cannot be increased. We must protest and react against this brutal pessimism; we must try to demonstrate that it is founded upon nothing" (1909, p. 101).

The children identified by Binet's test were to be helped, not indelibly labeled. Binet had definite pedagogical suggestions, and many were implemented. He believed, first of all, that special education must be tailored to the individual needs of disadvantaged children: it must be based on "their character and their aptitudes, and on the necessity for adapting ourselves to their needs and their capacities" (1909, p. 15). Binet recommended small classrooms of fifteen to twenty students, compared with sixty to eighty then common in public schools catering to poor children. In particular, he advocated special methods of education, including a program that he called "mental orthopedics":

What they should learn first is not the subjects ordinarily taught, however important they may be; they should be given lessons of will, of attention, of discipline; before exercises in grammar, they need to be exercised in mental orthopedics; in a word they must learn how to learn (1908, p. 257).

Binet's interesting program of mental orthopedics included a set of physical exercises designed to improve, by transfer to mental functioning, the will, attention, and discipline that Binet viewed as prerequisites for studying academic subjects. In one, called "*l'exercice des statues*," and designed to increase attention span, children moved vigorously until told to adopt and retain an immobile position. (I played this game as a kid in the streets of New York; we also called it "statues.") Each day the period of immobility would be increased. In another, designed to improve speed, children filled a piece of paper with as many dots as they could produce in the allotted time.

Binet spoke with pleasure about the success of his special classrooms (1909, p. 104) and argued that pupils so benefited had not only increased their knowledge, but their intelligence as well. Intelligence, in any meaningful sense of the word, can be augmented by good education; it is not a fixed and inborn quantity:

It is in this practical sense, the only one accessible to us, that we say that the intelligence of these children has been increased. We have increased what constitutes the intelligence of a pupil: the capacity to learn and to assimilate instruction.

1. The equation of "heritable" with "inevitable." To a biologist, heritability refers to the passage of traits or tendencies along family lines as a result of genetic transmission. It says little about the range of environmental modification to which these traits are subject. In our vernacular, "inherited" often means "inevitable." But not to a biologist. Genes do not make specific bits and pieces of a body; they code for a range of forms under an array of environmental conditions. Moreover, even when a trait has been built and set, environmental intervention may still modify inherited defects. Millions of Americans see normally through lenses that correct innate deficiencies of vision. The claim that IQ is so-many percent "heritable" does not conflict with the belief that enriched education can increase what we call, also in the vernacular, "intelligence." A partially inherited low IQ might be subject to extensive improvement through proper education. And it might not. The mere fact of its heritability permits no conclusion.

2. The confusion of within- and between-group heredity. The major political impact of hereditarian theories does not arise from the inferred heritability of tests, but from a logically invalid extension. Studies of the heritability of IQ, performed by such traditional methods as comparing scores of relatives, or contrasting scores of adopted children with both their biological and legal parents, are all of the "within-group" type—that is, they permit an estimate of heritability *within* a single, coherent population (white Americans, for example). The common fallacy consists in assuming that if heredity explains a certain percentage of variation among individuals within a group, it must also explain a similar percentage of the difference in average IQ between groups—whites and blacks, for example. But variation among individuals within a group and differences in mean values between groups are entirely separate phenomena. One item provides no license for speculation about the other.

A hypothetical and noncontroversial example will suffice. Human height has a higher heritability than any value ever proposed for IQ. Take two separate groups of males. The first, with an average height of 5 feet 10 inches, live in a prosperous American town. The second, with an average height of 5 feet 6 inches, are starving in a third-world village. Heritability is 95 percent or so in each place—meaning only that relatively tall fathers tend to have

tall sons and relatively short fathers short sons. This high within-group heritability argues neither for nor against the possibility that better nutrition in the next generation might raise the average height of third-world villagers above that of prosperous Americans. Likewise, IQ could be highly heritable within groups, and the average difference between whites and blacks in America might still only record the environmental disadvantages of blacks.

I have often been frustrated with the following response to this admonition: "Oh well, I see what you mean, and you're right in theory. There may be no necessary connection in logic, but isn't it more likely all the same that mean differences between groups would have the same causes as variation within groups." The answer is still "no." Within- and between-group heredity are not tied by rising degrees of probability as heritability increases within groups and differences enlarge between them. The two phenomena are simply separate. Few arguments are more dangerous than the ones that "feel" right but can't be justified.

Alfred Binet avoided these fallacies and stuck by his three principles. American psychologists perverted Binet's intention and invented the hereditarian theory of IQ. They reified Binet's scores, and took them as measures of an entity called intelligence. They assumed that intelligence was largely inherited, and developed a series of specious arguments confusing cultural differences with innate properties. They believed that inherited IQ scores marked people and groups for an inevitable station in life. And they assumed that average differences between groups were largely the products of heredity, despite manifest and profound variation in quality of life.

This chapter analyzes the major works of the three pioneers of hereditarianism in America: H. H. Goddard, who brought Binet's scale to America and reified its scores as innate intelligence; L. M. Terman, who developed the Stanford-Binet scale, and dreamed of a rational society that would allocate professions by IQ scores; and R. M. Yerkes, who persuaded the army to test 1.75 million men in World War I, thus establishing the supposedly objective data that vindicated hereditarian claims and led to the Immigration Restriction Act of 1924, with its low ceiling for lands suffering the blight of poor genes.

The hereditarian theory of IQ is a home-grown American

product. If this claim seems paradoxical for a land with egalitarian traditions, remember also the jingoistic nationalism of World War I, the fear of established old Americans facing a tide of cheap (and sometimes politically radical) labor immigrating from southern and eastern Europe, and above all our persistent, indigenous racism.

H. H. Goddard and the menace of the feeble-minded *Intelligence as a Mendelian gene*

GODDARD IDENTIFIES THE MORON

It remains now for someone to determine the nature of feeble-mindedness and complete the theory of the intelligence quotient.

—H. H. GODDARD, 1917, in a review of Terman, 1916

Taxonomy is always a contentious issue because the world does not come to us in neat little packages. The classification of mental deficiency aroused a healthy debate early in our century. Two categories of a tripartite arrangement won general acceptance: idiots could not develop full speech and had mental ages below three; imbeciles could not master written language and ranged from three to seven in mental age. (Both terms are now so entrenched in the vernacular of invectives that few people recognize their technical status in an older psychology.) Idiots and imbeciles could be categorized and separated to the satisfaction of most professionals, for their affliction was sufficiently severe to warrant a diagnosis of true pathology. They are not like us.

But consider the nebulous and more threatening realm of "high-grade defectives"—the people who could be trained to function in society, the ones who established a bridge between pathology and normality and thereby threatened the taxonomic edifice. These people, with mental ages of eight to twelve, were called *débile* (or weak) by the French. Americans and Englishmen usually called them "feeble-minded," a term mired in hopeless ambiguity because other psychologists used feeble-minded as a generic term for all mental defectives, not just those of high grade.

Taxonomists often confuse the invention of a name with the solution of a problem. H. H. Goddard, the energetic and crusading director of research at the Vineland Training School for Feeble-Minded Girls and Boys in New Jersey, made this crucial error. He devised a name for "high-grade" defectives, a word that became

entrenched in our language through a series of jokes that rivaled the knock-knock or elephant jokes of other generations. The metaphorical whiskers on these jokes are now so long that most people would probably grant an ancient pedigree to the name. But Goddard invented the word in our century. He christened these people "morons," from a Greek word meaning foolish.

Goddard was the first popularizer of the Binet scale in America. He translated Binet's articles into English, applied his tests, and agitated for their general use. He agreed with Binet that the tests worked best in identifying people just below the normal range—Goddard's newly christened morons. But the resemblance between Binet and Goddard ends there. Binet refused to define his scores as "intelligence," and wished to identify in order to help. Goddard regarded the scores as measures of a single, innate entity. He wished to identify in order to recognize limits, segregate, and curtail breeding to prevent further deterioration of an endangered American stock, threatened by immigration from without and by prolific reproduction of its feeble-minded within.

A UNILINEAR SCALE OF INTELLIGENCE

The attempt to establish a unilinear classification of mental deficiency, a rising scale from idiots to imbeciles to morons, embodies two common fallacies pervading most theories of biological determinism discussed in this book: the reification of intelligence as a single, measurable entity; and the assumption, extending back to Morton's skulls (pp. 82-101) and forward to Jensen's universal scaling of general intelligence (pp. 347-350), that evolution is a tale of unilinear progress, and that a single scale ascending from primitive to advanced represents the best way of ordering variation. The concept of progress is a deep prejudice with an ancient pedigree (Bury, 1920) and a subtle power, even over those who would deny it explicitly (Nisbet, 1980).

Can the plethora of causes and phenomena grouped under the rubric of mental deficiency possibly be ordered usefully on a single scale, with its implication that each person owes his rank to the relative amount of a single substance—and that mental deficiency means having less than most? Consider some phenomena mixed up in the common numbers once assigned to defectives of high grade: general low-level mental retardation, specific learning disa-

bilities caused by local neurological damage, environmental disadvantages, cultural differences, hostility to testers. Consider some of the potential causes: inherited patterns of function, genetic pathologies arising accidentally and not passed in family lines, congenital brain damage caused by maternal illness during pregnancy, birth traumas, poor nutrition of fetuses and babies, a variety of environmental disadvantages in early and later life. Yet, to Goddard, all people with mental ages between eight and twelve were morons, all to be treated in roughly the same way: institutionalized or carefully regulated, made happy by catering to their limits, and, above all, prevented from breeding.

Goddard may have been the most unobvious hereditarian of all. He used his unilinear scale of mental deficiency to identify intelligence as a single entity, and he assumed that everything important about it was inborn and inherited in family lines. He wrote in 1920 (quoted in Tuddenham, 1962, p. 491):

Stated in its boldest form, our thesis is that the chief determiner of human conduct is a unitary mental process which we call intelligence: that this process is conditioned by a nervous mechanism which is inborn: that the degree of efficiency to be attained by that nervous mechanism and the consequent grade of intellectual or mental level for each individual is determined by the kind of chromosomes that come together with the union of the germ cells: that it is but little affected by any later influences except such serious accidents as may destroy part of the mechanism.

Goddard extended the range of social phenomena caused by differences in innate intelligence until it encompassed almost everything that concerns us about human behavior. Beginning with morons, and working up the scale, he attributed most undesirable behavior to inherited mental deficiency of the offenders. Their problems are caused not only by stupidity *per se*, but by the link between deficient intelligence and immorality.* High intelligence not only permits us to do our sums; it also engenders the good judgment that underlies all moral behavior.

The intelligence controls the emotions and the emotions are controlled in proportion to the degree of intelligence. . . . It follows that if there is

*The link of morality to intelligence was a favorite eugenical theme. Thorndike (1940, pp. 264-265), refuting a popular impression that all monarchs are reprobrates, cited a correlation coefficient of 0.56 for the estimated intelligence vs. the estimated morality of 269 male members of European royal families!

little intelligence the emotions will be uncontrolled and whether they be strong or weak will result in actions that are unregulated, uncontrolled and, as experience proves, usually undesirable. Therefore, when we measure the intelligence of an individual and learn that he has so much less than normal as to come within the group that we call feeble-minded, we have ascertained by far the most important fact about him (1919, p. 272).

Many criminals, most alcoholics and prostitutes, and even the "ne'er do wells" who simply don't fit in, are morons: "We know what feeble-mindedness is, and we have come to suspect all persons who are incapable of adapting themselves to their environment and living up to the conventions of society or acting sensibly, of being feeble-minded" (1914, p. 571).

At the next level of the merely dull, we find the toiling masses, doing what comes naturally. "The people who are doing the drudgery," Goddard writes (1919, p. 246), "are, as a rule, in their proper places."

We must next learn that there are great groups of men, laborers, who are but little above the child, who must be told what to do and shown how to do it; and who, if we would avoid disaster, must not be put into positions where they will have to act upon their own initiative or their own judgment. . . . There are only a few leaders, most must be followers (1919, pp. 243-244).

At the upper end, intelligent men rule in comfort and by right. Speaking before a group of Princeton undergraduates in 1919, Goddard proclaimed:

Now the fact is, that workmen may have a 10 year intelligence while you have a 20. To demand for him such a home as you enjoy is as absurd as it would be to insist that every laborer should receive a graduate fellowship. How can there be such a thing as social equality with this wide range of mental capacity?

"Democracy," Goddard argued (1919, p. 237), "means that the people rule by selecting the wisest, most intelligent and most human to tell them what to do to be happy. Thus Democracy is a method for arriving at a truly benevolent aristocracy."

BREAKING THE SCALE INTO MENDELIAN COMPARTMENTS

But if intelligence forms a single and unbroken scale, how can we solve the social problems that beset us? For at one level, low intelligence generates sociopaths, while at the next grade, indus-

trial society needs docile and dull workers to run its machinery and accept low recompense. How can we convert the unbroken scale into two categories at this crucial point, and still maintain the idea that intelligence is a single, inherited entity? We can now understand why Goddard lavished so much attention upon the moron. The moron threatens racial health because he ranks highest among the undesirable and might, if not identified, be allowed to flourish and propagate. We all recognize the idiot and imbecile and know what must be done; the scale must be broken just above the level of the moron.

The idiot is not our greatest problem. He is indeed loathsome. . . . Nevertheless, he lives his life and is done. He does not continue the race with a line of children like himself. . . . It is the moron type that makes for us our great problem (1912, pp. 101-102).

Goddard worked in the first flourish of excitement that greeted the rediscovery of Mendel's work and the basic deciphering of heredity. We now know that virtually every major feature of our body is built by the interaction of many genes with each other and with an external environment. But in these early days, many biologists naïvely assumed that all human traits would behave like the color, size, or wrinkling of Mendel's peas: they believed, in short, that even the most complex parts of a body might be built by single genes, and that variation in anatomy or behavior would record the different dominant and recessive forms of these genes. Eugenicists seized upon this foolish notion with avidity, for it allowed them to assert that all undesirable traits might be traced to single genes and eliminated with proper strictures upon breeding. The early literature of eugenics is filled with speculations, and pedigrees laboriously compiled and fudged, about the gene for *Wanderlust* traced through the family lines of naval captains, or *the gene for temperament* that makes some of us placid and others domineering. We must not be misled by how silly such ideas seem today; they represented orthodox genetics for a brief time, and had a major social impact in America.

Goddard joined the transient bandwagon with a hypothesis that must represent an ultimate in the attempted reification of intelligence. He tried to trace the pedigrees of mental defectives in his Vineland School and concluded that "feeble-mindedness" obeyed Mendelian rules of inheritance. Mental deficiency must therefore

be a definite thing, and it must be governed by a single gene, undoubtedly recessive to normal intelligence (1914, p. 539). "Normal intelligence," Goddard concluded, "seems to be a unit character and transmitted in true Mendelian fashion" (1914, p. ix).

Goddard claimed that he had been compelled to make this unlikely conclusion by the press of evidence, not by any prior hope or prejudice.

Any theories or hypotheses that have been presented have been merely those that were suggested by the data themselves, and have been worked out in an effort to understand what the data seem to comprise. Some of the conclusions are as surprising to the writer and as difficult for him to accept as they are likely to be to many readers (1914, p. viii).

Can we seriously view Goddard as a forced and reluctant convert to a hypothesis that fit his general scheme so well and solved his most pressing problem so neatly? A single gene for normal intelligence removed the potential contradiction between a unit-linear scale that marked intelligence as a single, measurable entity, and a desire to separate and identify the mentally deficient as a category apart. Goddard had broken his scale into two sections at just the right place: morons carried a double dose of the bad recessive; dull laborers had at least one copy of the normal gene and could be set before their machines. Moreover, the scourge of feeble-mindedness might now be eliminated by schemes of breeding easily planned. One gene can be traced, located, and bred out. If one hundred genes regulate intelligence, eugenic breeding must fail or proceed with hopeless sloth.

THE PROPER CARE AND FEEDING (BUT NOT BREEDING) OF MORONS

If mental deficiency is the effect of a single gene, the path to its eventual elimination lies evidently before us: do not allow such people to bear children:

If both parents are feeble-minded all the children will be feeble-minded. It is obvious that such matings should not be allowed. It is perfectly clear that no feeble-minded person should ever be allowed to marry or to become a parent. It is obvious that if this rule is to be carried out the intelligent part of society must enforce it (1914, p. 561).

If morons could control their own sexual urges and desist for the good of mankind, we might permit them to live freely among us. But they cannot, because immorality and stupidity are inexorably

ably linked. The wise man can control his sexuality in a rational manner: "Consider for a moment the sex emotion, supposed to be the most uncontrollable of all human instincts; yet it is notorious that the intelligent man controls even this" (1919, p. 273). The moron cannot behave in so exemplary and abstemious a fashion:

They are not only lacking in control but they are lacking often in the perception of moral qualities; if they are not allowed to marry they are nevertheless not hindered from becoming parents. So that if we are absolutely to prevent a feeble-minded person from becoming a parent, something must be done other than merely prohibiting the marrying. To this end there are two proposals: the first is colonization, the second is sterilization (1914, p. 566).

Goddard did not oppose sterilization, but he regarded it as impractical because traditional sensibilities of a society not yet wholly rational would prevent such widespread mayhem. Colonization in exemplary institutions like his own at Vineland, New Jersey, must be our preferred solution. Only here could the reproduction of morons be curtailed. If the public balked at the great expense of building so many new centers for confinement, the cost could easily be recouped by its own savings:

If such colonies were provided in sufficient number to take care of all the distinctly feeble-minded cases in the community, they would very largely take the place of our present almshouses and prisons, and they would greatly decrease the numbers in our insane hospitals. Such colonies would save an annual loss in property and life, due to the action of these irresponsible people, sufficient to nearly, or quite, offset the expense of the new plant (1912, pp. 105-106).

Inside these institutions, morons could operate in contentment at their biologically appointed level, denied only the basic biology of their own sexuality. Goddard ended his book on the causes of mental deficiency with this plea for the care of institutionalized morons: "Treat them as children according to their mental age, constantly encourage and praise, never discourage or scold; and keep them happy" (1919, p. 327).

