NOTES TO THE TEXT OF

Facing Reality

INTRODUCTION

Page ix: *By* facts, *I mean what Senator Daniel Patrick Moynihan meant.*

The late senator and public intellectual Pat Moynihan usually gets credit for "Everyone is entitled to his own opinion but not to his own facts," which comes from an op-ed ("More than Social Security Was at Stake," *Washington Post*, January 18, 1983). Earlier versions are attributed to Bernard Baruch as far back as 1918 and in the 1940s, while James Schlesinger contributed variations in the 1970s. Moynihan himself once attributed the saying to Alan Greenspan. The website quoteinvestigator.com has the whole story.

Page ix: *By* reality*, I mean what the science fiction novelist Philip Dick meant.*

The Philip Dick quote is definitely in *I Hope I Shall Arrive Soon* (1987), but he may have used it in more than one book.

CHAPTER ONE
The American Creed Imperiled

Page 1: *It has been our fate as a nation not to have ideologies, but to be one.*

This quote is attributed to Richard Hofstadter in many books and articles, but none of them cite a source. Perhaps it was a spoken remark, not words he published. The earliest use of the quote I have been able to find is Hans Kohn, *American Nationalism: An Interpretive Essay* (1957), p. 13.

Page 2: *Gunnar Myrdal capitalized the term and marveled at the creed's continuing universality.*

The quote from Gunnar Myrdal is in the introduction to *An American Dilemma: The Negro Problem and Modern Democracy* (1944), p. xlviii.

Page 2: *The most dramatic single moment of that crusade, Martin Luther King's "I have a dream" speech on the Washington Mall on August 28, 1963, evoked the American creed from start to finish.*

The text of the speech is readily available online. I used the text at americanrhetoric.com, which also has a video of the event.

Page 3: *Some who voted for the bill had misgivings about a few provisions. Titles II and III, banning race discrimination in public accommodations and public facilities, entailed obvious restrictions on freedom of association. Title VII, on equal employment opportunity, made employers vulnerable to legal scrutiny if they didn't think in terms of groups. But in the floor debates and in the press, these provisions were described as one-time exceptions justified by the unique injustice done to African Americans.*

The indispensable book about the long-term consequences of the Civil Rights Act of 1964 is Christopher Caldwell, *The Age of Entitlement: America Since the 1960s* (2020). Caldwell argues convincingly that the act generated unintended consequences that deformed the American project beyond recognition, displacing American individualism with a sprawling legal regime to punish transgressions against groups.

Page 3: *As Hubert Humphrey, the Senate's leading liberal, put it when discussing the section on employment discrimination, the wording of the bill "does not limit the employer's freedom to hire, fire, promote, or demote for any reason – or no reason – as long as his action is not based on race, color, religion, national origin, or sex."*

The quotation comes from the *Congressional Record*: 100 Cong. Rec. 6549. It is quoted in Richard A. Epstein, *Forbidden Grounds: The Case against Employment Discrimination Laws* (1992), p. 161. To see how far we have traveled, try to imagine a leading Democratic politician in 2021 saying that employers have an underlying presumption of freedom to hire or fire "for any reason – or no reason."

The inclusion of a prohibition of job discrimination by sex in the 1964 Civil Rights Act got almost no attention at the time. It was introduced as an amendment late in the floor debate by Howard W. Smith, chairman of the powerful House Rules Committee – whether out of genuine conviction, as a cynical attempt to create problems for the bill's passage, or some of both, is unclear. It passed with virtually no floor debate.

Page 4: *The twenty-first century saw the growth of a new ideology that repudiated the American creed altogether.*

For a critical perspective on the evolution of the new ideology and its effects, see Helen Pluckrose and James A. Lindsay, *Cynical Theories: How Activist Scholarship Made Everything about Race, Gender, and Identity – and Why This Harms Everybody* (2020).

Page 6: *Within a week of his inauguration, President Biden signed four executive orders intended to promote "racial equity," promising that "we're going to make strides to end systemic racism, and every branch of the White House and the federal government will be part of that."*

The quotation comes from "Biden Signs Orders on Racial Equity, and Civil Rights Groups Press for More," *Washington Post*, January 26, 2021.

CHAPTER TWO
Multiracial America

Page 9: *Table 1 below shows the racial and ethnic breakdown of the American population as reported in the American Community Survey (ACS) for 2019.*

The ACS is a survey begun in 2006 that the Census Bureau mails to about 3.5 million people annually. It collects the supplementary demographic and economic information that used to be part of the decennial census. In addition to its annual reports, the Census Bureau publishes five-year aggregations of the ACS that provide data down to the zip code and census tract levels. ACS raw data (and decennial census data from earlier years) may be downloaded without cost from usa.ipums.org. For data aggregated by geography (towns, zip codes, census tracts, etc.), go to socialexplorer.com, which charges a subscription fee.

The East Asian category in Table 1 includes Chinese, Japanese, Okinawans, Koreans, and Taiwanese. The South

Asian category includes Asian Indians, Pakistanis, Bangladeshis, and Sri Lankans. The Southeast Asian category consists of Vietnamese, Cambodians, Thais, Laotians, Hmongs, Malaysians, Indonesians, and Burmese. Note that the total of 18.4 percent for Hispanics includes those who self-identify racially as Black, Asian, or Filipino/Pacific. The definition of Latin that I use (see below) excludes those groups, leaving the 17.9 percent Latin population that I give in Table 2 for the 2019 racial profile.

Page 10: *They have found that they can accurately calibrate people's mix of ancestral heritages, whether they are popularly understood as races or ethnicities, by examining patterns of genetic variants.*

I devote a chapter in my *Human Diversity: The Biology of Gender, Race, and Class* (2020) to this part of the story, which began in the early 1990s and was largely concluded within a decade after the sequencing of the genome in the early 2000s. The technique that geneticists used was statistical cluster analysis of large numbers of single nucleotide polymorphisms, or SNPs, the base pairs in the DNA sequence that can take on more than one form and thereby create human variation. The analysts don't pre-identify the races and then see whether they can reliably match them with SNPs. Rather, they instruct their software to cluster hundreds of thousands of SNPs statistically. The results get progressively more precise. When the software is told to create two clusters, it separates Africans from everyone else. At three clusters, peoples from Asia and the Americas split out. At four clusters, Amerindians break away from Asians. At five clusters, the peoples of Oceania split off, resulting in the five continental groupings. At six clusters, Central and South Asians split away from other Asians. At seven clusters, peoples of the Middle East split off from Europeans. New techniques developed more recently permit calculation of the proportions of various ancestral heritages that a person carries.

Page 11: *In a large study based on 23andMe data, [Whites] had a mean of 98.6 percent European ancestry, 0.2 percent Native American ancestry, and 0.2 percent African ancestry, with the rest being "Other."*

For the White admixture, see Katarzyna Bryc, Eric Y. Durand, et al., "The Genetic Ancestry of African Americans, Latinos, and European Americans Across the United States," *American Journal of Human Genetics* (2015).

For the Black admixture, see Soheil Baharian, Maxime Barakatt, Christopher R. Gignoux, et al., "The Great Migration and African-American Genomic Diversity," *PLoS Genetics* (May 27, 2016). The Health and Retirement Study used in the Baharian article permits nationally representative estimates.

The Bryc article also reports statistics on the genetic profile on Blacks, with an estimate of 72.3 percent African ancestry, notably lower than the 82.1 percent found in the Baharian study. However, the Bryc article cautions that the 23andMe sample has disproportionate numbers of people living in California and New York, which are regions where Blacks have customarily shown lower mean African ancestry than in other parts of the country. In addition, "participation in 23andMe is not free and requires online access, so therefore it is important to note that other social, cultural, or economic factors might interact to affect ancestry proportions of those individuals who choose to participate in 23andMe." Bryc, "Genetic Ancestry," pp. 48–49.

Page 11: *Self-identified Latinos can be of any race if their families came to the United States from Latin America.*

The definition of *Latins* in Chapters 3–6 departs slightly from the Census Bureau's definition. I define *Latins* as those who ethnically self-define in the ACS as Latino and racially self-identify as White, Native American, a combination of two or more races, or "Some Other Race Alone." In effect, this means that Latino Africans are classified as African, Latino Pacific Islanders as Pacific Islanders, and Latino Asians as Asians. I do this in the belief that those racial heritages usually trump a Latin cultural heritage, but whether that's correct is an argument we don't need to have because it wouldn't make any material difference to the analyses in *Facing Reality*. The numbers involved are too small. People who self-identify as Latino and as a member of a single race besides White, Amerindian, or "Other" amount to just 2.8 percent of the Latino population as defined by the Census Bureau and 0.5 percent of the total population.

Page 13: *All this means that it is problematic to lump Latinos into a single group when analyzing either cognitive ability or crime.*

The same issue is not nearly as relevant to White or Black ethnic subgroups (e.g., English Whites versus Italian Whites; Yoruba Blacks versus Mandinka Blacks) as it is to Latino ethnic subgroups. This is partly because a large majority of self-identified Whites and self-identified Blacks have lived in the United States for at least three generations, usually more, and are fully assimilated into American culture. Recent White and Black immigrants are a small percentage of the total.

In contrast, a large majority of all Latinos have immigrated to the United States in the last half century and especially in the last few decades. Many Latino neighborhoods are often like culturally distinct Irish, Italian, and Jewish ethnic neighborhoods in the nineteenth and early twentieth century. Today's Latino counterparts include culturally distinct Central American neighborhoods, Mexican neighborhoods, Puerto Rican neighborhoods, and Cuban neighborhoods.

Aggregating Latino subgroups is also problematic because they are genetically more distant from each other than Mandinka are from Yoruba or English are from Italians. Genetics does not enter the discussion in this book. Those who are curious are encouraged to look into David Reich's *Who We Are and How We Got Here: Ancient DNA and the New Science of the Past* (2019) or Chapters 7–9 of my *Human Diversity*.

Page 15: *As of the 1960 census, America was about 87 percent European, 11 percent African, something more than 1 percent Latin, and something less than 1 percent Asian.*

Since the book went to press, I have obtained better information for the 1960 census than I used to calculate the percentages for the text. The correct figures for the 1960 census are 85 percent European, 11 percent African, 3 percent Latin, 0.5 percent Asian/Pacific Islander, 0.3 percent Amerindian, and 0.1 percent "Other."

Page 17: *I am defining* big-city America *as urban areas with populations of 500,000 or more in a contiguous urban environment (which often does not correspond to the legal boundaries of the city).*

The 52 areas are Albuquerque, Atlanta, Austin, Baltimore, Boston, Buffalo, Charlotte, Chicago, Cincinnati, Cleveland, Colorado Springs, Columbus, Dallas–Fort Worth, Denver, Detroit, El Paso, Fresno, Houston, Indianapolis, Jacksonville, Kansas City, Las Vegas, Greater Los Angeles, Louisville, Memphis, Miami, Milwaukee, Minneapolis–Saint Paul, Nashville, New Orleans, New York, Norfolk–Virginia Beach, Oklahoma City, Omaha, Orlando, Philadelphia, Phoenix, Pittsburgh, Portland, Providence, Raleigh–Durham, Richmond, Sacramento, Saint Louis, Salt Lake City, San Antonio, San Diego, San Francisco–San Jose, Seattle–Tacoma, Tampa–St. Petersburg, Tucson, and Washington. These urban areas enfold many other cities with large populations. In all, the 52 urban areas include 84 of the 100 cities with the largest official populations in the Lower Forty-Eight.

The definition of "urban area" is not based on legal boundaries. Such boundaries seldom coincide with the contiguous urban space. In some cases (Boston, for example), the urban area sprawls across several other legally defined cities. In other cases (Tucson, for example), large areas that are legally part of the city are actually rural.

The big-city urban environment is characterized by a cluster of high-density zip codes, defined here as at least 10,000 persons per square mile, surrounded by urban residential areas (apartment buildings and rowhouses) with densities around 3,000 or more, which in turn are bordered by suburbs with residential densities that are typically in the 1,500–3,000 range.

If the population is large enough, even cities without an inner core of high-density zip codes clearly count as "urban." Almost half of the 52 areas I designated as urban have no high-density zip codes at all, but they include cities such as Kansas City, Las Vegas, and Cleveland, all of which should be classified as urban. The smaller the total population, the less urban the environment. Places like Little Rock, Des Moines, Dayton, or Wichita (none of which is among the 52 urban areas of big-city America) have urban downtowns, but the tall buildings occupy just a few blocks. As you move away from those blocks, the streetscape quickly turns into that of a small city, with low-rise businesses. Within a few more blocks, stand-alone single-family homes are already interspersed among the businesses.

No decision rule can unambiguously distinguish cities with a classic urban feel and lifestyle from those with the feel and lifestyle of a small city. My rule was to examine the densities of zip codes in all American cities with a listed population of 100,000 or more, using data on zip codes from the 2014–2018 ACS and plotted on maps using Tableau software. For each city, I identified all the contiguous zip codes with a density of 2,000 per square mile or higher, plus zip codes of at least 1,500 people per square mile that are bordered by a zip code of 2,000 or higher. A zip code with a density under 1,500 that was surrounded by zip codes with densities of 1,500 or more was also classified as being part of the urban area in question. The population of a designated urban area consists of the sum of the populations of all the zip codes that meet these requirements.

