sons 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 1915. He had been on the faculty of Harvard University since 1902. He was a superb organizer, and an eloquent promotor of 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 plusses 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, commandeered 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 grousing 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-

---

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, 1941, 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 moronity 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 11.34; 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 eugenical 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 bemused 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 acquirement, 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

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

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

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.

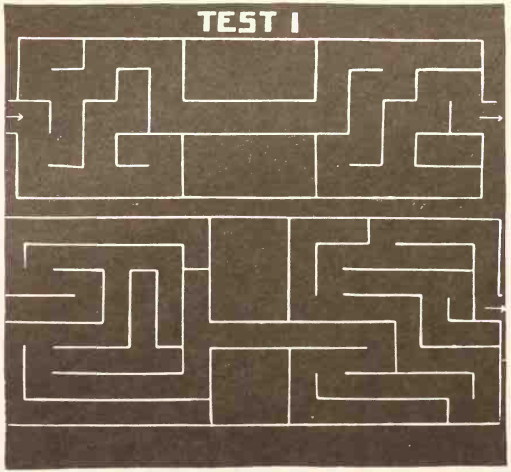
The idea of working fast must be impressed on the men during the maze test. Examiner and orderlies walk around the room, motioning to

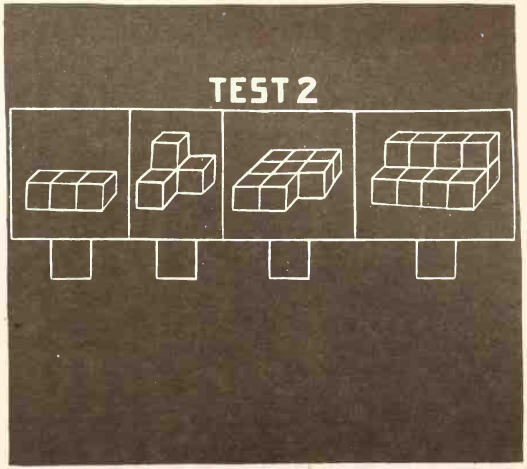
men who are not working, and saying, "Do it, do it, hurry up, quick." At the end of 2 minutes examiner says, "Stop! Turn over the page to test 2."

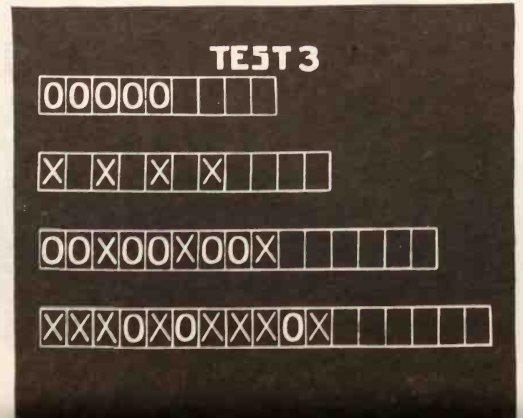
The examiner demonstrated test 2, cube counting, with three-dimensional models (my son had some left over from his baby days). Note that recruits who could not write numbers would receive scores of zero even if they counted all the cubes correctly. Test 3, the X-O series, will be recognized by nearly 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:

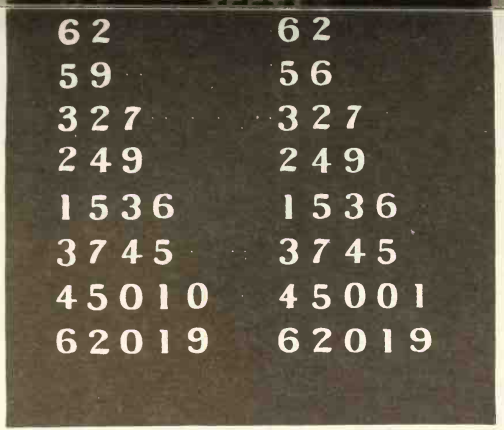
"This is test 6 here. Look. A lot of pictures." After everyone has found the place, "Now watch." Examiner points to hand and says to demonstrator, "Fix it." Demonstrator does nothing, but looks puzzled. Examiner points to the picture of the hand, and then to the place where the finger is missing and 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."