Preventing the immigration and propagation of morons

Once Goddard had identified the cause of feeble-mindedness in a single gene, the cure seemed simple enough: don't allow native

morons to breed and keep foreign ones out. As a contribution to the second step, Goddard and his associates visited Ellis Island in 1912 "to observe conditions and offer any suggestions as to what might be done to secure a more thorough examination of immigrants for the purpose of detecting mental defectives" (Goddard, 1917, p. 253).

As Goddard described the scene, a fog hung over New York harbor that day and no immigrants could land. But one hundred were about ready to leave, when Goddard intervened: "We picked out one young man whom we suspected was defective, and, through the interpreter, proceeded to give him the test. The boy tested 8 by the Binet scale. The interpreter said, 'I could not have done that when I came to this country,' and seemed to think the test unfair. We convinced him that the boy was defective" (Goddard, 1913, p. 105).

Encouraged by this, one of the first applications of the Binet scale in America, Goddard raised some funds for a more thorough study and, in the spring of 1913, sent two women to Ellis Island for two and a half months. They were instructed to pick out the feeble-minded by sight, a task that Goddard preferred to assign to women, to whom he granted innately superior intuition:

After a person has had considerable experience in this work, he almost gets a sense of what a feeble-minded person is so that he can tell one afar off. The people who are best at this work, and who I believe should do this work, are women. Women seem to have closer observation than men. It was quite impossible for others to see how these two young women could pick out the feeble-minded without the aid of the Binet test at all (1913, p. 106).

Goddard's women tested thirty-five Jews, twenty-two Hungarians, fifty Italians, and forty-five Russians. These groups could not be regarded as random samples because government officials had already "culled out those they recognized as defective." To balance this bias, Goddard and his associates "passed by the obviously normal. That left us the great mass of 'average immigrants.'" (1917, p. 244). (I am continually amazed by the unconscious statements of prejudice that slip into supposedly objective accounts. Note here that average immigrants are below normal, or at least not obviously normal—the proposition that Goddard was supposedly testing, not asserting a priori.)

Binet tests on the four groups led to an astounding result: 83 percent of the Jews, 80 percent of the Hungarians, 79 percent of the Italians, and 87 percent of the Russians were feeble-minded—that is, below age twelve on the Binet scale. Goddard himself was flabbergasted: could anyone be made to believe that four-fifths of any nation were morons? "The results obtained by the foregoing evaluation of the data are so surprising and difficult of acceptance that they can hardly stand by themselves as valid" (1917, p. 247). Perhaps the tests had not been adequately explained by interpreters? But the Jews had been tested by a Yiddish-speaking psychologist, and they ranked no higher than the other groups. Eventually, Goddard monkied about with the tests, tossed several out, and got his figures down to 40 to 50 percent, but still he was disturbed.

Goddard's figures were even more absurd than he imagined for two reasons, one obvious, the other less so. As a nonevident reason, Goddard's original translation of the Binet scale scored people harshly and made morons out of subjects usually regarded as normal. When Terman devised the Stanford-Binet scale in 1916, he found that Goddard's version ranked people well below his own. Terman reports (1916, p. 62) that of 104 adults tested by him as between twelve and fourteen years mental age (low, but normal intelligence), 50 percent were morons on the Goddard scale.

For the evident reason, consider a group of frightened men and women who speak no English and who have just endured an oceanic voyage in steerage. Most are poor and have never gone to school; many have never held a pencil or pen in their hand. They march off the boat; one of Goddard's intuitive women takes them aside shortly thereafter, sits them down, hands them a pencil, and asks them to reproduce on paper a figure shown to them a moment ago, but now withdrawn from their sight. Could their failure be a result of testing conditions, of weakness, fear, or confusion, rather than of innate stupidity? Goddard considered the possibility, but rejected it:

The next question is 'drawing a design from memory,' which is passed by only 50 percent. To the uninitiated this will not seem surprising since it looks hard, and even those who are familiar with the fact that normal children of 10 pass it without difficulty may admit that persons who have never had a pen or pencil in their hands, as was true of many of the immigrants, may find it impossible to draw the design (1917, p. 250).

Permitting a charitable view of this failure, what but stupidity could explain an inability to state more than sixty words, any words, in one's own language during three minutes?

What shall we say of the fact that only 45 percent can give 60 words in three minutes, when normal children of 11 years sometimes give 200 words in that time! It is hard to find an explanation except lack of intelligence or lack of vocabulary, and such a lack of vocabulary in an adult would probably mean lack of intelligence. How could a person live even 15 years in any environment without learning hundreds of names of which he could certainly think of 60 in three minutes? (1917, p. 251)

Or ignorance of the date, or even the month or year?

Must we again conclude that the European peasant of the type that immigrates to America pays no attention to the passage of time? That the drudgery of life is so severe that he cares not whether it is January or July, whether it is 1912 or 1906? Is it possible that the person may be of considerable intelligence and yet, because of the peculiarity of his environment, not have acquired this ordinary bit of knowledge, even though the calendar is not in general use on the continent, or is somewhat complicated as in Russia? If so what an environment it must have been! (1917, p. 250)

Since environment, either European or immediate, could not explain such abject failure, Goddard stated: "We cannot escape the general conclusion that these immigrants were of surprisingly low intelligence" (1917, p. 251). The high proportion of morons still bothered Goddard, but he finally attributed it to the changing character of immigration: "It should be noted that the immigration of recent years is of a decidedly different character from the early immigration. . . . We are now getting the poorest of each race" (1917, p. 266). "The intelligence of the average 'third class' immigrant is low, perhaps of moron grade" (1917, p. 243). Perhaps, Goddard hoped out loud, things were better on the upper decks, but he did not test these wealthier customers.

What then should be done? Should all these morons be shipped back, or prevented from starting out in the first place? Forshad-owing the restrictions that would be legislated within a decade, Goddard argued that his conclusions "furnish important considerations for future actions both scientific and social as well as legislative" (1917, p. 261). But by this time Goddard had softened his earlier harsh position on the colonization of morons. Perhaps there

were not enough merely dull workers to fill the vast number of frankly undesirable jobs. The moron might have to be recruited: "They do a great deal of work that no one else will do. . . . There is an immense amount of drudgery to be done, an immense amount of work for which we do not wish to pay enough to secure more intelligent workers. . . . May it be that possibly the moron has his place" (1917, p. 269).

Nonetheless, Goddard rejoiced in the general tightening of standards for admission. He reports that deportations for mental deficiency increased 350 percent in 1913 and 570 percent in 1914 over the average of the five preceding years:

This was due to the untiring efforts of the physicians who were inspired by the belief that mental tests could be used for the detection of feeble-minded aliens. . . . If the American public wishes feeble-minded aliens excluded, it must demand that congress provide the necessary facilities at the ports of entry (1917, p. 271).

Meanwhile, at home, the feeble-minded must be identified and kept from breeding. In several studies, Goddard exposed the menace of morosity by publishing pedigrees of hundreds of worthless souls, charges upon the state and community, who would never have been born had their feeble-minded forebears been debarred from reproduction. Goddard discovered a stock of paupers and ne'er-do-wells in the pine barrens of New Jersey and traced their ancestry back to the illicit union of an upstanding man with a supposedly feeble-minded tavern wench. The same man later married a worthy Quakeress and started another line composed wholly of upstanding citizens. Since the progenitor had fathered both a good and a bad line, Goddard combined the Greek words for beauty (*kallios*) and bad (*kakos*), and awarded him the pseudonym Martin Kallikak. Goddard's Kallikak family functioned as a primal myth of the eugenics movement for several decades.

Goddard's study is little more than guesswork rooted in conclusions set from the start. His method, as always, rested upon the training of intuitive women to recognize the feeble-minded by sight. Goddard did not administer Binet tests in pine-barren shacks. Goddard's faith in visual identification was virtually unbounded. In 1919 he analyzed Edwin Markham's poem "The Man With The Hoe":

Bowed by the weight of centuries he leans
Upon his hoe and gazes at the ground,
The emptiness of ages in his face
And on his back the burden of the world. . . .

Markham's poem had been inspired by Millet's famous painting of the same name. The poem, Goddard complained (1919, p. 239), "seems to imply that the man Millet painted came to his condition as the result of social conditions which held him down and made him like the clods that he turned over." Nonsense, exclaimed Goddard; most poor peasants suffer only from their own feeble-mindedness, and Millet's painting proves it. Couldn't Markham see that the peasant is mentally deficient? "Millet's Man With The Hoe is a man of arrested mental development—the painting is a perfect picture of an imbecile" (1919, pp. 239–240). To Markham's searing question: "Whose breath blew out the light within this brain?" Goddard replied that mental fire had never been kindled.

Since Goddard could determine degrees of mental deficiency by examining a painting, he certainly anticipated no trouble with flesh and blood. He dispatched the redoubtable Ms. Kite, soon to see further service on Ellis Island, to the pine barrens and quickly produced the sad pedigree of the kakos line. Goddard describes one of Ms. Kite's identifications (1912, pp. 77–78):

Used as she was to the sights of misery and degradation, she was hardly prepared for the spectacle within. The father, a strong, healthy, broad-shouldered man, was sitting helplessly in a corner. . . . Three children, scantily clad and with shoes that would barely hold together, stood about with drooping jaws and the unmistakable look of the feeble-minded. . . . The whole family was a living demonstration of the futility of trying to make desirable citizens from defective stock through making and enforcing compulsory education laws. . . . The father himself, though strong and vigorous, showed by his face that he had only a child's mentality. The mother in her filth and rags was also a child. In this house of abject poverty, only one sure prospect was ahead, that it would produce more feeble-minded children with which to clog the wheels of human progress.

If these spot identifications seem a bit hasty or dubious, consider Goddard's method for inferring the mental state of the departed, or otherwise unavailable (1912, p. 15):

Goddard recounts

By 1928 Goddard had changed his mind and become a latter-day supporter of the man whose work he had originally perverted, Alfred Binet. Goddard admitted, first of all, that he had set the upper limit of moronity far too high:

It was for a time rather carelessly assumed that everybody who tested 12 years or less was feeble-minded. . . . We now know, of course, that only a small percentage of the people who test 12 are actually feeble-minded—that is, are incapable of managing their affairs with ordinary prudence or of competing in the struggle for existence (1928, p. 220).

But genuine morons still abound at their redefined level. What shall be done with them? Goddard did not abandon his belief in their inherited mentality, but he now took Binet's line and argued that most, if not all, could be trained to lead useful lives in society:

The problem of the moron is a problem of education and training. . . . This may surprise you, but frankly when I see what has been made out of the moron by a system of education, which as a rule is only half right, I have no difficulty in concluding that when we get an education that is entirely right there will be no morons who cannot manage themselves and their affairs and compete in the struggle for existence. If we could hope to add to this a social order that would literally give every man a chance, I should be perfectly sure of the result (1928, pp. 223-224).

But if we let morons live in society, will they not marry and bear children; is this not the greatest danger of all, the source of Goddard's previous and passionate warnings?

Some will object that this plan neglects the eugenic aspect of the problem. In the community, these morons will marry and have children. And why not? . . . It may still be objected that moron parents are likely to have imbecile or idiot children. There is not much evidence that this is the case. The danger is probably negligible. At least it is not likely to occur any

5•2 Altered photographs of members of the Kallikak family living in poverty in the New Jersey pine barrens. Note how mouths and eyebrows are accentuated to produce an appearance of evil or stupidity. The effect is much clearer on the original photographs produced in Goddard's book.



oftener than it does in the general population.* I assume that most of you, like myself, will find it difficult to admit that the foregoing may be the true view. We have worked too long under the old concept (1928, pp. 223-224).

Goddard concluded (1928, p. 225) in reversing the two bulwarks of his former system:

1. Feeble-mindedness (the moron) is *not incurable* [Goddard's italics].
2. The feeble-minded do not generally need to be segregated in institutions.

"As for myself," Goddard confessed (p. 224), "I think I have gone over to the enemy."

Lewis M. Terman and the mass marketing of innate IQ

Without offering any data on all that occurs between conception and the age of kindergarten, they announce on the basis of what they have got out of a few thousand questionnaires that they are measuring the hereditary mental endowment of human beings. Obviously, this is not a conclusion obtained by research. It is a conclusion planted by the will to believe. It is, I think, for the most part unconsciously planted. . . . If the impression takes root that these tests really measure intelligence, that they constitute a sort of last judgment on the child's capacity, that they reveal "scientifically" his predestined ability, then it would be a thousand times better if all the intelligence testers and all their questionnaires were sunk without warning in the Sargasso Sea.

—WALTER LIPPMANN, in the course of a debate with Lewis Terman

Mass testing and the Stanford-Binet

Lewis M. Terman, the twelfth child in an Indiana farm family of fourteen, traced his interest in the study of intelligence to an itinerant book peddler and phrenologist who visited his home when he was nine or ten and predicted good things after feeling the bumps on his skull. Terman pursued this early interest, never doubting that a measurable mental worth lay inside people's heads. In his doctoral dissertation of 1906, Terman examined seven "bright" and seven "stupid" boys and defended each of his tests as a measure of intelligence by appealing to the standard catalogue of

* Do not read into this statement more than Goddard intended. He had not abandoned his belief in the heritability of morosity itself. Moron parents will have moron children, but they can be made useful through education. Moron parents, however, do not preferentially beget defectives of lower grade—idiots and imbeciles.

racial and national stereotypes. Of tests for invention, he wrote: "We have only to compare the negro with the Eskimo or Indian, and the Australian native with the Anglo-Saxon, to be struck by an apparent kinship between general intellectual and inventive ability" (1906, p. 14). Of mathematical ability, he proclaimed (1906, p. 29): "Ethnology shows that racial progress has been closely paralleled by development of the ability to deal with mathematical concepts and relations."

Terman concluded his study by committing both of the fallacies identified on p. 185 as foundations of the hereditarian view. He reified average test scores as a "thing" called general intelligence by advocating the first of two possible positions (1906, p. 9): "Is intellectual ability a bank account, on which we can draw for any desired purpose, or is it rather a bundle of separate drafts, each drawn for a specific purpose and unconvertible?" And, while admitting that he could provide no real support for it, he defended the innatist view (1906, p. 68): "While offering little positive data on the subject, the study has strengthened my impression of the relatively greater importance of endowment over training as a determinant of an individual's intellectual rank among his fellows."

Goddard introduced Binet's scale to America, but Terman was the primary architect of its popularity. Binet's last version of 1911 included fifty-four tasks, graded from pre nursery to mid-teen-age years. Terman's first revision of 1916 extended the scale to "superior adults" and increased the number of tasks to ninety. Terman, by then a professor at Stanford University, gave his revision a name that has become part of our century's vocabulary—the Stanford-Binet, the standard for virtually all "IQ" tests that followed.*

I offer no detailed analysis of content (see Block and Dworkin, 1976 or Chase, 1977), but present two examples to show how Terman's tests stressed conformity with expectation and downgraded original response. When expectations are society's norms, then do

* Terman (1919) provided a lengthy list of the attributes of general intelligence captured by the Stanford-Binet tests: memory, language comprehension, size of vocabulary, orientation in space and time, eye-hand coordination, knowledge of familiar things, judgment, likeness and differences, arithmetical reasoning, resourcefulness and ingenuity in difficult practical situations, ability to detect absurdities, speed and richness of association of ideas, power to combine the dissected parts of a form board or a group of ideas into a unitary whole, capacity to generalize from particulars, and ability to deduce a rule from connected facts.

the tests measure some abstract property of reasoning, or familiarity with conventional behavior? Terman added the following item to Binet's list:

An Indian who had come to town for the first time in his life saw a white man riding along the street. As the white man rode by, the Indian said—"The white man is lazy; he walks sitting down." What was the white man riding on that caused the Indian to say, "He walks sitting down."

Terman accepted "bicycle" as the only correct response—not cars or other vehicles because legs don't go up and down in them; not horses (the most common "incorrect" answer) because any self-respecting Indian would have known what he was looking at. (I myself answered "horse," because I saw the Indian as a clever ironist, criticizing an effete city relative.) Such original responses as "a cripple in a wheel chair," and "a person riding on someone's back" were also marked wrong.

Terman also included this item from Binet's original: "My neighbor has been having queer visitors. First a doctor came to his house, then a lawyer, then a minister. What do you think happened there?" Terman permitted little latitude beyond "a death," though he did allow "a marriage" from a boy he described as "an enlightened young eugenicist" who replied that the doctor came to see if the partners were fit, the lawyer to arrange, and the minister to tie the knot. He did not accept the combination "divorce and remarriage," though he reports that a colleague in Reno, Nevada, had found the response "very, very common." He also did not permit plausible but uncomplicated solutions (a dinner, or an entertaining married and making his will before he dies.)

But Terman's major influence did not reside in his sharpening or extension of the Binet scale. Binet's tasks had to be administered by a trained tester working with one child at a time. They could not be used as instruments for general ranking. But Terman wished to test everybody, for he hoped to establish a gradation of innate ability that could sort all children into their proper stations in life:

What pupils shall be tested? The answer is, all. If only selected children are tested, many of the cases most in need of adjustment will be over-

looked. The purpose of the tests is to tell us what we do not already know, and it would be a mistake to test only those pupils who are recognized as obviously below or above average. Some of the biggest surprises are encountered in testing those who have been looked upon as close to average in ability. Universal testing is fully warranted (1923, p. 22).

The Stanford-Binet, like its parent, remained a test for individuals, but it became the paradigm for virtually all the written versions that followed. By careful juggling and elimination,* Terman standardized the scale so that "average" children would score 100 at each age (mental age equal to chronological age). Terman also evened out the variation among children by establishing a standard deviation of 15 or 16 points at each chronological age. With its mean of 100 and standard deviation of 15, the Stanford-Binet became (and in many respects remains to this day) the primary criterion for judging a plethora of mass-marketed written tests that followed. The invalid argument runs: we know that the Stanford-Binet measures intelligence; therefore, any written test that correlates strongly with Stanford-Binet also measures intelligence. Much of the elaborate statistical work performed by testers during the past fifty years provides no independent confirmation for the proposition that tests measure intelligence, but merely establishes correlation with a preconceived and unquestioned standard.