Page 18: *The zip codes associated with Amerindian reservations are geographically large but sparsely populated, containing just 530,046 self-identified Amerindians.*

The number 530,046 is probably an exaggeration, though it is hard to be more precise than that. The self-identified Amerindian population doubled from 1990 to 2000 when the Census Bureau first allowed people to select multiple races. See Carolyn A. Liebler, Renuka Bhaskar, and Sonya R. Porter, "Joining, leaving, and staying in the American Indian/Alaska Native race category between 2000 and 2010," *Demography* (2016).

It should also be noted that Amerindian fertility rates have fallen substantially and have been lower than those for Europeans since 1980. See Sarah Cannon and Christine Percheski, "Fertility Change in the American Indian and Alaska Native Population, 1980–2010," *Demographic Research* (2017).

After Chapter 2, Amerindians do not enter the discussion in *Facing Reality*'s text because their numbers are too small to bear on the issues that the book is about. But of all the stories of America's races and ethnicities, that of contemporary Amerindians is in many ways the saddest.

Page 19: *Big-city America is authentically multiracial, far more so than the major cities of Europe or Asia.*

The United Kingdom comes closest, but it's not that close. As of the 2011 census, London was 60 percent White, 20 percent Asian (mostly South Asian), and 16 percent Black – quite diverse, but nothing approaching New York or Los Angeles. Birmingham (58 percent White) and Bradford (67 percent White) also have substantial minority populations, but again nothing close to the typical American big city. The rest of the UK's largest cities remain more than 80 percent White.

The only other major city in Europe that has a large non-European population is Paris. France doesn't maintain statistics by race, but as of 2013, 81.5 percent of the population of the Paris Region had been born in metropolitan France. Some portion of those people were not ethnically French; on the other hand, many of those born in former French colonies were ethnically French. The other major European cities are around 90 percent European or more. Asian cities have significant mixtures of different Asian ethnicities, but small proportions of Europeans, Latins, and Africans. The examples of genuinely multiracial cities outside the United States that I have been able to find are in Latin America, especially Brazil, and in Africa, especially South Africa.

CHAPTER THREE
Race Differences in Cognitive Ability

Page 20: *The charges of pseudoscience have many sources.*

In addition to the charges raised by Gould and Taleb, two others have been raised in many sources. The first is that IQ in general and race differences specifically are undermined by the Flynn effect, referring to the rise of IQ scores over time. The second is that IQ is really a measure of socioeconomic status – the children of the affluent get high scores because of the privileged environment in which they are raised, not because they are innately more intelligent than the children of poor families. Each of these arguments has been tested against the empirical record in detail.

The Flynn effect. The late James Flynn, a political philosopher who became an important scholar of cognitive ability, did not discover the phenomenon of secularly rising IQ – it was first observed in the 1930s – but he was instrumental in assembling evidence for its generalizability and bringing the phenomenon to public attention, so Richard Herrnstein and I decided to name it after him in *The Bell Curve*. The label stuck.

Here's how the Flynn effect works: In many countries and for tests administered over much of the twentieth century, it has been found that when a test that was standardized to a mean of 100 in year X is administered to a comparable population in year $X+10$, the mean IQ score for the new sample is somewhere near 103. Same test, but mean IQ has gone up. Are people really getting smarter? Is it some sort of psychometric artifact? Is it a cultural artifact?

In thinking about the meaning of the Flynn effect, we can start from a secure assumption: To the extent that the Flynn effect reflects genuine increases in cognitive ability, it cannot have been going on at the rate of 3 IQ points per decade for very long. Otherwise, the average American at the time of the Revolution would have been a moron. But I needn't go that far to make the point. If Americans had been gaining 3 IQ points per decade since 1950, the average American today would be in the 90th percentile of the distribution as of 1950. Nothing in everyday experience suggests that this is remotely the case.

The Flynn effect is apparently a little of everything – partly an effect of culture, partly one of age (particularly for adolescents), and partly an increase in cognitive ability. For an accessible description of the state of knowledge that also has citations of the technical literature, I recommend Russell Warne, *In the Know: Debunking 35 Myths about Human Intelligence* (2020), pp. 126–29. The balance of the evidence indicates that there has been some increment in cognitive ability, produced by a variety of causes associated with modernity. Among them are increased education, improved physical health, lower blood lead levels, better nutrition prenatally and in childhood, and lower smoking and drinking rates among pregnant women. Flynn himself saw a plausible link between the Flynn effect and the increasing cognitive demands of navigating daily life in the twentieth and twenty-first centuries. As Warne puts it, "As people went to school for longer periods of time and learned how to reason and think better, they were better able to think abstractly. The more complex environment ensured that they would have to use these skills in daily life."

The Flynn effect is a longitudinal phenomenon. It does not appear to have implications for the use of IQ scores at any given point in time. The predictive validity of IQ scores for classroom and workplace performance is comparable in studies conducted in the 1970s and in the 2010s, despite the intervening 40+ years of the Flynn effect. Nor does the Flynn effect appear to be directly relevant to race differences in IQ. A 2004 article by a team of Dutch methodologists found that "the nature of the Flynn effect is qualitatively different from the nature of B-W [Black-White] differences in the United States," for reasons that are highly technical but may be roughly summarized like this: With the B-W difference, analysis of the factor structure for a sample of Black test takers and a sample of White test takers indicates that the tests are measuring the same construct for both groups. The same is not true of the differences in IQ scores across two cohorts over time – the construct measured by the IQ test in 1970 is somewhat different from the construct measured in 1960, for example. More formally, "[t]he overall conclusion of the present paper is that factorial invariance with respect to cohorts is not tenable." Jelte M. Wicherts, C. V. Dolan, David J. Hessen, et al., "Are Intelligence Tests Measurement Invariant over Time? Investigating the Nature of the Flynn Effect," *Intelligence* (2004), p. 531.

Finally, I should mention that the Flynn effect has recently gone into reverse in some European countries. See Bernt Bratsburg and Ole Rogeberg, "Flynn Effect and Its Reversal Are Both Environmentally Caused," *Proceedings of the National Academy of Sciences* (2018). I have seen no evidence one way or the other for the United States.

Cognitive test scores reflect socioeconomic status (SES), not intelligence. "The SAT is a wealth test. You can tell how high a student scores by knowing how much money the parents have." That allegation, with numerous variants, is so common that it has become conventional wisdom.

It's not true in the sense that people think it's true – income per se doesn't buy high scores. The correlation between parental income and the SAT is modest, just .10 and .23 for the SAT in two large-sample studies. Counterintuitively, the correlation is higher for nationally representative studies in which the students have no incentive to prepare, usually in the .3 to .4 range, than for the SAT. See Russell Warne, *In the Know*, pp. 107–13, for citations and additional evidence. But even a correlation of .4 explains only 16 percent of the variance, so obviously a lot more goes into test scores than parental income.

The reality is that all *g*-loaded academic tests look as if they're wealth tests. It's inevitable. Parental IQ is correlated with children's IQ everywhere. In all advanced societies, income is correlated with IQ. Scores on academic achievement tests are always correlated with the test takers' IQ. Those three correlations guarantee that every standardized academic test shows higher average test scores as parental income increases. And it's not just tests. As educational scholar Rebecca Zwick put it, "[T]he studies reviewed here suggest that if we were to 'disqualify' any admissions criterion that reflected parents' income and education, we would have to eliminate high school grades, courses taken, teacher ratings, and participation in extracurricular activities along with admissions test scores." Rebecca Zwick, "Is the SAT a Wealth Test?" *Phi Delta Kappan* (Dec. 2002), p. 310.

Consider data from the National Longitudinal Survey of 1979, which contains thousands of cases with data on family income, the mother's IQ, and her children's performance on the math and reading tests of the Peabody Individual Achievement Test (PIAT) battery. In a multivariate analysis, a child from a family with an income of $400,000 – in the fabled 1 percent – with a mother who has a college degree but an IQ of 100 is predicted to be at the 68th percentile on the PIAT, equivalent to 107 in the IQ metric. A child in a family with an income of $40,000 – close to poverty – and a mother with only a high school diploma but an IQ of 135 is predicted to be in the 78th percentile on the PIAT, equivalent to 112 in the IQ metric. Put roughly, if you want high test scores and have a chance to choose your mother, take a poor mother with a high IQ over a rich one with an average IQ. (See Charles Murray, "Why the SAT Isn't a 'Student Affluence Test,'" *Wall Street Journal*, March 24, 2015.)

The larger point, which few people want to acknowledge, is that high-IQ children tend to come disproportionately from high-IQ parents, who also tend to be above average in income. This was not the case (or only very modestly so) in 1900, when the great majority of high-IQ children never went beyond high school. But it became increasingly true during the twentieth century, and especially from 1960 onward. Richard Herrnstein and I discussed the dynamics of this process in Chapters 1–4 of *The Bell Curve* and I updated it in Chapter 2 of *Coming Apart: The State of White America, 1960–2010* (2012). If elite colleges admitted students purely based on IQ, their student bodies would be populated even more densely with the offspring of the upper-middle class than they already are – not because their parents are rich, but because they are smart. No improvement in the SAT or IQ tests can do away with this underlying reality.

Page 23: *All this indicated a large racial difference [in the Project Talent results]. Exactly how large is uncertain, but it was around the equivalent of 19 to 23 IQ points.*

In 1977, a reanalysis of the Project Talent data included a table that enables an estimate of a European–African difference at 1.28 SDs, but this calculation assumes that the standard deviations for Europeans and Africans were the same as the total sample standard deviation, which is unlikely (standard deviations within subpopulations are usually smaller than the standard deviation for the total sample). See Lauress L. Wise, Donald H. McLaughlin, and Kevin J. Gilmartin, *The American Citizen: 11 Years after High School* (1977), pp. A-v and A-51.

A subsequent public database from Project Talent, downloadable at the ICPSR website (icpsr.umich.edu), yields a European–African difference of 1.50 SDs.

Both estimates of the gap have a built-in downward bias because the cognitive battery was administered to 15-year-olds enrolled in high school. As of 1960, many of the poorest students, disproportionately African, had left school after eighth grade – a consideration that presumably worked to reduce the observed European–African difference in the Project Talent sample.

Page 23: [*The Coleman Report*] *was a pivotal event in social science, representing the first important use of multivariate regression, a technique that has since become the workhorse of quantitative economic and sociological analysis.*

The introduction of computers revolutionized the social sciences. The statistical theory behind multivariate regression analysis had begun with Adrien-Marie Legendre and Carl Gauss in the early nineteenth century and was well developed by the mid-twentieth century, but until computers came along the computational load was too great. Even with the electromechanical calculatora available in the 1950s, a simple regression with a modest sample and a few independent variables took hours to complete, and the load increased nonlinearly with each additional independent variable. The computers of the 1960s were far less powerful than today's smartphones, but they nonetheless opened up a huge range of questions that social scientists could explore quantitatively. The analyses in the Coleman Report were an early illustration of what had become possible. For an excellent description of how this pioneering analysis was done, based on interviews with one of Coleman's principal research assistants, see Elizabeth Evitts, "Coleman Report Set the Standard for the Study of Public Education," *Johns Hopkins Magazine* (Winter 2016).

Page 23: *Subsequent analyses refined the results, finding that the European–African difference was about 15 points for ninth-graders and 18 points for twelfth-graders.*

The standards for reporting quantitative results were still developing when the Coleman Report was being prepared. All that I have been able to find in the report itself is a table on p. 20 that shows the means for each component of the cognitive test battery by race and the standard deviations for the entire sample. See James S. Coleman et al., *Equality of Educational Opportunity* (1966), downloadable at the Institute of Education Sciences website (eric.ed.gov). Subsequently, Arthur Jensen analyzed the massive database published in the Coleman Report's Supplemental Appendix, Sec. 9.10. The results I give in the text represent the average difference of African scores from European scores based on the European standard deviation. The difference was 1.06 SDs for 9th-graders and 1.20 SDs for 12th-graders. Arthur R. Jensen, *Bias in Mental Testing* (1980), Table 10.3. Larry Hedges and Amy Nowell put the twelfth-grade difference in the Coleman study at 1.18 SDs. See Larry V. Hedges and Amy Nowell, "Black-White Test Score Convergence since 1965," in *The Black-White Test Score Gap*, ed. Christopher Jencks and Meredith Phillips (1998), Table 51.

Page 24: *A notable exception was Arthur Jensen's 1969 article in the* Harvard Educational Review *arguing that educational programs were unlikely to close the gap because it was substantially genetic. But that was followed in 1972 by Christopher Jencks's* Inequality: A Reassessment of the Effect of Family and Schooling in America, *which made the case for environmental explanations of the gap.*

Arthur R. Jensen's article was titled "How Much Can We Boost IQ and Scholastic Achievement?" For an excellent summary of it and subsequent developments, see "The Persistence of Cognitive Inequality: Reflections on Arthur Jensen's 'Not Unreasonable Hypothesis' after Fifty Years," humanvarieties.org (December 22, 2019).

The formal citation of *Inequality* names the authors as

Christopher Jencks, Marshall Smith, Henry Acland, et al. The research was collaborative, but Jencks directed it and wrote the text. In informal references, authorship is usually assigned to Jencks alone.

