PAPERS

Racial "Box-Checking" and the Administrative State

DAVID E. BERNSTEIN*

Americans have grown accustomed to checking ethnic and racial boxes when applying for colleges, requesting a mortgage, filling out medical paperwork, and more. "Where do these boxes come from?," Justice Gorsuch asked in his concurring opinion in *Students for Fair Admissions*. "Bureaucrats. A federal interagency commission devised this scheme of classifications in the 1970s to facilitate data collection."[1] Indeed, the classifications reflected in those boxes are the product of an obscure bureaucratic process that reflected "a combination of amateur anthropology and sociology, interest group lobbying, incompetence, inertia, lack of public oversight, and happenstance."[2]

The federal Office and Management and Budget enacted Statistical Directive No. 15 in 1978 to create uniform racial and ethnic classifications so that data could be efficiently shared and compared across federal agencies.[3] The relevant classifications decided upon were American Indian, Asian or Pacific Islander, Black, Hispanic–the only ethnic, not racial classification–and White. Each classification came with an official, somewhat arbitrary definition.[4]

For example, "Hispanic" was defined as "of Spanish origin or culture," thus excluding Brazilians but including Spanish Americans.[5] South Asian Americans,

---

* Distinguished University Professor, Antonin Scalia Law School, George Mason University.

1. Students for Fair Admissions, Inc. v. President & Fellows of Harvard Coll., 600 U.S. 181, 291 (2023) (Gorsuch, J., concurring).

2. DAVID E. BERNSTEIN, CLASSIFIED: THE UNTOLD STORY OF RACIAL CLASSIFICATION IN AMERICA xi (2022).

3. Directives for the Conduct of Federal Statistical Activities, Directive No. 15, Race and Ethnic Standards for Federal Statistics and Administrative Reporting, 43 Fed. Reg. 19269 (May 4, 1978).

4. The categories, at *id*., were defined as follows:

> *American Indian or Alaskan Native*. A person having origins in any of the original peoples of North America and who maintains cultural identification through tribal affiliation or community recognition.
>
> *Asian or Pacific Islander*. A person having origins in any of the original peoples of the Far East, Southeast Asia, the Indian subcontinent, or the Pacific Islands. This area includes, for example, China, India, Japan, Korea, the Philippine Islands, and Samoa.
>
> *Black*. A person having origins in any of the black racial groups of Africa.
>
> *Hispanic*. A person of Mexican, Puerto Rican, Cuban, Central or South American or other Spanish culture or origin, regardless of race.
>
> *White*. A person having origins in any of the original peoples of Europe, North Africa, or the Middle East . . . .

5. *Id.*

like their "Asian" Middle Eastern counterparts, were originally slated to be in the White category. A last-minute lobbying campaign by a small Indian American organization resulted in South Asians being classified as Asian Americans.[6]

At the time, the classifications received very little public attention. No one seemed to anticipate the vast influence the classifications would come to have. OMB explicitly warned that the classifications were not to be used to determine eligibility for any government program, nor did they purport to be scientific or anthropological in nature.[7]

Those warnings and caveats have been ignored. The result has been that these classifications have had a profound effect on American life, especially on how Americans identify themselves and others. Identities that barely existed in 1977, such as "Hispanic" and "Asian American," are now mainstream. Identities not recognized by Directive 15, such as "Italian American" or "Chicano," have fallen into disuse.

The social influence of government racial classifications is constantly reinforced by public discussion of academic and other studies that rely on the Directive 15 classification scheme. Some uses of these classifications, such as by pollsters, have arisen as matter of custom. In many other situations researchers have little choice but to rely on government-collected data to do their research, because it would be too expensive to collect their own. Government-collected data, in turn, relies on the Directive 15 classifications. So, for example, if a researcher wants to undertake research on group educational achievement in the U.S., he will almost inevitably rely on Department of Education data, which is broken down by Directive 15 classification.

Imagine, for example, someone is researching educational achievement in Florida. Department of Education data for Cuban, Puerto Rican, Mexican and Venezuelan Americans in Florida is combined into the Hispanic classification, even though each of these groups has distinct cultural, economic, and other attributes, along with their own internal diversity.

A particularly diligent researcher may try to break up the data by subgroup by investigating which subgroups populate which schools, and then combine the individual school data to get a rough estimate of each subgroup's achievement. Most researchers, however, will not have the resources to do so, and in any event that strategy won't work for other statewide data, such as health. Similar problems would arise if trying to differentiate between data for "white" Italian and Jewish Americans, or Indian, Korean, Japanese, Vietnamese, and Chinese Americans in California, all classified as "Asian American."

A related problem is the changing internal demographics of the Directive 15 groups. Consider a researcher studying the socioeconomic status of the African American community. In comparing, say, 1970 to 2020 census data, one is

6. BERNSTEIN, *supra* note 2, at 90–91.

7. Race and Ethnic Standards for Federal Statistics and Administrative Reporting, 43 Fed. Reg. at 19269. 19260.

comparing apples to oranges. In 1970, very few African Americans were recent immigrants. Now, over 20% of the African American population is composed of first- and second-generation immigrants. In the late 1980s, the Census Bureau decided by fiat, without any public debate, that Caribbean and African black immigrants should inhabit the same classification as American descendants of slaves (ADOS) and that soon became the norm.[8]

Meanwhile, over time many more Americans the government classifies as African American/Black ("descended from one of the black races of Africa," according to Directive 15) are the product of recent interracial relationships. Between immigrants and the progeny of interracial couples, the data is now skewed toward showing more socioeconomic progress; the increasing population of black immigrants and children of interracial marriages have significantly higher educational achievement and incomes on average than do ADOS.[9]

Thanks to Directive 15, researchers are, essentially, combining the data for the Kamala Harrises and Barack Obamas of the world, children of immigrants and interracial marriage, with the data for individuals like Clarence Thomas who have two ADOS parents and whose ancestors have been in the U.S. for hundreds of years. Absent government statistical policies, researchers would be much less likely to combine such disparate groups.