Testing soon became a multimillion-dollar industry; marketing companies dared not take a chance with tests not proven by their correlation with Terman's standard. The Army Alpha (see pp. 222-252) initiated mass testing, but a flood of competitors greeted school administrators within a few years after the war's end. A quick glance at the advertisements appended to Terman's later book (1923) illustrates, dramatically and unintentionally, how all Terman's cautious words about careful and lengthy assessment (1919, p. 299, for example) could evaporate before strictures of cost and time when his desire to test all children became a reality (Fig. 5-3). Thirty minutes and five tests might mark a child for life, if schools adopted the following examination, advertised in Terman 1923, and constructed by a committee that included Thorndike, Yerkes, and Terman himself.

*This, in itself, is not finagling, but a valid statistical procedure for establishing uniformity of average score and variance across age levels.

Prepared under the auspices of the National Research Council

NATIONAL INTELLIGENCE TESTS

By M. E. HAGGERTY, L. M. TERMAN, E. L. THORNDIKE
G. M. WHIFFLE, and R. M. YERKES

THESSE tests are the direct result of the application of the army testing methods to school needs. They were devised in order to supply group tests for the examination of school children that would embody the greater benefits derived from the Binet and similar tests.

The effectiveness of the army intelligence tests in problems of classification and diagnosis is a measure of the success that may be expected to attend the use of the National Intelligence Tests, which have been greatly improved in the light of army experiences.

The tests have been selected from a large group of tests after a try-out and a careful analysis by a statistical staff. The two scales prepared consist of five tests each (with practice exercises), and either may be administered in thirty minutes. They are simple in application, reliable, and immediately useful for classifying children in Grades 3 to 8 with respect to intellectual ability. Scoring is unusually simple.

Either scale may be used separately to advantage. The reliability of results is increased, however, by reexamination with the other scale after an interval of at least a day.

Scale A consists of an arithmetical reasoning, a sentence completion, a logical selection, a synonym-antonym, and a symbol-digit test. Scale B includes a completion, an information, a vocabulary, an analogies, and a comparison test.

Scale A: Form 1. 12 pages. Price per package of 25 Examination Booklets, 2 Scoring Keys, and 1 Class Record \$1.45 net.

Scale A: Form 2. Same description. Same price.

Scale B: Form 1. 12 pages. Price per package of 25 Examination Booklets, Scoring Key, and Class Record \$1.45 net.

Scale B: Form 2. Same description. Same price.

Manual of Directions. Paper. 32 pages. Price 25 cents net.

Specimen Set. One copy of each Scale and Scoring Keys and Manual of Directions. Price 50 cents postpaid.

Experimental work financed by the General Education Board

by appropriation of \$25,000

WORLD BOOK COMPANY

YONKERS-ON-HUDSON, NEW YORK
2126 PRAIRIE AVENUE, CHICAGO

National Intelligence Tests for Grades 3-8

The direct result of the application of the army testing methods to school needs. . . . The tests have been selected from a large group of tests after a try-out and a careful analysis by a statistical staff. The two scales prepared consist of five tests each (with practical exercises) and either may be administered in thirty minutes. They are simple in application, reliable, and immediately useful in classifying children in Grades 3 to 8 with respect to intellectual ability. Scoring is unusually simple.

Binet, had he lived, might have been distressed enough by such a superficial assessment, but he would have reacted even more strongly against Terman's intent. Terman agreed with Binet that the tests worked best for identifying "high-grade defectives," but his reasons for so doing stand in chilling contrast with Binet's desire to segregate and help (1916, pp. 6-7):

It is safe to predict that in the near future intelligence tests will bring tens of thousands of these high-grade defectives under the surveillance and protection of society. This will ultimately result in curtailing the reproduction of feeble-mindedness and in the elimination of an enormous amount of crime, pauperism, and industrial inefficiency. It is hardly necessary to emphasize that the high-grade cases, of the type now so frequently overlooked, are precisely the ones whose guardianship it is most important for the State to assume.

Terman relentlessly emphasized limits and their inevitability. He needed less than an hour to crush the hopes and belittle the efforts of struggling, "well-educated" parents afflicted with a child of IQ 75.

Strange to say, the mother is encouraged and hopeful because she sees that her boy is learning to read. She does not seem to realize that at his age he ought to be within three years of entering high school. The forty-minute test has told more about the mental ability of this boy than the intelligent mother had been able to learn in eleven years of daily and hourly observation. For X is feeble-minded; he will never complete the grammar school; he will never be an efficient worker or a responsible citizen (1916).

Walter Lippmann, then a young journalist, saw through Terman's numbers to the heart of his preconceived attempt, and wrote in measured anger:

The danger of the intelligence tests is that in a wholesale system of education, the less sophisticated or the more prejudiced will stop when

5 • 3 An advertisement for mass mental testing using an examination written by, among others, Terman and Yerkes.

they have classified and forget that their duty is to educate. They will grade the retarded child instead of fighting the causes of his backwardness. For the whole drift of the propaganda based on intelligence testing is to treat people with low intelligence quotients as congenitally and hopelessly inferior.

Terman's technocracy of innateness

If it were true, the emotional and worldly satisfactions in store for the intelligence tester would be very great. If he were really measuring intelligence, and if intelligence were a fixed hereditary quantity, it would be for him to say not only where to place each child in school, but also which children should go to high school, which to college, which into the professions, which into the manual trades and common labor. If the tester would make good his claim, he would soon occupy a position of power which no intellectual has held since the collapse of theocracy. The vista is enchanting, and even a little of the vista is intoxicating enough. If only it could be proved, or at least believed, that intelligence is fixed by heredity, and that the tester can measure it, what a future to dream about! The unconscious temptation is too strong for the ordinary critical defenses of the scientific methods. With the help of a subtle statistical illusion, intricate logical fallacies and a few straggled *obiter dicta*, self-deception as the preliminary to public deception is almost automatic. — WALTER LIPPMAN, in a debate with Terman

Plato had dreamed of a rational world ruled by philosopher-kings. Terman revived this dangerous vision but led his corps of mental testers in an act of usurpation. If all people could be tested, and then sorted into roles appropriate for their intelligence, then a just, and, above all, efficient society might be constructed for the first time in history.

Dealing off the bottom, Terman argued that we must first restrain or eliminate those whose intelligence is too low for an effective or moral life. The primary cause of social pathology is innate feeble-mindedness. Terman (1916, p. 7) criticized Lombroso for thinking that the externalities of anatomy might record criminal behavior. Innateness, to be sure, is the source, but its direct sign is low IQ, not long arms or a jutting jaw:

The theories of Lombroso have been wholly discredited by the results of intelligence tests. Such tests have demonstrated, beyond any possibility of doubt, that the most important trait of at least 25 percent of our criminals is mental weakness. The physical abnormalities which have been found so common among prisoners are not the stigmata of criminality, but the physical accompaniments of feeble-mindedness. They have no diagnostic significance except in so far as they are indications of mental deficiency (1916, p. 7).

Feeble-minded people are doubly burdened by their unfortunate inheritance, for lack of intelligence, debilitating enough in itself, leads to immorality. If we would eliminate social pathology, we must identify its cause in the biology of sociopaths themselves—and then eliminate them by confinement in institutions and, above all, by preventing their marriage and the production of offspring.

Not all criminals are feeble-minded, but all feeble-minded persons are at least potential criminals. That every feeble-minded woman is a potential prostitute would hardly be disputed by anyone. Moral judgment, like business judgment, social judgment, or any other kind of higher thought process, is a function of intelligence. Morality cannot flower and fruit if intelligence remains infantile (1916, p. 11).

The feeble-minded, in the sense of social incompetents, are by definition a burden rather than an asset, not only economically but still more because of their tendencies to become delinquent or criminal. . . . The only effective way to deal with the hopelessly feeble-minded is by permanent custodial care. The obligations of the public school rest rather with the large and more hopeful group of children who are merely inferior (1919, pp. 132-133).

In a plea for universal testing, Terman wrote (1916, p. 12): "Considering the tremendous cost of vice and crime, which in all probability amounts to not less than \$500,000,000 per year in the United States alone, it is evident that psychological testing has found here one of its richest applications."

After marking the sociopath for removal from society, intelligence tests might then channel biologically acceptable people into professions suited for their mental level. Terman hoped that his testers would "determine the minimum 'intelligence quotient' necessary for success in each leading occupation" (1916, p. 17). Any conscientious professor tries to find jobs for his students, but few are audacious enough to tout their disciples as apostles of a new social order:

Industrial concerns doubtless suffer enormous losses from the employment of persons whose mental ability is not equal to the tasks they are expected to perform. . . . Any business employing as many as 500 or 1000 workers, as, for example, a large department store, could save in this way several times the salary of a well-trained psychologist.

Terman virtually closed professions of prestige and monetary reward to people with IQ below 100 (1919, p. 28a), and argued

that "substantial success" probably required an IQ above 115 or 120. But he was more interested in establishing ranks at the low end of the scale, among those he had deemed "merely inferior." Modern industrial society needs its technological equivalent of the Biblical metaphor for more bucolic times—the hewers of wood and drawers of water. And there are so many of them:

The evolution of modern industrial organization together with the mechanization of processes by machinery is making possible the larger and larger utilization of inferior mentality. One man with ability to think and plan guides the labor of ten or twenty laborers, who do what they are told to do and have little need for resourcefulness or initiative (1919, p. 276).

IQ of 75 or below should be the realm of unskilled labor, 75 to 85 "preeminently the range for semi-skilled labor." More specific judgments could also be made. "Anything above 85 IQ in the case of a barber probably represents so much dead waste" (1919, p. 288). IQ 75 is an "unsafe risk in a motorman or conductor, and it conduces to discontent" (Terman, 1919). Proper vocational training and placement is essential for those "of the 70 to 85 class." Without it, they tend to leave school "and drift easily into the ranks of the anti-social or join the army of Bolshevik discontents" (1919, p. 285).

Terman investigated IQ among professions and concluded with satisfaction that an imperfect allocation by intelligence had already occurred naturally. The embarrassing exceptions he explained away. He studied 47 express company employees, for example, men engaged in rote, repetitive work "offering exceedingly limited opportunity for the exercise of ingenuity or even personal judgment" (1919, p. 275). Yet their median IQ stood at 95 and fully 25 percent measured above 104, thus winning a place among the ranks of the intelligent. Terman was puzzled, but attributed such low achievement primarily to a lack of "certain emotional, moral, or other desirable qualities," though he admitted that "economic pressures" might have forced some "out of school before they were able to prepare for more exacting service" (1919, p. 275). In another study, Terman amassed a sample of 256 "hoboes and unemployed," largely from a "hobo hotel" in Palo Alto. He expected to find their average IQ at the bottom of his list; yet, while the hobo mean of 89 did not suggest enormous endowment, they still ranked above motormen, salesgirls, firemen, and

policemen. Terman suppressed this embarrassment by ordering his table in a curious way. The hobo mean was distressingly high, but hobos also varied more than any other group, and included a substantial number of rather low scores. So Terman arranged his list by the scores of the lowest 25 percent in each group, and sunk his hobos into the cellar.

Had Terman merely advocated a meritocracy based on achievement, one might still decry his elitism, but applaud a scheme that awarded opportunity to hard work and strong motivation. But Terman believed that class boundaries had been set by innate intelligence. His coordinated rank of professions, prestige, and salaries reflected the biological worth of existing social classes. If barbers did not remain Italian, they would continue to arise from the poor and to stay appropriately among them:

The common opinion that the child from a cultured home does better in tests solely by reason of his superior home advantages is an entirely gratuitous assumption. Practically all of the investigations which have been made of the influence of nature and nurture on mental performance agree in attributing far more to original endowment than to environment. Common observation would itself suggest that the social class to which the family belongs depends less on chance than on the parents' native qualities of intellect and character. . . . The children of successful and cultured parents test higher than children from wretched and ignorant homes for the simple reason that their heredity is better (1916, p. 115).

Fossil IQ's of past geniuses

Society may need masses of the "merely inferior" to run its machines, Terman believed, but its ultimate health depends upon the leadership of rare geniuses with elevated IQ's. Terman and his associates published a five-volume series on *Genetic Studies of Genius* in an attempt to define and follow people at the upper end of the Stanford-Binet scale.

In one volume, Terman decided to measure, retrospectively, the IQ of history's prime movers—its statesmen, soldiers, and intellectuals. If they ranked at the top, then IQ is surely the single measure of ultimate worth. But how can a fossil IQ be recovered without conjuring up young Copernicus and asking him what the white man was riding? Undaunted, Terman and his colleagues tried to reconstruct the IQ of past notables, and published a thick book (Cox, 1926) that must rank as a primary curiosity within a

Cox's dossiers are motley lists of childhood and youthful accomplishments, with an emphasis on examples of precocity. Since her method involved adding to the base figure of 100 for each notable item in the dossier, estimated IQ records little more than the volume of available information. In general, low IQ's reflect an absence of information, and high IQ's an extensive list. (Cox even admits that she is not measuring true IQ, but only what can be deduced from limited data, though this disclaimer was invariably lost in translation to popular accounts.) To believe, even for a moment, that such a procedure can recover the proper ordering of IQ among "men of genius," one must assume that the childhood of all subjects was watched and recorded with roughly equal attention. One must claim (as Cox does) that an absence of documented childhood precocity indicates a humdrum life not worth writing about, not an extraordinary giftedness that no one bothered to record.

Two basic results of Cox's study immediately arouse our strong suspicion that her IQ scores reflect the historical accidents of surviving records, rather than the true accomplishments of her geniuses. First, IQ is not supposed to alter in a definite direction during a person's life. Yet average A1 IQ is 135 in her study, and average A2 IQ is a substantially higher 145. When we scrutinize her dossiers (printed in full in Cox, 1926), the reason is readily apparent, and a clear artifact of her method. She has more information on her subjects as young adults than as children (A2 IQ records achievements during ages seventeen through twenty-six; A1 IQ marks the earlier years). Second, Cox published disturbingly low A1 IQ figures for some formidable characters, including Cervantes and Copernicus, both at 105. Her dossiers show the reason: little or nothing is known about their childhood, providing no data for addition to the base figure of 100. Cox established seven levels of reliability for her figures. The seventh, believe it or not, is "guess, based on no data."

As a further and obvious test, consider geniuses born into humble circumstances, where tutors and scribes did not abound to encourage and then to record daring feats of precocity. John Stuart Mill may have learned Greek in his cradle, but did Faraday or Bunyan ever get the chance? Poor children are at a double disadvantage; not only did no one bother to record their early years,

literature already studded with absurdity—though Jensen (1979, pp. 113 and 355) and others still take it seriously.*

Terman (1917) had already published a preliminary study of Francis Galton and awarded a staggering IQ of 200 to this pioneer of mental testing. He therefore encouraged his associates to proceed with a larger investigation. J. M. Cattell had published a ranking of the 1,000 prime movers of history by measuring the lengths of their entries in biographical dictionaries. Catherine M. Cox, Terman's associate, whittled the list to 282, assembled detailed biographical information about their early life, and proceeded to estimate two IQ values for each—one, called A1 IQ, for birth to seventeen years; the other, A2 IQ, for ages seventeen to twenty-six.

Cox ran into problems right at the start. She asked five people, including Terman, to read her dossiers and to estimate the two IQ scores for each person. Three of the five agreed substantially in their mean values, with A1 IQ clustering around 135 and A2 IQ near 145. But two of the raters differed markedly, one awarding an average IQ well above, the other well below, the common figure. Cox simply eliminated their scores, thereby throwing out 40 percent of her data. Their low and high scores would have balanced each other at the mean in any case, she argued (1926, p. 72). Yet if five people working in the same research group could not agree, what hope for uniformity or consistency—not to mention objectivity—could be offered?

Apart from these debilitating practical difficulties, the basic logic of the study was hopelessly flawed from the first. The differences in IQ that Cox recorded among her subjects do not measure their varying accomplishments, not to mention their native intelligence. Instead, the differences are a methodological artifact of the varying quality of information that Cox was able to compile about the childhood and early youth of her subjects. Cox began by assigning a base IQ of 100 to each individual; the raters then added to (or, rarely, subtracted from) this value according to the data provided.

* Jensen writes: "The average estimated IQ of three hundred historical persons . . . on whom sufficient childhood evidence was available for a reliable estimate was IQ 156. . . . Thus the majority of these eminent men would most likely have been recognized as intellectually gifted in childhood had they been given IQ tests" (Jensen, 1979, p. 113).

but they are also demoted as a direct result of their poverty. For Cox, using the favorite ploy of eugenicists, inferred innate parental intelligence from their occupations and social standing! She ranked parents on a scale of professions from 1 to 5, awarding their children an IQ of 100 for parental rank 3, and a bonus (or deficit) of 10 IQ points for each step above or below. A child who did nothing worth noting for the first seventeen years of his life could still score an IQ of 120 by virtue of his parent's wealth or professional standing.

Consider the case of poor Massena, Napoleon's great general, who bottomed out at 100 A1 IQ and about whom, as a child, we know nothing except that he served as a cabin boy for two long voyages on his uncle's ship. Cox writes (p. 88):

Nephews of battleship commanders probably rate somewhat above 100 IQ; but cabin boys who remain cabin boys for two long voyages and of whom there is nothing more to report until the age of 17 than their service as cabin boys, may average below 100 IQ.

Other admirable subjects with impoverished parents and meager records should have suffered the ignominy of scores below 100. But Cox managed to fudge and temporize, pushing them all above the triple-digit divide, if only slightly. Consider the unfortunate Saint-Cyr, saved only by remote kin, and granted an A1 IQ of 105: "The father was a tanner after having been a butcher, which would give his son an occupational IQ status of 90 to 100; but two distant relatives achieved signal martial honors, thus indicating a higher strain in the family" (pp. 90-91). John Bunyan faced more familial obstacles than his famous Pilgrim, but Cox managed to extract a score of 105 for him:

Bunyan's father was a brazier or tinker, but a tinker of recognized position in the village; and the mother was not of the squalid poor, but of people who were "decent and worthy in their ways." This would be sufficient evidence for a rating between 90 and 100. But the record goes further, and we read that notwithstanding their "meanness and inconsiderableness," Bunyan's parents put their boy to school to learn "both to read and write," which probably indicates that he showed something more than the promise of a future tinker (p. 90).