Page 24: *A few years later I read Arthur Jensen's* Bias in Mental Testing, *documenting that the major tests were not biased against minorities.*

See Arthur R. Jensen, *Bias in Mental Testing* (1980). It is a huge book – almost 800 pages in a large-format hardback. Jensen, whose 1969 article on the Black-White test score gap had been widely dismissed because the tests were allegedly biased against Blacks, was tacitly challenging critics to find any significant study bearing on bias that he hadn't already exhaustively examined. For an assessment of how well his magnum opus stood up, see Robert T. Brown, Cecil R. Reynolds, and Jean S. Whitaker, "Bias in mental testing since *Bias in Mental Testing*," *School Psychology Quarterly* (1999). The article's abstract reads in part:

> This paper summarizes the major conclusions from *Bias in Mental Testing* (BIMT) and evaluates writing on test bias published since BIMT. We conclude that empirical research to date consistently finds that standardized cognitive tests are not biased in terms of predictive and construct validity. Furthermore, continued claims of test bias, which appear in academic journals, the popular media, and some psychology textbooks, are not empirically justified. These claims of bias should be met with skepticism and evaluated critically according to established scientific principles. (p. 208)

Page 25: *"I suggest that when we give such parents vouchers, we will observe substantial convergence of black and* *white test scores in a single generation," I wrote, confident that I was right.*

See Charles Murray, *Losing Ground: American Social Policy 1950–1980* (1984), p. 224. My opinion subsequently changed because of one of those chance events that can alter a career's trajectory. In 1986, I was asked by two scholars of cognitive ability, Robert Gordon and Linda Gottfredson, to be on a panel discussing papers that each of them would be presenting at the annual convention of the American Psychological Association. Gordon's paper discussed the role of IQ in explaining White and Black differences in crime rates and Gottfredson's paper discussed the role of IQ in explaining White and Black differences in the labor market. A thoroughgoing skeptic when I began to read them, I was stunned that so much work had been done on the ways in which IQ interacted with social policy issues outside education. That experience triggered my interest in psychometric literature that eventually led to my collaboration with Richard Herrnstein on *The Bell Curve.*

Page 25: *During the 1980s, a number of new studies gave reason to think that things were getting better even without a school voucher program.*

The results of these studies were summarized in two chapters of Christopher Jencks and Meredith Phillips, eds., *The Black-White Test Score Gap* (1998). The two chapters are Larry V. Hedges and Amy Nowell, "Black-White Test Score Convergence since 1965" and David Grissmer, Ann Flanagan, and Stephanie Williamson, "Why Did the Black-White Score Gap Narrow in the 1970s and 1980s?"

The Hedges and Nowell study reported data for six of the large federally sponsored studies: the data for the Coleman Report, the 1979 cohort of the National Longitudinal Survey of Youth (NLSY-79), and the first four longitudinal studies in the Department of Education's series (NLS-72, HS&B-80, HS&B-82, and NELS-88). The authors' purpose was to maximize the comparability of scores across the six studies, so they limited their test composites to the reading and mathematics tests, plus a vocabulary test when it was available, and limited their samples to 12th-graders. They also limited the samples to persons still in school.

In my analysis in Chapter 3, I used the raw data for five of the six tests that now have downloadable databases (to my knowledge, there is no downloadable dataset for the Coleman Report). The website for downloading the Department of Education databases is maintained by the Inter-university Consortium for Political and Social Research (ICPSR), icpsr.umich.edu. The database for downloading the NLSY-79 is maintained by the Department of Labor, nlsinfo.org. My purpose was to report the most complete measure of cognitive ability available for each study based on all the participants. In the case of the NLSY-79, I used the entire sample rather than just the 12th-graders. In the case of NLS-72 and HS&B-80, I used a composite of all the cognitive subtests in the battery. The samples in my analyses used all persons for whom there were test scores, regardless of school enrollment at the time of the test.

These different aspects of the analyses produced some differences in the magnitude of the European–African difference estimated by Hedges and Nowell and by me. They are not large enough to materially affect the results in the text of *Facing Reality*. Both the timing of the convergence and its magnitude would have been the same if I had used the figures from Hedges and Nowell.

Page 25: *When Richard Herrnstein and I were writing* The Bell Curve *in the early 1990s, we included encouraging signs that the European–African test-score difference was diminishing, though we were worried about signs that the narrowing had stalled.*

Here is the relevant passage:

> As we reach the end of this discussion of convergence, we can imagine the responses of readers of varying persuasions. Many of you will be wondering why we have felt it necessary to qualify the good news. A smaller number of readers who specialize in mental testing may be wondering why we have given so much prominence to educational achievement trends and a scattering of IQ results that may be psychometrically ephemeral. The answer for everyone is that predicting the future on this issue is little more than guesswork at this point. We urge upon our readers a similar suspension of judgment.
> (Herrnstein and Murray, *The Bell Curve*, p. 295.)

Page 26: *Third, the persons in the sample in the study must have reached the onset of adolescence.... The reason for the age restriction is that my objective is to estimate mean race differences among adults. Race differences in cognitive ability increase significantly from infancy to childhood to adulthood for reasons that are disputed but aren't relevant to this book. As an empirical matter, the onset of puberty marks the point at which the size of the difference has stabilized.*

The increase in race differences from childhood until adolescence has been examined with regard to the European–African difference. I am not aware of studies of age-related changes in European–Latin or European–Asian differences. The situation regarding the European–African difference as of the late 1990s was summarized by Arthur Jensen in *The* g *Factor*:

> Between ages three and five years, which is before children normally enter school, the mean W-B IQ difference steadily increases. By five to six years of

> age, the mean difference is about 0.70σ (eleven IQ points), then approaches about 1σ during elementary school years, remaining fairly constant until puberty, when it increases slightly and stabilizes at about 1.2σ.
>
> (Jensen, *The* g *Factor*, p. 359.)

"W-B" stands for "White-Black" and σ is the Greek letter sigma, the mathematical symbol for standard deviation. As indicated in the text, Jensen's estimate of the adult difference (about 1.2 SDs) is too high for tests given since the mid-to-late 1980s. My estimate of a current difference of 0.85 SDs is probably too small (see below), but there's no evidence that the current difference approaches 1.2 SDs.

Results for pre-adolescents on the IQ standardizations. In the inventory of IQ standardizations used in *Facing Reality*, I have breakdowns for pre-adolescents on two Stanford-Binet standardizations (ages 7–11), three standardizations of the WISC (6–11) and four standardizations of the Woodcock-Johnson (6–12).

The mean European–African difference on those standardizations was 0.84 SDs, compared to 1.03 SDs for adolescents and adults for those same standardizations.

The only European–Latin breakdowns for pre-adolescents on the standardizations were for the most recent three standardizations of the Woodcock-Johnson. The mean for ages 6–12 was 0.57 SDs compared to a mean of 0.60 SDs for ages 13–65 on the same tests.

None of the European–Asian breakdowns for pre-adolescents had adequate sample sizes to report the results.

Results for pre-adolescents on the NAEP. For practical purposes, the pre-adolescent scores are for 9-year-olds (all 9-year-olds for the LTT and mostly 9-year-olds among the 4th-graders for the standard administration of the SAT).

The mean European–African difference for 9-year-olds was 0.87 SDs. The means for 13-year-olds and 17-year-olds both rounded to 0.94 SDs.

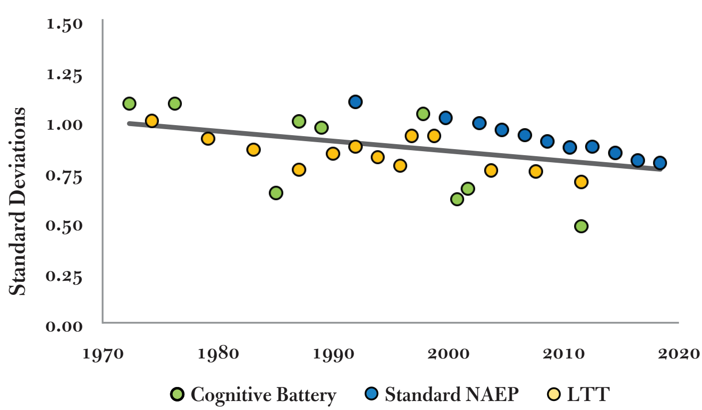
The mean European–Latin difference for 9-year-olds was 0.73 SDs compared to 0.71 SDs for 13-year-olds and 0.69 SDs for 17-year-olds.

The mean European–Asian difference for 9-year-olds was –0.17 SDs compared to –0.12 SDs for 13-year-olds and –0.06 SDs for 17-year-olds.

These results suggest the possibility that the European–Latin difference on mental tests does not increase with age and that the European–Asian difference increases with age (meaning that the Asian advantage increases). But these data do not make the case. I leave these possibilities for others to examine more closely with additional data.

Pre-adolescent trends over time. Graphing the European–Latin and European–African differences for pre-adolescents over time, as in the graphs in Chapter 3, produces scatterplots that mirror the Chapter 3 graphs. The plot of the

Eurropean–African Differences in Mental Test Scores Among Pre-Adolescents



European–African difference for pre-adolescents produces a pattern that is distinctively different from the one in Chapter 3. It is shown in the figure below.

A strong case can be made for a continuing decline in the size of the gap for European and African pre-adolescents. The three most recent IQ standardizations show smaller gaps than five of the previous six standardizations. The most recent Long-Term Trend (LTT) study in 2012 showed a slightly smaller gap (0.70 SDs) than the previous low in 1986–88 (0.76 SDs). The standard NAEP studies have shown a consistent decline since the first datapoint in 1992. The endpoints of the trendline are a fitted value of 0.99 SDs in 1972 and 0.76 SDs in 2019, a drop of 23 percent from the 1972 difference. The unambiguous reduction of the gap in the 2000s gives reason to think that the ambiguous reduction among adolescents and adults during the same period may be real and continuing. Time will tell.

Page 27: *The tests that meet these criteria are standardizations of the major IQ tests, large federally sponsored studies using cognitive test batteries with good measures of* g (“g-*loaded,” in the jargon*) *and the longitudinal assessments of academic achievement known as the National Assessment of Educational Progress.*

After *Facing Reality* was sent to the printers, I found some additional studies that were not included in the figures on pp. 34–37 and made alterations of the scores for the first three Woodcock-Johnson standardizations.

Additions to the inventory of cognitive tests. I discovered that the Program for International Student Assessment (PISA) and the Program for the International Assessment of Adult Competencies (PIAAC) included score breakdowns by race for the United States. I knew about the

PISA assessments – they figured prominently in my discussion of male-female differences in test scores in *Human Diversity* (Chapter 3) – but it hadn't occurred to me that results for the United States might be reported by race (few countries do so). I had not even heard of the PIAAC, which began in 2012–2014 and conducted a second survey in 2017. Both testing programs are especially valuable because of their rigorous psychometric standards for test items and the care with which the sampling strategy ensured the inclusion of teenagers not enrolled in school. For technical information and data see the NCES International Data Explorer at nces.ed.gov.

Amendments to the Woodcock-Johnson Scores. The IQ scores for the first three editions of the Woodcock-Johnson cognitive test in the scatterplots in Chapter 3 were based on the complete battery of cognitive tests in those three editions, as described in Charles Murray, "The Magnitude and Components of Change in the Black-White IQ Difference from 1920–1991: A Birth Cohort Analysis of the Woodcock-Johnson Standardizations," *Intelligence* (2007). The scores for WJ-IV (2012) provided by Riverside Insights were for the measure that the Woodcock-Johnson documentation labels *GIA* (General Intellectual Ability), based on a core subset of the cognitive tests. In the downloadable datafile of IQ standardizations, I have substituted the GIA score for the previous three editions of the Woodcock-Johnson standardizations to make the scores more comparable across editions.

The downloadable files of test scores include the results from PISA and PIAAC that were not used to estimate the current size of racial differences in Chapter 3. Their addition to the calculation has trivial effects on the estimate of the European–Latin difference during the 2010s (which increases from 0.62 SDs to 0.63 SDs) or the European–Asian difference (which shrinks from –0.30 SDs to –0.27

SDs). The estimate of the European–African difference increases from 0.85 SDs to 0.91 SDs. I noted in the text (p. 35) that my estimate of the European–African difference was probably too small. The effects of adding the PISA and PIAAC results are consistent with that expectation.

A contribution of the PISA data is to reinforce the evidence for the declining European–African difference during the 2000s that is observed in the NAEP scores. The European–African difference in the PISA assessments declined from 1.08 SDs in 2003 to 0.90 SDs in 2018. We cannot know whether PISA scores would also have mimicked the declining difference in the NEAP in the 1970s and early 1980s or the increasing difference from the late 1980s into the mid-1990s.

These alterations had minor visual effects on the scatterplots shown in Chapter 3. The 1987–2019 trendline slopes fractionally upward instead of fractionally downward, but even that change is so slight that it is easily missed. The augmented European–Latin and European–Asian test results are even less visually differentiated from the scatterplots in Chapter 3.

The data you need to recreate three figures in Chapter 3 showing European–African, European–Latin, and European–Asian differences in mental test scores are found in the downloadable file *Nationally Representative Studies.xlsx*. The file also includes a sheet with all the LTT means and SDs for all three age groups for Europeans, Africans, Latins, and Asians for all the administrations of the LTT and all the standard administrations of the NAEP from 1990 to 2019 when the math and reading tests were administered in the same year.

Page 27: *The IQ standardizations that have reported results by race (some have not) are…*

The IQ tests in the inventory are the fourth and fifth editions of the Stanford-Binet (SB), all four editions of Woodcock-Johnson (WJ), the second through fourth editions of the Wechsler Intelligence Scale for Children (WISC) restricted to persons 12–16 years old, the second through fourth editions of the Wechsler Adult Intelligence Scale (WAIS), and the first and only edition of the Kaufman Adolescent and Adult Intelligence Test (KAIT).

Missing Editions. The first editions of the SB, WISC, and WAIS are not included either because they did not have an African sample or because the number of Africans was too small to report. WISC-III (1989) had a Latin sample and WISC-IV (2002) had both a Latin and an Asian sample, but Pearson Inc., which has proprietary rights over the WISC data, declined my request for the scores for persons ages 12–16.

