Facing Reality

TWO TRUTHS ABOUT RACE IN AMERICA

Charles Murray

Encounte

New York · London

© 2021 by Cox and Murray, Inc.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of Encounter Books, 900 Broadway, Suite 601, New York, New York, 10003.

First American edition published in 2021 by Encounter Books, an activity of Encounter for Culture and Education, Inc., a nonprofit, tax-exempt corporation. Encounter Books website address: www.encounterbooks.com

Manufactured in the United States and printed on acid-free paper. The paper used in this publication meets the minimum requirements of ANSI/NISO Z39.48–1992 (R 1997) (*Permanence of Paper*).

FIRST AMERICAN EDITION

LIBRARY OF CONGRESS CATALOGING-IN-PUBLICATION DATA

Names: Murray, Charles A., author. Title: Facing Reality: Two Truths about Race in America / Charles Murray. Description: First American edition. | New York, New York: Encounter Books, 2021. | Includes bibliographical references and index. Identifiers: LCCN 2021000549 (print) | LCCN 2021000550 (ebook) | ISBN 9781641771979 (hardcover) | ISBN 9781641771986 (ebook) Subjects: LCSH: Racism—United States. | Race. | Intelligence levels— United States. | Crime and race-United States. | Discrimination in law enforcement—United States. | United States—Race relations. | United States-Social policy. Classification: LCC E184.A1 M8955 2021 (print) LCC E184.A1 (ebook) DDC 305.800973-dc23 LC record available at https://lccn.loc.gov/2021000549 LC ebook record available at https://lccn.loc.gov/2021000550





Figure 2 Arrests for Violent Offenses in the District of Columbia and Socioeconomic Status by Zip Code





Election of 2016



Note to the Reader

THE EMPIRICAL ASSERTIONS in *Facing Reality* are not complicated in themselves, and in a reasonable world they would not be controversial. They are facts that we must face. It shouldn't take long to read them, and it won't. You can read the main text of *Facing Reality* over the course of an evening. Maybe two.

story behind facts The the is occasionally complicated, however, and aspects of the facts are controversial for understandable reasons, but different readers will have dif-ferent reservations. Some of you will be comfortable accepting arrest data as quantitative evidence of criminal behavior but doubt that IQ tests tell us anything worth knowing. Others will be familiar with the basics of IQ but suspicious of anything the police tell us. The endnotes present additional evidence or further explanation of tech-nical issues. Standard documentation of sources, still more elaboration of technical issues, and downloadable databases have been posted online at encounterbooks.com/books/facing-reality. This companion www.blackhillsaudiobooks.com/ PDF is available at facingreality.

Table 1
America's Detailed Racial and Ethnic Profile as of 2019

	Non-Latino	Latino
White	60.0%	12.1%
Black	12.4%	0.4%
East Asian	2.4%	0.0%
South Asian	1.5%	0.0%
Filipino/Pacific Islander	1.1%	0.0%
Native American	0.7%	0.2%
Southeast Asian	0.6%	0.0%
Other Asian	0.1%	0.1%
Other Single Race	0.3%	4.7%
White & Black	0.7%	0.1%
White & Native American	0.5%	0.1%
White & Asian	0.5%	0.1%
Other Combination	0.8%	0.6%
Total	81.6%	18.4%

Nomenclature

Should I refer to the groups in Table 1 as races? As ethnicities? Both? The question is particularly loaded in an age when race is widely argued to be a social construct – an artificial classification that marginalizes minorities but lacks a meaningful genetic foundation. Geneticists have dealt with this problem by dispensing with both *race* and *ethnicity*, and instead using the word *population*. 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. Then, in 1965, Congress passed the Immigration and Nationality Act. America changed rapidly. Applying my nomenclature and rules for classifying Latins, the American population profile as of 2019 looked like this:

European	60.0%
Latin	17.9%
African	12.8%
Asian	5.7%
Amerindian	0.7%
Pacific Islander	0.2%
Other	2.8%

The European-African Difference

The race difference in cognitive ability that has caused by far the most controversy and angst since the 1960s is the difference between America's Africans and Europeans. It also is informed by the most data – fiftynine estimates, adding up the age breakdowns from the studies in my inventory. The figure below shows how the size of the dif-ference varied from 1972 to 2019.

The solid black circles in the plot indicate results from the average of math and reading tests (all but two are from NAEP), while the white circles indicate IQ standardiza-tions org-loaded tests in federal surveys.

The figure shows a major reduction in the difference during the 1970s and into the 1980s. Exactly when the narrowing stopped is open to interpretation. Regression analyses show that the trendline was nearly flat for tests conducted from 1983 through 2019. But 1987 saw the lowest mean for any year (0.69 SDs), so I will use 1987 to illustrate the change in the trendlines.

FACING REALITY



The European-Latin Difference

The European–Latin difference in mental test scores is consistently smaller than the European–African difference, but it remains substantial.

I have drawn a linear trendline to help organize your view of the history, but obviously the story is more compli-cated than that – notably, the difference increased during the 1990s, then resumed its decline in the early 2000s.

The fitted values for the European–Latin difference in 1972 and 2019 are 0.87 SDs and 0.63 SDs respectively – a reduction of 28 percent. The decline shows no signs of ending. I have no satisfactory explanation for the rise in the difference during the 1990s. A close examination of changes in the nature of the Latin immigrant pool over that period might give some clues.

The mean difference in the eleven test results during the 2010s is 0.62 SDs, and that will serve as my estimate of the current European–Latin difference.





European-Asian Difference in Mental Test Scores Expressed in Standard Deviations

This graph has something new: negative values on the vertical axis, which are needed to show when Asians have a higher mean test score than Europeans. The farther below zero, the larger the Asian advantage.

The generally small but irregular European–Asian difference into the 1990s is typical of the spread of results when sample sizes are small, as they were for some of those tests. But a more important explanation, albeit speculative, is that large-scale Asian immigration began in the 1970s with large numbers of Vietnamese refugees whose children were still being acculturated to America and used English as their second language when they took the NAEP in the 1980s. Subsequent Asian immigration has drawn heavily from highly educated East Asians and South Asians. The more consistent results from the mid-1990s onward could reflect a more consistently talented immigrant pool.

The trendline is based on 1995–2019, when the fitted values of the European–Asian difference went from 0.09 SDs to -0.39 SDs, a remarkable net change of nearly half a standard deviation in twenty-four years. The estimate of the current European–Asian difference, like the others, uses tests from the 2010s, which averaged -0.30 SDs.

Estimates of the Current Differences

Using tests administered during the 2010s, which in effect means NAEP, the mean European–African difference was 0.85 SDs, the mean European–Latin difference was 0.62 SDs, and the mean European–Asian difference was -0.30 SDs. Assuming a mean of 100 and SD of 15 for IQ, the corresponding equivalents in IQ points are 12.75, 9.30, and 4.50 respectively.

My estimate of European IQ is the mean of the four IQ standardizations from the 2000s, which works out to 103.35. Thus we are left with mean IQ estimates of 90.60 for Africans, 94.05 for Latins, and 107.85 for Asians.