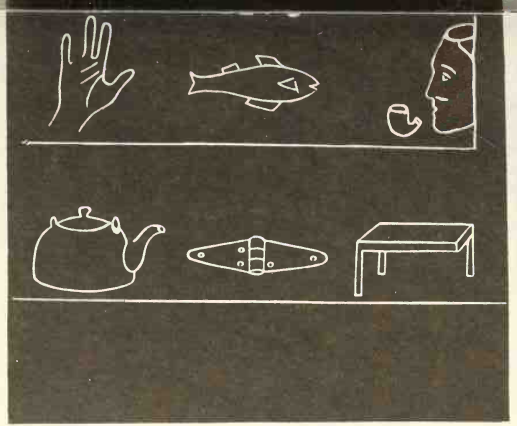

TEST 1

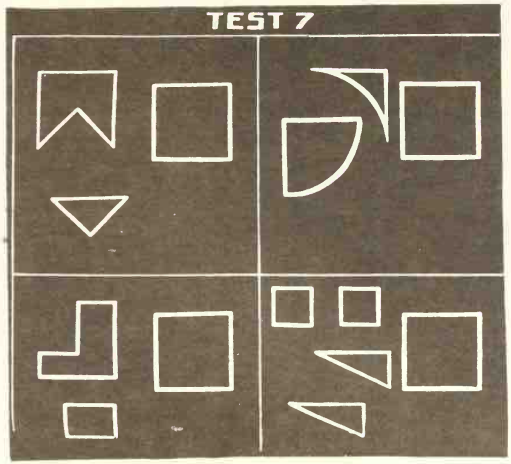

TEST 2


TEST 3
OOOOO
X X X X
OOXOOXOOX
XXXOXOXXXOX


TEST 4
1 2 3 4 5 6 7 8 9
3 1 2 3 2 1 2 1 3 4 7 5 4 1 6







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

The examination itself is also worth reprinting (Fig. 5.5). Best of luck with pig tails, crab legs, bowling balls, tennis nets, 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:

