10 The science of merit and the merit of science

Mental order and social order in early twentieth-century France and America

John Carson

On 16 February 2001, Richard C. Atkinson, president of the University of California, announced that he was proposing to abolish the SAT/ACT requirement for applicants seeking admission to any school in the university system. Atkinson justified his decision, the culmination of decades of controversy surrounding college aptitude tests and their role in American society, on the grounds that such tests were “not compatible with the American view on how merit should be defined and opportunities distributed” (Schemo 2001). Atkinson’s proposal marked a decisive shift in the understanding of the role of aptitude/intelligence tests in the American educational system. From their development at the turn of the century up to the 1960s, mental tests were promoted precisely as a means of defining merit scientifically and thereby ensuring that all who took them would be treated equally. To their advocates, testing was an invaluable agent of reform, able to move the distribution of opportunities away from the privileges of birth or money or power. It could produce, they believed, a system that preserved equality by providing objectivity and accountability while still allowing for the extraordinary heterogeneity of the locally administered system of US primary and secondary education (Lemann 1999).

Instruments such as the SAT, in their eyes, promised to help negotiate an issue of fundamental importance to American democracy in the twentieth century: how to distribute coveted and limited social goods such as educational opportunities in ways that would appear fair and equal even to those least successful in garnering rewards from the system. As Atkinson suggested, increasing public skepticism about the ability of intelligence tests to fulfill this function – to make the decisionmaking processes seem legitimate and fair by being based solely on merit – placed the tests in a precarious position à la à rs their continued utility for questions of college admissions.

This problem of how to satisfy demands for both equality and merit within a democratic political culture, and specifically the recourse to scientific objects and instruments and methods of quantification/classification to manage the tension between them, forms the subject of this essay. Adopting a co-productionist perspective, I will emphasize that “the ways in which we know and represent the world (both nature and society) are inseparable from the ways in which we choose to live in it”, as Sheila Jasanoff states in Chapter 1. I suggest that technologies of
merit such as intelligence tests are best understood not as invented apart from and then applied to the social realm, but rather as developed in tandem with other elements of the social order, together producing the reality in which all must operate (Jasanoﬀ 1987; 1990; 1999). To sustain this position, I examine the emergence and response to intelligence and its tests in the early twentieth century, first in their birthplace, France, and then in the United States. By so doing, I explore how two distinct political cultures recast their approaches to equality and merit according to the possibilities that the ability to measure intelligence provided. It is a story at once about the fashioning of a scientiﬁc instrument to ﬁt the needs of democratic political culture, and the fashioning of political cultures to ﬁt the ﬁndings of an instrument.

The modern intelligence test provoked little interest in France upon its formulation by French psychologists Alfred Binet and Théodore Simon in the period 1905–11. Conceived of, at best, as a supplement to other ways of analyzing human psychology, administrators and public ofﬁcials deemed it of slight relevance to a system of merit already thought to be well served by public education. Thus French culture experienced few of the paroxysms of worry over intelligence and its measurement that would be engendered in the United States. Deemed useful in certain limited contexts, intelligence assessment generated little enthusiasm within a society that had already institutionalized a means to resolve, or at least channel, frictions between the claims of meritocracy and democracy through a highly standardized, universal system of education. The ability of that system to both produce and justify social hierarchy – through its training of the political/technocratic elite – while still representing such a social hierarchy as potentially open to all, was critical to the French approach to linking equality and merit. Intelligence, whatever its ontological status in French psychology, had little role to play in justifying such determinations, and, as a consequence, its purview remained highly circumscribed.

France’s indifference to intelligence and intelligence tests contrasted markedly with the situation in the United States, where psychologists and administrators embraced both object and instrument and promoted them as a means of making a range of social decisions seem objective and fair. The most important early success of the American mental testers occurred soon after the Binet test arrived in America. During World War I, the exigencies of coping with mass mobilization made the US Army receptive to the initiative of a number of American psychologists to establish the nation’s ﬁrst large-scale intelligence testing program, for the purpose of classifying and sorting new recruits (Carson 1993). In the aftermath of the war, the results of that testing, particularly the widely reported ﬁnding that the average American soldier had the mental age of a thirteen-year-old, provoked an intense debate over the implications of intelligence for democracy. Waged in a number of journals, articles pitted leading mental testers such as Lewis M. Terman and Guy M. Whipple and their allies against a range of critics, including the educator William C. Bagley and such prominent public intellectuals as Walter Lippmann and John Dewey.

At its most basic, the controversy swirlled around the ﬁgure of thirteen years: did it imply that a signiﬁcant proportion of the American population was of limited intelligence and would remain so, or did it reveal ﬂaws in the testing process that belied the psychologists’ more exuberant claims about the knowledge they were producing? More broadly, issues of the nature of merit, equality and democratic citizenship were pushed to the fore. Could a democracy, some wondered, provide the same kinds of citizenship to all of its adult members if there were signiﬁcant individual differences in the ability to be, or even choose, a good leader? The proponents of intelligence testing, by and large, argued that what was fundamental to a democracy was equality of opportunity, which meant only that each person should be afforded educational opportunities commensurate with their abilities. The critics of mental testing, on the other hand, considered the hierarchies of merit that testing could create inimical to a democracy, and stressed instead an equality founded on uniform treatment of all. For both sides, questions of democratic citizenship were intimately linked to understandings of human nature, and buried within the debate was an argument as well over who, if anyone, had the right to decide what individuals were, in fact, capable of accomplishing (Callon forthcoming; Rabeharisoa and Callon this volume).

The contrast between the very different trajectories of intelligence and its tests in the United States and France reveals how culturally speciﬁc were the objectivities by which concepts such as “equality” and “merit” were made socially real in each nation. Drawing on a tradition of the celebration of individual rights derived from Britain, Americans since the eighteenth century had consistently turned to notions of nature and natural rights as one means of grounding and justifying political claims. Within American civic discourse, nature was framed as standing outside of society and as a counterweight to it, a source of rights independent of the vagaries of particular governments or social arrangements, and also a source of difference that could undercut claims to such rights (Rodgers 1987). Intelligence and its tests, in this context, could be seen as a seemingly natural means of determining what equality could and should mean, what differences were and were not real, independent of state or partisan inﬂuence. In France, on the other hand, republican discourse since the eighteenth century, while celebrating the need to replace tradition with reason and imbued as well with the presumption that nature was the ultimate source of rights, provided a far greater role for society as a whole (Nord 1995). Rousseau’s notion of the “general will” and a commitment to sovereignty lying not in individuals but in the nation, formed the core of French republican doctrine, and in practice it was not so much nature as the good of society that served as the justification for various attempts at reform or revolution. Within this political culture, where citizenship was conceived of as communal and corporate as well as individual, the state served as the primary vehicle to ensure equality for all citizens. It accomplished this primarily through the establishment of national institutions, theoretically competitively open to every citizen, whose purpose was to bring the cream of the nation to its service. The objectivities provided by intelligence and
its tests, in such a context, were not only superfluous, but might even have been the source of rival claims to elite privilege on the basis of criteria unmediated by the collective influence of the nation.