Rounding estimated IQ to the nearest whole number, here are the means and their percentiles in the national distribution that I will use for the rest of the book:

	Estimated Mean IQ	<i>z</i> -score Equivalent	Percentile Equivalent
European	103	0.20	58
African	91	-0.60	27
Latin	94	-0.40	34
Asian	108	0.53	70

The top figure shows how much overlap exists in the distributions. It is not a threatening picture. Yes, differ-ences exist, but it is also true that millions of Africans and Latins have higher cognitive ability than millions of Euro-peans and Asians. The top figure should also serve as an object lesson in the necessity of judging people as individ-uals, not members of groups. If you rely on the difference in means you are going to make a huge number of mistakes about individuals.

The bottom figure shows why race differences in cogni-tive ability nonetheless have consequences for the society as a whole. The differences in the raw numbers of individ-uals on the right-hand side of the bell curve become larger as IQ goes up. Among people of the four races with IQs of 100, 70 percent are European or Asian. For IQs of 115, 85 percent. For IQs of 125, 90 percent. For IQs of 140,96 percent.

Since the most prestigious, powerful, and highest paying jobs are so concentrated among people on the right-hand side of the distribution, a variety of important social and economic consequences are not just possible. They are inevitable. Chapter 5 takes up a few of the most important ones.

Two Perspectives on the Distribution of Cognitive Ability Across Races



The measure of interest here is the racial ratio of arrests for violent crimes, focusing on two comparisons: the number of African arrests per 100,000 people divided by the number of European arrests per 100,000, and the number of Latin arrests per 100,000 divided by the number of European arrests per 100,000. In all the ratios I present, the larger number is divided by the smaller, so the denomi-nator is always 1. To simplify the presentation, I will report just the numerator. For example, a ratio of 2.5 to 1 will be expressed as a ratio of 2.5. Table 2 below shows the African/ European and Latin/ European ratios for the combined vio-lent crimes. A separate table for murder follows presently.

Table 2 omits ratios involving Asians because very low Asian crime rates yielded absurdly large ratios in most of the thirteen cities. For example, Fort Lauderdale recorded just one arrest of an Asian for a violent offense from 2015 through 2019 among the 3,000 Asians who lived there during that period. But Asian arrest rates were exceptionally low even in most cities with large Asian populations. The lone exception to the rule was New where the Asian violent York. crime rate was marginally higher than the European one.

Table 2

Ratios of Violent Arrest Rates in Thirteen Cities

City	Years	Population	African/	Latin/
		in ooos	European	European
			Ratio	Ratio
New York NY	2006-2019	8,375	11.6	4.1
Los Angeles CA	2010-2019	3,921	9.0	2.4
Chicago IL	2014-2017	2,714	14.5	2.8
Washington DC	2013-2019	682	19.9	6.4
Baltimore MD	2014-2019	610	$5 \cdot 3$	
Tucson AZ	2011-2019	533	5.5	1.6
Lincoln NE	2013-2018	279	13.3	1.7
Chandler AZ	2013-2019	256	5.7	2.5
Fayetteville NC	2010-2019	205	4.0	
Fort Lauderdale	2015-2019	180	9.0	1.3
Charleston SC	2015-2019	139	10.2	
Asheville NC	2012-2019	90	5.2	
Urbana IL	1988-2014	37	11.3	1.2
Median			9.0	2.4
Mean			9.6	2.7
Mean weighted by	population		11.2	3.2

CHAPTER FOUR

Table 3 below shows the ratios for murder arrests for all thirteen cities in our analysis. Recall that Latin arrest rates were not available for Asheville, Baltimore, Charleston, or Fayetteville. 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.

The italicized ratios in the table indicate that the denominator (the European rate) is based on a sample of six or fewer murder arrests over the entire period covered by the database. In those cases, the ratio should simply be interpreted as "large" without attaching much importance to the specific number. The denominators for the rest of the cities were at least eleven murder arrests.

City	African/ European Ratio	Latin/ European Ratio
New York NY	18.1	5.5
Los Angeles CA	19.8	5.4
Chicago IL	21.6	3.9
Washington DC	84.9	10.4
Baltimore MD	6.3	
Tucson AZ	7.2	1.3
Lincoln NE	33.3	3.7
Chandler AZ	14.1	7.7
Fayetteville NC	8.7	
Fort Lauderdale FL	5.5	
Charleston SC	61.4	
Asheville NC	7.4	
Urbana IL	20.3	2.6
Median	18.1	4.7
Mean	23.7	5.1
Mean weighted by population	21.0	4.9

	Table 3	
Arrest Rates	for Murder in	Thirteen Cities

CHAPTER FOUR

New York City's Shootings Database

New York offers yet another way to triangulate. The New York Police Department has assembled a separate dataset of all shootings from 2006 to 2017 – not simply "shots fired," but shots that struck a human being. This dataset thus includes shootings that did not result in a fatality, and it's a big number: 81 percent of the 21,626 shootings in the NYPD database were nonfatal. By combining the shoot-ings database with the arrest database, it is possible to cre-ate another measure: the race of probable perpetrators in cases where the NYPD concluded they knew who did it but didn't have enough evidence for an arrest.

The table on the next page shows the results when we compare New York City ratios based on victims' reports of a suspect's race, arrests for violent crimes other than mur-der (i.e., rape, robbery, and aggravated assaults), arrests for murder, and shootings that did not result in an arrest. Once again, the arrest data are the most conservative estimate of the racial disproportions, with the single exception of Latin suspects in reported violent offenses.

The New York database of shootings is also useful as a counterweight to much of the rhetoric from the Black Lives Matter movement. Of course they matter, no matter what the race of the shooters in the New York database may be.

Measure	African/ European Ratio	Latin/ European Ratio
Suspects in reported violent offenses	14.8	3.9
Arrests for violent crimes other than murder	11.3	4.0
Arrests for murder	17.9	5.5
Suspected perpetrator of shootings not resulting in an arrest	rs 48.6	8.7

CHAPTER FOUR

Nonetheless, it is useful to take a quick look at race dif-ferences in arrests for property crime. They reinforce the effects of differences in violent crime, and to some extent they interact. Table 4 below shows the numbers for our thirteen cities.

Table 4Ratios of Property Offense Rates inThirteen Cities			
City	African/	Latin/	
	Ratio	Ratio	
New York NY	5.2	2.0	
Los Angeles CA	5.9	1.5	
Chicago IL	6.9	1.2	
Washington DC	10.2	2.8	
Baltimore MD	2.7		
Tucson AZ	2.5	0.9	
Lincoln NE	7.5	1.4	
Chandler AZ	3.9	1.8	
Fayetteville NC	1.7		
Fort Lauderdale FL	5.4	1.2	
Charleston SC	3.7		
Asheville NC	3.0		
Urbana IL	6.5	0.7	
Median ratio	5.2	1.4	
Mean ratio	5.0	1.5	
Mean ratio weighted by population	5.6	1.6	

]	Mean Expres	sed in <i>z</i> -Sc	ores	
	Test Year	European	African	Latin	Asian
SAT ACT	2020 2020	$\begin{array}{c} 0.25 \\ 0.24 \end{array}$	-0.59 -0.66	-0.39 -0.36	0.79 0.73
The national population		0.20	-0.60	-0.40	0.53

The scores for both the SAT and the ACT match closely with the national estimates for Europeans, Africans, and Latins presented in Chapter 3, while Asians score well above their national estimate. That is, the SAT and ACT differences in means are about the same or greater among college applicants than they are in the population as a whole.