Michael Faraday squeaked by at 105, overcoming the demerit of parental standing with snippets about his reliability as an errand

boy and his questioning nature. His elevated A2 IQ of 150 only records increasing information about his more notable young manhood. In one case, however, Cox couldn't bear to record the unpleasant result that her methods dictated. Shakespeare, of humble origin and unknown childhood, would have scored below 100. So Cox simply left him out, even though she included several others with equally inadequate childhood records.

Among other curiosities of scoring that reflect Cox and Terman's social prejudices, several precocious youngsters (Clive, Liebig, and Swift, in particular) were downgraded for their rebelliousness in school, particularly for their unwillingness to study classics. An animus against the performing arts is evident in the rating of composers, who (as a group) rank just above soldiers at the bottom of the final list. Consider the following understatement about Mozart (p. 129): "A child who learns to play the piano at 3, who receives and benefits by musical instruction at that age, and who studies and executes the most difficult counterpoint at age 14, is probably above the average level of his social group."

In the end, I suspect that Cox recognized the shaky basis of her work, but persisted bravely nonetheless. Correlations between rank in eminence (length of Cattell's entry) and awarded IQ were disappointing to say the least—a mere 0.25 for eminence vs. A2 IQ, with no figure recorded at all for eminence vs. A1 IQ (it is a lower 0.20 by my calculation). Instead, Cox makes much of the fact that her ten most eminent subjects average 4—yes only 4—A1 IQ points above her ten least eminent.

Cox calculated her strongest correlation (0.77) between A2 IQ and "index of reliability," a measure of available information about her subjects. I can imagine no better demonstration that Cox's IQ's are artifacts of differential amounts of data, not measures of innate ability or even, for that matter, of simple talent. Cox recognized this and, in a final effort, tried to "correct" her scores for missing information by adjusting poorly documented subjects upward toward the group means of 135 for A1 IQ and 145 for A2 IQ. These adjustments boosted average IQ's substantially, but led to other embarrassments. For uncorrected scores, the most eminent fifty averaged 142 for A1 IQ, while the least eminent fifty scored comfortably lower at 133. With corrections, the first fifty scored 160, the last fifty, 165. Ultimately, only Goethe and Voltaire scored

near the top both in IQ and eminence. One might paraphrase Voltaire's famous quip about God and conclude that even though adequate information on the IQ of history's eminent men does not exist, it was probably inevitable that the American hereditarians would try to invent it.

Terman on group differences

Terman's empirical work measured what statisticians call the "within-group variance" of IQ—that is, the differences in scores within single populations (all children in a school, for example). At best, he was able to show that children testing well or poorly at a young age generally maintain their ordering with respect to other children as the population grows up. Terman ascribed most of these differences to variation in biological endowment, without much evidence beyond an assertion that all right-minded people recognize the domination of nurture by nature. This brand of hereditarianism might offend our present sensibilities with its elitism and its accompanying proposals for institutional care and forced abstinence from breeding, but it does not, by itself, entail the more contentious claim for innate differences between groups.

Terman made this invalid extrapolation, as virtually all hereditarians did and still do. He then compounded his error by confusing the genesis of true pathologies with causes for variation in normal behavior. We know, for example, that the mental retardation associated with Down's syndrome has its origin in a specific genetic defect (an extra chromosome). But we cannot therefore attribute the low IQ of many apparently normal children to an innate biology. We might as well claim that all overweight people can't help it because some very obese individuals can trace their condition to hormonal imbalances. Terman's data on the stability of ordering in IQ within groups of growing children relied largely upon the persistently low IQ of biologically afflicted individuals, despite Terman's attempt to bring all scores under the umbrella of a normal curve (1916, pp. 65–67), and thus to suggest that all variation has a common root in the possession of more or less of a single substance. In short, it is invalid to extrapolate from variation within a group to differences between groups. It is doubly invalid to use the innate biology of pathological individuals as a basis for ascribing normal variation within a group to inborn causes.

At least the IQ hereditarians did not follow their craniological

forebears in harsh judgments about women. Girls did not score below boys in IQ, and Terman proclaimed their limited access to professions both unjust and wasteful of intellectual talent (1916, p. 72; 1919, p. 288). He noted, assuming that IQ should earn its monetary reward, that women scoring between 100 and 120 generally earned, as teachers or "high-grade stenographers," what men with an IQ of 85 received as motormen, firemen, or policemen (1919, p. 278).

But Terman took the hereditarian line on race and class and proclaimed its validation as a primary aim of his work. In ending his chapter on the uses of IQ (1916, pp. 19–20), Terman posed three questions:

Is the place of the so-called lower classes in the social and industrial scale the result of their inferior native endowment, or is their apparent inferiority merely a result of their inferior home and school training? Is genius more common among children of the educated classes than among the children of the ignorant and poor? Are the inferior races really inferior, or are they merely unfortunate in their lack of opportunity to learn?

Despite a poor correlation of 0.4 between social status and IQ, Terman (1917) advanced five major reasons for claiming that "environment is much less important than is original endowment in determining the nature of the traits in question" (p. 91). The first three, based on additional correlations, add no evidence for innate causes. Terman calculated: 1) a correlation of 0.55 between social status and teachers' assessments of intelligence; 2) 0.47 between social status and school work; and 3) a lower, but unstated,* correlation between "age-grade progress" and social status. Since all five properties—IQ, social status, teacher's assessment, school work, and age-grade progress—may be redundant measures of the same complex and unknown causes, the correlation between any additional pair adds little to the basic result of 0.4 between IQ and social status. If the 0.4 correlation offers no evidence for innate causes, then the additional correlations do not either.

The fourth argument, recognized as weak by Terman himself

*It is annoyingly characteristic of Terman's work that he cites correlations when they are high and favorable, but does not give the actual figures when they are low and still favorable to his hypothesis. This ploy abounds in Cox's study of posthumous genius and in Terman's analysis of IQ among professions, both discussed previously.

(1916, p. 98), confuses probable pathology with normal variation, and is therefore irrelevant, as discussed above: feeble-minded children are occasionally born to rich or to intellectually successful parents.

The fifth argument reveals the strength of Terman's hereditarian convictions and his remarkable insensitivity to the influence of environment. Terman measured the IQ of twenty children in a California orphanage. Only three were "fully normal," while seventeen ranged from 75 to 95. The low scores cannot be attributed to life without parents, Terman argues, because (p. 99):

The orphanage in question is a reasonably good one and affords an environment which is about as stimulating to normal mental development as average home life among the middle classes. The children live in the orphanage and attend an excellent public school in a California village.

Low scores must reflect the biology of children committed to such institutions:

Some of the tests which have been made in such institutions indicate that mental subnormality of both high and moderate grades is extremely frequent among children who are placed in these homes. Most, though admittedly not all of these, are children of inferior social classes (p. 99).

Terman offers no direct evidence about the lives of his twenty children beyond the fact of their institutional placement. He is not even certain that they all came from "inferior social classes." Surely, the most parsimonious assumption would relate low IQ scores to the one incontestable and common fact about the children—their life in the orphanage itself.

Terman moved easily from individuals, to social classes, to races. Distressed by the frequency of IQ scores between 70 and 80, he lamented (1916, pp. 91-92):

Among laboring men and servant girls there are thousands like them. . . . The tests have told the truth. These boys are ineducable beyond the merest rudiments of training. No amount of school instruction will ever make them intelligent voters or capable citizens. . . . They represent the level of intelligence which is very, very common among Spanish-Indian and Mexican families of the Southwest and also among negroes. Their dullness seems to be racial, or at least inherent in the family stocks from which they came. The fact that one meets this type with such extraordinary frequency among Indians, Mexicans, and negroes suggests quite forcibly

that the whole question of racial differences in mental traits will have to be taken up anew and by experimental methods. The writer predicts that when this is done there will be discovered enormously significant racial differences in general intelligence, differences which cannot be wiped out by any scheme of mental culture. Children of this group should be segregated in special classes and be given instruction which is concrete and practical. They cannot master abstractions, but they can often be made efficient workers, able to look out for themselves. There is no possibility at present of convincing society that they should not be allowed to reproduce, although from a eugenic point of view they constitute a grave problem because of their unusually prolific breeding.

Terman sensed that his arguments for innateness were weak. Yet what did it matter? Do we need to prove what common sense proclaims so clearly?

After all, does not common observation teach us that, in the main, native qualities of intellect and character, rather than chance, determine the social class to which a family belongs? From what is already known about heredity, should we not naturally expect to find the children of well-to-do, cultured, and successful parents better endowed than the children who have been reared in slums and poverty? An affirmative answer to the above question is suggested by nearly all the available scientific evidence (1917, p. 99).

Whose common sense?

Terman recants

Terman's book on the Stanford-Binet revision of 1937 was so different from the original volume of 1916 that common authorship seems at first improbable. But then times had changed and intellectual fashions of jingoism and eugenics had been swamped in the morass of a Great Depression. In 1916 Terman had fixed adult mental age at sixteen because he couldn't get a random sample of older schoolboys for testing. In 1937 he could extend his scale to age eighteen; for "the task was facilitated by the extremely unfavorable employment situation at the time the tests were made, which operated to reduce considerably the school elimination normally occurring after fourteen" (1937, p. 30).

Terman did not explicitly abjure his previous conclusions, but a veil of silence descended upon them. Not a word beyond a few statements of caution do we hear about heredity. All potential rea-

sions for differences between groups are framed in environmental terms. Terman presents his old curves for average differences in IQ between social classes, but he warns us that mean differences are too small to provide any predictive information for individuals. We also do not know how to partition the average differences between genetic and environmental influences:

It is hardly necessary to stress the fact that these figures refer to mean values only, and that in view of the variability of the IQ within each group the respective distributions greatly overlap one another. Nor should it be necessary to point out that such data do not, in themselves, offer any conclusive evidence of the relative contributions of genetic and environmental factors in determining the mean differences observed.

A few pages later, Terman discusses the differences between rural and urban children, noting the lower country scores and the curious finding that rural IQ drops with age after entrance to school, while IQ for urban children of semiskilled and unskilled workers rises. He expresses no firm opinion, but note that the only hypotheses he wishes to test are now environmental:

It would require extensive research, carefully planned for the purpose, to determine whether the lowered IQ of rural children can be ascribed to the relatively poorer educational facilities in rural communities, and whether the gain for children from the lower economic strata can be attributed to an assumed enrichment of intellectual environment that school attendance bestows.

Autres temps, autres moeurs.

R. M. Yerkes and the Army Mental Tests: IQ comes of age

Psychology's great leap forward

Robert M. Yerkes, about to turn forty, was a frustrated man in 1902. He had been on the faculty of Harvard University since his profession. Yet psychology still wallowed in its reputation as a "soft" science, if a science at all. Some colleges did not acknowledge its existence; others ranked it among the humanities and placed psychologists in departments of philosophy. Yerkes wished, above all, to establish his profession by proving that it could be as

rigorous a science as physics. Yerkes and most of his contemporaries equated rigor and science with numbers and quantification. The most promising source of copious and objective numbers, Yerkes believed, lay in the embryonic field of mental testing. Psychology would come of age, and gain acceptance as a true science worthy of financial and institutional support, if it could bring the question of human potential under the umbrella of science:

Most of us are wholly convinced that the future of mankind depends in no small measure upon the development of the various biological and social sciences. . . . We must . . . strive increasingly for the improvement of our methods of mental measurement, for there is no longer ground for doubt concerning the practical as well as the theoretical importance of studies of human behavior. We must learn to measure skillfully every form and aspect of behavior which has psychological and sociological significance (Yerkes, 1917a, p. 111).

But mental testing suffered from inadequate support and its own internal contradictions. It was, first of all, practiced extensively by poorly trained amateurs whose manifestly absurd results were giving the enterprise a bad name. In 1915, at the annual meeting of the American Psychological Association in Chicago, a critic reported that the mayor of Chicago himself had tested as a moron on one version of the Binet scales. Yerkes joined with critics in discussions at the meeting and proclaimed: "We are building up a science, but we have not yet devised a mechanism which anyone can operate" (quoted in Chase, 1977, p. 242).

Second, available scales gave markedly different results even when properly applied. As discussed on p. 166, half the individuals who tested in the low, but normal range on the Stanford-Binet, were morons on Goddard's version of the Binet scale. Finally, support had been too inadequate, and coordination too sporadic, to build up a pool of data sufficiently copious and uniform to compel belief (Yerkes, 1917b).

Wars always generate their retinue of camp followers with ulterior motives. Many are simply scoundrels and profiteers, but a few are spurred by higher ideals. As mobilization for World War I approached, Yerkes got one of those "big ideas" that propel the history of science: could psychologists possibly persuade the army to test all its recruits? If so, the philosopher's stone of psychology might be constructed: the copious, useful, and uniform body of

numbers that would fuel a transition from dubious art to respected science. Yerkes proselytized within his profession and within government circles, and he won his point. As Colonel Yerkes, he presided over the administration of mental tests to 1.75 million recruits during World War I. Afterward, he proclaimed that mental testing "helped to win the war." "At the same time," he added, "it has incidentally established itself among the other sciences and demonstrated its right to serious consideration in human engineering" (quoted in Kevles, 1968, p. 581).

Yerkes brought together all the major hereditarians of American psychometrics to write the army mental tests. From May to July 1917 he worked with Terman, Goddard, and other colleagues at Goddard's Training School in Vineland, New Jersey.

Their scheme included three types of tests. Literate recruits would be given a written examination, called the Army Alpha. Illiterates and men who had failed Alpha would be given a pictorial test, called the Army Beta. Failures in Beta would be recalled for an individual examination, usually some version of the Binet scales. Army psychologists would then grade each man from A to E (with pluses and minuses) and offer suggestions for proper military placement. Yerkes suggested that recruits with a score of C- should be marked as "low average intelligence—ordinary private." Men of grade D are "rarely suited for tasks requiring special skill, forethought, resourcefulness or sustained alertness." D and E men could not be expected "to read and understand written directions."

I do not think that the army ever made much use of the tests. One can well imagine how professional officers felt about smart-assed young psychologists who arrived without invitation, often assumed an officer's rank without undergoing basic training, commanded a building to give the tests (if they could), saw each recruit for an hour in a large group, and then proceeded to usurp an officer's traditional role in judging the worthiness of men for various military tasks. Yerkes's corps encountered hostility in some camps; in others, they suffered a penalty in many ways more painful: they were treated politely, given appropriate facilities, and then ignored.* Some army officials became suspicious of Yerkes's

* Yerkes continued to complain throughout his career that military psychology had not achieved its due respect, despite its accomplishments in World War I. During World War II the aging Yerkes was still grouching and arguing that the Nazis were

intent and launched three independent investigations of the testing program. One concluded that it should be controlled so that "no theorist may . . . ride it as a hobby for the purpose of obtaining data for research work and the future benefit of the human race" (quoted in Kevles, 1968, p. 577).

Still, the tests did have a strong impact in some areas, particularly in screening men for officer training. At the start of the war, the army and national guard maintained nine thousand officers. By the end, two hundred thousand officers presided, and two-thirds of them had started their careers in training camps where the tests were applied. In some camps, no man scoring below C could be considered for officer training.

But the major impact of Yerkes's tests did not fall upon the army. Yerkes may not have brought the army its victory, but he certainly won his battle. He now had uniform data on 1.75 million men, and he had devised, in the Alpha and Beta exams, the first mass-produced written tests of intelligence. Inquiries flooded in from schools and businesses. In his massive monograph (Yerkes, 1921) on *Psychological Examining in the United States Army*, Yerkes buried a statement of great social significance in an aside on page 96. He spoke of "the steady stream of requests from commercial concerns, educational institutions, and individuals for the use of army methods of psychological examining or for the adaptation of such methods to special needs." Binet's purpose could now be circumvented because a technology had been developed for testing all pupils. Tests could now rank and stream everybody; the era of mass testing had begun.

Results of the army tests

The primary impact of the tests arose not from the army's lackadaisical use of scores for individuals, but from general propaganda that accompanied Yerkes's report of the summary statistics (Yerkes, 1921, pp. 553-875). E. G. Boring, later a famous psychol-

ogist upstaging America in their proper use and encouragement of mental testing for military personnel. "Germany has a long lead in the development of military psychology. . . . The Nazis have achieved something that is entirely without parallel in military history. . . . What has happened in Germany is the logical sequel to the psychological and personnel services in our own Army during 1917-1918" (Yerkes, 1921, p. 209).

ogist himself but then Yerkes's lieutenant (and the army's captain), selected one hundred sixty thousand cases from the files and produced data that reverberated through the 1920s with a hard hereditarian ring. The task was a formidable one. The sample, which Boring culled himself with the aid of only one assistant, was very large; moreover, the scales of three different tests (Alpha, Beta, and individual) had to be converted to a common standard so that racial and national averages could be constructed from samples of men who had taken the tests in different proportions (few blacks took Alpha, for example).

From Boring's ocean of numbers, three "facts" rose to the top and continued to influence social policy in America long after their source in the tests had been forgotten.

1. The average mental age of white American adults stood just above the edge of morosity at a shocking and meager thirteen. Terman had previously set the standard at sixteen. The new figure became a rallying point for eugenicists who predicted doom and lamented our declining intelligence, caused by the unconstrained breeding of the poor and feeble-minded, the spread of Negro blood through miscegenation, and the swamping of an intelligent native stock by the immigrating dregs of southern and eastern Europe. Yerkes* wrote:

It is customary to say that the mental age of the average adult is about 16 years. This figure is based, however, upon examinations of only 62 persons; 32 of them high-school pupils from 16-20 years of age, and 30 of them "business men of moderate success and of very limited educational advantages." The group is too small to give very reliable results and is furthermore probably not typical. . . . It appears that the intelligence of the principal sample of the white draft, when transmuted from Alpha and Beta exams into terms of mental age, is about 13 years (13.08) (1921, p. 785).

Yet, even as he wrote, Yerkes began to sense the logical absurdity of such a statement. An average is what it is; it cannot lie three years below what it should be. So Yerkes thought again and added:

We can hardly say, however, with assurance that these recruits are three years mental age below the average. Indeed, it might be argued on *I doubt that Yerkes wrote all parts of the massive 1921 monograph himself. But he is listed as the only author of this official report, and I shall continue to attribute its statements to him, both as shorthand and for want of other information.

extrinsic grounds that the draft itself is more representative of the average intelligence of the country than is a group of high-school students and business men (1921, p. 785).