The Binet-Simon scale in France: a solution without a problem

Ironically, in view of its eventual uptake, the story of the modern intelligence test begins in France. When Alfred Binet commenced work in 1904 on the first version of his measuring scale, published in 1905, the direct impetus was his appointment to a ministerial commission on children lagging in school (Binet and Simon 1905b; Binet and Simon 1916). The scale began, as Binet and his colleague Théodore Simon explained at the Fifth International Congress of Psychology in Rome (1905), as a new means of diagnosing idiocy, imbecility and feeblemindedness (débilité), and distinguishing these subnormal types of intelligence from normal minds lacking sufficient training (Binet and Simon 1905a: 508). Their initial goal was to replace the “arbitrary” classificatory methods of doctors and educators with a procedure for defining degrees of intellectual deficit that was more objective, precise, and above all scientific. “It is a hackneyed remark”, Binet declared, “that the definitions, thus far proposed, for the different states of subnormal intelligence, lack precision” (Binet and Simon 1916: 10). The thrust of Binet’s work in the 1905 scale was to remedy this lack of precision by creating a series of tasks that would differentiate unambiguously between the normal mind and the three pathological classifications of intelligence enumerated above.

In subsequent years, Binet and Simon’s conception of the possible applications of the scale and the social roles it could fulfill altered, and as it did, the instrument itself evolved correspondingly (Reuchlin 1968: 390). Almost immediately, French physicians proved hostile to a diagnostic technique that apparently challenged their authority; a reaction that pushed Binet to reorient his intelligence work away from medical diagnosis and “toward practical and social questions”, as he declared publicly in 1908 (Binet 1908: 5). By the scale’s 1911 revision, Binet and Simon had substantially transformed the notion of intelligence embedded within it. In place of the meaning common in French psychology before their efforts—intelligence as the composite name for a diversity of faculties of interest for diagnosing intellectual pathologies—they had substituted the notion of intelligence as a singular, quantifiable entity applicable to the entire population (Gould 1996: 150–152; Tuddenham 1963: 490; Wolf 1969: 235–236).

The scale equated normal intelligence functionally with the ability to make judgments, comparisons and decisions in line with broadly accepted cultural norms, and statistically with the mean performance of its sample population on the items in the test. In the process, Binet and Simon extended the potential social applications of the scale well beyond simple classification of the feebleminded, instead envisioning it as a way of objectively ranking entire populations and allocating resources to meet a variety of needs:

As the range of these applications suggests, by 1911 Binet and Simon had conceptualized the metric scale primarily as a mechanism for assisting institutions in the management of individuals. Having removed much of the clinical feel of the 1905 version, they had rendered Binet-Simon intelligence instead as something quantifiable and unidimensional, a social technology advertised as providing impartial mediation between the rationalizing imperatives of various social institutions and the levels of ability of the citizenry. Binet and Simon’s representation of intelligence as universally distributed and of fundamental significance in assessing human potential allowed them to argue that use of the scale could bring objectivity and accountability to a range of administrative practices and social decisions. Its employment, they suggested, could shift the basis for action from subjective choice to scientific determination (Wise 1988).

The potential ramifications of such a social technology, if truly enlisted to perform all of the tasks that Binet and Simon envisioned for it, might have been far reaching and profound. However, the actual uses to which the metric scale was put in France turned out to be rather modest. Binet’s sudden death in 1911 at the age of fifty-four deprived the scale of its most important champion, and Binet’s Sorbonne laboratory, symbolic home of his activities, was entrusted not to Binet’s disciple Simon, but to a younger rival, Henri Piéron (Piéron 1939; 1965). While most French psychologists acknowledged that the Binet-Simon scale was a measuring instrument of some practical value, few found it of more than limited relevance to their own research programs. The orientation of French psychological investigation in the early twentieth century was overwhelmingly toward clinical studies of individual pathology or laboratory experimentation on basic psychological functions, not toward the practical applications of psychological science to social problems (Brooks 1993; Carroy and Plas 1996; Danziger 1965; Reuchlin 1978).

Even those psychologists, such as Piéron and his wife Marguerite, who did produce studies of intelligence and its development during the 1910s and 1920s, employed a focus significantly different from Binet’s. In one form or another, they sought to investigate the complexities of intelligence and to understand it as a multi-faceted phenomenon, whether ultimately singular or compound in composition. For them, it was not Binet’s metric scales of intelligence that served as a model, but his 1903 study of his daughters, L’Etude experimentale de l’intelligence (Binet 1903). As Henri Piéron noted in 1927, “(t)hough we always employ the
same word, intelligence, for the aptitude to solve problems, it is still necessary to understand that under this term the mental action may be quite different, depending on the nature of the problems to be solved" (Piéron 1929: 178-179). In other words, intelligence was not one thing but many, a sentiment echoed by Benjamin Bourdon in 1926 when he argued that proficiency in one area, even to the level of genius, did not insure ability in any other domain (Bourdon 1926).

Such a stance did not mean that the connection of intelligence with the practical needs of the individual and the state disappeared completely in France. The Piérons, along with the physiologist Henri Laugier and psychologist Jean-Marie Laby, devoted significant time and energy to the development of the field of occupational testing (orientation professionnelle), and the Binet-Simon scale was an important component in their repertoire of investigative tools. In addition, the psychiatrist Georges Huyer used the intelligence scale extensively in his studies of abnormal children – in particular juvenile delinquents who were required to have psychological examinations before appearance in court – but again the Binet-Simon test was only one of a number of ways in which these individuals were assessed (Schneider 1989; 1992). As historian William Schneider has observed, recourse to a single, global measurement of intelligence as a primary means of social decisionmaking did not even become a serious public issue in France until the mid-1930s, when the Popular Front suggested a nationwide survey to determine the percentage of retarded and abnormal children requiring special educational services. Postponed with the defeat of the leftists in the late 1930s, the proposal was resuscitated with the advent of the Vichy government in 1940 – spurred on by its own interest in the health of the nation and the family – but only finally reached conclusion under the Fourth Republic in 1954, with the report "The intellectual level of school-age children". And even at this point, the role for global intelligence remained limited: a means of categorizing certain intellectual deficits that were deemed of particular relevance to French educators and state administrators.

In the end, what is most apparent about this story of the emergence of intelligence and its tests in France is how circumscribed the claims for them came to be. When French psychologists or psychiatrists needed to investigate the nature of an individual's intellect, as we have seen, they used not one assessment technique but a whole battery of them, in which Binet-Simon intelligence might be, at most, a single element. What is more, rarely did important voices in French culture suggest that intelligence measures might stand in for, or even substantially contribute to, culturally sanctioned methods of judging individual merit, outside the diagnosis of particular pathological conditions. As important as the issue of merit was in Third Republic France, it was addressed primarily through a different type of examination, the concours, whose purpose was not to identify individual potential (or deficit), but to choose at each level a cadre of high achievers for advanced training to meet the needs of state and society.

Since at least Napoleon III's Second Empire, republicans had routinely included demands for rule by the most able in their plans for the reconstruction of the French polity along more democratic lines. Instead of intelligence, however, at the heart of the French republican vision of merit lay the system of education, with its institutional commitment to equality and excellence through competition and selection. One of the central institutions of Third Republic culture, and both manufacturer and employer of many of the Republic's most ardent proponents, the educational system had already developed by the early nineteenth century a highly elaborated system for winnowing the "best" from the school-age population and directing them to special institutions of higher learning (Clark 1973; Mayeur 1981; Smith 1982; Weisz 1983). A series of competitive examinations (the concours) and ever more elite schools, culminating in the Ecole Normale Supérieure or Ecole Polytechnique and the agrégation, served to select and make available to the French state la crème de la crème, a group of extremely well trained young people whose very success in negotiating the system defined them as the most talented in the nation (Clark 1973; Smith 1982; Shinn 1980).