Applications to Professional Schools

Table 5 below summarizes recent test scores for students who want to go into one of the professions by way of a medical degree, law degree, MBA, or PhD. For practical purposes, everyone who wants to get into one of these pro-grams takes the Medical College Admission Test (MCAT), the Law School Admission Test (LSAT), or the Graduate Record Examinations (GRE). I show the z-score of the average applicant on the major test used for the profes-sional field that the applicant wants to enter. *STEM* refers to science, technology, engineering, and mathematics.

The average difference between Europeans and Africans on these tests was larger than the IQ difference in the gen-eral population. The same was true of the Asian and African comparison. In contrast, Latins taking these tests had somewhat *smaller* differences with Europeans and Asians than Latins in the general population.

In terms of percentiles, Africans are in the bottom quar-tile of test scores for all the admissions tests except for those heading to business school or graduate school in education. The Latin scores are all in the second quartile.

Table 5 Race Differences in Admissions Tests for Professional Training

Mean Expressed in z-Scores

	European	African	Latin	Asian
Medical School	0.15	-0.91	-0.63	0.25
Law School	0.20	-0.88	-0.44	0.20
MBA Programs	0.14	-0.65	-0.31	0.09
STEM PhD Programs				
Physical Sciences	0.03	-0.99	-0.46	0.41
Life Sciences	0.08	-0.80	-0.36	0.32
Engineering	0.09	-1.02	-0.55	0.32
Other PhD Programs				
Education	0.20	-0.59	-0.35	0.10
Social &	0.15	-0.72	-0.38	0.06
Behavioral Sciences				
Humanities	0.11	-0.82	-0.56	0.01
The national population	0.20	-0.60	-0.40	0.53

20

CHAPTER FIVE

test scores of the applicants but also the test scores of those who are accepted and matriculate. Here are the results for 2019:

	Mean z-Scores			
	European	African	Latin	Asian
Applicants Matriculants	$0.15 \\ 0.09$	-0.91 -0.89	-0.63 -0.82	$0.25 \\ 0.37$

Not much happened between application and admission to affect the differences. The European–African difference was reduced slightly while the European–Latin difference increased. The Asian differences from the other three races all increased.

U.S. Medical Licensing Exam (USMLE). This examination is required for admission into most residency programs. "Step 1" of the USMLE measures whether the test taker can "understand and can apply important concepts of the sciences basic to the practice of medicine." It consists of seven 60-minute blocks administered over an eighthour period. In effect, it is an exit test from medical school. Below are the Step 1 z-scores for 10,541 applicants to residency programs during 2014–2015 at the Zucker School of Medicine at Hofstra/Northwell on Long Island, New York.

		Mean z-Scores				
	European	African	Latin	Asian		
Applicants	0.15	-0.48	-0.30	0.04		

CHAPTER FIVE

IQ Differences in Ordinary Jobs

Comparatively few people are physicists, physicians, attor-neys, or work in other high-prestige occupations, which means that even a sample of 20,203 includes too few in such occupations to provide analyzable samples by race for them. But the three studies do have enough people in more normal occupations to do so, and the sample weights used by the studies enable us to reach estimates that are representative of the national population, so I can return to using the IQ metric.

Table 6 below shows mean IQs and the sizes of the race differences for nine familiar occupations ranging from cognitively demanding (accountant) to a low-skill job (janitor or building cleaner). They are a selection from a larger set of occupations for which data are presented in the online documentation. I have ordered the occupations by the European mean IQ from highest to lowest.

Race Differences in IQ Within Occupations						
		Mean IQ		Race Differences in SDs European– European–		
Occupation	European	African	Latin	African	Latin	
Accountants	111	100	104	0.96	0.60	
K-12 teachers	\$ 110	95	101	1.35	0.76	
Registered nurses	109	94	105	1.49	0.42	
Social worker	s 105	95	93	0.93	1.09	
Retail sales workers	102	89	93	1.17	0.80	
Childcare workers	102	83	85	1.55	1.34	
Secretaries & AAs	102	90	93	0.96	0.72	
Vehicle mechanics	94	83	87	0.85	0.57	
Janitors & bldg. cleane	92 ers	79	82	1.10	0.78	
Median	102	90	93	1.10	0.76	
Mean	103	90	94	1.15	0.79	

Table 6
Race Differences in IQ Within Occupations

22

Table 7 below expands the power of the analysis by grouping occupations that are filled by people with similar IQs. For example, the combined samples do not have enough African or Latin physicians to provide reliable estimates, but combining them with other occupations that draw from people with mean IQs above 115 – college teachers and lawyers, for example – increases the samples to usable sizes.

The IQ groupings in Table 7 are based on the European mean. Categories below 100 are all blue-collar jobs, ranging from unskilled to highly skilled labor. The 100–104 category includes occupations for which the European mean was at least 100 and less than 105. Examples are preschool teachers, police, and electricians. Examples of occupations in the 105–109 category are registered nurses, secretaries, and social workers. Examples of occupations in the 110–114 category are accountants, clergy, computer programmers, and engineers. Examples of occupations in the 115+ category are physicians, dentists, lawyers, and college teachers.

The mean IQs of Africans and Latins increase with each higher category, just as the European means do, but the size

Race Diffe	rences in Oc	Tal cupations	ole 7 Grouped	d by the Euro	pean Mean	
IQ Group	Mean			Differe Standard I	Difference in Standard Deviations	
	European	African	Latin	European– African	European- Latin	
90-94	93	82	85	0.98	0.67	
95-100	97	85	88	1.02	0.70	
100 - 105	103	88	93	1.15	0.77	
105 - 109	107	95	97	1.02	0.83	
110 - 114	113	99	104	1.21	0.73	
115+	118	105	112	1.11	0.53	



I project that reaction onto you (which of course you may not have had) based on my experience in the quarter century since *The Bell Curve* appeared. Too many of the conversations I've had resemble one that I recently saw depicted on the Internet, as shown here.

The experience of most White readers of this book - disproportionately college-educated and upper middle class - includes many Black and Latino acquaintances who correspond to the dots on the right-hand tail of that graphic. For example, suppose that your personal experi-ence has consisted of life as a White in an upper-middle-class American suburb. Your Black, Latino, and Asian neighbors have been as smart, engaging, and helpful as your White neighbors. The bell curve of your personal experience does not involve mean differences in cognitive ability or crime rates. It is natural to think your experience invalidates the data about group differences in means. The mind insists on generalizing. But when mean differences between groups are real, it is absolutely essential to resist generalizing; it is essential to accept the reality of docu-mented group differences but to insist on thinking of and treating every person as an individual.