I can provide European–African differences for the 12–16 age groups for WISC-R, WISC-III, and WISC-IV because those numbers were provided to me courtesy of the late James Flynn, who had calculated age breakdowns using the raw standardization data. For WISC-V, Pearson declined my request to obtain race breakdowns for participants ages 12–16. Pearson has published the race breakdown for the entire WISC-V sample (ages 6–16). The European differences with Africans, Latins, and Asians in the WISC-V were +0.81 SDs, +0.64 SDs, and −0.35 SDs respectively. See Lawrence G. Weiss, Donald H. Saklofske, James A. Holdnack, and Aurelio Prifitera, *WISC-V Assessment and Interpretation: Scientist-Practitioner Perspectives* (2016), Table 5.3. In the previous three editions of the WISC, the European–African differences for ages 12–16 were substantially larger than the differences for ages 6–11. If the WISC-V followed that pattern (I do not know if it did), the European–African difference for the 12–16 age group probably exceeded 0.9 SDs. If that didn't happen, perhaps the people at Pearson will tell us.

Race breakdowns of standardization data for two IQ test batteries were unavailable for the two standardizations of the Differential Ability Scales, DAS-1 (1988) and DAS-2 (2005), and for the two standardizations of the Reynolds Intellectual Assessment Scales, RAIS-1 (2000) and RAIS-2 (2013). In the former case, the publisher declined my request; in the latter, the publisher did not acknowledge it. If these missing results show a reduction in the European–African difference, perhaps the publication of *Facing Reality* will encourage their disclosure.

Two large federal studies with mental test data are excluded from my inventory:

The High School Longitudinal Study of 2009 (*HSLS-09*). The HSLS-09 is the most recent in the series of longitudinal studies sponsored by the National Center for Education Statistics that began with the NLS-72. I omit its test results because the only test administered to the sample was a mathematics test. It was administered twice, when the students were in 9th and 11th grades. The European–African difference on the math test was 0.70 SDs in the first administration and 0.67 SDs in the second administration.

National Longitudinal Study of Adolescent to Adult Health (*ADD Health*). The ADD Health study administered the Peabody Picture Vocabulary Test (PPVT) to a sample drawn from grades 7–12 in 1995 and again in 2001–2002. The PPVT is a well-established measure of verbal skills but ADD Health did not include a measure of math or visuospatial skills. The European–African difference on the PPVT was 1.03 SDs in the first administration and 0.97 SDs in the second administration.

The other missing data is for the 2017 administration of the PIACC. I was able to find the mean scores by race, but not the standard deviations. See "Highlights of 2017 Results," nces.ed.gov/surveys/piaac/current_results.asp.

The European means in 2017 were 283 points on the literacy index and 269 on the numeracy index. The 2017 European–African difference was 39 points on the literacy index and 53 points on the numeracy index. The European–African difference was essentially unchanged from 2012 to 2014, increasing by 2 points on the literacy index and shrinking by 3 points on the numeracy index. The European–Latin difference shrank substantially, by 17 points on the literacy index and 10 points on the numeracy index. I was unable to find any breakdowns by age (and hence by birth year) for the 2017 administration. The public use database for the 2017 administration does not include the final scale scores for the numeracy and literacy indexes. Presumably they will eventually be available through the International Data Explorer, but they weren't as of May 2021.

Page 28: (*NB: A year in this discussion always refers to the year in which the cognitive tests were administered, not the year when results were published.*)

The testing for a standardization often extends over a period of more than one year and predates publication by at least one year and up to as many as five. For the administration of the tests for SB-4, SB-5, WISC-R, WISC-III, WISC-IV, WAIS-R, and WAIS-III, I used the testing year assigned by William T. Dickens and James Flynn in "Black Americans Reduce the Racial IQ Gap: Evidence from Standardization Samples," *Psychological Science* (2006). For Project Talent, EEOS, WJ-1, WJ-2, WJ-3, NLSY-79, and NLSY-97, I was able to identify the actual dates of data collection. In the case of overlap into two years, I assigned the year with the most months falling within the period. In the case of overlap of more than two years, I assigned the year in which the midpoint fell. For WAIS-IV and WISC-V, I was unable to determine when the standardization samples were tested. I assigned the year two years prior to publication as the testing year. For NLS-72, NELS-88, HS&B-80, HS&B-82, and ELS-02, the year given in the label is the year in which the tests were administered.

Page 28: *The cognitive tests they employed were administered in 1980 and 1998 respectively. Both studies used the Armed Forces Qualification Test (AFQT), a highly* g*-loaded test battery.*

Technical issues regarding NLSY-79 can be found in Appendix 2 of *The Bell Curve.* The psychometric properties of the AFQT, one of the most highly *g*-loaded written tests in use, are described in Appendix 3 of *The Bell Curve.*

Converting the AFQT score to an IQ score is complicated by the age range of the testees – from 15–23 for NLSY-79 and from 12–17 for NLSY-97, meaning from 9th-graders to college graduates in the former case and 7th-graders to 12th-graders in the latter case. I scored both versions of the NLSY using the same method:

Step 1. Compute AFQT scores from the raw subtest scores.

Step 2. Using the sample weights provided by the NLSY, prepare separate score distributions for each birth year.

Step 3. Convert each subject's rank in his or her birth year population ("population" being the sum of the sample weights for that birth year) into percentiles. This step is necessitated by the leftward skew in the NLSY metric. See the discussion on pp. 595–96 in *The Bell Curve.*

Step 4. Assign each subject the *z*-score associated with that percentile in a normal distribution.

Step 5. Convert the *z*-scores to the IQ metric, with a mean of 100 and SD of 15.

These procedures yielded a European–African difference of +1.23 SDs for the entire NLSY-79 sample and +1.02 SDs for the entire NLSY-97 sample.

Hedges and Nowell, in "Black-White Test Score Convergence since 1965," limited their analysis of NLSY-79 to 12th-graders and reported a European–African difference of +1.15 SDs (Table 51). Economist Derek Neal used a different method for dealing with the age issue – transforming test scores into deviations from the average score among persons born in the same two-month interval. He reported a composite European–African difference for persons ages 15–17 of +1.13 SDs for NLSY-79. On the NLSY-97, he reported a difference on the composite of +0.86 SDs for persons ages 13–14 and of +0.94 SDs for persons ages 15–17. Derek Neal, "Why Has the Black-White Skill Convergence Stopped?" in *Handbook of the Economics of Education*, ed. Eric Hanushek and Finis Welch (2006), Table 4.

Pages 29–30: *Rather than engage in the statistical assumptions that would have been necessary to combine the NAEP reading and math scores, I computed the race differences using the known means and standard deviations (see below) for the reading and math scores separately and then used the mean of those two differences to represent the race difference.*

Here's how the calculation was done, using the European–African difference for the example.

Step 1. Calculate the pooled standard deviation for Europeans and Africans on the math test. That calculation requires taking the square root of the squared standard deviations of the two samples weighted by their respective sample sizes. The NAEP doesn't reveal much about sample sizes except to assure us that they are large. Since the samples are designed to be nationally representative, I created proxy samples of 10,000 based on the national proportions of the population represented by each race in that testing year as interpolated from decennial censuses or, from 2006 onward, from the American Community Survey. For example, Europeans were 63.45 percent of the population in 2011 and Africans were 12.40 percent, so my European "sample" for computing the pooled standard deviations in 2011 was 6,345 and my African "sample" was 1,240.

Step 2. Determine the European–African difference on the math test expressed in standard deviations (the European mean minus the African mean divided by the pooled standard deviation).

Step 3. Repeat the process for the reading test for that year.

Step 4. Add the two differences and divide by two.

There is a more accurate way of estimating the race difference by making alternative plausible assumptions about the correlation of the reading and math scores if both tests were administered to the same sample. I conducted two versions of that exercise, one assuming a correlation of .60 and the other a correlation of .75. But it is a less transparent way of presenting the differences and tends to produce larger race differences (for a legitimate reason) than a simple average of the two *z* scores. I decided that presenting them would raise questions about statistical legerdemain without having any substantive effect on the conclusions I draw.

Page 30: *If no other issues were involved, I would have included all of the administrations in the analysis, but some had to be omitted for reasons explained in the online documentation.*

From the first administration of an LTT test in 1971 to the end of the 1980s, the math test and reading test were never administered in the same year. From 1990 through 2012, they were always administered in the same year. By showing the combined tests for 1973–1975, 1978–1980, 1982–1984, and 1986–1988, we at least have a good sense of the trends in the NAEP in the 1979s and 1980s – given the glacial pace at which NAEP scores changed from 1973 to 1988 there is no reason to think that combining scores obtained two years apart is misleading. (The national means for 13-year-olds over that period changed by 3.0 points on the math test and 1.6 points on the reading test. For 17-year-olds, the corresponding changes were 2.0 and 4.5 points.)

Two additional datapoints might have been gained by adding math and reading tests on the regular NAEP 1994–1996 and 1998–2000, but we already have good data from the five administrations of the LTT in the 1990s plus the same-year administration of the regular NAEP in 1992. I doubt if combining two tests from nearby years is misleading, but the change in math score for 8th-graders from 1990 to 2000 on the regular NAEP was 10.6 points. Furthermore, results from those two additional datapoints are very similar to the LTT results from 1994, 1996, and 1999, so there seems no justification to muddy the waters even a little. For the record, the results that I omitted from the table are shown below.

If you're wondering why I sometimes refer to 8th- and 12th-graders and sometimes to 13- and 17-year-olds, it's because the LTT samples are based on age and the regular NAEP is based on school year.

		Mean Difference in SDs		
Grade	Tests	European–African	European–Latin	European–Asian
8th	1994 reading & 1996 math	1.01	0.85	0.08
8th	1998 reading & 2000 math	0.87	0.83	0.19
12th	1994 reading & 1996 math	0.89	0.63	0.10
12th	1998 reading & 2000 math	0.91	0.70	0.20

For tests prior to 2003, the ethnicity variable was "race/ethnicity used to report trends, school-reported." For tests from 2003 to 2010, it was "race/ethnicity allowing multiple responses, student-reported." For tests from 2011 to 2019, it was "race/ethnicity using 2011 guidelines."

Page 33: *The online documentation provides interested readers with downloadable files containing all the data used to prepare the scatterplots and associated data that enable more complicated analyses. I discuss them in the online documentation.*

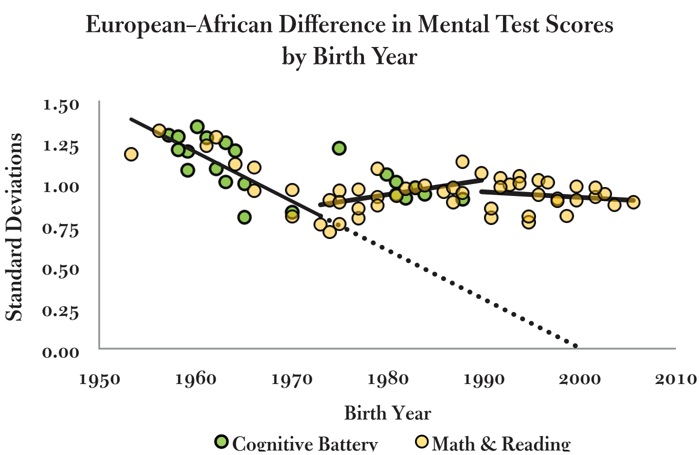
The graphs in Chapter 3 showing the European–Latin and European–Asian differences are reasonably straightforward. There may have been a tendency for the European–Latin difference to increase in the 1990s, but the size of the increase was small, and the overall picture is that of a shallow but continuing shrinking of the difference. The size of the European–Asian difference until the mid-1990s was all over the map, but since then the picture has been a substantial and continuing increase of the Asian advantage. The only real question of importance relative to the content of *Facing Reality* is whether alternative ways of interpreting the data can give reason to conclude that the European–African advantage is shrinking again and can plausibly be expected to continue shrinking past the previous low in the late 1980s.

A Different Perspective: Birth Year Instead of Test Year. The scatterplots in Chapter 3 are based on test year and are known in the jargon as *period* analyses. Each dot represents a result obtained from a test administered at a given point in time. An alternative is to conduct *cohort* analyses in which each dot represents a result obtained among people who were born in a given year (or narrow range of years).

There's a good argument for supplementing the period analyses with cohort analyses, plotting the European–African difference using the year of birth of the test takers as the horizontal axis. The environment in which African children spent their formative years changed significantly from the end of World War II through the 1970s. Education for African children in the North was far better than education had been for their parents in the rural South before the Great Migration. Education for African children in the South also improved – it still wasn't good, but it was better than it had been early in the century. The Civil Rights Movement in the 1950s, the civil rights legislation of the 1960s, the increase in African American urban crime from the mid-1960s onward, the accelerating decrease in Black two-parent families from the mid-1960s onward, and other social and economic trends presumably had effects, good and bad, that impinged on the environmental component of the European–African difference in test scores (whatever that proportion might be). These considerations suggest plotting the European–African difference using the year of birth of the test takers as the horizontal axis.