Although the percentage of primary-age children enrolled in the French educational system continued to increase throughout the nineteenth century, this expansion had little effect on the system's pyramidal structure or its ability to generate sufficient numbers of "superior" products to meet the needs of state and society. Rather, consumed with their battle with the Catholic church over control of education (and convinced that public instruction was the bulwark of the republic), leading republicans of the left and right – all products of the lycées and grandes écoles – viewed the educational system as the guarantor of the triumph of talent over tradition. Thus few in positions of authority perceived any real need for new mechanisms of selection and classification within most areas of French life, and they routinely rejected as superfluous additional methods for selection, even one purporting to measure something as fundamental as intelligence. For them, merit and equality had to be seen as byproducts of a system potentially open to all, and not lodged in a faculty presumably present from birth.

In the United States, on the other hand, intelligence was quickly seized upon and promoted by psychologists as a major constituent of merit. The systems of social sorting that had worked tolerably well during the antebellum period proved largely unable to cope with the powerful transformations reshaping late nineteenth-century American society and culture. Large-scale urbanization and industrialization, rapid shifts between prosperity and depression, unprecedented labor unrest, massive immigration from eastern and southern Europe, and the emancipation of millions of formerly enslaved African Americans changed the nature of community life. A multitude of new languages and cultural practices were introduced into urban areas increasingly segregated by race/ethnicity and class, and new practices of control and exclusion were fashioned. At the same time, worries about cultural degeneration and the deleterious influence of the abnormal, and particularly the subnormal, spread widely in the United States, anxieties both fueled by and fueling interest in eugenics (Gravens 1978; Degler 1991; Kevles 1986). Faced with the emergence of mass society, leaders in education and business increasingly viewed methods of selection – be it for higher
education, governmental bureaucracies or industrial jobs – that had previously relied on personal familiarity, family status, community connections or craft guild hierarchies, as antiquated and out of place.

As a consequence, space opened for newer approaches to selection, inflected less by the older Victorian language of character than by seemingly more constant Progressive formulations steeped in the idiom of science and engineering. Civil service reform, advocated by Progressives as a way of replacing the regime of politics and spoils with one of "ability", was typical of turn-of-the-century moves to redefine merit. Reformers extolled transparency, objective measures and equal opportunities for all, seeking to replace backroom patronage and subjective assessments that to them were symbols of corruption (Ingraham 1995; Shepard 1884). In addition, there were a number of significant changes in the nature of education. Higher education expanded enormously, spurred by the Morrill Act (1862) for land-grant colleges and by the philanthropy of Gilded Age robber barons such as Rockefeller, Carnegie and Stanford for private universities (Reuben 1996; Rudolph 1962; Veysey 1965). At the same time, publicly funded systems of secondary education in urban and suburban areas grew even more rapidly, meaning that the number of Americans, both men and women, with post-primary educations soared, with the result that such qualifications seemed less selective than they once had seemed. Finally, the piecemeal nature of the American education system, characterized by local control of public schools and universities and the existence of a wide variety of private institutions, meant that no approach to classification or selection based on uniform curricular standards or a singular ideal of public service was likely to develop. Within this cultural/social context, so different in important ways from France, psychologists and administrators met the arrival of the Binet-Simon scale not with indifference, but with enthusiasm, touching off a vogue for testing that had powerful and far-reaching consequences.

Building an American intelligence: a techno-scientific solution

The 1905 Binet-Simon scale arrived in the United States in 1908, one of the spoils of a research junket to Europe undertaken by American psychologist Henry H. Goddard of the New Jersey Training School for Feebleminded Girls and Boys in Vineland (Goddard 1908; Zenderland 1998). Goddard commenced experimenting with the scale soon after his return, but found it of limited value. When Binet published the 1908 revised Binet-Simon, Goddard at first hesitated to use it, recalling later that it had "seemed impossible to grade intelligence in that way. It was too easy, too simple" (Goddard 1916: 5). Intrigued by the possibility, however, Goddard finally administered the revised instrument to residents at his school, and reported himself to be amazed by the results. The scale, he declared, provided accurate diagnoses of the mental levels of all of the children he had examined (Goddard 1910: 389). What had taken Goddard and the staff months to determine through long exposure to the subjects, the Binet-Simon was able to reveal in a single testing session. It would soon be clear that Goddard's experience was by no means unique. By 1916, when Lewis Terman completed the Stanford-Binet, his version of the Binet-Simon intelligence scale, intelligence was already something of an industry in American psychology. Articles either about the Binet-Simon scale or research data generated by the scale filled professional journals; numerous rival versions of the scale competed for clientele and professional dominance; and a great deal of hand-wringing was in evidence about the improper use of the technology by those deemed "ill trained" to apply it appropriately (Goddard 1916; Terman 1913; Zenderland 1998). American psychologists were infatuated with the test; they adopted with few reservations what their French counterparts found either uninteresting or problematic (Schneider 1992; Wolf 1973). In the process, however, they adopted wholesale the version of intelligence – singular, hierarchical, unidimensional – built into the Binet-Simon instrument, as well as a vision of how the scale might best be deployed (Kitson 1916).

When American psychologists confronted the new psychological instrument, it appeared to them to fit the prescriptions of psychological science for objectively produced quantitative data, and as well to have application to a number of areas of fundamental concern to American culture. At first, however, in the prewar years, the response of psychologists in the US to the potentials of the scale was not so entirely different from that of their French colleagues. Although far more enthusiastic about its possibilities, the first generation of mental testers often started out by assuming that the range of direct applications would be relatively limited, involving mostly the detection of pathological conditions associated with feeblemindedness or failure in school. Goddard, a psychologist at an institution for the feebleminded, was in this sense typical of the American psychologists who initially experimented with intelligence testing; almost all worked either with school children or those deemed of limited cognitive capacity.

Had intelligence and its tests remained tied to psychological research projects and a limited set of clinical applications, it is difficult to imagine that they would have had any broad impact on American culture, for all of their resonance with Progressive ideology and middle-class social anxieties (Lunbeck 1994). However, in the latter part of the 1910s two events propelled intelligence and its tests into national prominence. First, in 1916 Lewis Terman completed his Stanford-Binet Intelligence Scale, the articulation of intelligence that would quickly come to dominate the growing field of mental testing research and practice (Chapman 1988; Gould 1996; Minton 1998). Standardized on almost 1,000 California school children, the Stanford-Binet constituted the most complete revision of the Binet-Simon scale for an American population then undertaken, and was deemed technically superior in every sense, at least for white middle-class children. One of Terman's most important innovations was to introduce the concept of the "intelligence quotient" (IQ), a ratio of mental age to chronological age originally proposed by the German psychologist William Stern in 1912 and designed to produce a measure of intelligence that was independent of the examinee's age (Stern 1914: 80; Terman 1919: 8–9). This quantity, Terman
asserted, “has been found in the large majority of cases to remain fairly constant”, an opinion he used to buttress his conclusion that “[t]here is nothing in one’s equipment, with the exception of character, which rivals IQ in importance” (Terman et al. 1917: 10). With IQ, Terman had fully transformed intelligence into a standardized, quantifiable characteristic applicable to the entire range of human minds. Whatever variation individual intellects of the same IQ might manifest, and even whatever growth they might sustain, were rendered invisible in the process of producing a Stanford-Binet intelligence quotient. Its primary function was to create a linear index of relative brightness that could encompass the idiot, the genius and, most notably, everyone in between, whether child or adult, male or female, white or black. Especially with Terman’s packaging, which emphasized the innate hereditary nature of intelligence and its overwhelming significance in determining an individual’s life course, IQ became a characteristic of potentially immense significance, relevant to social and personal decisionmaking well beyond the confines of the psychological clinic.