Notes

The endnotes that follow are only a fraction of the material that is available to the curious at encounterbooks.com/books/facing-reality. The notes I include here are restricted to material that readers might need to answer their most immediate questions. Some of that material cites specific sources, but most of the documentation of sources in this book is reserved for the online documentation.

Facing Reality is not a formally academic text, so I have taken the liberty of streamlining my citation format to fit the way scholars actually do their research these days. They no longer acquire a specific volume of a technical journal from the library stacks and look up an article by using the issue number and page numbers in the citation. They type the title of the article (even a portion of it is usually enough) and perhaps the surname of one of the authors into their Internet browser and hit "return." So I have used an abbreviated form of article citation, with everything needed to find the source. I include the page number for citations of direct quotes. In the case of Internet sources, I avoid linking to a specific page because they so quickly go out of date, instead trying to give you a link that will get you to the correct website with enough additional information to let you search for the appropriate page.

NOTES TO CHAPTERS 1-2

CHAPTER 1: THE AMERICAN CREED IMPERILED

 Samuel P. Huntington's Who Are We? The Challenges to America's National Identity (2005), a brilliant book, is the best recent source about the American creed. The quotation is taken from p. 46. For descriptions of how thoroughly the American creed (though not yet called that) permeated life in the early nineteenth century, see Francis Grund, The Americans in Their Moral, Social, and Political Relations (1837) and, of course, Alexis de Tocqueville, Democracy in America (1838). For a description of how the creed persisted at the end of the nineteenth century, see a two-volume study, James Bryce, The American Commonwealth (1903). Bryce, a leading British scholar of American society at the time, reflected on "certain dogmas or maxims which are in so far fundamental that ... one usually strikes upon them when sinking a shaft, so to speak, in an American mind" – dogmas and maxims that amounted to the American creed (pp. 536–37).

During World War I, the U.S. House of Representatives passed a resolution titled "The American's Creed" written by William Tyler Page. It conveys elements of the creed as I describe it but is by no means an official definition. To my knowledge, there isn't one.

CHAPTER 2: MULTIRACIAL AMERICA

 In 2001, the Gallup polling organization found that the average American estimated that 32 percent of the population was Black. The correct answer was 12.3 percent. The same poll found that the average American thought that 29 percent of the population was Latino. The correct answer was 12.5 percent. Joseph Carroll, "Public Overestimates U.S. Black and Latino Populations," Gallup News (June 4, 2001).

Gallup has not asked those questions since 2001, but a 2013 survey by the Center for American Progress indicated that the American public still overestimated the size of minority populations. Most Blacks and Latinos thought that minorities combined already made up half the population or more, while non-Latino Whites estimated that minorities were 48 percent of the population. Only Asians (who gave an estimate of 43 percent) were even close to the correct answer, which in 2013 was 37 percent. See Ruy Teixeira and John Halpin, "Building an All-In Nation: A View

NOTES TO CHAPTERS 2-3

from the American Public," Center for American Progress (October 22, 2013).

- 2. You may have a larger question about the substitution of *European* for *White*: Is it still accurate, given recent immigration of peoples from North Africa and the Middle East who in the old terminology are classified as Caucasian? Part of the answer lies in the 23 andMe results showing that self-identified Whites had 98.6 percent European ancestry. Additional evidence may be found in the combined ACS surveys for 2014–2018, which reveal that 95.1 percent of non-Latin Whites who answered the "Ancestry 1" question specified a European ancestry while only 1.6 percent gave a North African or Middle Eastern ancestry. *Europeans* seems reasonable as a way of identifying the way that the overwhelming majority of non-Latin Whites think of themselves.
- 3. It is possible that the percentage of Latins in 1960 is understated. The limited information in the 1960 census is frustrating. There is no "Latino" category either as part of the "What race are you?" question or as a separate "What ethnicity are you?" question. A separate table shows country of birth for the foreign-born, reporting that 1,735,992 people were born in Mexico, but the numbers in the table for race are inconsistent with the numbers in the table on country of birth for the foreign-born. But my estimate could be too low by half and we would still be talking about only 3 or 4 percent of the population who were not European or African.

CHAPTER 3: RACE DIFFERENCES IN COGNITIVE ABILITY

- If you want to know the whole story and are reasonably knowledgeable about statistics, go to Arthur Jensen's magisterial *The* g *Factor: The Science of Mental Ability* (1998). For more recent and readable discussions of what IQ tests measure, the thorough version is Russell Warne, *In the Know: Debunking 35 Myths about Human Intelligence* (2020). A short, breezy, but scientifically accurate account is Stuart Ritchie, *Intelligence: All That Matters* (2015).
- 2. The importance of self-esteem, so enthusiastically assumed by educators from the 1970s through the end of the century, flunked empirical attempts to demonstrate its causal role in academic performance or other achievements. A single comprehensive review article dismantled self-esteem's reputation among scholars, partly because of its massive documentation and partly because the lead

author had previously been an open advocate of the importance of self-esteem. See Roy Baumeister, Jennifer D. Campbell, Joachim I. Krueger, and Kathleen D. Vohs, "Does High Self-Esteem Cause Better Performance, Interpersonal Success, Happiness, or Healthier Lifestyles?" *Psychological Science in the Public Interest* (2003).

"Stereotype threat" enjoyed a similar vogue from 1995 through the next twenty years. The concept was introduced by Claude M. Steele and Joshua Aronson in "Stereotype Threat and the Intellectual Test Performance of African Americans," *Journal of Personality and Social Psychology* (1995). It was seized upon so uncritically that by 2003, just eight years after its debut, it was already covered in two-thirds of introductory psychology textbooks.

Since 2015, its reputation has been battered by a series of failures to replicate the effects seen in early studies and by evidence of "publication bias" - the tendency of scholars to fail to publish negative results. Two of the most rigorous critiques leave little room for the advocates of stereotype threat to make their case: Paulette C. Flore and Jelte M. Wicherts, "Does Stereotype Threat Influence Performance of Girls in Stereotyped Domains? A Meta-Analysis," Journal of School Psychology (2015); and Oren Shewach, Paul R. Sackett, and Sander Quint, "Stereotype Threat Effects in Settings with Features Likely Versus Unlikely in Operational Test Settings: A Meta-Analysis," Journal of Applied Psychology (2019). The former, coauthored by one the world's most highly regarded quantitative social science methodologists (Jelte Wicherts), concluded that "based on the small average effect size in our meta-analysis, which is most likely inflated due to publication bias, we would not feel confident to proclaim that stereotype manipulations will harm mathematic performance of girls in a systematic way." (p. 41). The latter article, written by a team of psychologists at the University of Minnesota, concluded, "Based on the result of the focal analysis, operational and motivational subsets, and publication bias analyses, we conclude that the burden of proof shifts back to those that claim that stereotype threat exerts a substantial effect on standardized test takers." (p. 1529).