If 13.08 is the white average, and everyone from mental age 8 through 12 is a moron, then we are a nation of nearly half-morons. Yerkes concluded (1921, p. 791): "It would be totally impossible to exclude all morons as that term is at present defined, for there are under 13 years 37 percent of whites and 89 percent of negroes."

2. European immigrants can be graded by their country of origin. The average man of many nations is a moron. The darker peoples of southern Europe and the Slavs of eastern Europe are less intelligent than the fair peoples of western and northern Europe. Nordic supremacy is not a jingoistic prejudice. The average Russian has a mental age of 11.94; the Italian, 11.01; the Pole, 10.74. The Polish joke attained the same legitimacy as the moron joke—indeed, they described the same animal.

3. The Negro lies at the bottom of the scale with an average mental age of 10.41. Some camps tried to carry the analysis a bit further, and in obvious racist directions. At Camp Lee, blacks were divided into three groups based upon intensity of color; the lighter groups scored higher (p. 531). Yerkes reported that the opinions of officers matched his numbers (p. 742):

All officers without exception agree that the negro lacks initiative, displays little or no leadership, and cannot accept responsibility. Some point out that these defects are greater in the southern negro. All officers seem further to agree that the negro is a cheerful, willing soldier, naturally subservient. These qualities make for immediate obedience, although not necessarily for good discipline, since petty thieving and venereal disease are commoner than with white troops.

Along the way, Yerkes and company tested several other social prejudices. Some fared poorly, particularly the popular eugenic notion that most offenders are feeble-minded. Among conscientious objectors for political reasons, 59 percent received a grade of A. Even outright disloyals scored above the average (p. 803). But other results buoyed their prejudices. As camp followers themselves, Yerkes's corps decided to test a more traditional category of colleagues: the local prostitutes. They found that 53 percent (44 percent of whites and 68 percent of blacks) ranked at age ten or

below on the Goddard version of the Binet scales. (They acknowledge that the Goddard scales ranked people well below their scores on other versions of the Binet tests.) Yerkes concluded (p. 808):

The results of Army examining of prostitutes corroborate the conclusion, attained by civilian examinations of prostitutes in various parts of the country, that from 30 to 60 percent of prostitutes are deficient and are for the most part high-grade morons; and that 15 to 25 percent of all prostitutes are so low-grade mentally that it is wise (as well as possible under the existing laws in most states) permanently to segregate them in institutions for the feeble-minded.

One must be thankful for small bits of humor to lighten the reading of an eight-hundred-page statistical monograph. The thought of army personnel rounding up the local prostitutes and sitting them down to take the Binet tests amused me no end, and must have betrayed the ladies even more.

As pure numbers, these data carried no inherent social message. They might have been used to promote equality of opportunity and to underscore the disadvantages imposed upon so many Americans. Yerkes might have argued that an average mental age of thirteen reflected the fact that relatively few recruits had the opportunity to finish or even to attend high school. He might have attributed the low average of some national groups to the fact that most recruits from these countries were recent immigrants who did not speak English and were unfamiliar with American culture. He might have recognized the link between low Negro scores and the history of slavery and racism.

But scarcely a word do we read through eight hundred pages of any role for environmental influence. The tests had been written by a committee that included all the leading American hereditarians discussed in this chapter. They had been constructed to measure innate intelligence, and they did so by definition. The circularity of argument could not be broken. All the major findings received hereditarian interpretations, often by near miracles of special pleading to argue past a patent environmental influence. A circular issued from the School of Military Psychology at Camp Greenleaf proclaimed (do pardon its questionable grammar): "These tests do not measure occupational fitness nor educational attainment; they measure intellectual ability. This latter has been

shown to be important in estimating military value" (p. 424). And the boss himself argued (Yerkes, quoted in Chase, 1977, p. 249):

Examinations Alpha and Beta are so constructed and administered as to minimize the handicap of men who because of foreign birth or lack of education are little skilled in the use of English. These group examinations were originally intended, and are now definitely known, to measure native intellectual ability. They are to some extent influenced by educational attainment, but in the main the soldier's inborn intelligence and not the accidents of environment determines his mental rating or grade in the army.

A critique of the Army Mental Tests

THE CONTENT OF THE TESTS

The Alpha test included eight parts, the Beta seven; each took less than an hour and could be given to large groups. Most of the Alpha parts presented items that have become familiar to generations of test-takers ever since: analogies, filling in the next number in a sequence, unscrambling sentences, and so forth. This similarity is no accident; the Army Alpha was the granddaddy, literally as well as figuratively, of all written mental tests. One of Yerkes's disciples, C. C. Brigham, later became secretary of the College Entrance Examination Board and developed the Scholastic Aptitude Test on army models. If people get a peculiar feeling of déjà-vu in perusing Yerkes's monograph, I suggest that they think back to their own College Boards, with all its attendant anxiety.

These familiar parts are not especially subject to charges of cultural bias, at least no more so than their modern descendants. In a general way, of course, they test literacy, and literacy records education more than inherited intelligence. Moreover, a schoolmaster's claim that he tests children of the same age and school experience, and therefore may be recording some internal biology, didn't apply to the army recruits—for they varied greatly in access to education and recorded different amounts of schooling in their scores. A few of the items are amusing in the light of Yerkes's assertion that the tests "measure native intellectual ability." Consider the Alpha analogy: "Washington is to Adams as first is to . . ."

But one part of each test is simply ludicrous in the light of Yerkes's analysis. How could Yerkes and company attribute the low

scores of recent immigrants to innate stupidity when their multiple-choice test consisted entirely of questions like:

Crisco is a: patent medicine, disinfectant, toothpaste, food product

The number of a Kaffir's legs is: 2, 4, 6, 8

Christy Mathewson is famous as a: writer, artist, baseball player, comedian

I got the last one, but my intelligent brother, who, to my distress, grew up in New York utterly oblivious to the heroics of three great baseball teams then resident, did not.

Yerkes might have responded that recent immigrants generally took Beta rather than Alpha, but Beta contains a pictorial version of the same theme. In this complete-a-picture test, early items might be defended as sufficiently universal: adding a mouth to a face or an ear to a rabbit. But later items required a rivet in a pocket knife, a filament in a light bulb, a horn on a phonograph, a net on a tennis court, and a ball in a bowler's hand (marked wrong, Yerkes explained, if an examinee drew the ball in the alley, for you can tell from the bowler's posture that he has not yet released the ball). Franz Boas, an early critic, told the tale of a Sicilian recruit who added a crucifix where it always appeared in his native land to a house without a chimney. He was marked wrong.

The tests were strictly timed, for the next fifty were waiting by the door. Recruits were not expected to finish each part; this was explained to the Alpha men, but not to Beta people. Yerkes wondered why so many recruits scored flat zero on so many of the parts (the most telling proof of the tests' worthlessness—see pp. 244-247). How many of us, if nervous, uncomfortable, and crowded (and even if not), would have understood enough to write anything at all in the ten seconds allotted for completing the following commands, each given but once in Alpha, Part 1?

Attention! Look at 4. When I say "go" make a figure 1 in the space which is in the circle but not in the triangle or square, and also make a figure 2 in the space which is in the triangle and circle, but not in the square. Go.

Attention! Look at 6. When I say "go" put in the second circle the right answer to the question: "How many months has a year?" In the third circle do nothing, but in the fourth circle put any number that is a wrong answer to the question that you have just answered correctly. Go.

INADEQUATE CONDITIONS

Yerkes's protocol was rigorous and trying enough. His examiners had to process men rapidly and grade the exams immediately, so that failures could be recalled for a different test. When faced with the added burden of thinly veiled hostility from the brass at several camps, Yerkes's testers were rarely able to carry out more than a caricature of their own stated procedure. They continually compromised, backtracked, and altered in the face of necessity. Procedures varied so much from camp to camp that results could scarcely be collated and compared. The whole effort, through no fault of Yerkes's beyond impracticality and overambition, became something of a shambles, if not a disgrace. The details are all in Yerkes's monograph, but hardly anyone ever read it. The summary statistics became an important social weapon for racists and eugenicists; their rotten core lay exposed in the monograph, but who looks within when the surface shines with such a congenial message.

The army mandated that special buildings be supplied or even constructed for Yerkes's examinations, but a different reality prevailed (1921, p. 61). The examiners had to take what they could get, often rooms in cramped barracks with no furnishings at all, and inadequate acoustics, illumination, and lines of sight. The chief tester at one camp complained (p. 106): "Part of this inaccuracy I believe to be due to the fact that the room in which the examination is held is filled too full of men. As a result, the men who are sitting in the rear of the room are unable to hear clearly and thoroughly enough to understand the instructions."

Tensions rose between Yerkes's testers and regular officers. The chief tester of Camp Custer complained (p. 111): "The ignorance of the subject on the part of the average officer is equalled only by his indifference to it." Yerkes urged restraint and accommodation (p. 155):

The examiner should strive especially to take the military point of view. Unwarranted claims concerning the accuracy of the results should be avoided. In general, straightforward commonsense statements will be found more convincing than technical descriptions, statistical exhibits, or academic arguments.

As friction and doubt mounted, the secretary of war polled commanding officers of all camps to ask their opinion of Yerkes's tests. He received one hundred replies, nearly all negative. They were, Yerkes admitted (p. 43), "with a few exceptions, unfavorable to psychological work, and have led to the conclusion on the part of various officers of the General Staff that this work has little, if any, value to the army and should be discontinued." Yerkes fought back and won a standoff (but not all the promotions, commissions, and hirings he had been promised); his work proceeded under a cloud of suspicion.

Minor frustrations never abated. Camp Jackson ran out of forms and had to improvise on blank paper (p. 78). But a major and persistent difficulty dogged the entire enterprise and finally, as I shall demonstrate, deprived the summary statistics of any meaning. Recruits had to be allocated to their appropriate test. Men illiterate in English, either by lack of schooling or foreign birth, should have taken examination Beta, either by direct assignment, or indirectly upon failing Alpha. Yerkes's corps tried heroically to fulfill this procedure. In at least three camps, they marked identification tags or even painted letters directly on the bodies of men who failed—a ready identification guide for further assessment (p. 73, p. 76): "A list of D men was sent within six hours after the group examination to the clerk at the mustering office. As the men appeared, this clerk marked on the body of each D man a letter P" (indicating that the psychiatrist should examine them further).

But standards for the division between Alpha and Beta varied substantially from camp to camp. A survey across camps revealed that the minimum score on an early version of Alpha varied from 20 to 100 for assignment to further testing (p. 476). Yerkes admitted (p. 354):

This lack of a uniform process of segregation is certainly unfortunate. On account of the variable facilities for examining and the variable quality of the groups examined however, it appeared entirely impossible to establish a standard uniform for all camps.

C. C. Brigham, Yerkes's most zealous votary, even complained (1921):

The method of selecting men for Beta varied from camp to camp, and sometimes from week to week in the same camp. There was no established criterion of literacy, and no uniform method of selecting illiterates.

The problem cut far deeper than simple inconsistency among camps. The persistent logistical difficulties imposed a systematic bias that substantially lowered the mean scores of blacks and immigrants. For two major reasons, many men took only Alpha and scored either zero or next to nothing, not because they were innately dumb, but because they were illiterate and should have taken Beta by Yerkes's own protocol. First, recruits and draftees had, on average, spent fewer years in school than Yerkes had anticipated. Lines for Beta began to lengthen and the entire operation threatened to clog at this bottleneck. At many camps, unqualified men were sent in droves to Alpha by artificial lowering of standards. Schooling to the third grade sufficed for Alpha in one camp; in another, anyone who said he could read, at whatever level, took Alpha. The chief tester at Camp Dix reported (p. 72): "To avoid excessively large Beta groups, standards for admission to examination Alpha were set low."

Second, and more important, the press of time and the hostility of regular officers often precluded a Beta retest for men who had incorrectly taken Alpha. Yerkes admitted (p. 472): "It was never successfully shown, however, that the continued recalls . . . were so essential that repeated interference with company maneuvers should be permitted." As the pace became more frantic, the problem worsened. The chief tester at Camp Dix complained (pp. 72-73): "In June it was found impossible to recall a thousand men listed for individual examination. In July Alpha failures among negroes were not recalled." The stated protocol scarcely applied to blacks who, as usual, were treated with less concern and more contempt by everyone. Failure on Beta, for example, should have led to an individual examination. Half the black recruits scored D— on Beta, but only one-fifth of these were recalled and four-fifths received no further examination (p. 708). Yet we know that scores for blacks improved substantially when the protocol was followed. At one camp (p. 736), only 14.1 percent of men who had scored D— on Alpha failed to gain a higher grade on Beta.

The effects of this systematic bias are evident in one of Boring's

the form, and had nothing at stake. (One friend later suggested that I should have required names—and posted results—as just a small contribution to simulating the anxiety of the original.)

I knew before I started that internal contradictions and a priori prejudice thoroughly invalidated the hereditarian conclusions that Yerkes had drawn from the results. Boring himself called these conclusions "preposterous" late in his career (in a 1962 interview, quoted in Kevles, 1968). But I had not understood how the Draconian conditions of testing made such a thorough mockery of the claim that recruits could have been in a frame of mind to record anything about their innate abilities. In short, most of the men must have ended up either utterly confused or scared shitless.

The recruits were ushered into a room and seated before an examiner and demonstrator standing atop a platform, and several orderlies at floor level. Examiners were instructed to administer the test "in a genial manner" since "the subjects who take this examination sometimes sulk and refuse to work" (p. 163). Recruits were told nothing about the examination or its purposes. The examiner simply said: "Here are some papers. You must not open them or turn them over until you are told to." The men then filled in their names, age, and education (with help for those too illiterate to do so). After these perfunctory preliminaries, the examiner plunged right in:

Attention. Watch this man (pointing to demonstrator). He (pointing to demonstrator again) is going to do here (tapping blackboard with pointer) what you (pointing to different members of the group) are to do on your papers (here examiner points to several papers that lie before men in the group, picks up one, holds it next to the blackboard, returns the paper, points to demonstrator and the blackboard in succession, then to the men and their papers). Ask no questions. Wait till I say "Go ahead!" (p. 163).

By comparison, Alpha men were virtually inundated with information (p. 157), for the Alpha examiner said:

Attention! The purpose of this examination is to see how well you can remember, think, and carry out what you are told to do. We are not looking for crazy people. The aim is to help find out what you are best fitted to do in the Army. The grade you make in this examination will be put on your qualification card and will also go to your company commander. Some of the things you are told to do will be very easy. Some you may find

experiments with the summary statistics. He culled 4,893 cases of men who had taken both Alpha and Beta. Converting their scores to the common scale, he calculated an average mental age of 10.775 for Alpha, and a Beta mean of 12.158 (p. 655). He used only the Beta scores in his summaries; Yerkes procedure worked. But what of the myriads who should have taken Beta, but only received Alpha and scored abysmally as a result—primarily poorly educated blacks and immigrants with an imperfect command of English—the very groups whose low scores caused such a hereditarian stir later on?

DUBIOUS AND PERVERSE PROCEEDINGS: A PERSONAL TESTIMONY

Academicians often forget how poorly or incompletely the written record, their primary source, may represent experience. Some things have to be seen, touched, and tasted. What was it like to be an illiterate black or foreign recruit, anxious and befuddled at the novel experience of taking an examination, never told why, or what would be made of the results; expulsion, the front lines? In 1968 (quoted in Kevles), an examiner recalled his administration of Beta: "It was touching to see the intense effort . . . put into answering the questions, often by men who never before had held a pencil in their hands." Yerkes had overlooked, or consciously bypassed, something of importance. The Beta examination contained only pictures, numbers, and symbols. But it still required pencil work and, on three of its seven parts, a knowledge of numbers and how to write them.

Yerkes's monograph is so thorough that his procedure for giving the two examinations can be reconstructed down to the choreography of motion for all examiners and orderlies. He provides facsimiles in full size for the examinations themselves, and for all explanatory material used by examiners. The standardized words and gestures of examiners are reproduced in full. Since I wanted to know in as complete a way as possible what it felt like to give and take the test, I administered examination Beta (for illiterates) to a group of fifty-three Harvard undergraduates in my course on biology as a social weapon. I tried to follow Yerkes's protocol scrupulously in all its details. I feel that I reconstructed the original situation accurately, with one important exception: my students knew what they were doing, didn't have to provide their names on

men who are not working, and saying, "Do it, do it, hurry up, quick." At the end of 2 minutes examiner says, "Stop! Turn over the page to test 2."

The examiner demonstrated test 2, cube counting, with three-dimensional models (my son had some left over from his baby days). Note that recruits who could not write numbers would receive scores of zero even if they counted all the cubes correctly. Test 3, the X-O series, will be recognized by nearly everyone today as the pictorial version of "what is the next number in the sequence." Test 4, digit symbols, required the translation of nine digits into corresponding symbols. It looks easy enough, but the test itself included ninety items and could hardly be finished by anybody in the two minutes allotted. A man who couldn't write numbers was faced with two sets of unfamiliar symbols and suffered a severe additional disadvantage. Test 5, number checking, asked men to compare numerical sequences, up to eleven digits in length, in two parallel columns. If items on the same line were identical in the two columns, recruits were instructed (by gestures) to write an X next to the item. Fifty sequences occupied three minutes, and few recruits could finish. Again, an inability to write or recognize numbers would make the task virtually impossible.

Test 6, pictorial completion, is Beta's visual analogue of Alpha's multiple-choice examination for testing innate intelligence by asking recruits about commercial products, famous sporting or film stars, or the primary industries of various cities and states. Its instructions are worth repeating:

"This is test 6 here. Look. A lot of pictures." After everyone has found the place, "Now watch." Examiner points to hand and says to demonstrate, "Fix it." Demonstrator does nothing, but looks puzzled. Examiner points to the picture of the hand, and then to the place where the finger is missing and says to demonstrator, "Fix it; fix it." Demonstrator then draws in finger. Examiner says, "That's right." Examiner then points to fish and place for eye and says, "Fix it." After demonstrator has drawn missing eye, examiner points to each of the four remaining drawings and says, "Fix them all." Demonstrator works samples out slowly and with apparent effort. When the samples are finished examiner says, "All right. Go ahead. Hurry up!" During the course of this test the orderlies walk around the room and locate individuals who are doing nothing, point to their pages and say, "Fix it. Fix them," trying to set everyone working. At the end of 3 minutes examiner says, "Stop! But don't turn over the page."

hard. You are not expected to make a perfect grade, but do the very best you can. . . . Listen closely. Ask no questions.