Second, the advent of World War I afforded American psychologists the opportunity to demonstrate the relevance of intelligence measures to more mainstream arenas than the asylum or programs for the educationally lagging. Harvard psychologist Robert M. Yerkes assembled a group of mental testers—including Terman and Goddard—to aid the war effort by providing the Army with an efficient way of classifying the millions of new recruits it would need to mobilize for the war. These psychologists developed a new method of administering mental tests by groups, in the form of Army Alpha (for literates) and Army Beta (for English-language illiterates), and examined over 1.75 million soldiers. The results were then used as one means of sorting recruits into various categories of military usefulness—ranging from officer candidates to those deemed unfit for frontline duty—and of justifying those decisions (Carson 1993; Kevles 1968). The enormous legitimacy given to intelligence testing by this program, along with the publicity focused on the finding that the average American soldier had a mental age of thirteen and thus that a large percentage were “feebleminded” or worse, transformed an endeavor that had existed mainly on the margins of American culture to one that seemed right in the center. Confronted with scientific evidence that seemed to confirm their worst fears about the declining quality of the American population, leaders in many sectors of American society—education, industry, government—turned to tests of “intelligence” to aid their personnel processes.

By the end of the war, the version of intelligence and its instruments promulgated by the military testers had emerged victorious. New multiple-choice tests began to supplant the Stanford-Binet as the most common technology of intelligence assessment, and became commercialized commodities sold by a number of publishing houses or newly founded companies such as Psychological Corporation (Sokal 1981). At the same time, the greatest growth in the new intelligence industry occurred in applications to school and business, where leaders turned to intelligence testing as an objective, efficient and credible means of differentiating students, workers, and applicants for employment or admissions. The nature of intelligence reified in Army Alpha and these postwar measures perpetuated the Stanford-Binet model: a unitary, global entity, biological in origin and hereditary, that allowed all human beings to be ranked on a single scale, and presumed that intelligence was a prime factor in success in virtually all human endeavors. Intelligence in this form could explain, in almost Darwinian terms, why some individuals were at the top of the social/occupational hierarchy, and others were at the bottom.

This was not the only conception of intelligence jockeying for position in early twentieth-century American culture. But its embodiment in an easily commodified and disseminated technology of display, the mental test, that fit the needs of the efficiency-seeking bureaucracies of Progressive-era America, provided an enormous advantage to the unitary understanding of the human mind. By making visible fine grades of intellectual difference—whether revealed or created—tying those distinctions to particular social consequences—class and occupation—both of which were represented as linear gradations, and promising that profits could be made from psychological assessments, the intelligence instrument produced a reality that proved difficult to dispute, especially in a political culture preoccupied with balancing demands for merit and equality. In addition, because intelligence was translated into technologies—forms and mental tests—from which first army personnel and then large segments of the civilian population simply could not escape, both the examinees and those using the testing information were encouraged to think in concrete terms about intelligence, what it might mean, what its importance might be, and how it might be used. In reaction to this insertion of intelligence and its tests into the social landscape, and to the broader claims advanced by social scientists to reorganize the polity according to similar objective methods and rationalized procedures, questions began to emerge about the nature of democratic citizenship.7

Grand ambitions or monumental hubris?

Scarcely six months after the armistice ending World War I, Joseph Kimkont Hart of Reed College used the pages of the education journal School and Society to wonder about the fate of democracy (Hart 1919). It was a rather unlikely question to ponder at the end of America’s triumphal success in Europe, where the nation, with a minimum loss of American lives, had seemingly made good on its claim to make the world safe for democracy. Hart was writing during the first moments of the “Red Scare” of 1919–1920, in which a wave of strikes and a rash of bombings ignited a wholesale crusade against Bolsheviks, socialists, anarchists, labor organizers, foreigners, blacks, and anyone else who could be painted sufficiently “red” or “other” to be deemed a threat to the American way of life (Painter 1987). But this internal menace was not the immediate source of Hart’s anxieties. Rather, he worried about the lesson of the war itself, and specifically about how the spectacular successes of science in wartime would translate into an America at peace. Spurred on by the experience of the war, with its vast marshaling of material and
manpower and its heady application of expertise to the management of the economy, Hart imagined a social/political order not just open to the authority of science, but subservient to and transformed by it (Camfield 1969; Kevels 1993; Yerkes 1920). In Hart's new, postwar America, science was to be the final arbiter, dispassionately settling social questions and finally dispelling local prejudices by bringing objectivity and impartial reason to what had previously been hidden behind custom and corruption. Democracy itself required such a move in Hart's view, for only science, and particularly social science, could legitimately establish the boundaries within which a true democracy could operate:

Without science there can be no democracy, but only old prejudicial social forms, degenerating into autocracy, again....In the future, all crucial action of a social nature must be determined by scientific investigation, rather than by customs, and men must be brave enough to fight for these things, even to the losing of their – jobs!

(Hart 1919: 256–257)

Although Hart's rather hyperbolic plea that scientists risk even their careers to extend the authority of their findings to every aspect of social life may have fallen on deaf ears, his overall ambitions for the social sciences certainly did not. In the postwar period, the social sciences in America flourished as never before, and in no field was this more true than psychology, where many psychologists, having made the Army safe for intelligence, returned from their military duties at the end of World War I determined to carry the gospel of science and mental testing to the public at large. As Yerkes remarked in a letter to Abraham Flexner, president of the General Education Board (GEB) of the Rockefeller Foundation, in 1919, "[a]lready we are bombarded by requests from public school men for our army mental tests in order that they may be used in school systems" (Yerkes 1919b).

Psychologists' first serious foray into bringing science to the service of postwar democracy was the creation of the National Intelligence Tests (NIT) in 1919, a joint product of two one-time rivals but wartime colleagues, Terman and Yerkes (along with Melvin E. Haggerty, Edward L. Thorndike and Guy M. Whipple). With funding from the General Education Board of the Rockefeller Foundation in place, Yerkes et al. met over the course of 1919 to design a new test, the NIT, modeled on the military's Army Alpha examination, but modified to fit the capabilities of a school-age population (Yerkes and Terman 1919). Completed by the winter, the NIT was a collection of group-administered, multiple-choice instruments designed to rank the entire American school population on a single scale. Once adopted, its creators contended, a wholesale transformation could be wrought in the nature of American education. No longer need students of varying abilities be grouped in the same classroom; no longer need all students be subjected to the same curriculum; and no longer need every student be prepared for the same future. Rather, as B. R. Buckingham put it in 1921 in an editorial in the *Journal of Educational Research:

Our educational and intelligence tests permit us to ascertain the capacities of pupils far more accurately than ever before. Thus, the teacher becomes a guide and director. He is still a trainer of youth but he selects one to be trained in this way and another in that. ... Instead of prescribing the same treatment for all, he will become the expert diagnostian. On the basis of mental ability he will reclassify children, and because of their special abilities, he will further subdivide them.