The figure below shows what happens when the horizontal axis is based on the birth year of the persons being tested. It includes estimates of the European–African difference by separate birth year for the subjects in the

NLSY-79 (1957 to 1964) and the NLSY-97 (1980 to 1984). For the three Woodcock-Johnson standardizations for which I possess the raw data, I used groups of birth years that had combined African samples of at least 199 (1956–1962, 1963–1967, 1968–1973, and 1974–1985). These breakdowns are shown in the scatterplot by the rounded mean of all the birth dates for the Africans in that grouping: 1959, 1965, 1970, and 1981 respectively. For the PIAARC, I used the median birth year for the spans 1948–1958, 1959–1968, 1969–1978, 1979–1988, and 1989–1997. The WAIS African sample sizes by age group were too small to permit disaggregation by birth year, so the WAIS results, which are based on ages 16–74, are not represented in the graph.



The figure for the European–African difference in Chapter 3, organized by test year, suggests an increase in the difference for tests administered during the 1990s and a subsequent resumption of a declining difference in the 2000s. The figure above using birth year as the horizontal axis suggests a parallel pattern. Upon visual examination, it appears that a substantial decline in the European–African difference occurred for subjects born from the mid-1950s to the early 1970s, followed by an increase for subjects born from the mid-1970s through the 1980s, followed by a partial recovery for subjects born from around 1990 until the mid-2000s. How much confidence can we put in the reality of these trends?

More sophisticated methods exist (and I invite other researchers to apply them to the downloadable datafiles), but a simple set of regressions sets out a starting point. I define the three periods as 1956–1973, the birth year with the smallest mean difference; 1973–1990, the birth year with the subsequent largest mean difference; and 1990 to 2006. Coincidentally, the three periods are nearly the same length – 18, 18, and 17 years respectively. The trendlines shown in the figure above represent the results when the European–African difference is regressed on the birth year in the three periods. The downward trend from 1956 to 1973 was substantial and statistically highly significant ($p < .000$), with fitted values of the European–African difference that went from 1.39 SDs in 1953 to 0.81 SDs in 1973. If that trend had continued, the difference would have reached zero in 2001.

The upward trend from 1973 to 1990 was shallower but also reached statistical significance ($p = .029$). The fitted values for those years rose from 0.88 SDs to 1.02 SDs.

The downward trend from 1990 to 2006 was shallower yet and fell short of even a loose definition of statistical significance ($p = .431$). The fitted values fell from 0.96 SDs to 0.90 SDs.

This analysis doesn't dispose of the issue. If the same regression is conducted for the entire period 1974–2006, the regression coefficient turns negative, albeit by only one ten-thousandth of an SD per year. I interpret this as reason to question the reality of the apparent rise in the difference from 1973 through the 1980s and the decline from 1990 onward. The less ambitious conclusion is that the European–African difference has been effectively unchanged for subjects born since the early 1970s.

These additional technical issues stand in the way of confident statements that go beyond the bare minimum.

- Most of the data from the period of a shrinking European–African difference during the 1970s and early 1980s are based on *g*-loaded test batteries; most of the data during the 1990s and 2000s are based on math and reading achievement tests.

- The regular NAEP tests do not include persons not in school when the test is administered. This is obviously a significant problem because of race differences in school dropout among 17-year-olds. But it could also be a problem because of race differences in absenteeism among 13-year-olds who are still legally obligated to attend school.

- Are the samples of racial populations representative? Samples for IQ standardizations are typically stratified to reflect the racial distribution, educational attainment, and to some degree the geographic distribution of the general population, but this leaves room for unrepresentative sampling of subgroups. In the case of racial subgroups, a pertinent question is whether the sampling procedure adequately represents residents in low-income big-city neighborhoods. Another possibility is that the selection of such participants may have changed over time. The NAEP results are less vulnerable to such artifacts. The NAEP program is specifically tasked with obtaining samples that are not only racially representative, but representative within different types of schools. See "NCES Handbook of Survey Methods" at nces.ed.gov.

- Have IQ tests been getting easier? Easier tests mean smaller group differences (to see why, imagine a test in which everyone gets all the items correct – no group differences whatsoever). In this regard, the periodic re-standardizations of IQ tests to a mean of 100 and a standard deviation of 15 give confidence that a given score for a standardization in 1970 marks the same point in the distribution as a standardization in 2010 if the test batteries and their weights in calculating the full-scale IQ score remain comparable. The extent to which those provisos are true varies across IQ tests, for a legitimate reason: the designers of the test are trying to improve them with each new edition. The technical manuals for IQ standardizations usually report psychometric information that allows analysis of changes in content and weighting, but such changes have had unknown (or unreported) effects on results by race.
- Have the NAEP tests been getting easier? The NAEP program has tried to provide a consistent yardstick with the LTT series, but NAEP functions in a politicized environment. Test makers have been subjected to continual criticism based on race differences in test scores, and those criticisms create pressures. The problem is not that people who write the items deliberately make changes that they think will make the test easier for minorities. Rather, they are trying to avoid any conceivable racial bias in an item – which can mean mistakenly rejecting items that are not racially biased at all but are more cognitively challenging than the alternatives that replace them.

> This will tend to decrease the *g*-loading of the test, which will in turn tend to decrease the European–African difference (because of the direct statistical relationship that has been found between *g*-loadings and the magnitude of the European–African difference). In addition, the standard administrations of the NAEP are not constrained to be consistent with previous tests, but instead are supposed to reflect contemporary developments in the curriculum which, to put it gently, have not been increasing the intellectual challenges of the curriculum.

The inflation of scores in the SAT as a cautionary tale. These observations point to potential problems. I have been unable to find explorations of them conducted by independent psychometricians (i.e., not contractors for the NAEP program or publishers of IQ tests) for either the IQ standardizations or the NAEP. However, we can be confident that such issues have had major effects on the SAT.

The SAT's approach to scoring has changed radically over the years, both in the metric (the "recentering" in the 1995) and content (e.g., dropping the antonyms section in 1994 and the analogies section in 2005). Consider what happened to the coveted scores of 700+ on the SAT verbal and math tests since the mid-1990s.

In 1993, the last year before the changes of the mid-1990s, the drop in SAT scores from the late 1960s onward meant that only 1.0 percent of all test takers still scored in the 700s on the verbal test. That percentage more than quadrupled from 1993 to 1996 (4.4 percent). The revisions of the SAT in 2016–2017 led to another increase, from 4.5 percent in 2015 to 6.6 percent in 2018. For the math test, the percentage scoring in the 700s was almost unaffected by the changes in the mid-1990s, rising from 5.1 percent in

1993 to 5.4 percent in 1996. But by 2015 the percentage had drifted up to 7.1 percent. After the reforms of 2016–2017, the percentage in 2018 rose to 9.5 percent.

To grasp how much easier the SAT had gotten, recall that the original design of the SAT was based on a mean of 500 and a standard deviation of 100, meaning that scores of 700+ would represent those at least 2 SDs above the mean, which demarcates the top 2.3 percent in a normal distribution – in the IQ metric, people with IQs of 130+. The percentages scoring in the 700s in the verbal and math tests respectively are nearly triple and more than quadruple the percentages that the original SAT was designed to produce.

The actual scores are only part of the story. These increases in the percentages scoring in the 700s happened over a period when the number of test takers expressed as a percentage of the nation's 17-year-olds rose from 29 percent to 54 percent. That kind of expansion of the test-taking pool should have *lowered* the percentage of students scoring in the 700s by a lot. The proportion of the nation's students with IQs of 130+ who took the SAT in 1993 was already extremely high. The expansion of the test-taking pool by 84 percent from 1993 to 2018 had to be overwhelmingly among those with IQs under 130. And yet the percentages getting scores in the 700s nonetheless climbed. The SAT score data come from the annual reports of the College Board (the reports from 1996 to 2020 are downloadable at collegeboard.org).

Multivariate analyses. Some readers will be wondering what happens when period effects and cohort effects are considered in combination. There is no confident way to know from the dataset of test scores. The correlation between birth year and test year is so high, .85, that multicollinearity makes the coefficients unstable. For the record, regressing the European–African difference on birth year and test year indicated that all the change was associated with birth year. But little faith should be put in that result.

Page 36: *By now the evidence has piled up and is conclusive. On average, Asians outscore Europeans, Africans, and Latins.*

Almost all the available measures underestimate the Asian mean to some degree because researchers have routinely grouped peoples from Asia with peoples from the Philippines, Hawaii, and other Pacific islands. That's potentially a problem because the mean IQ of Asians in general is substantially higher than the mean for the Pacific peoples. But the proportion of Pacific peoples in the Asian-Pacific combination has usually been small and the understatement of the Asian mean has been correspondingly small.

It is sometimes alleged that the apparent Asian advantage is inflated by intensive test preparation among East Asians and widespread cheating on the SAT. There is indeed evidence of systematic schemes for obtaining and sharing answers to SAT questions, especially on tests administered outside the United States. But the advantage of Asian Americans over other groups is roughly the same for the NAEP as for the SAT, and both test prep and cheating are not plausible explanations. The NAEP is a zero-stakes test – the score makes no difference to the student's academic record or college applications – so there's no incentive either to prepare or to cheat.

Page 38: *My estimate of European IQ is the mean of the four IQ standardizations from the 2000s, which works out to 103.35.*

The four standardizations were for ages 12–16 on the WISC-IV, ages 12–23 on Stanford-Binet 5th edition (SB-5), WAIS-IV, and Woodcock-Johnson 4th edition (WJ-4) for ages 13–65. I used the IQ standardizations instead of tests from the 2010s because the IQ tests are much better measures of *g* than the math and reading tests that accounted for all the tests in the 2010s except one. And whereas the sample sizes for Africans, Latins, and Asians in IQ standardizations are often too small to provide reliable estimates, the sample sizes for Europeans on these four tests ranged from 646 to 2,644. Alongside these reasons, the near-identical estimates of the four European means is striking, bunched from 103.2 to 103.9.

CHAPTER FOUR
Race Differences in Violent Crime

Page 47: *This chapter is exclusively about the most serious crimes, called* index offenses *by the FBI – the ones used to create the violent crime index and property crime index included in the FBI's annual report,* Crime in the United States.

The offenses in the violent crime index are murder, rape, robbery, and aggravated assault. The following definitions are taken from the 2018 edition of the FBI's annual *Crime in the United States*, available online at the FBI website.

Murder. "The willful (nonnegligent) killing of one human being by another. The classification of this offense is based solely on police investigation as opposed to the determination of a court, medical examiner, coroner, jury, or other judicial body. The UCR Program does not include the following situations in this offense classification: deaths caused by negligence, suicide, or accident; justifiable homicides; and attempts to murder or assaults to murder, which are classified as aggravated assaults."

Rape. "Penetration, no matter how slight, of the vagina or anus with any body part or object, or oral penetration by a sex organ of another person, without the consent of the victim. Attempts or assaults to commit rape are also included in the statistics presented here; however, statutory rape and incest are excluded."

Robbery. "The taking or attempting to take anything of value from the care, custody, or control of a person or persons by force or threat of force or violence and/or by putting the victim in fear."

Aggravated Assault. "An unlawful attack by one person upon another for the purpose of inflicting severe or aggravated bodily injury. The UCR Program further specifies that this type of assault is usually accompanied by the use of a weapon or by other means likely to produce death or great bodily harm. Attempted aggravated assault that involves the display of – or threat to use – a gun, knife, or other weapon is included in this crime category because serious personal injury would likely result if the assault were completed."

Burglary. "The unlawful entry of a structure to commit a felony or theft. To classify an offense as a burglary, the use of force to gain entry need not have occurred. The UCR Program has three subclassifications for burglary: forcible entry, unlawful entry where no force is used, and attempted forcible entry. The UCR definition of 'structure' includes an apartment, barn, house trailer, or houseboat when used as a permanent dwelling, office, railroad car (but not automobile), stable, or vessel (i.e., ship)."

Larceny-theft. "The unlawful taking, carrying, leading, or riding away of property from the possession or constructive possession of another. Examples are thefts of bicycles, thefts of motor vehicle parts and accessories, shoplifting, pocket-picking, or the stealing of any property or article that is not taken by force, violence, or fraud. Attempted larcenies are included in offense totals. Embezzlement, confidence games, forgery, check fraud, etc., are excluded."

Motor Vehicle Theft. "The theft or attempted theft of a motor vehicle. A motor vehicle is defined in the UCR Program as a self-propelled vehicle that runs on land surfaces and not on rails. Examples of motor vehicles include sport utility vehicles, automobiles, trucks, buses, motorcycles, motor scooters, all-terrain vehicles, and snowmobiles. Motor vehicle theft does not include farm equipment, bulldozers, airplanes, construction equipment, or watercraft such as motorboats, sailboats, houseboats, or jet skis. The taking of a motor vehicle for temporary use by persons having lawful access is excluded from this definition."

Arson. "Any willful or malicious burning or attempting to burn, with or without intent to defraud, a dwelling house, public building, motor vehicle or aircraft, personal property of another, etc."

Page 49: *Differences* [*in crime rates*] *at the national level are substantially understated, for reasons explained in the note.*

Endnote 2 for Chapter 4 explains why aggregating crimes by race to the national level inherently tends to understate race differences in crime rates at the local level. You should be aware that even this inferior way of calculating disproportions in arrest rates is impossible given the way that the UCR data are presented. The UCR's count of arrests by racial group is based on the participating police agencies. In 2018 the arrest data covered only 247.8 million out of the national population of 327.2 million – 76 percent of the population, heavily weighted toward urban areas. What are the racial percentages of the 247.8 million people in the covered population – the denominators for calculating a meaningful number? The published UCR information doesn't include those crucial numbers.