3. If you have come to *Facing Reality* thinking that Stephen J. Gould demolished the concept of g forty years ago with *The Mismeasure* of *Man* (1981), you should know that the scientific consensus about that book is that Gould had attacked psychometrics as it stood in the 1930s and systematically misrepresented psychometrics

as it stood when he was writing in the 1970s. See Arthur Jensen, "The Debunking of Scientific Fossils and Straw Persons," Contemporary Education Review (1982); Bernard D. Davis, "Neo-Lysenkoism, IQ, and the Press," Public Interest (1983); and John B. Carroll, "Reflections on Stephen Jay Gould's The Mismeasure of Man (1981): A Retrospective Review," Intelligence (1995). A 2019 analysis by a team of psychologists concluded as follows: "Given Gould's pervasively incorrect statements in The Mismeasure of Man about the Army Beta, factor analysis, the place of intelligence testing in the immigration debates of the 1920s, the biological basis for intelligence, and the questions regarding Gould's analysis of Morton's work, we wonder whether there is any section of The Mismeasure of Man that is factually accurate." Russell T. Warne, Jared Z. Burton, Aisa Gibbons, and Daniel A. Melendez, "Stephen Jay Gould's Analysis of the Army Beta Test in The Mismeasure of Man: Distortions and Misconceptions Regarding a Pioneering Mental Test," *Journal of Intelligence* (2019), p. 18. Emphasis in the original.

Or you may think that Nassim Taleb has more recently proved that, as he titled his article, "IQ is largely a pseudoscientific swindle," medium.com/incerto (January 1, 2019). Several responses are available on the Internet. One of the first but also most direct is Sean Last, "Nassim Taleb on IQ," archive.ph/PCvgk (January 8, 2019). Jonatan Pallesen gives a highly technical response in "Taleb is wrong about IQ," jsmp.dk (June 15, 2019), but that's unavoidable in dealing with some of Taleb's statistical assertions. James Thompson has an accessible series of articles about Taleb's arguments. The first three are "Swanning About: Fooled by Algebra?" "In the Wake of the Swan," and "The Intelligent Investor," all available at the *Unz Review* (unz.com).

4. Howard Gardner's theory of multiple intelligences is orthogonal to the psychometrics literature. His presentation of the theory in *Frames of Mind: The Theory of Multiple Intelligences* (1983) is a fascinating discussion of human talents, but he has never tried to prove statistically that his "intelligences" can be distinguished from g or from personality traits. Gardner offered this amusing and I think correct observation in a 2018 interview: "I have never been able to reconstruct when I made the fateful decision not to call these abilities, talents, or gifts, but rather to call them 'intelligences.' Because if I had called them anything else, I would not be well known in different corners of the world and journalists like you

wouldn't come to interview me. It was picking the word 'intelligence' and pluralizing it." Liz Mineo, " 'The Greatest Gift You Can Have Is a Good Education, One That Isn't Strictly Professional," *Harvard Gazette* (May 9, 2018).

5. The results from all the known studies of African intelligence through the 1950s are reported in Audrey M. Shuey, The Testing of Negro Intelligence (1966). The first large sample used to calculate African cognitive ability was assembled during World War I, when the U.S. Army used two IO tests for incoming recruits, one designed for literate recruits and the other designed for illiterate or non-English-speaking recruits. Robert Yerkes found a European-African difference of 1.16 SDs, reported in John C. Loehlin, Gardner Lindzey, and J. N. Spuhler, Race Differences in Intelligence (1975), but little confidence can be attached to that number. Even a bad elementary education is associated with a substantial increase in cognitive ability over a population with no elementary education. As of World War I, 70 percent of all Blacks still lived in the rural South, where most African children got only the most rudimentary education or no education at all. There is reason to believe that this population was underrepresented among those draftees who reached the point of being administered the Army Alpha and Army Beta tests. See Jeanette Keith, Rich Man's War, Poor Man's Fight: Race, Class, and Power in the Rural South During the First World War (2004).

The caution with which one must approach the World War I data is accentuated by the data from World War II. The European-African difference on the Army General Classification Test for inductions in 1944–1945 has been put at 1.52 SDs (Loehlin, *Race Differences*). This represents the scores of men born from 1925 to 1927 and is very close to the 1.59 SD difference observed among the Woodcock-Johnson subjects born in the 1920s. See 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).

How could the European–African difference in cognitive ability have risen from 1.16 SDs to 1.52 SDs in 20 years? The simplest explanation is that the World War II testing produced a more accurate estimate of the European–African difference than did the World War I testing.

6. The phrase "g-loaded" traces its roots to the English psychologist

Charles Spearman's seminal 1904 article, "'General Intelligence,' Objectively Determined and Measured," in the *American Journal* of *Psychology*. Spearman observed that students' test scores across unrelated subjects were statistically correlated. His explanation was that the correlations reflected a general underlying mental ability, which he called g for general. The most powerful tool for measuring g is a statistical method called factor analysis. When the subtests in an IQ test battery are factor-analyzed, the first factor always explains a much larger proportion of the variation across the subtests than any other. That first factor is g. The g-loading of a subtest is its correlation with the overall measure of g. I sometimes use the phrase more loosely to signify a test that is a good measure of g, meaning that its correlation with the overall measure of g would be high if it were a subtest in an IQ test battery.

- 7. Scores on academic achievement tests are affected by the test taker's education - you can't score well on a reading comprehension test unless you know how to read. But reading tests such as the ones for the NAEP or the SAT don't ask how much the student has learned with questions (for example) about how Middlemarch fits into the history of the English novel. The tests present passages of text and ask about the meaning of those self-contained passages, which calls directly upon the test taker's ability to figure things out. The same principle applies to the NAEP math tests: the items are devised so that they presume only the math courses that are normally required of 4th-, 8th-, and 12th-graders. The test items require the students to figure out the answers from information contained in the items themselves. Composites of the major standardized tests of math and reading skills are thus g-loaded, but they don't provide as good a measure of g as a more comprehensive set of cognitive subtests.
- 8. The group differences are likely to increase as highly g-loaded tests are added because of another of Charles Spearman's hypotheses: the higher the g-loading of a subtest, the greater the size of group differences in IQ. The hypothesis has subsequently been confirmed through an extensive literature. See Arthur Jensen, *The* g *Factor: The Science of Mental Ability* (1998), pp. 369–402.
- 9. The proper method for combining separate measures depends on what you're trying to measure. Sometimes it is appropriate to add them (for example, if you are trying to measure a family's aggregate income you just add wages and dividend income); sometimes to

take the mean (for example, of judges' ratings in a gymnastics competition); and sometimes – as in the case of measuring cognitive ability – to add them after taking into account what they share. Verbal and mathematics ability are expressions of the same mental ability to some extent but also contribute independently to mental ability. The size of the add-on to a simple mean depends on the correlation between the two test scores. If the correlation were a perfect 1.0, the correct composite score would be the mean of the two scores because the two tests are imparting exactly the same information about mental ability. The lower the correlation between the tests, the more that each test is contributing independently to the measure of cognitive ability. As an empirical matter, correlations of reading and math tests in different studies cluster in the .60 to .75 range, which means a modest but nontrivial add-on to the mean.