(Buckingham 1921: 139)

Where once the single-room schoolhouse had stood as an icon of the commitment to primary education for all, now American psychologists proposed a new, more modern substitute: the multi-tracked high school, in which the results produced by impersonal mechanisms of assessment could be translated into objective systems of classification and separate educational destinies, all justified as a form of equal treatment by representing children as more different than the same, and less susceptible to molding than to sieving. In an article entitled "The mental rating of school children", in fact, Yerkes went so far as to suggest that children be seen as different and grouped separately according to their level of intelligence as early as kindergarten, and after fifth grade be sent off onto distinct educational tracks: professional, for the high intelligence group A children; industrial, for the medium intelligence group B children; and manual, for the low intelligence group C children (Yerkes 1919a).

The demands of educational efficiency and the ideals of democracy themselves required, these psychologists believed, that the schools be transformed according to the dictates of science. "I believe that the real meaning of democracy", University of Michigan psychologist Guy M. Whipple noted in 1922, "is properly safeguarded in the notion of 'equity of opportunity,' and if any nation is destined to perish it is that one which fails to provide the best possible educational training for those of its rising generation that show promise of educational leadership" (Whipple 1922: 602). Yerkes contended as well that equality of opportunity was the sole true form of democratic education, suggesting that only ability grouping would allow "the free intermingling of children of the various [class] strata in any given intelligence section" (Yerkes 1919a). Though few noticed the parallel, in many respects Whipple, Yerkes, et al. championed a version of education and democracy not far different from that institutionalized in France. Where the French used competitive examinations and exclusion of most from the upper echelons to create a system that was elite-dominated but technically open to all, these American psychologists proposed an approach that remained inclusive, but could still identify and sanction an elite. Both proclaimed the democratic bases of their educational visions on the grounds of equality of opportunity, and both justified the differential provisioning of educational resources on the belief that all people were decidedly not created equal, and that social progress demanded that these differences be acknowledged and acted upon.

Needless to say, this was not a vision shared by everyone in postwar America. Some more traditionally minded Americans simply rejected it out of hand,
skeptical about the pretensions of "experts" and more comfortable with traditional educational structures and well established pedagogical approaches. For them, the common school was a potent symbol of American democracy and its commitment to equality as uniform treatment of all, and any attempt to restructure it in the name of science and human difference posed a threat to basic values that had few offsetting compensations. Others, more self-consciously modern, however, could not slough off the claims of science nor the cult of opportunity so easily. Walter Lippmann, John Dewey, William C. Bagley, and other Progressive intellectuals and educators were as committed as the psychologists to objective methods and the reform of democratic institutions according to the dictates of empirical fact, and shared as well the belief that opportunity and access lay at the heart of the American concept of equality. While they were troubled by various features of the vision put forward by Terman, Yerkes, et al., of a new educational system, and by implication a re-fashioned democratic citizenry, they could not simply reject it out of hand. Rather, in the pages of magazines and journals ranging from the Saturday Evening Post and the Atlantic Monthly to School and Society and the Journal of Educational Research, they formulated challenges to specific factual claims and interpretations, so that the findings of psychological science might be domesticated within their own conceptions of democracy and scientific objectivity.

"The sky is falling": race, democracy, and the IQ

Although the United States had survived World War I largely unscathed, in the immediate postwar period many Americans remained concerned about the nation’s future. Troubled by the enormous influx of new immigrants and uneasy about the social/cultural transformations that had accelerated with the century’s end, members of the old elite, especially, feared that the war marked not the triumph of civilization, but another moment in its precipitous decline (Lears 1981). When the news broke about the average mental age of the American soldier, many were not so much shocked as confirmed in the worst of their suspicions, and seized on this “fact” as a golden opportunity to ring the alarm and publicly decry the state of the American republic. Echoing worries about national degeneration and decline that had been prevalent for the preceding three decades, for example, Cornelia James Cannon, wife of noted Harvard physiological Walter B. Cannon, opined in the pages of the Atlantic Monthly that “the lower grade man is material unusable in a democracy” (Cannon 1922: 154). George B. Cotten, president of Colgate University, followed suit, suggesting in School and Society that “we have never had a true democracy, and the low level of the intelligence of the people will not permit of our having one” (Cotten 1922: 479). Perhaps most inflammatory were the claims put forth by Harvard-educated Boston lawyer Lothrop Stoddard in The Revolt Against Civilization (1922), in which he marshaled the testing data as part of a eugenicist and frankly racist portrait of civilization under siege:

Against these assaults of inferiority; against the cleverly led legions of the degenerate and the backward; where can civilization look for its champions? Where but in the slender ranks of the racially superior — those “A” and “B” stocks which, in America for example, we know to-day [because of the World War I Army testing data] constitute barely 13% of the population? It is this “thin red line” of rich, untainted blood which stands between us and barbarism and chaos. There alone lies our hope. Let us not deceive ourselves by prating about “government”, “education”, “democracy”: our laws, our constitutions, our very sacred books, are in the last analysis mere paper barriers, which will hold only so long as there stands behind them men and women with the intelligence to understand and the character to maintain them.

(Stoddard 1925: 106)

Like his friend and fellow Ivy League graduate Madison Grant, whose own racial call to arms The Passing of the Great Race (1916) had been a best seller, Stoddard easily wove together fears of degeneracy, horrors of miscegenation, visions of corporeal and racial purity, nightmares of race war, images of primitive savagery, social Darwinist renderings of evolution, and skepticism about education and democracy, all to render vivid the image of a beleaguered aristocracy of red-blooded intellect — “A” and “B” men — upon whose powers to repress and procreate rested the future of civilization (Bederman 1993). Florid though the account surely was, its core drew on the popular pronouncements of much more distinguished and professional scholars. In fact, one of Stoddard’s chief sources was Scottish-born Harvard psychology professor William McDougall, whose Is America Safe for Democracy? (1921) also seized on the Army testing data to buttress dire conclusions about the biological warrant for democratic politics.

McDougall laced his tract with much of the new psychological knowledge — especially the latest findings from mental testing and eugenics research — in order to establish (along the lines of Herbert Spencer) that civilization in the form of modern industrial urban life was growing ever more complex. As a result, he argued, “the demand for A and B men steadily increases”, while the supply inexorably diminished (McDougall 1921: 168). Promoted as an investigation into “the influence of anthropological constitution on the destinies of nations”, McDougall’s account was concerned above all with race, understood in ethno-national terms as much as in broad color-based distinctions. McDougall’s goal was to preserve the presumed apex of humanity, Nordic stock, from the deleterious effects of degeneration from within, symbolized by the procreative menace of the feebleminded, and degradation from without, symbolized above all by the specter of black/white miscegenation. Indeed, throughout his text, data on white/“colored” differences in IQ and other racial characteristics — derived from at times tortured interpretations of the Army testing results — loomed large, serving as the touchstone for his arguments about the inferiority of non-Nordic Europeans, and anchoring his conclusion that only vigorously enforced policies
of positive and negative eugenics could halt America’s decline and make it again ready for robust, white-dominated democracy (McDougall 1921: 51–58).