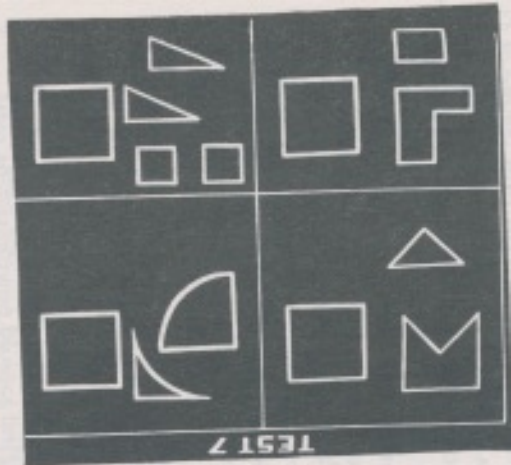
The extreme limits imposed upon the Beta examiner's vocabulary did not only reflect Yerkes's poor opinion of what Beta recruits might understand by virtue of their stupidity. Many Beta examinees were recent immigrants who did not speak English, and instruction had to be as pictorial and gestural as possible. Yerkes advised (p. 163): "One camp has had great success with a 'window seller' as demonstrator. Actors should also be considered for the work." One particularly important bit of information was not transmitted: examinees were not told that it was virtually impossible to finish at least three of the tests, and that they were not expected to do so.

Atop the platform, the demonstrator stood in front of a blackboard roll covered by a curtain; the examiner stood at his side. Before each of the seven tests, the curtain was raised to expose a sample problem (all reproduced in Figure 5-4), and examiner and demonstrator engaged in a bit of pantomime to illustrate proper procedure. The examiner then issued an order to work, and the demonstrator closed the curtain and advanced the roll to the next sample. The first test, maze running, received the following demonstration:

Demonstrator traces path through first maze with crayon, slowly and hesitatingly. Examiner then traces second maze and motions to demonstrator to go ahead. Demonstrator makes one mistake by going into the blind alley at upper left-hand corner of maze. Examiner apparently does not notice what demonstrator is doing until he crosses line at end of alley; then examiner shakes his head vigorously, says "No-no," takes demonstrator's hand and traces back to the place where he may start right again. Demonstrator traces rest of maze so as to indicate an attempt at haste, hesitating only at ambiguous points. Examiner says "Good." Then holding up blank, "Look here," and draws an imaginary line across the page from left to right for every maze on the page. Then, "All right. Go ahead. Do it (pointing to men and then to books). Hurry up."

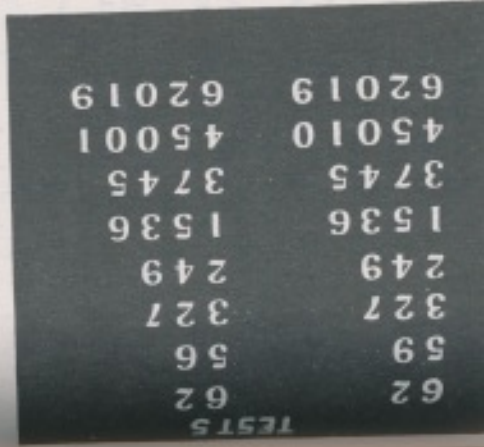
This paragraph may be naively amusing (some of my students thought so). The next statement, by comparison, is a bit diabolical.

The idea of working fast must be impressed on the men during the maze test. Examiner and orderlies walk around the room, motioning to



TEST 7

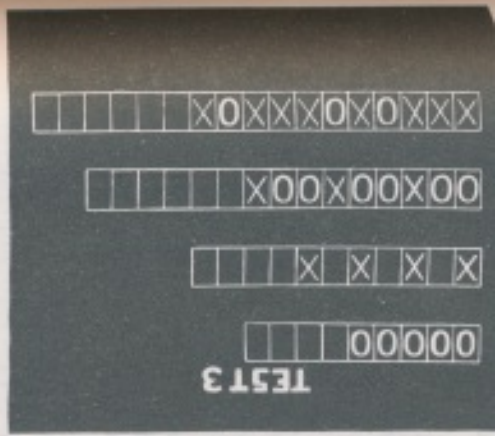
5•4 The blackboard demonstrations for all seven parts of the Beta test. From Yerkes, 1921.



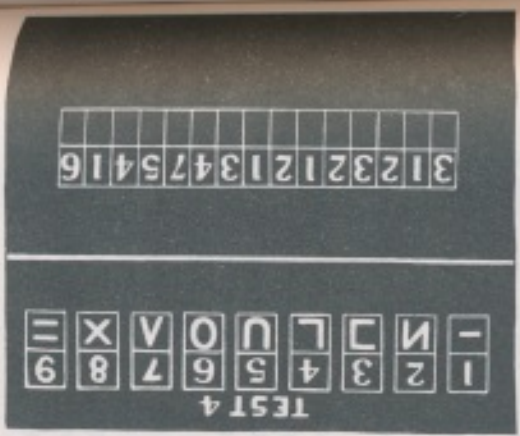
TEST 5



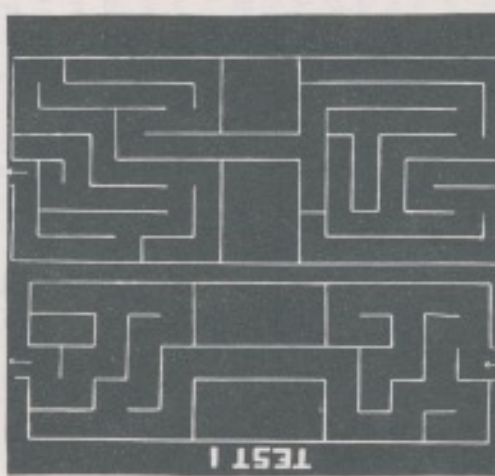
TEST 6



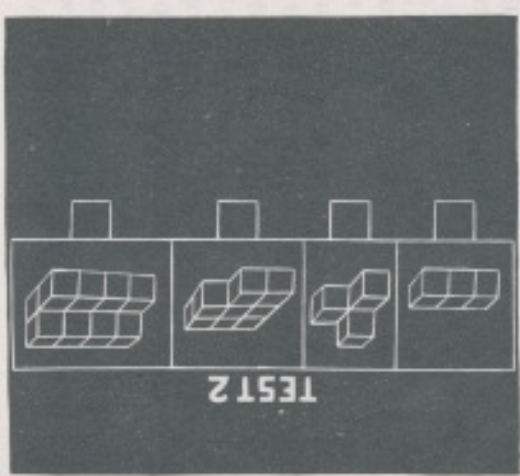
TEST 3



TEST 4



TEST 1



TEST 2

The examination itself is also worth reprinting (Fig. 5-5). Best of luck with pig tails, crab legs, bowling balls, tennis rackets, and the Jack's missing diamond, not to mention the phonograph horn (a real stumper for my students). Yerkes provided the following instructions for grading:

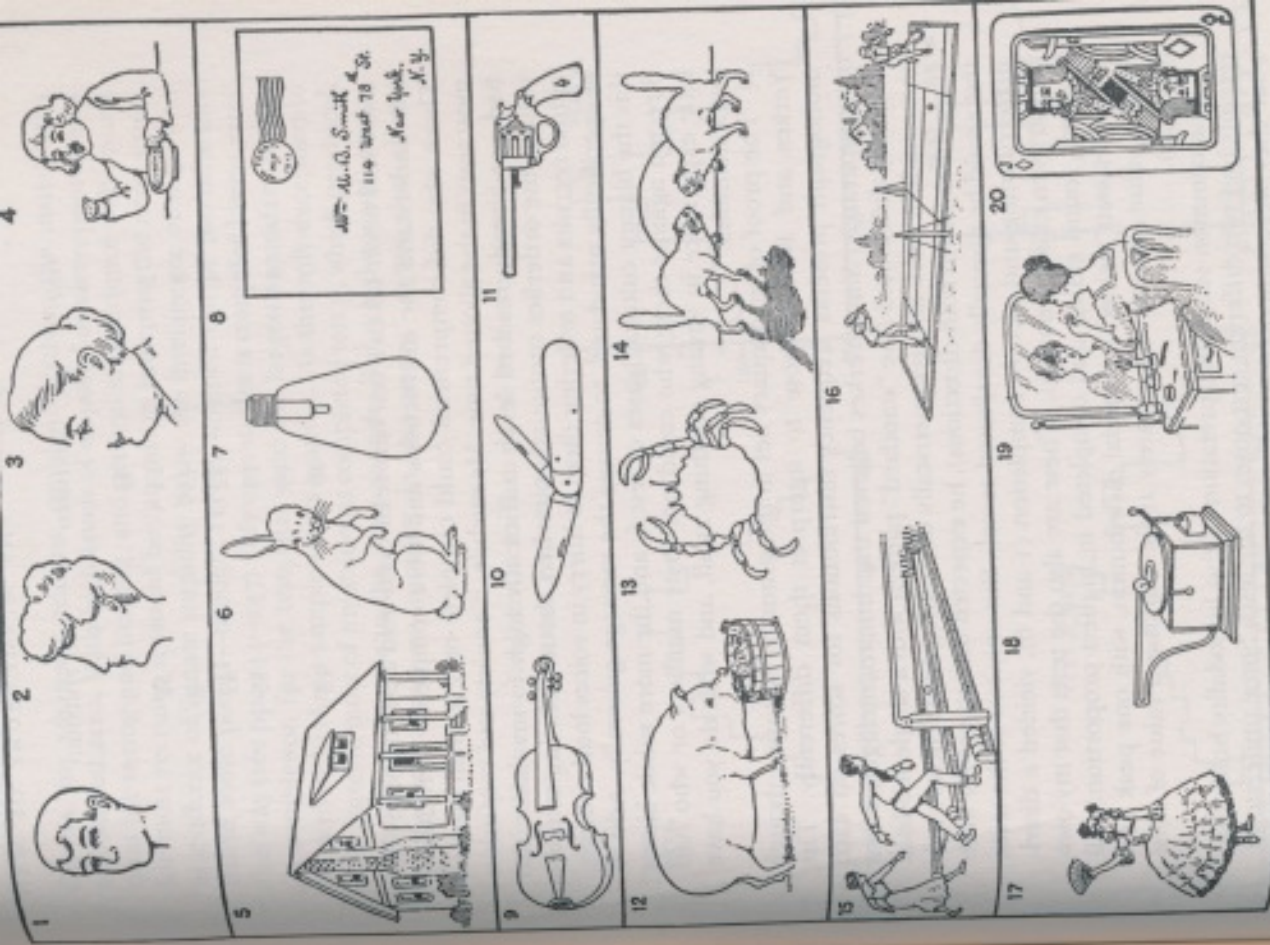
Rules for Individual Items

- Item 4.—Any spoon at any angle in right hand receives credit. Left hand, or unattached spoon, no credit.
- Item 5.—Chimney must be in right place. No credit for smoke.
- Item 6.—Another ear on same side as first receives no credit.
- Item 8.—Plain square, cross, etc., in proper location for stamp, receives credit.
- Item 10.—Missing part is the rivet. Line of "ear" may be omitted.
- Item 13.—Missing part is leg.
- Item 15.—Ball should be drawn in hand of man. If represented in hand of woman, or in motion, no credit.
- Item 16.—Single line indicating net receives credit.
- Item 18.—Any representation intended for horn, pointing in any direction, receives credit.
- Item 19.—Hand and powder puff must be put on proper side.
- Item 20.—Diamond is the missing part. Failure to complete hilt on sword is not an error.

The seventh and last test, geometrical construction, required that a square be broken into component pieces. Its ten parts were allotted two and a half minutes.

I believe that the conditions of testing, and the basic character of the examination, make it ludicrous to believe that Beta measured any internal state deserving the label intelligence. Despite the plea for geniality, the examination was conducted in an almost frantic rush. Most parts could not be finished in the time allotted, but recruits were not forewarned. My students compiled the following record of completions on the seven parts (see p. 242). For two of the tests, digit symbols and number checking (4 and 5), most students simply couldn't write fast enough to complete the ninety and fifty items, even though the protocol was clear to all. The third test with a majority of incompletes, cube counting (number 2), was too difficult for the number of items included and the time allotted.

In summary, many recruits could not see or hear the examiner:



5 • 5 Part six of examination Beta for testing innate intelligence.

TEST	FINISHED	NOT FINISHED
1	44	9
2	21	32
3	45	8
4	12	41
5	18	35
6	49	4
7	40	13

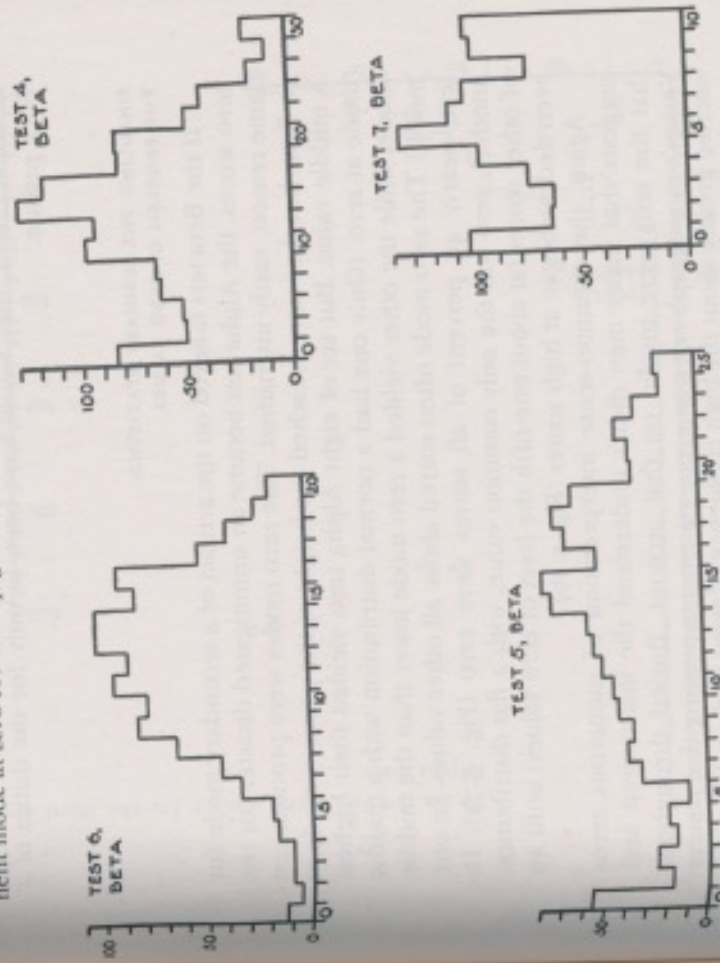
some had never taken a test before or even held a pencil. Many did not understand the instructions and were completely befuddled. Those who did comprehend could complete only a small part of most tests in the allotted time. Meanwhile, if anxiety and confusion had not already reached levels sufficiently high to invalidate the results, the orderlies continually marched about, pointing to individual recruits and ordering them to hurry in voices loud enough, as specifically mandated, to convey the message generally. Add to this the blatant cultural biases of test 6, and the more subtle biases directed against those who could not write numbers or who had little experience in writing anything at all, and what do you have but a shambles.

The proof of inadequacy lies in the summary statistics, though Yerkes and Boring chose to interpret them differently. The monograph presents frequency distributions for scores on each part separately. Since Yerkes believed that innate intelligence was normally distributed (the "standard" pattern with a single mode at some middle score and symmetrically decreasing frequencies away from the mode in both directions), he expected that scores for each test would be normally distributed as well. But only two of the tests, maze running and picture completion (1 and 6), yielded a distribution even close to normal. (These are also the tests that my own students found easiest and completed in highest proportion.) All the other tests yielded a bimodal distribution, with one peak at a middle value and another squarely at the minimum value of zero (Fig. 5-6).

The common-sense interpretation of this bimodality holds that recruits had two different responses to the tests. Some understood what they were supposed to do, and performed in varied ways.

Others, for whatever reasons, could not fathom the instructions and scored zero. With high levels of imposed anxiety, poor conditions for seeing and hearing, and general inexperience with testing for most recruits, it would be fatuous to interpret the zero scores as evidence of innate stupidity below the intelligence of men who made some points—though Yerkes wormed out of the difficulty this way (see pp. 244–247). (My own students compiled lowest rates of completion for the tests that yield the largest secondary modes at zero in Yerkes's sample—tests 4 and 5. As the only exception to this pattern, most of my students completed test 3, which produced a strong zero mode in the army sample. But 3 is the visual analog of "what is the next number in this series," a test that all my

5 • 6 Frequency distributions for four of the Beta tests. Note the prominent mode at zero for tests 4, 5, and 7.



students have taken more times than they care to remember.)

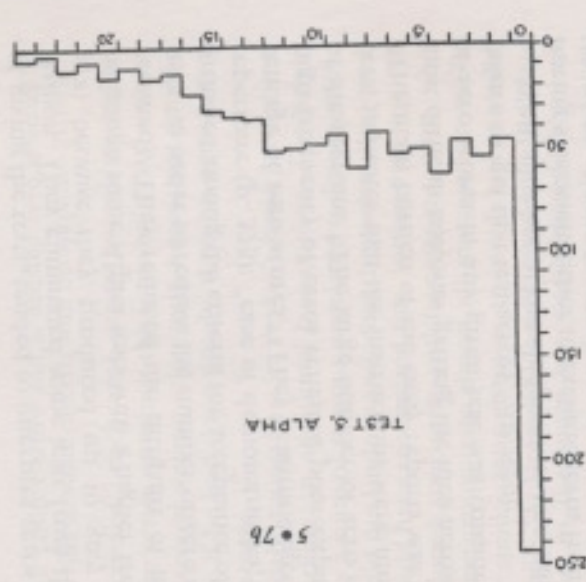
Statisticians are trained to be suspicious of distributions with multiple modes. Such distributions usually indicate inhomogeneity in the system, or, in plainer language, different causes for the different modes. All familiar proverbs about the invisibility of mixing apples and oranges apply. The multiple modes should have guided Yerkes to a suspicion that his tests were not measuring a single entity called intelligence. Instead, his statisticians found a way to redistribute zero scores in a manner favorable to hereditarian assumptions (see next section).