The UCR's European arrest percentage is especially uninterpretable. It includes both Europeans and the 66 percent of Latins who self-identify racially as White. As the data from the thirteen cities demonstrate, Latins have much higher arrest rates than Europeans. Therefore the proportion of violent crimes attributed to Whites in the FBI statistics would be considerably lower than 59 percent if it were restricted to non-Latin Whites, and presumably lower still if the data from the missing 24 percent of the population were included. Why? Violent crime tends to be rarer in rural areas, towns, and small cities than in large cities. Including the missing 24 percent would probably increase the estimated size of the European population out of proportion to the augmented number of European arrests for violent crimes.

The same considerations mean that the Latin/European violent crime ratio is understated in the UCR data. The non-Hispanic rate includes the 39.7 million self-identified Blacks out of the 267.4 million people who qualify as non-Hispanic by the FBI's definition (14.8 percent). The arrest rate for Hispanics includes those who self-identify as both Black and Hispanic, but they number just 1.2 million out of the 59.8 million Hispanics (2.0 percent).

Page 50: *I found thirteen police departments that have posted downloadable databases of arrests by race.*

Cincinnati has released its database of *reported* offenses and it includes the race of the reported suspect, but it is not part of the analysis because the city has not released arrest data. For the record, Cincinnati's African/European ratio for the suspect in reports of crime was 14.8. The Latin/European ratio was just 1.2.

Page 50: *The measure of interest here is the racial ratio of arrests for violent crimes.*

In addition to identifying the arrests that qualify as arrests for index crimes under the UCR criteria, I had to decide how to calculate the overall arrest ratios when the database covered multiple years.

Aggregate rates. Option A was to obtain the rates per 100,000 for each population by dividing the aggregate number of arrests over those years by the mean population of the city over those same years and use those rates to calculate the ratio.

Year-by-year rates. Option B was to calculate the *rates* for Africans, Latins, Asians, and Europeans for each year separately and then calculate the mean of the rates and use those means to calculate the ratio.

Year-by-year ratios. Option C was to calculate the African/European and Latin/European ratios for each year separately and take the average of the ratios over the years covered by the database.

The three methods produced results that are nearly interchangeable. Rounded to the nearest whole number, 63 of the 66 ratios (6 for each of 9 cities, 3 for the other 4) were identical. That result was not foreordained. If there had been major changes in the ethnic profiles over time or if the ratios had shown large year-to-year changes within cities, the three methods could have produced different results. But as it turned out, the changes in the ethnic populations of cities over the time periods for these data were not large, so taking the average population over several years, as option A requires, did not introduce significant error. The rates and ratios moved within remarkably narrow ranges, so large outliers did not make options B or C conspicuously different from one another. With no good reason to choose among them, I report the mean of all three options based on unrounded calculations.

Below are the URLs for the databases that worked as of August 2020. As of May 2021, they did not work for Baltimore, Chandler, or Charleston, nor was I able to find new URLs that accessed arrest data for those cities.

Asheville NC: data-avl.opendata.arcgis.com/datasets/38bd fa06548a45bc9c89c7dddcfc5f31_0

Baltimore MD (adult arrests only): data.baltimorecity.gov/Public-Safety/Arrests-with-race-data/gsmy-ijkf

Charleston SC: data-charleston-sc.opendata.arcgis.com/

Chandler AZ: data.chandlerpd.com/catalog/arrest-bookings/

Chicago IL: home.chicagopolice.org/statistics-data/public-arrest-data/

Fayetteville NC: data.fayettevillenc.gov/search?tags=Arrests

Fort Lauderdale FL: fortlauderdale.data.socrata.com/Police/Arrests/d443-fnye

Lincoln NE: opendata.lincoln.ne.gov/datasets/lpd-arrests-and-citations-2018 and other single-year arrest files from 2013–2017 at the same site.

Los Angeles CA: data.lacity.org/A-Safe-City/Arrest-Data-from-2010-to-2019/yru6-6re4

New York City NY: data.cityofnewyork.us/Public-Safety/NYPD-Arrests-Data-Historic-/8h9b-rp9u

Urbana IL: data.urbanaillinois.us/Police/Urbana-Police-Arrests-Since-1988/afbd-8beq. Urbana stopped recording the race of the arrestee after 2015

Washington DC (adult arrests only): mpdc.dc.gov/node/1379551

Tucson AZ: gisdata.tucsonaz.gov/datasets/tucson-police-arrests-2018-open-data and gisdata.tucsonaz.gov/datasets/tucson-police-arrests-2019-open-data

Page 50: *Table 2 omits ratios involving Asians because very low Asian crime rates yielded absurdly large ratios in most of the thirteen cities.*

I explain why the Asian ratios are not given but say nothing about Amerindians. Only four of the thirteen cities had large enough Amerindian populations to warrant calculating ratios. The four were New York (32,887), Los Angeles (26,660), Tucson (16,776), and Chicago (3,885). I cannot be confident that the Amerindian offense data for the three megalopolises are accurate for the same reason that the Census Bureau is not confident about the real size of the Amerindian population. Many Americans who continue to self-identify as Amerindian have so much European ancestry that they don't "look" Amerindian and don't have Amerindian names. When Los Angeles reports only twenty arrests of Amerindians for violent offenses and three for property offenses over a ten-year period, it must be remembered that many people who check the "Native American" box on the Census Bureau's questionnaire would, if arrested, look unambiguously European to the arresting officer and would probably be so classified in the arrest records if the arresting officer did not explicitly ask for racial self-identification. The same consideration attaches to the Chicago and New York data. Tucson is somewhat different. Amerindians are a familiar and reasonably clearly identified population in many parts of the Southwest and the Mountain West. I therefore hereby report the Amerindian/European ratios for Tucson – 2.9 for violent offenses and 1.0 for property offenses – but with reservations about their accuracy.

Page 55: *Table 3 below shows the ratios for murder arrests for all thirteen cities in our analysis.*

Murder is so rare that only the largest cities have enough European murder arrests to enable the year-by-year methods used to calculate overall violent crime rates and property crime rates. The rate for a given race used to calculate the ratios in Table 3 represents the total number of murder arrests over the years covered by the database divided by the mean population of that race during the years covered by the database.

Page 55: *The Latin/European entry for Fort Lauderdale is empty because no Latin was arrested for murder in that city during the five years covered by the arrest data.*

For calculating the median, Fort Lauderdale was considered to be at the bottom. Zero was entered as the Latin/European ratio in calculating the unweighted mean. The weighted mean is based on the aggregated raw data for all nine cities reporting Latin data, with zeros entered as the number of murder arrests for Fort Lauderdale.

Page 57: *This doesn't mean that members of the public always accurately identify the race of the perpetrators (though their accuracy rate is high), but the police haven't made the judgment.*

For the accuracy of victims' identification of the offender's race, see Michael Hindelang, "Variation in Sex-Race-Age-Specific Rates of Offending," *American Sociological Review* (1981).

Page 58: *We can carry this analysis another step by limiting the zip codes to ones where Africans and Latins* combined *constitute less than half the population.*

The 104 zip codes in which Africans and Latins combined constituted less than 50 percent of the population reported a total of 7,324 alleged violent index crimes including the race of both the victim and a reported suspect. The table below shows the crosstabulations.

To illustrate how the numbers in the text were calculated, the percentage of African victims reporting an African

Reports of Violent Offenses by Race of Suspect and Victim

Suspect	Victim				
	European	African	Latin	Asian	Total
European	653	68	146	111	978
African	1,313	1,345	922	691	4,271
Asian	436	201	794	260	1,691
Latin	40	23	33	288	384
Total	2442	1637	1895	1350	7324

Sample consists of New York City zip codes in which Africans plus Latins are a minority of the total population.

perpetrator is 1345 ÷ 1637 = .82. If you want to do calculations based on rates rather than raw numbers, the populations of those 104 zip codes consisted of 2,224,418 Europeans, 283,176 Africans, 964,203 Latins, and 766,498 Asians.

CHAPTER FIVE
First-Order Effects of Race Differences in Cognitive Ability

Page 65: *The consistent findings about cognitive ability and job performance that apply most directly to group differences in cognitive ability are these.*

A good one-source survey of the literature is Deniz Ones, S. Dilchert, and Chockalingam Viswesvaran, "Cognitive Abilities," in *The Oxford Handbook of Personnel Assessment and Selection*, ed. Neal Schmitt (2014), but it is not easily accessible online and the discussion is also technical. A more accessible summary of the evidence for the basics is Russell Warne, *In the Know*, Chapters 23 and 25. See also Herrnstein and Murray, *The Bell Curve*, Chapter

3 – the technical findings presented in it are still valid 27 years after its publication. For evidence that cognitive ability has the same relationship across different races, see Nathan Kuncel, Deniz Ones, and Paul Sackett, "Individual Differences as Predictors of Work, Educational, and Broad Life Outcomes," *Personality and Individual Differences* (2020).

Page 66: *Rules of thumb are that the correlations between IQ scores and job productivity for low-complexity jobs are seldom lower than .2; for medium-complexity jobs, they are seldom lower than .4; for high-complexity jobs, they are seldom below .5.*

I drew these rules of thumb from Table 3 in Kuncel et al., "Individual Differences as Predictors." It shows the operational validity coefficients for all the meta-analyses they examined.

Page 66: [*It is*] *a straightforward matter to calculate the dollar value of hiring someone with an IQ of 100 versus someone with an IQ of 115.*

Richard Herrnstein and I gave an example of such a calculation in *The Bell Curve*, p. 83.

Page 69: *We published the school-by-school information in* The Bell Curve.

The table showing those results is on page 452. All of Chapter 19 is devoted to affirmative action in higher education – its rationale, administration, and effects. Chapter 20 is devoted to affirmative action in the workplace.

Page 69: *We don't have a current version of the Red Book to work with, but testimony in the recent case charging Harvard with discrimination against Asian applicants included evidence that the same profile of test scores, GPA,* *and extracurricular activities that gave an Asian applicant a 25 percent chance of admission gave an African applicant a 95 percent chance of admission and a Latin a 77 percent chance.*

The material in this testimony is available online as an NBER Working Paper: Peter S. Arcidiacono, Josh Kinsler, and Tyler Ransom, "Asian American Discrimination in Harvard Admissions," National Bureau of Economic Research (2020).

Page 69: *Even without a Red Book, it is easy to make guesses on what then happens throughout the system. We can use a combination of two indicators, both of which are available in the* U.S. News *rankings of universities: the percentages of African and Latin students in the undergraduate student body, and the SAT scores for the 25th, 50th, and 75th percentiles of admitted students.*

The annual college rankings of the *U.S. News and World Report* can be found online at usnews.com/best-colleges.

Page 70: *Table of SAT and ACT data.*

SAT data are published annually, accessible online at research.collegeboard.org. Through 2016, the report was called the "College-Bound Seniors Total Group Profile Report." The current title is "SAT Suite of Assessments Annual Report: Total Group." The SAT data in the text were taken from the table titled "Race/Ethnicity" in the first group of tables in the report, "SAT Participation and Performance."

The ACT data are also published annually, accessible online at act.org. The title is "The ACT Profile Report – National." The ACT data in the text were taken from Table 2.3 of the report for the 2020 graduating class. The national standard deviation of 5.9 is taken from the ACT website report of National Norms for ACT Test Scores based on the 2018, 2019, and 2020 graduating classes.

I do not present time-series data for the race differences in the SAT because the combination of changes in the test and changes in the test-taking pool makes comparisons over time close to impossible without complex statistical modeling. A strong case can be made that the European–African difference reached its low point around 1990, the same time that the NAEP differences and IQ standardizations reached their low point, but making that case is a laborious process.

Page 71: *Many good universities below the top 50 have no African students with scores as high as the 1300s but some European and Asian students with scores in the 1400s and 1500s.*

The calculation of the numbers of Africans and Latins with various SAT scores discussed in the text requires an estimate of the numbers of Africans and Latins in each college class. The *U.S. News* tables show the total number of undergraduates at each school and racial percentages in the undergraduate student body. Drawing on other databases, I determined that the number of freshmen in elite universities is close to 25 percent of the total undergraduate enrollment. One could reasonably expect that percentage to be higher because of attrition over the course of four years of college. But elite colleges have low dropout rates (once you get in, they try hard to keep you) and a long waiting list of students who want to transfer in, thereby compensating for dropouts. I calculated the number of Africans in the freshmen classes for 2019 as (undergraduate enrollment ÷ 4) × percent of Africans in the student body, and the number of Latins similarly. Insofar as Africans and Latins tend to have higher dropout rates over the course of four years than

Europeans and Asians in elite colleges – I can't be more precise than that because of the fragmentary data that are released – my estimates of the numbers of African and Latin freshmen may be low if the transfers from other colleges who filled the vacancies were predominantly European and Asian.

Page 71: *For practical purposes, everyone who wants to get into one of these programs takes the Medical College Admission Test* (*MCAT*)*, the Law School Admission Test* (*LSAT*)*, or the Graduate Record Examinations* (*GRE*).

The sources for the tests in Table 8 and the subsequent pages were:

GRE: ETS, "A Snapshot of the Individuals Who Took the GRE General Test July 2013–June 2019."

MCAT: AAMC, "2019 FACTS," Table A-18, found at aamc.org.

LSAT: LSAC (Law School Admission Council), "LSAT Performance with Regional, Gender, and Racial/Ethnic Breakdowns: 2007–2008 Through 2011–2012 Testing Years."

USMLE: Myia Williams, Eun Ji Kim, Karalyn Pappas, et al., "The Impact of United States Medical Licensing Exam (USMLE) Step 1 Cutoff Scores on Recruitment of Underrepresented Minorities in Medicine: A Retrospective Cross-Sectional Study," *Health Science Reports* (2020).