McDougall's, and for that matter Stoddard's, intertwining of what they took to be the perils and potentials of modernity around the notions of the fragility of "civilization" and the preservation of the "race" (through careful tending of its most meritorious biological specimens), as historian Daniel J. Kevles has pointed out, articulated a set of hopes and worries common among segments of the middle-class white population in the US at the turn of the century (Kevles 1986: 72, 76; Bederman 1995; Newman 1999; Pernick 1996). Nonetheless, for all of the cultural resonance of these positions, more critical responses by prominent public intellectuals were not long in coming. In his address before the Society of College Teachers of Education on 27 February 1922, William Bagley of Teachers College Columbia led the charge, by providing an impassioned rebuttal of a number of these arguments, especially as they applied to education (Ravitch 2000). Rejecting what he termed belief in "educational determinism", or the primacy of innate mental ability, and the theory of aristocracy which it implied, Bagley instead celebrated the power of education to expand the intelligence of the common man and championed the provisioning of the same basic education for all. "If education is to save civilization", Bagley declared, "it must lift the common man to new levels, and not so much to new levels of industrial efficiency as to new levels of thinking and feeling" (Bagley 1922: 380). In Bagley’s view, education was at least as important a contributor to an individual’s overall intelligence as all other factors combined, and its relationship to the achievement of a just and peaceful society overwhelming (Bagley 1925). While he did not reject the concept of general intelligence, nor the value of intelligence testing for particular purposes, he did strongly denounce the vision of a society ruled by an intellectual elite chosen virtually from birth. Only mass education, American-style, Bagley contended, could truly enable the maintenance of a society open to the voices of all (Bagley 1925).

A few months later Bagley’s cause was joined by Walter Lippmann, who in a series of six articles in the New Republic took Stoddard and the mental testing community to task for what he argued were their shoddy procedures in generating and interpreting the Army test results (Lippmann 1922b). Lippmann argued that the figure of thirteen years for the average mental age was on the face of it absurd; IQ tests were mechanisms for classifying not instruments for measuring; predictions about school performance had little relevance to success in life; intelligence itself was an ill-defined concept within psychology; and there was little evidence that intelligence tests measured an innate heritable trait. He then concluded that however useful IQ examinations might be to accomplish specific classifications in specific settings, they failed to measure anything like pure intelligence, while according mental testers inordinate social power. “If the intelligence test”, he proposed, “really measured the unchangeable hereditary capacity of human beings, as so many assert, it would inevitably evolve from an administrative convenience into a basis for hereditary caste” (Lippmann 1922b). The following year, developing his arguments further in an article in the Century Magazine, Lippmann turned to the Army testing data relied on by McDougall to establish the intellectual inferiority of “coloreds”, and by pointing out the extreme regional variations in IQ regardless of race and their correspondence to the quality of the local school systems, made an impassioned defense of education and the possibilities of a democracy open to all. In one of his salvos at Terman, Lippmann revealed clearly the emotional basis of his reaction to the testers:

I hate the impudence of a claim that in fifty minutes you can judge and classify a human being's predestined fitness in life. I hate the pretentiousness of that claim. I hate the abuse of scientific method which it involves. I hate the sense of superiority which it imposes.

(Lippmann 1923b)

Observations such as these struck a nerve with a number of New Republic readers, among them John Dewey, who responded to Lippmann’s articles by questioning not so much the existence of individual differences but rather their limitation to any single construct such as intelligence (Dewey 1922a; 1922b). In his own reflections in the New Republic, Dewey contended that the essence of democracy was radical individuality, the belief that each person encompassed a unique set of attributes and that the duty of education was to allow those talents to flourish. “Democracy will not be democracy”, Dewey observed, “until education makes its chief concern to release distinctive aptitudes in art, thought and companionship. At present the intellectual obstacle in the way is the habit of classification and quantitative comparisons” (Dewey 1922b: 63). Mental testing was ill conceived, he continued along lines that French psychologists had argued, because it tried to hammer complicated human beings into simple administrative boxes, thus producing a society at odds with the goals of true “civilization”. The fetish for numbers, statistics and quantitative categories, Dewey argued, was an artifact of “our mechanical, industrialized civilization” and produced a “reverence for mediocrity, for submergence of individuality in mass ideals and creeds” that was inimical to both true education and true democracy (Dewey 1922b: 61). Dewey rejected notions of superiority and inferiority, whether applied to races or individuals, on the grounds that, while morally equal, human beings were otherwise incommensurable; each had to be appreciated in his or her own unique way.

These attacks by Bagley, Lippmann and Dewey on the testing community and its instruments did not go unchallenged; sarcastic responses from Terman and more considered replies from a number of other psychologists soon filled the pages of popular and professional journals. Terman turned first to Bagley. In the pages of the Journal of Educational Research he argued that Bagley’s refusal to concede the significance of differences in individual native mental endowment was in essence a denial of the truths of science and a return to superstition, and that Bagley was actually imperiling rather than protecting democracy (Terman 1922–3). Terman preached the need to adapt the curriculum to the individual
needs of the child, and not vice-versa, as most efficient for both child and society. He then went on to lambaste Bagley for failing to understand how intelligence testing could aid in producing a truly egalitarian democracy, one in which opportunity could flourish through identification of the most able, regardless of their class backgrounds, who would then receive the education most fitted to their abilities.

Terman was, if anything, even more dismissive of Lippmann, suggesting in his New Republic rejoinder that Lippmann lacked the expertise to judge the psychologists' work and that Lippmann's own understanding of intelligence was laughably naive (Terman 1922–3; Minton 1988). Characterizing (or was it caricaturing?) Lippmann as asserting that “the essential thing about a democracy is not equality of opportunity, as some foolish persons think, but equality of mental endowment”, Terman again celebrated the use of mental tests to “sift the schools for superior talent in order to give it a chance to make the most of itself, in whatever stratum of society it may be found” (Terman 1922–3: 117). Terman simply swept aside most of Lippmann's technical criticisms of the Army testing procedures and results, although he did allow himself to “explain” the controversial thirteen-year figure for the average mental age of the Army recruits by conceding that there was some disagreement within the professional community over the exact age where adult intelligence began. Most significantly, where Lippmann had argued passionately that the number of high-grade “A” and “B” men was a function of the time allotted to complete the test, and that more than the 13% per cent that Stoddard had made famous would have scored in those ranges if they had been given sufficient opportunity to complete each task, Terman countered that timing had little effect on a person's overall ranking, and that more time would simply have shifted the scale, without changing its meaning. In essence, Terman saw the proportion of most intelligent as fixed, and used relative test performance to identify them; Lippmann, on the other hand, considered the absolute level of performance itself as critical, and so contended that the most intelligent were all those who exhibited proficiency to a certain level, and not simply the top n per cent. Both conceded that mental tests could reveal a kind of merit, but understood that merit in decidedly different ways.

For the next two or three years, insults continued to fly and a variety of positions continued to be debated, with each commentator in one way or another wrestling with the implications of the possible existence of innate differences in mental ability for American society as a whole. In the end, the debate did not so much get resolved as drop from a boil to a simmer. While the frequency of articles in the popular and semi-professional press about intelligence and its tests remained high throughout the 1920s, after 1925 few authors accorded particular attention to the issue of intelligence and democracy. Rather, articles about the use, or misuse, of intelligence tests in particular (most typically educational) situations predominated, with the appropriateness of intelligence testing itself largely assumed. In this, the press reflected the American cultural landscape writ large. By the mid-1920s, a range of social decisionmaking systems routinely made recourse to assessments of intelligence, using them as aids in determining whether or not an individual should be hired by a firm, placed in an asylum for the feebleminded, declared by the courts to be of diminished capacity, or assigned to the academic track in the public high school.