10. Think of the standard deviation (SD) as a measure of spread-outness in a distribution of scores. If you're measuring height and your sample consists of five-year-olds, the SD of height is going to be compressed into a range of a few inches. If your sample includes everyone from infants to adults, the spread of heights will include everyone from newborns to NBA players, and the SD will be much larger.

The equation for computing a standard deviation of a population is this:

$$S \neq \frac{\Sigma(X,\overline{X})^2}{N}$$

where S = the standard deviation of a sample

 Σ means "sum of"

X =each value in the data set

 \overline{X} = mean of all values in the data set

 \mathcal{N} = number of values in the data set

Just as the average of a set of numbers is its sum divided by the number of numbers, the standard deviation represents the average *difference* between the numbers in a set and the mean for that set. That's not precisely correct – note the need to square the differences and eventually take the square root – but it's a convenient way to remember the basic concept.

11. Any set of scores can be converted to *z*-scores, but a normal distribution is necessary for accurate conversion of *z*-scores to percentiles.

- 12. Recent years have seen much work on establishing IQ estimates for countries around the world, but that is no help in trying to estimate ethnic subgroup means for U.S. populations. The United States does not draw a nationally representative sample of immigrants from anywhere. If you want to get a sense of how much work has been done and some of the controversies surrounding estimates of national IQ, visit the National IQ Dataset, a site that maintains a curated inventory of the studies of national IQ and posts discussions of many of the technical issues. As of February 2020, I found the releases of the data set at viewoniq.org/?page_id=9, and the discussions at viewoniq.org/.
- 13. I've been writing about interventions to raise IO for a long time. For a review of the state of knowledge as of the early 1990s, see Chapter 17 of Richard Herrnstein and Charles Murray, The Bell Curve: Intelligence and Class Structure in American Life (1994). For an update through the first half of the 2000s, see Chapter 2 of Charles Murray, Real Education: Four Simple Rules for Bringing American Education Back to Reality (2008). For a discussion that incorporates recent developments in the understanding of heritability and the nonshared environment as they relate to early interventions, see Chapter 13 of Charles Murray, Human Diversity: The Biology of Gender, Race, and Class (2020). I will not try to summarize the many findings here, but one is so uncontested that it deserves mention: For the vast majority of U.S. children, low scores on cognitive tests are not changed by remedial courses. Much more intensive efforts are required, and even those have produced disappointing results. A few experimental pre-K programs have produced statistically significant gains on exit tests, but these results suffer from fadeout. A good optimistic review of the pre-K literature is Greg J. Duncan and Katherine Magnuson, "Investing in Preschool Programs," Journal of Economic Perspectives (2013). A good skeptical review of the evidence is Grover J. Whitehurst, "Does State Pre-K Improve Children's Achievement?" Brookings (July 12, 2018).
- 14. Each item in a major test is scrutinized for evidence that it is unusually hard for some group – i.e., not just harder for that group than for another group, but harder in relation to other items in the same test. For example, suppose that the group difference runs between 0.6 and 0.8 SDs for 19 out of 20 items on an IQ subtest but is 1.5 SDs on the remaining one. That would be a red flag that

NOTES TO CHAPTERS 3-4

something other than a difference in g is affecting the result on that item.

Once a draft of the test passes scrutiny on the individual items, it can be tested for measurement invariance. The technique of choice is multiple-group confirmatory factor analysis. Cognitive tests have complex structures. There are a variety of ways to determine whether the structures are the same for different groups by measuring not only whether the factor structures themselves are similar, but also whether the factor loadings, intercepts, and residual variances are similar across groups. Together, these constitute strong evidence that the test is measuring the same construct for different groups. All the current versions of the major tests are known to be factor invariant.

CHAPTER 4: RACE DIFFERENCES IN VIOLENT CRIME

- 1. The exact bases for a legal arrest vary from jurisdiction to jurisdiction, but three are common to all:
 - > The police officer personally observed a crime taking place.
 - > The police officer at the scene has probable cause to believe that the arrestee recently committed or is about to commit a crime.
 - > The police officer has an arrest warrant issued by a judge.

"Probable cause" means that the police officer has specific, tangible evidence within the officer's knowledge that would lead a reasonable person to think the arrestee is guilty.

From a social scientist's perspective, an arrest by the police has several advantages as evidence of race differences in crime. Most police do not make arrests lightly, for reasons both professional and practical. With misdemeanors, a warning can often solve a problem without getting the culprit enmeshed in the criminal justice system. With felonies, police are aware how easily a prosecution can get thrown out if there's a problem with "probable cause." Even a completely legal arrest involves a lot of paperwork. From the social scientist's perspective, these disincentives to arrest people give the arrest extra credibility as evidence that a meaningful offense occurred and that the police arrested a plausible suspect.

Another reason that arrests for violent offenses are especially valuable for assessing race differences in criminal behavior is that

they are largely insulated from the usual problems that raise public concern. Much of the criticism of the police involves situations that give police officers considerable latitude in deciding whether to intervene – examples are stop-and-frisk, arrests for minor offenses as part of "broken windows" policing, enforcement of vice laws, and enforcement of drug laws. Given credible evidence that a murder, rape, robbery, or aggravated assault has been committed, there is much less police discretion about whether to act.

It's not just the seriousness of the crimes that sets arrests for violent crimes apart; it's also the circumstances under which those crimes get into the statistics. Police seldom catch offenders in the act of murdering, raping, robbing, or assaulting. Far more commonly, the crimes occur and then come to the attention of the police. The potential for the police to manufacture violent offenses for Africans or Latins while overlooking them for Europeans is constrained.

2. The proper calculation of the arrest rate for any group is to divide the number of arrests by the total number of people in that group in the jurisdiction where the arrests occurred. The *ratio* of two rates is one divided by another. If the Latin arrest rate for some crime is 50 per 100,000 and the Asian arrest rate for that crime is 20 per 100,000, the Latin/Asian ratio of rates is $50 \div 20$, or 2.5 to 1.

Just about every arrest rate by race you have ever seen in print or on television has been presented in terms of the percentage of arrests of a given race relative to that race's percentage of the national population. If Latins are 18 percent of the population but they account for 27 percent of the arrests for violent crime (as in the FBI statistics for 2019), it looks as if the disproportion amounts to $27 \div 18$, or 1.50, meaning that Latins were arrested 50 percent more often than their proportion of the population would predict.

The problem is that Latins – and Africans and Asians as well – are not spread evenly around the country. To see why this is important, imagine a nation with a majority population of 1,000,000 and a minority population of 50,000. A member of the majority was arrested 5,000 times last year. A member of the minority was arrested 2,000 times.

The total number of arrests in this nation last year was 7,000. So a newspaper story reads: "Minority accounts for 29 percent of the arrests despite being less than 5 percent of the population. That's a ratio of 6 to 1." That is the usual arithmetic behind reports of race differences in arrest rates. It's not arithmetically wrong, but how meaningful is it?