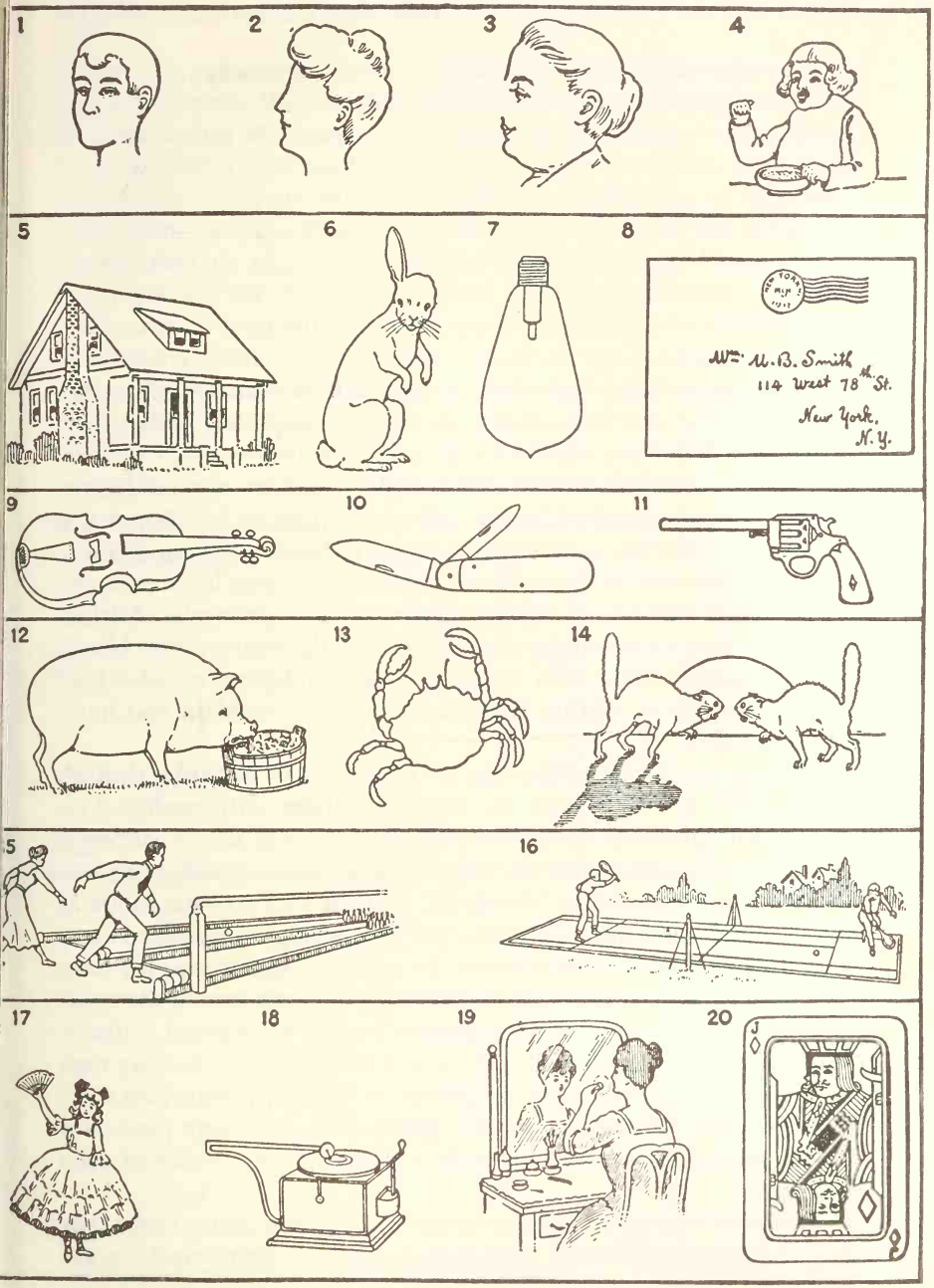
Rules for Individual Items

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

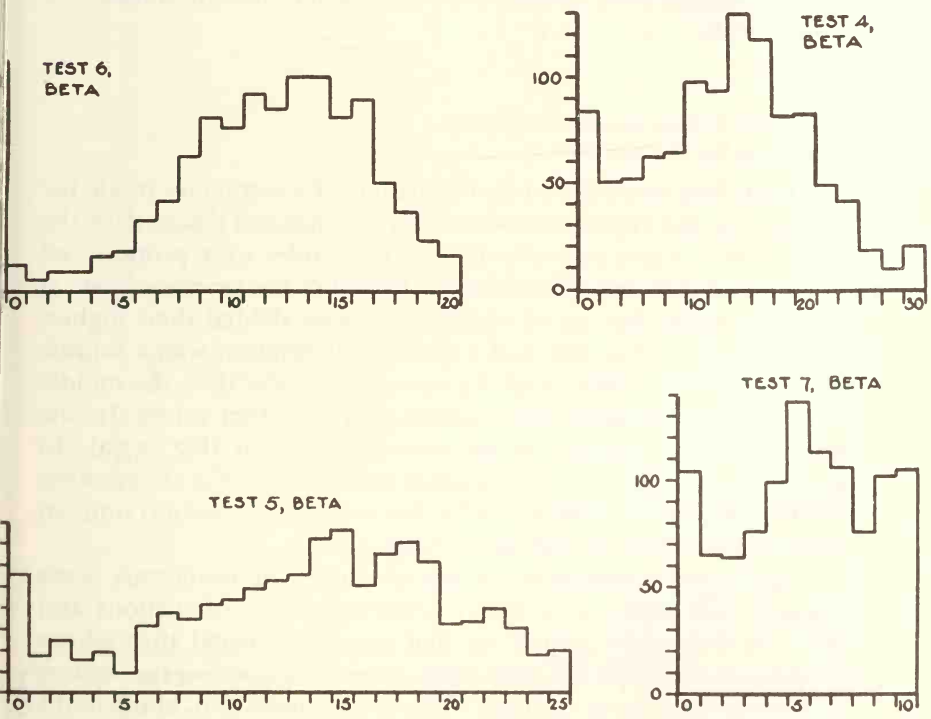
some had never taken a test before or even held a pencil. Many did not understand the instructions and were completely befuddled. Those who did comprehend could complete only a small part of most tests in the allotted time. Meanwhile, if anxiety and confusion had not already reached levels sufficiently 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 inadvisibility 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).