A tempest in a teapot? Domesticating the natural order

The furor over intelligence and democracy in the early 1920s makes visible a critical moment at which the issues of merit, intelligence and its tests were transforming aspects of American democracy, and American democracy was defining and shaping features of intelligence and its implications. To turn first to one of the most obvious aspects of this transformation, whatever position the major contributors to the conversation about intelligence and democracy took, by its conclusion all admitted that intelligence mattered. They might not have agreed on what it was, whether IQ tests measured it, whether it could be acquired or improved, or even whether it was the most significant human attribute, but for the most part each conceded that it was something important and worth arguing about (Carson 1994; Danziger 1997).

For many Americans, the significance of intelligence lay first and foremost in its place in discussions of race, and specifically as a means of describing how various racial and ethnic groups differed from one another. Part and parcel of the development of scientific racism, intelligence in this usage became so well entrenched that even those most assiduously attempting to dismantle the prejudices surrounding their particular group — be it women or African Americans or Irish Americans — were much less likely to argue that differentiation according to intelligence was inherently wrong than to contend that, given adequate opportunities, their group would prove equal in intellectual capacity to that of white middle-class males.14 In addition, the host of transformations, from urbanization to the rise of industrial corporate capitalism, that produced elements of a mass society in the United States concomitantly opened space for new methods to regulate, administer and make sense of what was becoming, in the eyes of many, a nation of immigrants and strangers. Following World War I, for example, the new urban high schools and expanding corporate bureaucracies saw in ability grouping a means of sorting quickly the flood of new students and job applicants and of organizing large-scale education or industry according to one of the buzzwords of the era, efficiency.15 This move was facilitated by the creation of new positions for educational psychologists in the rapidly growing school systems and industrial psychologists in expanding corporate America. When combined with the commodification of intelligence itself as an object packaged into standardized tests sold by a variety of private companies, such changes allowed intelligence to permeate large sectors of the culture and to become part of the everyday experience of millions of Americans, from World War I veterans, to children given the Stanford-Binet by their school psychologists, to job applicants assessed as part of the hiring process. Under such conditions the importance of intelligence was proclaimed by the very pervasiveness of its measurement technology.
If all participants agreed that intelligence mattered, most were equally certain that science was significant as well. However, much controversy over new mental tests engendered, few critics chose to reject their use entirely. Neither Bagley nor Lippmann, for example, suggested that mental testing lacked value in its use in schools should simply be precluded. Indeed, Lippmann was careful to specify a number of ways in which he thought intelligence tests could contribute usefully to school and society, and even Dewey conceded that there were certain practical situations where classification was appropriate and the tests might prove helpful (Dewey 1922b). Whatever else was in dispute, the relevance of scientific findings and pronouncements to issues of public policy and even to the nature of American democracy was largely accepted. Bagley and Lippmann may have strongly disagreed with the conclusions that Terman and Whipple drew about how education should be organized and democracy should be understood, but they did not claim that drawing such conclusions and subjecting public institutions to the light of science were in themselves inappropriate. Whether the public embraced or despised their work, once psychologists made credible the claim that testing might strengthen American commitments to equality and efficiency, simply ignoring it proved difficult at best. Rather, the task of sorting out rival claims and agendas as to the implications of such a project pushed both public institutions and the practices of psychologists themselves to make responses and concessions, restructing important aspects of both.

Finally, it is important to underscore the limits to scientific authority that evolved in the course of the debate and thus to make clear how fully the resolution involved repositionings on all sides. Perhaps the best way to explore this issue is to ask, who won? Unlike the French situation, where the answer was clear cut — mental testing found little resonance among either psychologists or educators — the American situation was more ambiguous. Lacking France’s well institutionalized national educational system, relatively homogeneous population, and rather clearly demarcated and deeply entrenched class system, and embracing a more participatory model of democracy, American society responded to intelligence and its tests in a variety of at times contradictory ways. From the vantage point of the end of the 1920s, or even the end of the twentieth century, one would be hard pressed to say that either the testers or their critics had completely prevailed. Scientific authority did not simply triumph, or if it did, its triumph was not simple; rather, the two sides reached a fairly complicated set of accommodations. Psychologists gained extensive powers to categorize and manage those deemed marginal, especially the feebleminded, and institutions often carried out even terribly coercive practices such as sterilizations on the basis of test results. Administrators established the ability-grouped, multi-tracked high school throughout the United States, and the results of mental tests were often a key criterion determining who was placed where. At least during the 1920s a number of companies used intelligence tests as aids in deciding who to employ and where to place them, especially in the context of hiring entry-level white-collar workers. And many colleges and universities turned to intelligence testing as part of their process of admitting students and advising them on possible academic majors.

The critics of testing, however, also achieved some important results: no state or private agency ever put into place a system of testing, classifying, and then preparing children for particular career trajectories based solely on the results of intelligence measures. No university ever used mental tests alone to decide admissions. Public officials never turned to intelligence tests as important gate keepers even for immigration or access to voting, where some were actively constructing systems of restriction. And many Americans — from recruits making snide remarks about the Army testing, to fundamentalists celebrating Christian over secular values, to Deweyites committed to radical individuality — continued to embrace more complicated understandings of merit than those put forward by Terman, Yerkes and Whipple, including ones that celebrated characteristics other than brains (Ryan 1997). In addition, Lippmann, Dewey, Bagley, et al. created a rhetoric of doubt about psychological instruments that facilitated the raising of questions in any particular instance of their use, especially around the issue of the meaning of statistical findings for individual cases. Moreover, numerous Americans — including such intellectuals as Randolph Bourne or Franz Boas and most leaders of America’s racial and ethnic minority communities — remained ambivalent or skeptical about claims for the innate and heritable nature of intelligence and its implications for racial or ethnic groups, with many more abandoning such beliefs by the end of the decade. More generally, while intelligence testing did have important effects in a number of areas at the level of administrative practice, the broader social vision of the “determinists” was largely rejected. Their highly rationalized and hyper-efficient “brave new world” — in which each citizen would be slotted into his or her occupation through the objective determinations of psychological experts — found few takers (Ryan 1997).

The partiality of this outcome must be emphasized. Viewed from a distance, what is most striking about the controversy over intelligence and democracy in the United States is that the testers and their opponents produced a space for debate, one in which notions of democracy, equality and merit were contested and re-formed. While scientific expertise was an important constituent of the conversations, its claims were as much open to dispute and revision as those from any other source. As such, this settlement thwarted the grand ambitions, if not the monumental hubris, of the American intelligence testers, who imagined a social world shorn of debate and strife through the certainties and efficiencies of their science.

**Conclusion: intelligence, democracy, and co-productions of merit**

The dispute over the place of intelligence and its tests in American culture was closely linked with analyses of the nature of American democracy and how notions such as equality, citizenship and merit should be understood in light of the findings of the tests. Why? Why should a new scientific procedure, intelligence
testing, and the data that it generated have initiated such extensive soul-searching about the nature and future of democracy among members of the American intellectual elite? How did the scientific and political come to be so intimately linked? And, to return to the comparative aspects of the story, why did a similar conversation over the proper representation of democracy fail to erupt in France? What was it about these cultures that made their settlements of the tensions between democracy and merit so different, both in terms of the realities they created and the relative robustness of their particular solutions?