If every jurisdiction in the country has precisely the national percentages of majority and minority, it's meaningful. But as it happens, my fictional nation consists entirely of rural countryside with the exception of a single city inhabited by 50,000 of the majority and all 50,000 of the minority. So the city experienced 250 arrests of members of the majority last year (its fair share of the 5,000 arrests of the majority) and all 2,000 of the arrests of the minority. That means the minority accounted for 89 percent of the arrests in the city, or a ratio of 2,000 to 250, which equals 8 to 1.

For everyone outside the city, the 6 to 1 national ratio is meaningless in terms of their own lives. If an arrest occurs where they live, the probability that it was committed by a member of the minority is not 89 percent. It is zero. The city is the only place in the country where race differences in arrest rates are relevant, and the relevant ratio is 8 to 1.

Such problems are common when statistics are aggregated. Suppose you wanted to investigate gender discrimination in a university's faculty and based your conclusions on all the departments combined. The patterns you observe in the university as a whole could be radically different from those in either the physics department or the sociology department. The kinds of mistakes I'm describing are sometimes grouped under the label "ecological fallacy."

3. For a list of cities and counties with Open Data sites, see data.gov or https://opendatainitiative.github.io/. I found the thirteen cities by checking each of the 200 largest American cities for an Open Data file of arrest records and conducted additional searches that identified a few smaller jurisdictions with downloadable arrest records. A fourteenth city, Dallas, has released its arrest records but they have been purged of all arrests for murder and rape, making the Dallas statistics incomparable with those of the other thirteen.

Getting from the raw data in the downloadable databases to arrest rates required all the usual steps involved in cleaning databases plus a crucial additional one: identifying the arrests that qualify as arrests for index crimes under the UCR criteria.

The Chicago database of arrests had a variable that explicitly classified arrests as index crimes under the FBI definition. For the other twelve cities, I used variables that classified arrests according

to the legal definitions used in their jurisdiction. For most crimes, this was not a problem. Arrests for homicides in all thirteen cities discriminated between murder and various degrees of manslaughter. Arrests for serious assaults were usually described with the word *aggravated* or by other details that qualified the assault as aggravated (e.g., *ADW*, meaning assault with a deadly weapon). For databases that classified a crime by 1st degree, 2nd degree, or 3rd degree, I looked up the legal definition for that jurisdiction to determine which categories would qualify under the FBI definition.

The crime that proved hardest to classify confidently was theft. The official FBI definition does not specify a dollar value, implying that even the most minor thefts could qualify. The arrest databases almost always had subcategories of theft defined by the dollar value. I designated a given arrest as an index theft using a combination of minimum value (usually \$500) and police classification of the theft as a felony.

I am sure that my decision rules did not perfectly identify the set of arrests that each police department submitted to the UCR, but I am confident that the overlap is extremely high. Moreover, there is no reason to think that the misclassified arrests would tend to overestimate or underestimate arrests by race.

- 4. Fayetteville broke out Latins as a separate category, but the Fayetteville database also included the name of each arrestee. Upon examination, it was apparent that many people with Latin names had not been classified as Latin. I'm not referring to borderline names that might possibly be Latin, but rather to names such as Gonzalez or Gomez. I attempted to prepare corrected numbers of Latin arrests, but going solely by names is too inaccurate, so I decided not to report a Latin rate for Fayetteville. I have no idea what the correct number might be.
- 5. Stewart D'Alessio and Lisa Stolzenberg, "Race and the Probability of Arrest," *Social Forces* (2003), p. 1381. A little history will help set the context of the quotation.

One of the first systematic analyses of crime, based on delinquency among the entire birth cohort of males born in Philadelphia in 1945, found that the rate of contacts of juveniles with the police was 139.9 per 1,000 for non-Whites and 9.2 per 1,000 for Whites – a ratio of 15.2. See Marvin E. Wolfgang, Robert M. Figlio, and Thorsten Sellin, *Delinquency in a Birth Cohort* (1972). That was an unpopular finding in the 1970s, when the conventional

wisdom among sociologists was that race differences in crime were an artifact. Their position was supported by self-report studies of criminal behavior that found only minor race differences. See Jay R. Williams and Martin Gold, "From Delinquent Behavior to Official Delinquency," *Social Forces* (1972); and Martin Gold and David Reimer, "Changing Patterns of Delinquent Behavior among Americans 13 through 16 Years Old," *Crime and Delinquency Literature* (1975).

Then in 1978, the criminologist Michael Hindelang published a landmark study in which he compared arrest data to victimization surveys. His conclusion:

These [victimization] data for rape, robbery, and assault are generally consistent with official data on arrestees and support the differential involvement hypothesis. Some evidence of differential selection for criminal justice processing is found; however, most of the racial disproportionality in arrest data is shown by victimization survey data to be attributable to the substantially greater involvement of blacks in the common law personal crimes of rape, robbery, and assault. These results suggest that traditional admonitions against using arrest data as an index of involvement in these crimes may be overly cautious. (Hindelang, "Race and Involvement in Common Law Personal Crimes," *American Sociological Review* (1978), p. 93.)

Hindelang's findings guided criminologists for the next thirty years, with some 160 studies citing him to justify the use of arrest data as a surrogate measure of criminal offending. The D'Alessio and Stolzenberg study in 2003 took advantage of a new database, the National Incident-Based Reporting System, to reinvestigate the issue with a more rigorous research design than was possible with the data that Hindelang had had to work with. The authors used multivariate logistic regression to calculate the probability of arrest after taking several independent variables into account. The independent variables included (among others), the race, sex, and age of both the victim and the offender, whether the victim was injured, whether a deadly weapon was used, the relationship between the victim and the offender, and the location of the crime. The objective of the analysis was to test whether, given that the race of the offender was identified, Black citizens have a higher

NOTES TO CHAPTERS 4-5

probability of being subjected to arrest than White citizens. The quotation in the text continued with this conclusion: "These findings suggest that the disproportionately high arrest rate for black citizens is most likely attributable to differential involvement in reported crime rather than to racially biased law enforcement practices." That's where the science on this issue still stands as I write.

- 6. According to Africans who reported crimes, the African/European ratio of alleged perpetrators was 82.9 and the Latin/European ratio was 8.3. According to Latins who reported crimes, the African/European ratio of alleged perpetrators was 14.9 and the Latin/European ratio was 26.3.
- 7. The emphasis that police departments put on identifying "hot spots" plus the ease with which police can use a mapping app to get GPS coordinates has led many departments to include GPS coordinates for the arrest as part of the arrest record. These in turn can be reverse-geocoded to yield the zip code in which the arrest fell. I should add that care must be taken in curating these data apparently, arresting officers often record the GPS coordinates of the police station where the suspect was booked rather than the location of the crime, and local jails generate a disproportionate number of charges of inmates that are recorded as arrests. The reverse geocoding for my databases was done by Texas A&M Geoservices (https://geoservices.tamu.edu/).