Patient Complaints in California: Patrick Rogers, "Demographics of Disciplinary Action by the Medical Board of California," California Research Bureau (2017).

Pass Rates for Bar Examinations in California: George Farkas, "Discrepancies by Race and Gender in Attorney Discipline by the State Bar of California: An Empirical Analysis," State Bar of California (2019).

Pages 75–80: *IQ Differences in Ordinary Jobs*

The database. The data for the analysis of IQ differences in ordinary jobs combined three longitudinal databases with information on the adult occupations of persons who were administered a highly *g*-loaded battery of cognitive tests in their teens or early twenties. The three databases come from the Department of Education's National Longitudinal Survey of 1772 (NLS-72), the 1979 cohort of the Department of Labor's National Longitudinal Surveys of Youth (NLSY-79), and the 1997 cohort of the National Longitudinal Surveys of Youth (NLSY-97).

NLS-72 administered an IQ-like battery of tests to a nationally representative sample of 22,652 high school seniors in 1972, 94 percent of whom were born in 1953 and 1954. They were 17–18 years old when they were tested. The occupational data come from the 1986 follow-up, when almost all of them were 32–33 years old.

NLSY-79 consisted of 12,686 persons born from 1958 to 1964. They were 16–22 years old when they were tested using the Armed Forces Qualification Test (AFQT) in 1980. The occupational data come from the 2004 follow-up, when they were ages 40–46.

NLSY-97 consisted of 8,984 persons born from 1980 to 1984. They were 12–17 years old when they were tested with the AFQT in 1997. The occupational data come from the 2017 follow-up, when they were ages 33–37.

The members of all three studies were thus of an age when almost all had completed their educations and entered their careers. The combined studies had 20,203 persons with complete data on IQ and occupation: 12,909 Europeans, 4,235 Africans, 2,462 Latins, 180 Asians, and 417 others. The occupations for NLS-72 used the 1970 version of the Census Occupational Code while both cohorts of the NLSY used the 2002 version. For many occupations, there is no exactly corresponding code for the 1970 and 2002 Occupational Codes. Table 9 is limited to well-known occupations for which the definitions were effectively identical.

As in the analysis of race differences in IQ in Chapter 3, the analysis of IQ differences in occupations is based on conservative choices that tend to underestimate the magnitude of the differences, and for the same reason: to forestall (as much as possible) charges that the analysis inflated race differences. First, I applied a strict criterion of educational attainment. Second, I used sample weights to calculate the differences. Each of these steps requires some elaboration.

A criterion for educational attainment. Participants in surveys such as the NLS-72, NLSY-79, and NLSY-97 sometimes exaggerate when reporting their educational attainment and occupations. Sometimes answers are inadvertent errors, either by an interviewer or by a participant who is recording information into a computer. This becomes a measurement issue when the self-reported educational attainment is plainly inconsistent with the self-reported occupation, as in the case of someone with only a high school diploma who reports being an architect. That person may be a technician working in collaboration with architects but is unlikely to be an actual architect. Often the inconsistencies are ambiguous, however. Someone with only a high school diploma might have acquired enough on-the-job training or night-school courses to be employed as an accountant. This is unlikely if the person has an IQ of 80; plausible with an IQ of 125. What about someone with an IQ of 136 who reports having a bachelor's degree and being a lawyer? It seems more likely that education has been misreported or miscoded than that the person is lying or mistakenly thinks he's a lawyer.

There's no coding system that eliminates all errors, but some sensible culling is better than none. My rule of thumb for flagging errors was to ask if it is plausible that a determined and able person could be working in a given occupation with a given level of education. Specifically, this meant flagging all people who did not meet the following conditions:

- At least a high school diploma for persons in management positions in education, health, finance, and government; specialists in banking, insurance, and taxes; computer programmers, financial analysts, statisticians, accountants, teachers and instructors not specifically classified; technical occupations related to STEM or health; social workers, police, and librarians.
- At least an AA or some college for optometrists, pharmacists, registered nurses, and other licensed health practitioners.
- At least a bachelor's degree for architects, engineers, physical and life scientists, social scientists, college teachers, and K–12 teachers.
- A professional degree for lawyers, physicians, and dentists.

For persons with missing data on educational attainment, I flagged persons with a measured IQ more than 30 points lower than the mean IQ of all persons in that occupation, limited to occupations with a mean IQ of 100 or higher. (I assumed that persons of almost any IQ could plausibly work in occupations with a mean IQ of less than 100 because of the importance of noncognitive skills and attributes for such occupations plus the possibility of measurement error in the IQ score.)

These decision rules flagged 266 persons with occupations that were questionable. The occupations with the most flags were engineers (76), K–12 teachers (36), registered nurses (28), college teachers (15), health technicians (14), and lawyers (11). It would be possible to reclassify many of these cases accurately by assuming the actual occupation was the next step down – engineering technician instead of engineer, practical nurse instead of registered nurse, and so forth, but 266 out of a sample of 20,203 is small enough that I simply coded these cases as missing in calculating means by race. This has almost no effect on the European means – the discarded cases constitute a trivial proportion of the persons in those occupations. It has the effect of slightly increasing the Latin and African means for a few occupations.

The use of sample weights. All three studies used sample weights that enable the calculation of nationally representative estimates, which raised a question: How to combine cases across studies when the sample weights for each study were different? One option was to assume that all three weighting systems were equally accurate and create a proxy sample based on the weights. The second option was to compute the means and standard deviations for each occupation and race separately for each study, then pool the means and standard deviations across the three studies. The third option was to ignore the weights, combine the cases from the three studies, and determine the unweighted means and standard deviations. The first option involved the most ambitious assumption: that the three weighting schemes were done extremely well in all three studies. The second option required the less ambitious assumption that the weights were internally accurate within each study separately.

The results are so similar for all three methods that the interpretation is unchanged no matter which is used. The numbers in Table 6 and Table 7 are the results from the method that pooled separate calculations for each study, chosen because it took the weights into account but involved a less sweeping assumption than the proxy sample method. The results also show somewhat smaller race differences within occupations than the method using sample weights, which led me to report those results rather than the third option of using unweighted data (following my default choice: given two results that show different magnitudes of race differences, report the smaller differences). The race differences in SDs expressed in Table 6 in the text are based on the pooled standard deviations for persons within that occupation.

The unweighted results. An argument can be made that the unweighted combined cases provide a more accurate picture for the question at issue than the method used for Tables 6 and 7. Sample weights are important for questions involving estimates of the number of people within an occupation – the number of African accountants, for example. The question for Chapter 5 is the mean IQ of those accountants. Everyone who is an accountant self-selected into that occupation, which obviates much of the usefulness of sample weights. It's a complicated question. For an analogous problem, use of sample weights in regression analyses, see Gary Solon, Steven J. Haider, and Jeffrey Wooldridge, "What Are We Weighting For?" *Journal of Human Resources* (2015). In the specific case of the database I am using, combining unweighted data across the three surveys has a straightforward advantage when the numbers of minority occupants of a given occupation are small for all three surveys. For example, if each of the three surveys has 15 Africans in a given occupation, using those samples of 15 to calculate weighted means is putting too big a burden on small samples. A total of 45 across all three surveys is large enough to make the mean interpretable.

With that in mind, the table below, based on combined unweighted data, shows a much wider range of occupations than Table 6 in the text. The minimum sample size of 25 produces an estimate of mean IQ that is too small to take literally but large enough to give a sense of the situation. The median African sample size for the occupations in the table is 53 and the mean is 65.

Race Differences in IQ Within Occupations (Unweighted)

	Mean IQ		Race Differences in SDs	
	African	Latin	European–African	European–Latin
Business Management				
Managers of office departments	99	98	0.62	0.72
Accountants	99	104	1.19	0.73
HR & labor relations specialists	97		1.27	
Insurance, loan & tax specialists	96	97	0.95	0.88
Supervisors of administrative staff	90	95	1.30	0.79
Secretaries & AAs	89	93	1.22	0.91
Clerks	88	91	1.19	0.90
Customer service reps	91	92	1.00	0.92
Receptionists	90	92	1.01	0.92
Sales				
Supervisors of sales workers	90	91	1.17	1.07
Retail sales workers	89	93	1.26	0.86
Cashiers	85	85	1.16	1.13
STEM				
Computer programmers & analysts	98	107	1.28	0.54
Technicians: Engineering & science	95		0.82	

Health				
Registered nurses	94	105	1.52	0.47
Health technicians	91	92	1.17	1.10
Health assistants & aides	84	88	1.46	1.08
Education				
K–12 teachers	96	101	1.32	0.76
Preschool teachers	87		1.45	
Teachers' assistants	85	88	1.41	1.26
Government				
Social workers	95		0.98	
Low-level government workers	91	94	1.59	1.28
Protective Services				
Police & detectives	94		0.89	
Security guards	85		1.38	
Construction / Trades				
Vehicle mechanics	82	88	0.90	0.51
Installers & repairers	90	93	0.82	0.57
Carpenters	84	88	1.23	0.90
Other blue-collar crafts	81	84	1.25	0.94
Transportation				
Truck & bus drivers	82	81	0.90	0.88
Deliverymen & routemen	84	86	1.14	0.95
Industrial Production				
Supervisors of production workers	89		0.78	
Machine setters & operators	80		1.15	
Machinists, tool & die makers, welders	82		1.28	
Assemblers	84		0.80	
Garment & textile workers	78		1.55	
Inspectors, sorters & weighers	88		0.61	
Low-skill manual labor	81	86	1.14	0.68

Hospitality Services				
Supervisors: Food prep & serving	88		1.01	
Chefs & cooks	83	83	1.06	0.89
Food prep workers	80	84	0.98	0.62
Bartenders & waitstaff	86	92	0.99	0.52
Maintenance Services				
Janitors & building cleaners	79	82	1.03	0.77
Maids & housekeepers	78		1.52	
Grounds workers	80		0.88	
Personal Services				
Personal appearance workers	81		1.21	
Childcare workers	81	85	1.64	1.35
Personal care & fitness aides	84		1.41	

Note: Differences in SDs are calculated within occupations.

Page 78: *For unskilled occupations, a substantial part of the difference is a statistically predictable phenomenon. It occurs when almost all of the population is "smart enough" to do a particular job, the races have different IQ means, and employers also value noncognitive qualifications such as reliability.*

In a perfectly nonracist world where everyone is hired on the basis of their individual qualifications and the assessment of those qualifications is error-free, what race differences in IQ for the same job would we observe in the same workplace? The intuitive answer is "none," but that's wrong unless the *only* qualification given any weight is IQ. Under the reasonable assumption that IQ is not the only important job qualification, a completely fair hiring process will produce some race differences in mean IQ if the two races have different means. See William Dickens and Thomas

Kane, "Racial Test Score Difference as Evidence of Reverse Discrimination: Less Than Meets the Eye," *Industrial Relations* (1999). Their logic is that employers value a package of attributes, of which IQ is one, in making their selections, and each of these attributes has a lower-bound cutoff point. Employers treat the employment decision as a choice among people whose packages are roughly the same. Dickens and Kane lay out the mathematics that systematically produce European and African test score differences under those circumstances. The European–African differential persists, they argue, even as the lower-bound IQ cutoff rises.

I am persuaded that their treatment of the decision-making process applies to cognitively undemanding jobs. If you're hiring janitors, many packages of qualities could lead you to choose one candidate over another who has a substantially higher IQ. For even moderately cognitively demanding jobs, however, I doubt that many employers think in terms of a cutoff above which job candidates are equally qualified. Certainly the evidence about the relationship of job productivity to cognitive ability supports a continuing relationship as IQ increases within occupations as well as between them. My position is that as jobs become more cognitively complex, it's not just that the lower-bound cutoff rises; so does the value that the employer places on increments in IQ above that cutoff. Thus I expect that a nonracist job market will produce a substantial mean IQ difference between Europeans and Africans if the job is hospital orderly, a modest difference if the job is registered nurse, and virtually no difference if the job is oncologist.

Page 79: *This is not the place to describe the murky jurisprudence surrounding the use of tests in employment decisions* (*the online documentation has a summary*).

From the outset, Title VII of the Civil Rights Act of 1964 had an inescapable problem. It was intended to require that employers hire and promote based on merit rather than race, but how were the regulatory agencies or courts to decide that merit and not racial bias was behind the employer's behavior? One way to solve that problem is to use objective criteria to hire and promote. An obvious candidate for an objective measure is a score on a reliable and valid standardized test.

The first landmark court decision involving such tests was *Griggs v. Duke Power Co.* in 1971. The Supreme Court unanimously held that Title VII imposed a "job-related" requirement on all hiring tests that had "disparate impact." An IQ test was impermissible even if it was a valid measure of cognitive ability and IQ was reliably related to job performance. In the Court's words, "Congress has placed on the employer the burden of showing that any given requirement must have a manifest relationship to the employment in question." (*Griggs v. Duke Power*, 401 U.S. 424 at 432). Employers had to provide evidence of what became known in legal circles as the *business necessity* of a test. The Equal Employment Opportunity Commission (EEOC) subsequently interpreted this as requiring that any test used for hiring must be validated *for that particular employer.* Such validation is extremely expensive and, even if the employer was willing to bear that expense, the EEOC established a track record of rejecting such validations. The most famous case involved the New York City Police Department, which went through extraordinary efforts to develop a test for hiring police trainees that would meet the scrutiny required by *Griggs* and the subsequent EEOC guidelines. Its use was invalidated in *Guardians Association v. Civil Service Commission* (1983).