Oh yes, was anyone wondering how my students fared? They did very well of course. Anything else would have been shocking, since all the tests are greatly simplified precursors of examinations they have been taking all their lives. Of fifty-three students, thirty-one scored A and sixteen B. Still, more than 10 percent (six of fifty-three) scored at the intellectual borderline of C; by the standards of some camps, they would have been fit only for the duties of a buck private.

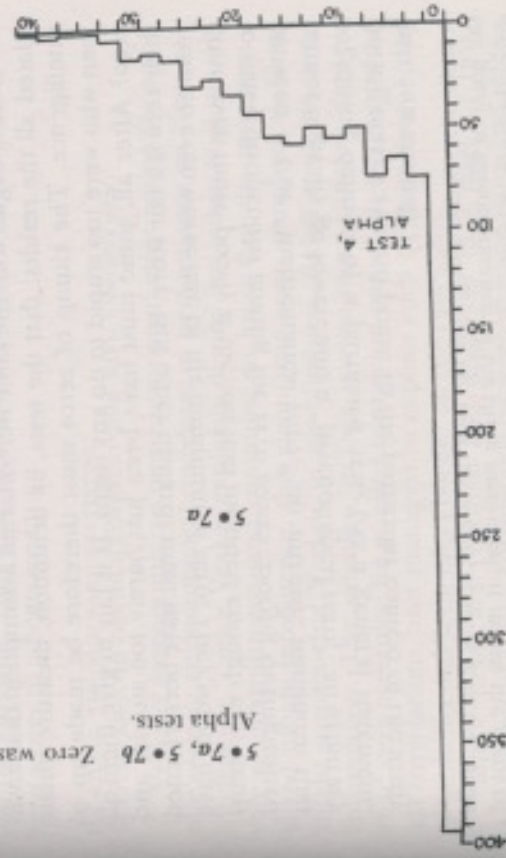
FINAGLING THE SUMMARY STATISTICS: THE PROBLEM OF ZERO VALUES

If the Beta test faltered on the artifact of a secondary mode for zero scores, the Alpha test became an unmitigated disaster for the same reason, vastly intensified. The zero modes were pronounced in Beta, but they never reached the height of the primary mode at a middle value. But six of eight Alpha tests yielded their highest mode at zero. (Only one had a normal distribution with a middle mode, while the other yielded a zero mode lower than the middle mode.) The zero mode often soared above all other values. In one test, nearly 40 percent of all scores were zero (Fig. 5.7a). In another, zero was the only common value, with a flat distribution of other scores (at about one-fifth the level of zero values) until an even decline began at high scores (Fig. 5.7b).

Again, the common-sense interpretation of numerous zeros suggests that many men didn't understand the instructions and that the tests were invalid on that account. Buried throughout Yerkes's monograph are numerous statements proving that testers worried greatly about the high frequency of zeros and, in the midst



5 • 7a, 5 • 7b Zero was by far the most common value in several of the Alpha tests.



of giving the tests, tended to interpret zeros in this common-sense fashion. They eliminated some tests from the Beta repertoire (p. 372) because they produced up to 30.7 percent zero scores (although some Alpha tests with a higher frequency of zeros were retained). They reduced the difficulty of initial items in several tests "in order to reduce the number of zero scores" (p. 341). They included among the criteria for acceptance of a test within the Beta repertoire (p. 373): "ease of demonstration, as shown by low percentage of zero scores." They acknowledged several times that a high frequency of zeros reflected poor explanation, not stupidity of the recruits: "The large number of zero scores, even with officers, indicates that the instructions were unsatisfactory" (p. 340). "The main burden of the early reports was to the effect that the most difficult task was 'getting the idea across.' A high percentage of zero scores in any given test was considered an indication of failure to 'get that test across'" (p. 379).

With all these acknowledgments, one might have anticipated Boring's decision either to exclude zeros from the summary statistics or to correct for them by assuming that most recruits would have scored some points if they had understood what they were supposed to do. Instead, Boring "corrected" zero scores in the opposite way, and actually demoted many of them into a negative range.

Boring began with the same hereditarian assumption that invalidated all the results: that the tests, by definition, measure innate intelligence. The clump of zeros must therefore be made up of men who were too stupid to do any items. Is it fair to give them all zero? After all, some must have been just barely too stupid, and their zero is a fair score. But other dullards must have been rescued from an even worse fate by the minimum of zero. They would have done even more poorly if the test had included enough easy items to make distinctions among the zero scores. Boring distinguished between a true "mathematical zero," an intrinsic minimum that cannot logically go lower, and a "psychological zero," an arbitrary beginning defined by a particular test. (As a general statement, Boring makes a sound point. In the particular context of the army tests, it is absurd):

A score of zero, therefore, does not mean no ability at all; it does not mean the point of discontinuance of the thing measured; it means the point of discontinuance of the instrument of measurement, the test. . . . The indi-

vidual who fails to earn a positive score and is marked zero is actually thereby given a bonus varying in value directly with his stupidity (p. 622). Boring therefore "corrected" each zero score by calibrating it against other tests in the series on which the same man had scored some points. If he had scored well on other tests, he was not doubly penalized for his zeros; if he had done poorly, then his zeros were converted to negative scores.

By this method, a debilitating flaw in Yerkes's basic procedure was accentuated by tacking an additional bias onto it. The zeros only indicated that, for a suite of reasons unrelated to intelligence, vast numbers of men did not understand what they were supposed to do. And Yerkes should have recognized this, for his own reports proved that, with reduced confusion and harassment, men who had scored zero on the group tests almost all managed to make points on the same or similar tests given in an individual examination. He writes (p. 406): "At Greenleaf it was found that the proportion of zero scores in the maze test was reduced from 28 percent in Beta to 2 percent in the performance scale, and that similarly zero scores in the digit-symbol test were reduced from 49 to 6 percent."

Yet, when given an opportunity to correct this bias by ignoring or properly redistributing the zero scores, Yerkes's statisticians did just the opposite. They exacted a double penalty by demoting most zero scores to a negative range.

FINAGLING THE SUMMARY STATISTICS:
GETTING AROUND OBVIOUS CORRELATIONS WITH ENVIRONMENT

Yerkes's monograph is a treasure-trove of information for anyone seeking environmental correlates of performance on "tests of intelligence." Since Yerkes explicitly denied any substantial causal role to environment, and continued to insist that the tests measured innate intelligence, this claim may seem paradoxical. One might suspect that Yerkes, in his blindness, didn't read his own information. The situation, in fact, is even more curious. Yerkes read very carefully; he puzzled over every one of his environmental correlations, and managed to explain each of them away with arguments that sometimes border on the ridiculous.

Minor items are reported and dispersed in a page or two. Yerkes found strong correlations between average score and infestation with hookworm in all 4 categories:

	INFECTED	NOT INFECTED
White Alpha	94.38	118.50
White Beta	45.38	53.26
Negro Alpha	34.86	40.82
Negro Beta	22.14	26.09

These results might have led to the obvious admission that state of health, particularly in diseases related to poverty, has some effect upon the scores. Although Yerkes did not deny this possibility, he stressed another explanation (p. 81): "Low native ability may induce such conditions of living as to result in hookworm infection."

In studying the distribution of scores by occupation, Yerkes conjectured that since intelligence brings its own reward, test scores should rise with expertise. He divided each job into apprentices, journeymen, and experts and searched for increasing scores between the groups. But he found no pattern. Instead of abandoning his hypothesis, he decided that his procedure for allocating men to the three categories must have been flawed (pp. 831-832):

It seems reasonable to suppose that a selection process goes on in industry which results in a selection of the mentally more alert for promotion from the apprentice stage to the journeyman stage and likewise from the journeyman stage to the expert. Those inferior mentally would stick at the lower levels of skill or be weeded out of the particular trade. On this hypothesis one begins to question the accuracy of the personnel interviewing procedure.

Among major patterns, Yerkes continually found relationships between intelligence and amount of schooling. He calculated a correlation coefficient of 0.75 between test score and years of education. Of 348 men who scored below the mean in Alpha, only 1 had ever attended college (as a dental student), 4 had graduated from high school, and only 10 had ever attended high school at all. Yet Yerkes did not conclude that more schooling leads to increasing scores per se; instead, he argued that men with more innate intelligence spend more time in school. "The theory that native intelligence is one of the most important conditioning factors in continuance in school is certainly borne out by this accumulation of data" (p. 780).

Yerkes noted the strongest correlation of scores with schooling in considering the differences between blacks and whites. He made a significant social observation, but gave it his usual innatist twist (p. 760):

The white draft of foreign birth is less schooled; more than half of this group have not gone beyond the fifth grade, while one-eighth, or 12.5 percent, report no schooling. Negro recruits though brought up in this country where elementary education is supposedly not only free but compulsory on all, report no schooling in astonishingly large proportion.

Failure of blacks to attend school, he argued, must reflect a disinclination based on low innate intelligence. Not a word about segregation (then officially sanctioned, if not mandated), poor conditions in black schools, or economic necessities for working among the impoverished. Yerkes acknowledged that schools might vary in quality, but he assumed that such an effect must be small and cited, as primary evidence for innate black stupidity, the lower scores of blacks when paired with whites who had spent an equal number of years in school (p. 773):

The grade standards, of course, are not identical all over the country, especially as between schools for white and for negro children, so that "fourth-grade schooling" doubtless varies in meaning from group to group, but this variability certainly cannot account for the clear intelligence differences between groups.

The data that might have led Yerkes to change his mind (had he approached the study with any flexibility) lay tabulated, but unused, within his monograph. Yerkes had noted regional differences in black education. Half the black recruits from Southern states had not attended school beyond the third grade, but half had reached the fifth grade in Northern states (p. 760). In the North, 25 percent completed primary school; in the South, a mere 7 percent. Yerkes also noted (p. 734) that "the percentage of Alphas is very much smaller and the percentage of Betas very much larger in the southern than in the northern group." Many years later, Ashley Montagu (1945) studied the tabulations by state that Yerkes had provided. He confirmed Yerkes's pattern: the average score on Alpha was 21.31 for blacks in thirteen Southern states, and 39.90 in nine Northern states. Montagu then noted that average black scores for the four highest Northern states (45-31) exceeded

the white mean for nine Southern states (43-94). He found the same pattern for Beta, where blacks of six Northern states averaged 34.63, and whites of fourteen Southern states, 31.11. Hereditarians had their pat answer, as usual: only the best Negroes had been smart enough to move North. To people of good will and common sense an explanation in terms of educational quality has always seemed more reasonable, especially since Montagu also found such high correlations between a state's expenditure for education and the average score of its recruits.

One other persistent correlation threatened Yerkes's hereditarian convictions, and his rescuing argument became a major social weapon in later political campaigns for restricting immigration. Test scores had been tabulated by country of origin, and Yerkes noted the pattern so dear to the hearts of Nordic supremacists. He divided recruits by country of origin into English, Scandinavian, and Teutonic on one side, and Latin and Slavic on the other, and stated (p. 699): "the differences are considerable (an extreme range of practically two years mental age)"—favoring the Nordics, of course.

But Yerkes acknowledged a potential problem. Most Latins and Slavs had arrived recently and spoke English either poorly or not at all; the main wave of Teutonic immigration had passed long before. According to Yerkes's protocol, it shouldn't have mattered. Men who could not speak English suffered no penalty. They took Beta, a pictorial test that supposedly measured innate ability independent of literacy and language. Yet the data still showed an apparent penalty for unfamiliarity with English. Of white recruits who scored E in Alpha and therefore took Beta as well (pp. 382-383), speakers of English averaged 101.6 in Beta, while nonspeakers averaged only 77.8. On the individual performance scale, which eliminated the harassment and confusion of Beta, native and foreign-born recruits did not differ (p. 403). (But very few men were ever given these individual tests, and they did not affect national averages.) Yerkes had to admit (p. 395): "There are indications to the effect that individuals handicapped by language difficulty and illiteracy are penalized to an appreciable degree in Beta as compared with men not so handicapped."

Another correlation was even more potentially disturbing. Yerkes found that average test scores for foreign-born recruits rose consistently with years of residence in America.

YEARS OF RESIDENCE	AVERAGE MENTAL AGE
0-5	11.29
6-10	11.70
11-15	12.53
16-20	13.50
20-	13.74

Didn't this indicate that familiarity with American ways, and not innate intelligence, regulated the differences in scores? Yerkes admitted the possibility, but held out strong hope for a hereditarian salvation (p. 704):

Apparently then the group that has been longer resident in this country does somewhat better* in intelligence examination. It is not possible to state whether the difference is caused by the better adaptation of the more thoroughly Americanized group to the situation of the examination or whether some other factor is operative. It might be, for instance, that the more intelligent immigrants succeed and therefore remain in this country, but this suggestion is weakened by the fact that so many successful immigrants do return to Europe. At best we can but leave for future decision the question as to whether the differences represent a real difference of intelligence or an artifact of the method of examination.

The Teutonic supremacists would soon supply that decision: recent immigration had drawn the dregs of Europe, lower-class Latins and Slavs. Immigrants of longer residence belonged predominantly to superior northern stocks. The correlation with years in America was an artifact of genetic status.

The army mental tests could have provided an impetus for social reform, since they documented that environmental disadvantages were robbing from millions of people an opportunity to develop their intellectual skills. Again and again, the data pointed to strong correlations between test scores and environment. Again and again, those who wrote and administered the tests invented tortuous, ad hoc explanations to preserve their hereditarian prejudices.

How powerful the hereditarian biases of Terman, Goddard, and Yerkes must have been to make them so blind to immediate

*Note how choice of language can serve as an indication of bias. This 2.5 year difference in mental ages (13.74-11.29) only represents "somewhat better" performance. The smaller (but presumably hereditary) difference of 2 years between Nordic-Teutonic and Latin-Slav groups had been described as "considerable."

circumstances! Terman seriously argued that good orphanages precluded any environmental cause of low IQ for children in them. Goddard tested confused and frightened immigrants who had just completed a grueling journey in steerage and thought he had captured innate intelligence. Yerkes badgered his recruits, obtained proof of confusion and harassment in their large mode of zero scores, and produced data on the inherent abilities of racial and national groups. One cannot attribute all these conclusions to some mysterious "temper of the times," for contemporary critics saw through the nonsense as well. Even by standards of their own era, the American hereditarians were dogmatists. But their dogma wafted up on favorable currents into realms of general acceptance, with tragic consequences.

Political impact of the army data

CAN DEMOCRACY SURVIVE AN AVERAGE MENTAL AGE OF THIRTEEN?

Yerkes was troubled by his own figure of 13.08 as an average mental age for the white draft. It fit his prejudices and the eugenical fears of prosperous old Americans, but it was too good to be true, or too low to be believed. Yerkes recognized that smarter folks had been excluded from the sample—officers who enlisted and "professional and business experts that were exempted from draft because essential to industrial activity in the war" (p. 785). But the obviously retarded and feeble-minded had also been culled before reaching Yerkes's examiners, thereby balancing exclusions at the other end. The resulting average of 13 might be a bit low, but it could not be far wrong (p. 785).

Yerkes faced two possibilities. He could recognize the figure as absurd, and search his methods for the flaws that engendered such nonsense. He would not have had far to look, had he been so inclined, since three major biases all conspired to bring the average down to his implausible figure. First, the tests measured education and familiarity with American culture, not innate intelligence—and many recruits, whatever their intelligence, were both woefully deficient in education and either too new to America or too impoverished to have much appreciation for the exemplary accomplishments of Mr. Mathewson (including an e.r.a. of 1.14 in 1909). Second, Yerkes's own stated protocol had not been followed. About two-thirds of the white sample took Alpha, and their high fre-

quency of zero scores indicated that many should have been retested in Beta. But time and the indifference of the regular brass conspired against it, and many recruits were not reexamined. Finally, Boring's treatment of zero values imposed an additional penalty on scores already (and artificially) too low.

Or Yerkes could accept the figure and remain a bit puzzled. He opted, of course, for the second strategy:

We know now approximately from clinical experience the capacity and mental ability of a man of 13 years mental age. We have never heretofore supposed that the mental ability of this man was the average of the country or anywhere near it. A moron has been defined as anyone with a mental age from 7 to 12 years. If this definition is interpreted as meaning anyone with a mental age less than 13 years, as has recently been done, then almost half of the white draft (47.3 percent) would have been morons. Thus it appears that feeble-mindedness, as at present defined, is of much greater frequency of occurrence than had been originally supposed.

Yerkes's colleagues were disturbed as well. Goddard, who had invented the moron, began to doubt his own creation: "We seem to be impaled on the horns of a dilemma: either half the population is feeble-minded; or 12 year mentality does not properly come within the limits of feeble-mindedness" (1919, p. 352). He also opted for Yerkes's solution and sounded the warning cry for American democracy:

If it is ultimately found that the intelligence of the average man is 13—instead of 16—it will only confirm what some are beginning to suspect; viz., that the average man can manage his affairs with only a moderate degree of prudence, can earn only a very modest living, and is vastly better off when following directions than when trying to plan for himself. In other words, it will show that there is a fundamental reason for many of the conditions that we find in human society and further that much of our effort to change conditions is unintelligent because we have not understood the nature of the average man (1919, p. 236).

Unfortunate 13 became a formula figure among those who sought to contain movements for social welfare. After all, if the average man is scarcely better than a moron, then poverty is fundamentally biological in origin, and neither education nor better opportunities for employment can alleviate it. In a famous address, entitled "Is America safe for democracy?", the chairman of Harvard's psychology department stated (W. McDougall, quoted in Chase, 1977, p. 226):

The results of the Army tests indicate that about 75 percent of the population has not sufficient innate capacity for intellectual development to enable it to complete the usual high school course. The very extensive testing of school-children carried on by Professor Terman and his colleagues leads to closely concordant results.