CHAPTER 5: FIRST-ORDER EFFECTS OF RACE DIFFERENCES IN COGNITIVE ABILITY

 In the 2020 SAT Suite of Assessments Annual Report, downloadable at collegeboard.org, the table titled "SAT Participation and Performance: Score Distributions by Subgroup" gives the percentage of test takers who scored 1400–1600 by race and the total number of test takers, providing an estimate of the number of test takers in the 1400–1600 range by race. It is a very imprecise estimate for Africans because the College Board rounded percentages to the nearest whole point. The 1 percent reported for African students could be anywhere from a maximum of 1.49 percent to a minimum of 0.5 percent – 50 percent too small or 50 percent too high. I assumed 1.0 percent in making my calculations.

The College Board also reported the mean and SD for each race. Combining these two sets of information and applying the

mathematics of a normal distribution, it was possible to reach estimates for numbers of students with SAT scores of 1500 or higher.

2. The online documentation has more on this issue, but to give you a sense of how quickly the pool of approximately 900 African applicants with scores of 1500+ would be depleted, just six schools in the US News top 25 universities for 2020 – Harvard, Yale, Penn, Columbia, Duke, and Johns Hopkins – had about 950 African freshmen in 2019. As for the approximately 3,300 Latins with such scores, 19 of the 25 top-rated universities had 3,360 Latin freshmen. In all, the top 25 universities had about 3,650 African freshmen and 8,650 Latins. We can be sure that virtually every African or Latin applicant with a 1500+ score was admitted with a full financial ride if they applied to one of those schools, and that an extremely high proportion of them accepted.

To give you a sense of how many well-regarded schools are left with few or no African or Latin applicants with scores of even 1200 (about the 76th percentile for the 2020 SAT) or above, the numbers of African and Latin applicants with scores of 1200+ were about 20,900 and 68,300 respectively. The numbers of African and Latin freshmen in the top 100 *US News* universities were 22,440 and 56,881 respectively. In other words, given perfect topdown matches of university rank with African and Latin SAT scores, the top 100 universities could have soaked up more than all of the African students with scores of 1200+ and 83 percent of Latin applicants with scores of 1200+. More than two-thirds of all high school graduates who enter four-year colleges do not attend one of those top 100. (In 2019, 438,000 students entered the top 100 as first-year students while total first-year enrollment for all four-year colleges and universities was about 1.4 million.)

3. The evidence for the summary statements in the text is given below. Sources are provided in the online documentation.

Pass Rates for Bar Examinations. A large-scale study in the 1990s found that the pass rates for persons taking a bar exam for the first time were 92 percent for Europeans, 81 percent for Asians, 75 percent for Latins, and 61 percent for Africans. The only more recent data I have been able to find, for a 2020 administration of the California bar exam, found much lower pass rates of 52 percent for Europeans, 31 percent for Latins, and 5 percent for Africans.

Client Complaints about Attorneys. A 2019 study of complaints lodged with the California Bar Association among attorneys

admitted to the bar from the 1990s to 2009 found the following percentages of attorneys who had been the subject of ten or more formal complaints and who had been disciplined with temporary or permanent suspension of their license to practice law.

	California Attorneys			
	European	African	Latin	Asian
Attorneys with 10+ Complaints Lod	2.8% ged	7.2%	5.0%	2.0%
Attorneys Placed on Probation	0.6%	1.9%	1.2%	0.4%

The African/European ratios for 10+ complaints and suspended licenses were 2.5 and 3.0 respectively and the corresponding Latin/European ratios were 1.8 and 1.9. The African/Asian ratios for 10+ complaints and suspended licenses were 3.5 and 4.3 respectively and the corresponding Latin/Asian ratios were 2.5 and 2.8. Note that Europeans had more investigations than Asians. The European/Asian ratios for 10+ complaints and suspensions were 1.4 and 1.5 respectively.

Board Certifications for Medical Specialties. Board certification in a medical specialty is not associated with large differences in quality of care, but physicians have strong professional incentives to become board-certified. Some patients searching for a new physician use board certification as one of their criteria. Many hospitals and managed-care organizations require board certification. It's not a demanding standard – more than 80 percent of physicians in specialties are board-certified – but *not* getting certification despite the normal incentives to do so is a negative indicator.

A study of all U.S. medical school graduates from 1997 to 2000 followed them through eight years after graduation. The study reported the percentage of physicians practicing in eight specialties who had obtained board certification broken down by White, Asian/Pacific, and "underrepresented minorities," defined as Blacks, Hispanics, and Native Americans. Aggregating across all eight specialties, 11.1 percent of Whites and 12.1 percent of Asian/

Pacific physicians were not board-certified, compared to 21.9 percent of underrepresented minorities.

Patient Complaints about Physicians. The Medical Board of California is the state agency for licensing and regulating physicians. A study of complaints, investigations, and discipline from July 2003 to June 2013 found the following proportions by race:

	California Physicians			
	European	African	Latin	Asian
Physicians with Complaints	28.0%	43.0%	36.5%	24.7%
Physicians with Investigations	6.7%	11.7%	9.7%	5.3%
Physicians Disciplined	1.0%	1.5%	1.9%	0.8%

These three indicators are roughly analogous to reports of crimes, arrests, and sentences in Chapter 4, with investigations probably being a better indicator than complaints or discipline for the same reason that arrests are a better indicator than reported crimes or sentences. Reported complaints and investigations follow the familiar ordering high to low – Africans, Latins, Europeans, and Asians. The African/European ratio was 1.7 and the Latin/European ratio was 1.4. Europeans had more investigations than Asians. The European/Asian ratio was 1.3.

Pass Rates on the Certified Public Accountant (CPA) Examination. The CPA exam is a sixteen-hour test administered in four separate sections. Passing the CPA exam is not required to work as an accountant, but is necessary for anyone who hopes to rise in the field. Most people take the exam from their late 20s through early 30s. The table below shows three results for CPA candidates from 2005 to 2016.

The European/African pass ratio was 2.3. The European/Latin ratio was 1.5. For the two types of dropout – those who did not reattempt the exam after failing and those who stopped after the first of the four sections – the African/European ratios were 2.5

CPA Candidates, 2005-2016

	European	African	Latin
Passed all four sections	47.7%	20.2%	31.9%
Dropped after first attempt	7.7%	19.2%	15.0%
Dropped after first section	2.0%	4.6%	3.8%

and 2.4 respectively. The corresponding Latin/European ratios were 1.9 and 2.0.

High-Stakes Ratings of K-12 Teachers. Michigan enacted a law mandating high-stakes ratings of teachers in 2011. From 2011 to 2015, 2.2 percent of European teachers were rated "minimally effective" or "ineffective" compared to 4.6 percent of Latin teachers and 7.5 percent of African teachers. This amounts to an African/ European ratio of 3.4 and a Latin/European ratio of 2.1 The same pattern applied to teachers who received multiple ratings of "minimally effective" or "ineffective" over the five-year period. The African/European ratio was 3.2 and the Latin/European ratio was again 2.1.