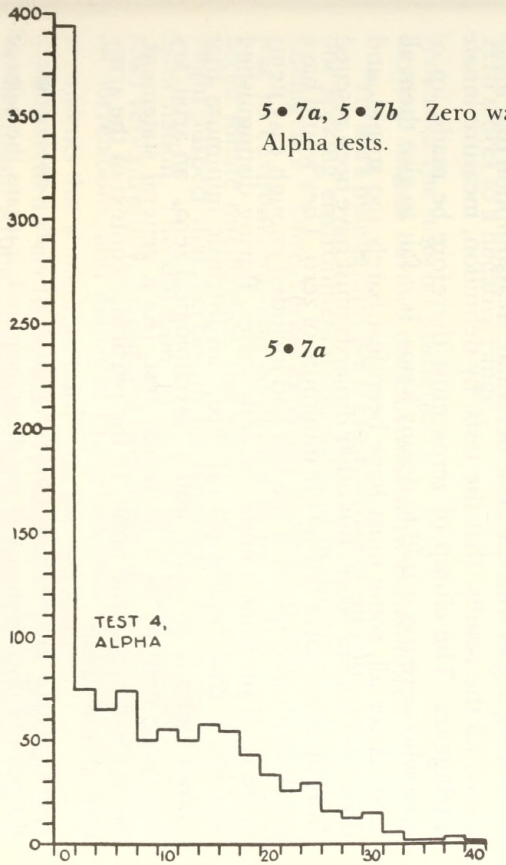
Oh yes, was anyone wondering how my students fared? They did very well of course. Anything else would have been shocking, since all the tests are greatly simplified precursors of examinations they have been taking all their lives. Of fifty-three students, thirty-one scored A and sixteen B. Still, more 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

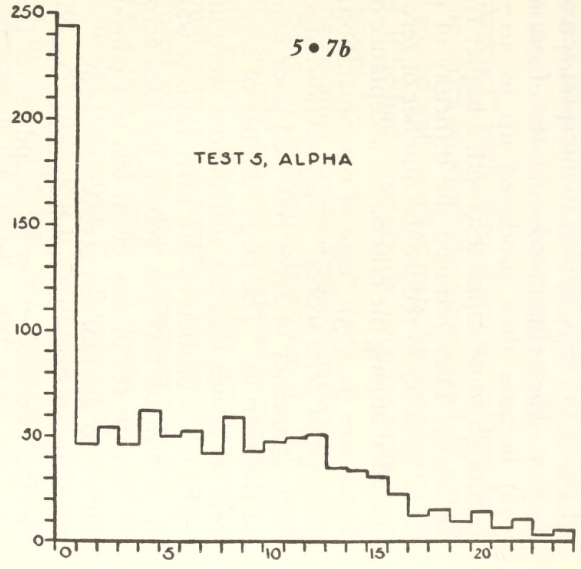
If the Beta test faltered on the artifact of a secondary mode for zero scores, the Alpha test became an unmitigated disaster for the same reason, vastly intensified. The zero modes were pronounced in Beta, but they never reached the height of the primary mode at a middle value. But six of eight Alpha 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.



5 • 7a



5 • 7b

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. 811): "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 fitted 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. 263):

> 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. Eugenicists 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 eugenical 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.