The next critical case was *Wards Cove Packing Co. v. Atonio* (1989). The case did not involve a test per se. Rather, the Supreme Court softened the requirement for a hiring criterion from *business necessity* to *business justification.* Congress passed legislation in 1991 that overturned some aspects of the *Wards Cove* decision, but the environment for using tests had been eased considerably.

Employers who want to use tests must still anticipate expensive legal trouble. Another landmark case, *Ricci v. DeStefano* (2009), illustrates the continuing hazards. The New Haven Fire Department developed a test for promotion to management positions. When it was administered in 2003, no Blacks qualified for the available promotions. New Haven's mayor, John DeStefano, Jr., declined to fill the appointments for fear that New Haven would be liable to a lawsuit under the "disparate impact" criterion of antidiscrimination law. The firefighters who qualified for promotion (19 Whites and one Latino) sued on grounds that they were victims of racial discrimination. The case went all the way to the Supreme Court, which ruled that the city of New Haven had failed to establish any "genuine dispute" about the examination's lawfulness. But Mayor DeStefano would surely also have faced a lawsuit if he had permitted the promotions – after all, the case was so fiercely contended that it was fought all the way to the Supreme Court, which then divided five to four. Hence my use of the word *murky* to describe the situation that still confronts employers who want to use standardized tests.

Page 81: *To illustrate, I'll use the cohort of young Americans ages 25–29, the age at which the potential candidates for such jobs are coming out of law schools, medical schools, business schools, and graduate STEM departments. In 2019, there were 23.2 million Americans in that age group. About 228,000 people in that age group can be expected to have IQs of 135 or higher.*

This statement is based on a simulation of the distributions of IQ for Europeans, Africans, Latins, and Asians, which in turn is based on the estimates of racial means shown in Chapter 3 and described in the notes for the chapter. To conduct the simulation, I needed to estimate the SDs for each race. I used the NAEP tests conducted during the 2010s and the IQ standardization and *g*-loaded federal surveys conducted from 1998 onward, reaching estimated SDs of 14.1, 13.8, 14.1, and 15.0 for Europeans, Africans, Latins, and Asians respectively. To create the simulation, I used the DRAWNORM function in Stata v. 15. The number of observations entered into the instructions were 100 percent of the population ages 25–59 for each race as given in the 2019 ACS.

Page 84: *Objective measures of job performance and subjective ratings of job performance show roughly similar differences.*

This is the common overall finding from the two studies. Insofar as ratings suffer from potential bias, it does not appear that the bias exaggerates race differences in performance. From the first study:

> Our results do not support the position that subjective measures have more potential for bias than objective measures. Instead, we found the opposite. This is important because J. K. Ford et al. (1986) noted that some researchers (not necessarily including themselves) have called for the increased use of objective measures to minimize Black–White differences based on the implicit assumption that objective measures are less prone to bias than subjective measures. Our results are more consistent with a position that there may be some pressure to minimize ethnic group differences on raters (e.g., Mobley, 1982).
>
> (P. L. Roth, A. I. Huffcutt, and P. Bobko, "Ethnic Group Differences in Measures of Job Performance: A New Meta-Analysis," *Journal of Applied Psychology* (2003), p. 702.)

Page 85: *A large 1989 study of performance ratings among Army enlisted personnel found that Europeans had a modest advantage over Africans on measures of task proficiency and job effort, but there was little difference on measures of discipline and an African advantage on measures of military bearing.*

See Elaine D. Pulakos, Leonard A. White, Scott H. Oppler, and Walter C. Borman, "Examination of Race and Sex Effects on Performance Ratings," *Journal of Applied Psychology* (1989).

Page 86: *Among accountants, race differences in the pass rate for the Certified Public Accountant exam are commensurate with the race differences in cognitive ability.*

See Arya Espahbodi, Linda Espahbodi, Reza Espahbodi, Rosemary Walker, and G. Thomas White, "Opportunity, Ethnicity, Gender, and CPA Exam Performance," (in press), Table 3.

Page 86: *In the legal profession, the race differences in pass rates for the bar exam are commensurate with race differences in cognitive ability. So are differences in the percentage of attorneys who have been the subject of repeated complaints in California.*

This will supplement the summary in the endnotes in the text. The table below is adapted from Table 1 in Stephen P. Klein and Roger Bolus, "The Size and Source of Differences in Bar Exam Passing Rates among Racial and Ethnic Groups," *The Bar Examiner* (1997).

In 1998, the Law School Admission Council published a national study of 23,086 students who entered American law schools in the fall of 1991 and were followed through the results from their first bar exam. The table below is adapted from Table 6 in Linda F. Wightman, "LSAC

Bar Exam Passing Rates by Race During the 1970s, 1980s, and 1990s

First-Time Test Takers

	California 1994–96	Colorado 1990–97	New Mexico 1976–80	New Jersey 1988–89
Europeans	78%	86%	86%	83%
Africans	47%	54%		53%
Latins	58%	71%	52%	
Asians	70%	80%		

All Test Takers

	New York 1992	Florida 1991
Europeans	82%	76%
Africans	37%	46%
Latins	49%	
Asians	53%	

National Longitudinal Bar Passage Study," Law School Admission Council (1998).

Reports of bar exam results since 2000 are hard to find. I have found only one, in the February 2020 California Bar Examination General Statistics Report posted at TaxProf blog (taxprof.typepad.com/files/feb-2020-ca-bar.pdf). The pass rates for first-time test takers were 51.7 percent for Europeans, 30.6 percent for Latins, and 5.0 percent for Africans. The overall pass rate of just 26.8 percent of all applicants was an all-time low, so this one result may be anomalous. But insofar as it shows a much larger African–European difference than the 1990s results, it does not represent evidence that the gaps in the earlier studies have closed. Current efforts in several states to make bar examinations easier so that the numbers of minorities who pass will increase imply a continuing large difference in pass

	Pass Rate	Number
Europeans	91.9%	19,285
Africans	61.4%	1,368
Latins	74.6%	1,046
Asians	80.8%	961
Amerindians	66.4%	107
Others	83.1%	319

rates. See Sam Skolnik, "Bar Exams May Soon Be Easier to Pass, as States Eye Changes," *Bloomberg Law* (March 29, 2021).

For race differences in disciplinary action against attorneys, see George Farcas, "Discrepancies by Race and Gender in Attorney Discipline by the State Bar of California: An Empirical Analysis," The State Bar of California (2019).

Page 86: *In the medical profession, race differences in board certification for a medical specialty are commensurate with race differences in cognitive ability. So are differences in investigations of complaints filed against physicians, and in disciplinary action by the state medical board of California.*

For differences in board certification, see Donna B. Jeffe and Dorothy A. Andriole, "Factors Associated with American Board of Medical Specialties Member Board Certification Among US Medical School Graduates," *JAMA* (Sept. 7, 2011), Tables 1–4.

For differences in complaints, see Patrick Rogers, *Demographics of Disciplinary Action by the Medical Board of California (2003–2013)*, California Research Bureau (Jan. 2017).

Page 86: *For K–12 teachers, race differences among those rated "minimally effective" or "ineffective" in Michigan* *were commensurate with race differences in cognitive ability.*

See Steven Drake, Amy Auletto, and Joshua M. Cowen, "Grading Teachers: Race and Gender Differences in Low Evaluation Ratings and Teacher Employment Outcomes," *American Educational Research Journal* (2019), Tables 2 and 3.

CHAPTER SIX
First-Order Effects of Race Differences in Violent Crime

Page 93: *Raj Chetty of Harvard and his colleagues have conducted extremely detailed geographic analyses of upward socioeconomic mobility down to the level of city blocks.*

With the cooperation of the Internal Revenue Service, Prof. Chetty assembled a database of tens of millions of 1040 and other tax forms that enable analyses of geographic variation in income at an unparalleled level of detail. Two of the most important technical products of that work are Raj Chetty, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez, "Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States," *Quarterly Journal of Economics* (2014); and Raj Chetty, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter, "Race and Economic Opportunity in the United States: An Intergenerational Perspective," *Quarterly Journal of Economics* (2020). Chetty has set up a website, opportunityatlas.com, that allows both scholars and the public to explore his extraordinary database.

Page 94: *The academic analyses of the results so far suggest that this initiative is producing the same unintended outcomes that have characterized previous efforts.*

For a technical analysis, see Alan Sage, Mike Langen, and Alex Van de Minne, "Where Is the Opportunity in Opportunity Zones?" Brookings (Feb. 2021). For less technical commentary on experience to date, see Sophie Quinton, "Black Businesses Largely Miss Out on Opportunity Zone Money," Pew (2020); Adam Looney, "Will Opportunity Zones Help Distressed Residents or Be a Tax Cut for Gentrification?" Brookings (2018); Stan Veuger, "'Opportunity Zones' Don't Actually Work," American Enterprise Institute (2020); and Brett Theodos, Eric Hangren, Jorge González, and Brady Meixell, "An Early Assessment of Opportunity Zones for Equitable Development Projects," Urban Institute (2020).

Page 95: *The job of a police patrol officer – a cop – in an urban setting is unique.*

The literature on policing is extensive. It may be hard to believe that a book written almost fifty years ago can be relevant to policing issues today, but James Q. Wilson's *Varieties of Police Behavior: The Management of Law and Order in Eight Communities* (1968) remains a classic worth reading. More recent books giving the cop's-eye view are Edward Conlon, *Blue Blood* (2004); Pat McCarthy, *Chicago Street Cop* (2016); and almost anything by Joseph Wambaugh, nonfiction or fiction.

Page 97: *Now think in terms of frequency distributions of the amount of force that police use.*

For a recent discussion of the use of force by police, see Seth W. Stoughton, Jeffrey J. Noble, and Geoffrey P. Alpert, *Evaluating Police Uses of Force* (2020). For the effects of anti-police activism, see Heather Mac Donald, *The War on Cops* (2017).

Page 98: *I have argued elsewhere that the differences between big-city America and everywhere else are the real cultural fault line that has polarized the nation.*

See Chapter 12 of Charles Murray, *By the People: Rebuilding Liberty Without Permission* (2015).

Page 103: *Toward the end of his career, James Q. Wilson, who for decades was one of America's leading scholars of crime and policing, captured the essence of the problem posed by race and crime better than I can.*

James Q. Wilson, "Crime," in *Beyond the Color Line: New Perspectives on Race and Ethnicity in America*, ed. Abigail Thernstrom and Stephan Thernstrom (2002), pp. 115–17. In addition to *Varieties of Police Behavior*, Wilson's *Thinking About Crime* (1972) is a classic that transformed the national dialogue about the role of incarceration in reducing crime, and his article "Broken Windows: Police and Neighborhood Safety," *Atlantic* (March 1982), coauthored with George Kelling, transformed police responses to low-level crime in many major cities.

CHAPTER SEVEN
If We Don't Face Reality

Page 110: *Jonah Goldberg has described the fragility of the American system by comparing it to a garden hacked out of a tropical jungle.*

Thinking back over fifty years of citing sources, I can recall no comparable situation: I clearly remember Jonah Goldberg's using the metaphor of the garden hacked out of the jungle. I read it in one of his weekly online columns known as "The GFile." Goldberg remembers having written it. Neither of us has succeeded in tracking down the text of that particular GFile or the date when it was published. For a book-length treatment of related themes, see Goldberg's *Suicide of the West: How the Rebirth of Tribalism, Nationalism, and Socialism Is Destroying American Democracy* (2018).

Page 115: *In 2001, Gallup's pollsters began asking the question, "Would you say relations between whites and blacks are very good, somewhat good, somewhat bad, or very bad."*

See Mohamed Younis, "Most Blacks Rate Race Relations with Whites as Bad," Gallup (2019).

Page 116: *Much of that change had nothing to do with race relations or identity politics, but with the alienation of middle-class and working-class Whites from the coastal elites. I have written about that alienation at length.*

See Murray, *Coming Apart*, especially Chapters 3, 4, and 17.

Page 118: *Since 1958, the Gallup polling organization has periodically asked Americans how much they trust the federal government to do what is right.*

See "Public Trust in Federal Government Near Historic Lows for More than a Decade," Pew Research Center (Sept. 14, 2020).

DOWNLOADABLE FILES

In the course of writing *Facing Reality*, I assembled databases that have potential for exploring important social policy issues than I could not fully explore in the book. I invite other scholars or simply interested readers to download them and conduct their own explorations.

Nationally Representative Studies.xlsx

This Excel file includes separate sheets as follow:

Inventory of g*-Loaded Studies*. This sheet includes means, SDs, and sample sizes by race for all of the IQ standardizations and large federal surveys using *g*-loaded test batteries, broken down by age groups when appropriate.

Inventory of M&R Studies. This sheet includes means, SDs, and sample sizes by race for all of the studies administering math and reading tests, broken down by age groups when appropriate.

Detailed NAEP Data. This sheet includes means, SDs, and sample sizes for the NAEP math and reading tests separately, by race and for all three age groups in the LTT and by grade for the standard NAEP administrations.

Combined Longitudinal Studies.xlsx

This file consists of 20,203 observations combining the NLS-72, NLSY-79, and NLSY-97 samples. It contains variables for IQ, educational attainment, occupation, and other demographic and socioeconomic indicators.

Violent Crime by Zip Code.xlsx

This file contains a single sheet concatenating the data on individual arrests for violent offenses for the cities reporting zip code (New York City, Los Angeles, Washington, Chandler, Fayetteville, Fort Lauderdale, and Tucson).