In an inaugural address as president of Colgate University, G. G. Cutten proclaimed in 1922 (quoted in Cravens, 1978, p. 224): "We cannot conceive of any worse form of chaos than a real democracy in a population of average intelligence of a little over 13 years."

Again, a catchy, numerical "fact" had risen to prominence as the discovery of objective science—while the fallacies and finagling that thoroughly invalidated it remained hidden in the details of an eight-hundred-page monograph that the propagandists never read.

THE ARMY TESTS AND AGITATION TO RESTRICT IMMIGRATION:
BRIGHAM'S MONOGRAPH ON AMERICAN INTELLIGENCE

The grand average of thirteen had political impact, but its potential for social havoc was small compared with Yerkes's figures for racial and national differences; for hereditarians could now claim that the fact and extent of group differences in innate intelligence had finally, once and for all, been established. Yerkes's disciple C. C. Brigham, then an assistant professor of psychology at Princeton University, proclaimed (1923, p. xx):

We have here an investigation which, of course, surpasses in reliability all preceding investigations, assembled and correlated, a hundred fold. These army data constitute the first really significant contribution to the study of race differences in mental traits. They give us a scientific basis for our conclusions.

In 1923 Brigham published a book, short enough and stated with sufficient baldness (some would say clarity) to be read and used by all propagandists. *A Study of American Intelligence* (Brigham, 1923) became a primary vehicle for translating the army results on group differences into social action (see Kamin, 1974 and Chase, 1977). Yerkes himself wrote the foreword and praised Brigham for his objectivity:

The author presents not theories or opinion but facts. It behooves us to consider their reliability and their meaning, for no one of us as a citizen can afford to ignore the menace of race deterioration or the evident rela-

tions of immigration to national progress and welfare (in Brigham, 1923, p. vii).

Since Brigham derived his "facts" on group differences entirely from the army results, he had first to dismiss the claim that Yerkes's tests might not be pure measures of innate intelligence. He admitted that Alpha might mingle the impact of education with native ability, for it did require literacy. But Beta could only record unadulterated innate intelligence: "Examination Beta involves no English, and the tests cannot be considered as educational measures in any sense" (p. 100). In any case, he added for good measure, it scarcely matters whether the tests also record what Yerkes had called "the better adaptation of the more thoroughly Americanized group to the situation of the examination" (p. 93), since (p. 96):

If the tests used included some mysterious type of situation that was "typically American," we are indeed fortunate, for this is America, and the purpose of our inquiry is that of obtaining a measure of the character of our immigration.* Inability to respond to a "typically American" situation is obviously an undesirable trait.

Once he had proved that the tests measure innate intelligence, Brigham devoted most of his book to dispelling common impressions that might threaten this basic assumption. The army tests had, for example, assessed Jews (primarily recent immigrants) as quite low in intelligence. Does this discovery not conflict with the notable accomplishments of so many Jewish scholars, statesmen, and performing artists? Brigham conjectured that Jews might be more variable than other groups; a low mean would not preclude a few geniuses in the upper range. In any case, Brigham added, we probably focus unduly on the Jewish heritage of some great men because it surprises us: "The able Jew is popularly recognized not only because of his ability, but because he is able and a Jew" (p. 190). "Our figures, then, would rather tend to disprove the popular belief that the Jew is highly intelligent" (p. 190).

But what about the higher scores of Northern vs. Southern blacks? Since Yerkes had also shown that Northern blacks, on average, attended school for several more years than their Southern counterparts, didn't the scores reflect differences in education

* In all other parts of the book, he claims that his aim is to measure and interpret innate differences in intelligence.

more than inborn ability? Brigham did not deny a small effect for education (p. 191), but he presented two reasons for attributing the higher scores of Northern blacks primarily to better biology: first, "the greater admixture of white blood" among Northern blacks; second, "the operation of economic and social forces, such as higher wages, better living conditions, identical school privileges, and a less complete social ostracism, tending to draw the more intelligent negro to the north" (p. 192).

Brigham faced the greatest challenge to hereditarianism on the issue of immigration. Even Yerkes had expressed agnosticism—the only time he considered a significant alternative to inborn biology—on the causes of steadily increasing scores for immigrants who had lived longer in America (see p. 251). The effects were certainly large, the regularity striking. Without exception (see chart on p. 251), each five years of residency brought an increase in test scores, and the total difference between recent arrivals and the longest residents was a full two and a half years in mental age.

Brigham directed himself around the appalling possibility of environmentalism by arguing in a circle. He began by assuming what he intended to demonstrate. He denied the possibility of environmental influence a priori, by accepting as proven the highly controversial claim that Beta must measure unadulterated innate intelligence, whatever Alpha may be doing with its requirement of literacy. The biological basis of declining scores for recent immigrants can then be proven by demonstrating that decrease on the combined scale is not an artifact of differences in Alpha only:

The hypothesis of growth of intelligence with increasing length of residence may be identified with the hypothesis of an error in the method of measuring intelligence, for we must assume that we are measuring native or inborn intelligence, and any increase in our test score due to any other factor may be regarded as an error. . . . If all members of our five years of residence groups had been given Alpha, Beta, and individual examinations in equal proportions, then all would have been treated alike, and the relationship shown would stand without any possibility of error (p. 100).

If the differences between residence groups are not innate, Brigham argued, then they reflect a technical flaw in constructing the combined scale from varying proportions of Alphas and Betas; they cannot arise from a defect in the tests themselves, and therefore cannot, by definition, be environmental indicators of increasing familiarity with American customs and language.

Brigham studied the performances of Alphas and Betas, found that differences between residence groups persisted among the Betas, and proclaimed his counter-intuitive hypothesis of decreasing innate intelligence among more recent immigrants. "We actually find," he proclaimed (p. 102), "that the gain from each type of examination [both Alpha and Beta] is about the same. This indicates, then, that the five years of residence groups are groups with real differences in native intelligence, and not groups laboring under more or less of a linguistic and educational handicap."

Instead of considering that our curve indicates a growth of intelligence with increasing length of residence, we are forced to take the reverse of the picture and accept the hypothesis that the curve indicates a gradual deterioration in the class of immigrants examined in the army, who came to this country in each succeeding 5 year period since 1902 (pp. 110-111). . . . The average intelligence of succeeding waves of immigration has become progressively lower (p. 155).

But why should recent immigrants be more stupid? To resolve this conundrum, Brigham invoked the leading theorist of racism in his day, the American Madison Grant (author of *The Passing of the Great Race*), and that aging relic from the heyday of French craniometry, Count Georges Vacher de Lapouge. Brigham argued that the European peoples are mixtures, to varying degrees, of three original races: 1) Nordics, "a race of soldiers, sailors, adventurers, and explorers, but above all, of rulers, organizers, and aristocrats . . . feudalism, class distinctions, and race pride among Europeans are traceable for the most part to the North." They are "domineering, individualistic, self-reliant . . . and as a result they are usually Protestants" (Grant, quoted in Brigham, p. 182); 2) Alpines, who are "submissive to authority both political and religious, being usually Roman Catholics" (Grant, in Brigham, p. 183), and whom Vacher de Lapouge described as "the perfect slave, the ideal serf, the model subject" (p. 183); 3) Mediterraneans, of whom Grant approved, given their accomplishments in ancient Greece and Rome, but whom Brigham despised because their average scores were even slightly lower than the Alpines.

Brigham then tried to assess the amount of Nordic, Alpine, and Mediterranean blood in various European peoples, and to calculate the army scores on this scientific and racial basis, rather than from the political expedient of national origin. He devised the following

figures for average intelligence: Nordic, 13.28; Alpine, 11.67; Mediterranean, 11.43.

The progressive decline of intelligence for each five-year residency group then achieved its easy, innatist explanation. The character of immigration had changed markedly during the past twenty years. Before then, arrivals had been predominantly Nordic; since then, we have been inundated by a progressively increasing number of Alpines and Mediterraneans, as the focus of immigration shifted from Germany, Scandinavia, and the British Isles to the great unwashed of southern and eastern Europe—Italians, Greeks, Turks, Hungarians, Poles, Russians, and other Slavs (including Jews, whom Brigham defined racially as "Alpine Slavs"). Of the inferiority of these recent immigrants, there can be no doubt (p. 202):

The Fourth of July orator can convincingly raise the popular belief in the intellectual level of Poland by shouting the name of Kosciusko from a high platform, but he cannot alter the distribution of the intelligence of the Polish immigrant.

But Brigham realized that two difficulties still stood before his innatist claim. He had proved that the army tests measured inborn intelligence, but he still feared that ignorant opponents might try to attribute high Nordic scores to the presence of so many native speakers of English in the group.

He therefore divided the Nordic group into native speakers from Canada and the British Isles, who averaged 13.84, and "non-English speakers," primarily from Germany, Holland, and Scandinavia, who averaged 12.97. Again, Brigham had virtually proved the environmentalist claim that army tests measured familiarity with American language and customs; but again, he devised an innatist fudge. The disparity between English and non-English Nordics was half as large as the difference between Nordics and Mediterraneans. Since differences among Nordics could only represent the environmental effects of language and culture (as Brigham admitted), why not attribute variation between European races to the same cause? After all, the so-called non-English Nordics were, on average, more familiar with American ways and should have scored higher than Alpines and Mediterraneans on this basis alone. Brigham called these men "non-English" and used

them as a test of his language hypothesis. But, in fact, he only knew their country of origin, not their degree of familiarity with English. On average, these so-called non-English Nordics had been in America far longer than the Alpines or Mediterraneans. Many spoke English well and had spent enough years in America to master the arcana of bowling, commercial products, and film stars. If they, with their intermediary knowledge of American culture, scored almost a year below the English Nordics, why not attribute the nearly two-year disadvantage of Alpines and Mediterraneans to their greater average unfamiliarity with American ways? It is surely more parsimonious to use the same explanation for a continuum of effects. Instead, Brigham admitted environmental causes for the disparity within Nordics, but then advanced innatism to explain the lower scores of his despised southern and eastern Europeans (pp. 171-172):

There are, of course, cogent historical and sociological reasons accounting for the inferiority of the non-English speaking Nordic group. On the other hand, if one wishes to deny, in the teeth of the facts, the superiority of the Nordic race on the ground that the language factor mysteriously aids this group when tested, he may cut out of the Nordic distribution the English speaking Nordics, and still find a marked superiority of the non-English speaking Nordics over the Alpine and Mediterranean groups, a fact which clearly indicates that the underlying cause of the nativity differences we have shown is race, and not language.

Having met this challenge, Brigham encountered another that he couldn't quite encompass. He had attributed the declining scores of successive five-year groups to the decreasing percentage of Nordics in their midst. Yet he had to admit a troubling anachronism. The Nordic wave had diminished long before, and immigration for the two or three most recent five-year groups had included a roughly constant proportion of Alpines and Mediterraneans. Yet scores continued to drop while racial composition remained constant. Didn't this, at least, implicate language and culture? After all, Brigham had avoided biology in explaining the substantial differences between Nordic groups; why not treat similar differences among Alpines and Mediterraneans in the same way? Again, prejudice annihilated common sense and Brigham invented an implausible explanation for which, he admitted, he had no direct evidence. Since scores of Alpines and Mediterraneans had

been declining, the nations harboring these miscreants must be sending a progressively poorer biological stock as the years wear on (p. 178):

The decline in intelligence is due to two factors, the change in the races migrating to this country, and to the additional factor of the sending of lower and lower representatives of each race.

The prospects for America, Brigham groused, were dismal. The European menace was bad enough, but America faced a special and more serious problem (p. xxi):

Running parallel with the movements of these European peoples, we have the most sinister development in the history of this continent, the importation of the negro.

Brigham concluded his tract with a political plea, advocating the hereditarian line on two hot political subjects of his time: the restriction of immigration and eugenical regulation of reproduction (pp. 209-210):

The decline of American intelligence will be more rapid than the decline of the intelligence of European national groups, owing to the presence here of the negro. These are the plain, if somewhat ugly, facts that our study shows. The deterioration of American intelligence is not inevitable, however, if public action can be aroused to prevent it. There is no reason why legal steps should not be taken which would insure a continuously progressive upward evolution.

The steps that should be taken to preserve or increase our present intellectual capacity must of course be dictated by science and not by political expediency. Immigration should not only be restrictive but highly selective. And the revision of the immigration and naturalization laws will only afford a slight relief from our present difficulty. The really important steps are those looking toward the prevention of the continued propagation of defective strains in the present population.

As Yerkes had said of Brigham: "The author presents not theories or opinions but facts."

THE TRIUMPH OF RESTRICTION ON IMMIGRATION

The army tests engendered a variety of social uses. Their most enduring effect surely lay in the field of mental testing itself. They were the first written IQ tests to gain respect, and they provided essential technology for implementing the hereditarian ideology

that advocated, contrary to Binet's wishes, the testing and ranking of all children.

Other propagandists used the army results to defend racial segregation and limited access of blacks to higher education. Cornelia James Cannon, writing in the *Atlantic Monthly* in 1922, noted that 89 percent of blacks had tested as morons and argued (quoted in Chase, 1977, p. 269):

Emphasis must necessarily be laid on the development of the primary schools, on the training in activities, habits, occupations which do not demand the more evolved faculties. In the South particularly . . . the education of the whites and colored in separate schools may have justification other than that created by race prejudice. . . . A public school system, preparing for life young people of a race, 50 percent of whom never reach a mental age of 10, is a system yet to be perfected.

But the army data had their most immediate and profound impact upon the great immigration debate, then a major political issue in America. Restriction was in the air, and would have occurred without scientific backing. (Consider the wide spectrum of support that limitationists could muster—from traditional craft unions fearing multitudes of low-paid laborers, to jingoists and America firsters who regarded most immigrants as bomb-throwing anarchists and who helped make martyrs of Sacco and Vanzetti.) But the timing, and especially the peculiar character, of the 1924 Restriction Act clearly reflected the lobbying of scientists and eugenicists, and the army data formed their most powerful battering ram (see Chase, 1977; Kamin, 1974; and Ludmerer, 1972).

Henry Fairfield Osborn, trustee of Columbia University and president of the American Museum of Natural History, wrote in 1923, in a statement that I cannot read without a shudder when I recall the gruesome statistics of mortality for World War I:

I believe those tests were worth what the war cost, even in human life, if they served to show clearly to our people the lack of intelligence in our country, and the degrees of intelligence in different races who are coming to us, in a way which no one can say is the result of prejudice. . . . We have learned once and for all that the negro is not like us. So in regard to many races and subraces in Europe we learned that some which we had believed possessed of an order of intelligence perhaps superior to ours [read Jews] were far inferior.

Congressional debates leading to passage of the Immigration Restriction Act of 1924 frequently invoked the army data. Eugenists lobbied not only for limits to immigration, but for changing its character by imposing harsh quotas against nations of inferior stock—a feature of the 1924 act that might never have been implemented, or even considered, without the army data and eugenicist propaganda. In short, southern and eastern Europeans, the Alpine and Mediterranean nations with minimal scores on the army tests, should be kept out. The eugenicists battled and won one of the greatest victories of scientific racism in American history. The first restriction act of 1921 had set yearly quotas at 3 percent of immigrants from any nation then resident in America. The 1924 act, following a barrage of eugenicist propaganda, reset the quotas at 2 percent of people from each nation recorded in the 1890 census. The 1890 figures were used until 1930. Why 1890 and not 1920 since the act was passed in 1924? 1890 marked a watershed in the history of immigration. Southern and eastern Europeans arrived in relatively small numbers before then, but began to predominate thereafter. Cynical, but effective, "America must be kept American," proclaimed Calvin Coolidge as he signed the bill.

BRIGHAM RECANTS

Six years after his data had so materially affected the establishment of national quotas, Brigham had a profound change of heart. He recognized that a test score could not be reified as an entity inside a person's head:

Most psychologists working in the test field have been guilty of a naming fallacy which easily enables them to slide mysteriously from the score in the test to the hypothetical faculty suggested by the name given to the test. Thus, they speak of sensory discrimination, perception, memory, intelligence, and the like while the reference is to a certain objective test situation (Brigham, 1930, p. 159).

In addition, Brigham now realized that the army data were worthless as measures of innate intelligence for two reasons. For each error, he apologized with an abjectness rarely encountered in scientific literature. First, he admitted that Alpha and Beta could not be combined into a single scale as he and Yerkes had done in producing averages for races and nations. The tests measured dif-

ferent things, and each was internally inconsistent in any case. Each nation was represented by a sample of recruits who had taken Alpha and Beta in differing proportions. Nations could not be compared at all (Brigham, 1930, p. 164):

As this method of amalgamating Alphas and Betas to produce a combined scale was used by the writer in his earlier analysis of the Army tests as applied to samples of foreign born in the draft, that study with its entire hypothetical superstructure of racial differences collapses completely.

Secondly, Brigham acknowledged that the tests had measured familiarity with American language and culture, not innate intelligence:

For purposes of comparing individuals or groups, it is apparent that tests in the vernacular must be used only with individuals having equal opportunity to acquire the vernacular of the test. This requirement precludes the use of such tests in making comparative studies of individuals brought up in homes in which the vernacular of the test is not used, or in which two vernaculars are used. The last condition is frequently violated here in studies of children born in this country whose parents speak another tongue. It is important, as the effects of bilingualism are not entirely known. . . . Comparative studies of various national and racial groups may not be made with existing tests. . . . One of the most pretentious of these comparative racial studies—the writer's own—was without foundation (Brigham, 1930, p. 165).

Brigham paid his personal debt, but he could not undo what the tests had accomplished. The quotas stood, and slowed immigration from southern and eastern Europe to a trickle. Throughout the 1930s, Jewish refugees, anticipating the holocaust, sought to emigrate, but were not admitted. The legal quotas, and continuing eugenicist propaganda, barred them even in years when inflated quotas for western and northern European nations were not filled. Chase (1977) has estimated that the quotas barred up to 6 million southern, central, and eastern Europeans between 1924 and the outbreak of World War II (assuming that immigration had continued at its pre-1924 rate). We know what happened to many who wished to leave but had nowhere to go. The paths to destruction are often indirect, but ideas can be agents as sure as guns and bombs.