To its promoters in the United States, measurements of intelligence promised simultaneously to reveal one of the fundamental characteristics of an individual’s nature and to allow social decisions about that person to be made according to seemingly objective and neutral criteria. It was a social technology that, in their rendering, could provide both equal treatment and accountability, and one perfectly suited to the new Progressive demands for a coordinated response to America’s rapidly changing social landscape (Campbell 1995; Keller 1994; Skocpol 1992). The supporters touted a democracy founded on belief in human differences as fundamental, in equality of opportunity and in the primacy of demands for social efficiency. To its critics, however, the vogue of intelligence threatened to undercut the very premise of American democracy by naturalizing a social hierarchy and substituting the norms of a particular group, the mental testers, for those of the nation as a whole. They too celebrated democracy and the importance of equality and accountability, but their democracy was one emphasizing the malleability of human nature, a common cultural heritage and a thoroughgoing commitment to social mobility. Their critiques of the tests were in part technical, that the instruments failed to provide the degree of legitimation that they proclaimed, and in part more fundamental, that any technology that threatened the American ideal of the liberal, self-directing citizen capable of personal growth and transformation was intrinsically problematic.

The clash over the claims of the mental testers, examined from this perspective, was largely a struggle over who should have the power to define what was and was not equal, democratic and fair. It was a political argument, and an argument as well over exactly how politics should be done in an age of human science. What gave this argument such purchase was that, by promising to provide a technology able to determine merit objectively and to organize the polity according to “natural” criteria, psychologists readily connected their new instruments with long-standing concerns about how to structure a republic along lines that maintained both democracy and efficient allocation of social resources. If successfully established, mental testing would almost invisibly naturalize particular definitions of merit and particular determinations of who should have access to what social goods, producing results little more controversial than those provided by the French system of competitive examinations, the concours.

These “obvious” connections between intelligence and merit that the American mental testers both fashioned and traded on and that even most critics broadly accepted, appear in a much different guise when developments in the United States are contrasted with those in other cultures. French psychologists and public officials, as we have seen, did not conceive of intelligence and its tests in the same way as their American counterparts, and the relevance of these objects to questions of merit was simply not admitted within French intellectual and administrative culture. Even to French psychologists, measurements of intelligence were at most one part of assessing and understanding the capabilities and deficits of particular individuals. No less obsessed with merit than their American counterparts, and much more entranced by the possibility of technocratic solutions to social problems, republicans in France nonetheless felt little need to stabilize their ways of establishing merit by recourse to natural objects such as intelligence. Rather, given a republican vision in which the active intervention of the state was deemed critical to the maintenance of the nation and its citizens, French republicans by and large looked to the government to actively select and mold the next generation of elite leaders. The free play of “natural” talents, for them, was much less significant an ideal than meeting the needs of the nation and its citizens through a system of training open to all.

The differences in how intelligence and merit were constituted within these two political cultures clearly reveals the complex and local nature of the interrelations of science, politics and society. Although the intelligence test itself moved relatively frictionlessly from one culture to another, the meanings of intelligence and the worries it engendered certainly did not. Rather, they were intertwined with particular cultural needs and made possible particular ways of shaping aspects of both the natural and social orders. France, with its heritage of absolutism and centralization, created a national system of education, pyramidal in structure, to ensure that the “best” reached the top of the administrative and educational hierarchies. A national curriculum, competitive examinations, and the eventual commitment to universal primary education, all insured that equality and merit would emerge as the seemingly natural results of a properly operating system. The United States, on the other hand, while also committed to equality and merit, has to this day eschewed such a centrally driven, systematic approach to the making of an elite. Deep skepticism about an interventionist state, especially in the realm of directly molding the citizenry, coupled with a pervasive ideology of personal liberty and radical individualism and a decentralized political structure, have all combined to make the French approach to sorting and grading people untenable at best. Preferring to provide equal opportunity and to let “natural” endowments flower as they will, Americans have turned instead to other methods, including recourse to naturalized objects such as intelligence and its tests, in order to understand and assess human beings. Quantified, unidimensional and hierarchical — intelligence as developed by American mental testers was constructed so as to fulfill demands for accountability, equality and merit in an American context in which plurality could flourish, state intervention remain minimal, and coordination of diversity without centralized control be privileged. The merit of science, in the American case, did not fit the same needs in France; correspondingly, the science of merit flourished in the one context, and withered in the other.
Notes

1 I would like to thank Tom Broman, Ellen Herman, members of the University of Michigan History Department non-tenured faculty colloquium, and especially Sheila Jasanoff for their very helpful comments on this article.

2 Michel Callon, among others, has been examining the role of lay people in scientific debate and the constitution of scientific fact. He has seen this phenomenon as relatively recent in origin. However, in this article it will be clear that contestations by lay people of the "truths" of the human sciences are long standing, reflective, I believe, of their sense of having their own form of expertise in such matters.

3 Stephen Jay Gould points out that Binet, at numerous times, insisted that intelligence was "not a single, scalable thing like height", and that the scale was only intended to be used with possibly backward children. Theda H. Wolf and Read D. Tuddenham also emphasize that Binet never committed to viewing intelligence as a single mental faculty, preferring to see what he was measuring as a complex of mental functions expressed in a set of externalized behaviors. Nevertheless, as constructed, the Binet-Simon scale did produce a singular measurement and was designed to be broadly administered.

4 Indeed, even the convulsive Dreyfus Affair (1898–1899) can be seen as, in part, a struggle between the claims of talent and prerogatives of tradition (Johnson 1966; Mayeur and Rebérioux 1987; Nord 1995).

5 Schneider makes much this same point (1992: 128). A recent example of the French connection between merit and the pyramidal educational system is the current controversy over introducing a form of affirmative action into the selection process for one of the grandes écoles, the Institut d'Etudes Politiques (Sciences Po) (Daley 2001).


8 For a related approach from the vantage point of biology, see Wiggam 1922.

9 For a strikingly similar post-World War II argument, see Popper 1950.

10 It was also in the immediate postwar period that another approach to the psychological, Freudianism, spread widely, especially among the intellectual and upper middle classes in the United States. On this phenomenon, see Buhle 1998; Burnham 1967; 1968; Caplan 1998; Hale 1995.

11 On the modern American high school, see Angus and Mirel 1999; Cremin 1988; Nasaw 1979; Tyack 1974. On mental testing in the school system, see Chapman 1988; Fass 1980; Resnick 1982; Williams 1986.

12 However, for a more jaundiced appreciation of mass democracy produced at almost the same time, see Lippmann 1922a.

13 On publication rates in the popular press about intelligence and its tests, see Hart's analysis of articles indexed in the Reader's Guide to Periodical Literature for the period 1905–1930, Table 3 in Hart 1933.

14 From this perspective, W. E. B. Du Bois' famous 1903 essay extolling the importance of the "talented tenth" was only one instance among many in which an author more or less took for granted distinctions in individual intelligence, while resoundingly rejecting claims about group inferiority (Du Bois 1903).

15 As Elizabeth Frasier succinctly put it in an article in the Saturday Evening Post about the new kinds of jobs available for sectors of the working and lower middle classes, "Sheer brawn, youth, quickness no longer count all. It needs something else to get by. And that something is gray matter. Brains" (Frazier 1923: 133). On the rise of white-collar work in America, see Chandler 1977; Trachtenberg 1982; Zunz 1990.

16 For an excellent examination of politics and the human sciences in the post-World War II era, see Herman 1995. And for an interesting case study of the tangle of science and politics, see Jasanoff 1992.

17 This somewhat challenges Bruno Latour's emphasis on the immutability of his immutable mobiles (Latour 1987).
States of Knowledge
The co-production of science and social order

Edited by Sheila Jasanoff