The Directive 15 classifications also quickly became affirmative action categories. In particular, the Directive 15 classifications are universally used in higher education, and, with occasional slight modifications, in government contracting for eligibility for Minority Business Enterprise preferences.

Affirmative action programs were originally created primarily to bring ADOS into the economic and educational mainstream, with other groups sometimes thrown in as an afterthought. But thanks to the Directive 15 minority classification, and large-scale immigration from Asia and Latin America since the 1970s, most of those eligible for minority business enterprise preferences today are post-1965 Hispanic or Asian American immigrants and their descendants. And given the broad way the classifications are defined, a self-described Hispanic, Native American, or Asian American applicant may be someone who has only a distant ancestor from that group.

To take an extreme example, the Small Business Administration determined that a Sephardic Jew was entitled to Hispanic status and thus minority business enterprise preferences because his ancestors had fled Spain centuries ago, though he otherwise had no discernable ties to Hispanic culture or the Spanish language.[10] In the educational realm, first- and second- generation immigrants and bi-racial students increasingly dominate the black student population at elite schools, meaning they get most of the benefit of affirmative action preferences

8. BERNSTEIN, *supra* note 2, at 91.

9. *See generally* KEVIN BROWN, BECAUSE OF OUR SUCCESS: THE CHANGING RACIAL AND ETHNIC ANCESTRY OF BLACKS ON AFFIRMATIVE ACTION (2014).

10. Rothschild-Lynn Legal & Fin. Servs., SBA No. MSBE-94-10-13-4 (Apr. 12, 1995).

for black students originally intended for ADOS. A 2007 study concluded suggests that over 40% of the black students at Ivy League colleges were first- or second-generation immigrants, compared to 13% of the African American population at the time.[11]

Directive 15 has also had profound (and negative) effects on biomedical research. In the late 1990s, Congress mandated that the National Institutes of Health and the FDA require biomedical researchers to recruit subjects and analyze data by race.[12] Rather than undertake a process to determine what sort of classifications would make sense in the context of biomedical research, the agencies simply required researchers to use the Directive 15 classifications.[13] This means that the Directive 15 classifications are now, in practice, used as pseudo-scientific racial categories in biomedical research. This policy has encouraged a pernicious racial essentialism in scientific research and medical practice that has no scientific grounding.

The policy of using Directive 15 classifications as a "weak surrogate" for genetic lineage and environmental influences[14] has also led to a huge waste of research resources. Use of the Directive 15 classifications means that biomedical researchers are mixing individuals with little genetic commonality into a single research classification. For example, researchers can satisfy the requirement to use "Asian American" subjects by recruiting Chinese Americans living near their research lab. But the data acquired is of no plausible scientific or sociological relevance for other groups in the Asian American classification, such as Filipino or Pakistani Americans. Filipinos, Pakistanis, and Chinese-descended individuals on average have no more in common genetically than any two random individuals from around the world. Indeed, the data likely would not even be useful for Chinese Americans, because it will be reported in the study as "Asian Americans," and other researchers will not know which subgroup(s) of Asian Americans participated in the study.

Even the "African American" classification for medical research subjects provides data of limited, at best, value. This problem is acute even for the ADOS population. The average black American is approximately 73% African by genetic origin;[15] self-identified African Americans range from zero to 100% African heritage. African DNA is itself extremely diverse. So even if the

---

11. Douglas S. Massey, et al., *Black Immigrants and Black Natives Attending Selective Colleges and Universities in the United States*, 113 AM. J. EDUC. 243 (2007).

12. *See* BERNSTEIN, *supra* note 2, at 141.

13. *Id.*

14. Francis S. Collins, *What We Do and Don't Know About 'Race', 'Ethnicity', Genetics and Health at the Dawn of the Genome Era*, 36 NATURE GENETICS S15 (2004).

15. Lizzie Wade, *Genetic Study Reveals Surprising Ancestry of Many Americans: Some African-Americans, European Americans, and Latinos Carry Genes That Don't Match Their Self-identified Ethnicities*, SCIENCE, (Dec. 18, 2014), https://www.science.org/content/article/genetic-study-reveals-surprising-ancestry-many-americans#:~:text=The%20average%20African%2DAmerican%20genome%2C%20for%20example%2C%20is%2073.2,American%20Journal%20of%20Human%20Genetics [https://perma.cc/H5C4-5J7W].

Directive 15 classifications were restricted to ADOS, any data collected would be of dubious salience to any given African American patient.

Making matters even more complex, however, is that the African American data also includes recent voluntary immigrants from Africa and their descendants. Somali and Ethiopian Americans, for example, are often more closely related genetically to Arabs than to black Africans.[16]

The result is garbage in, garbage out. A researcher could satisfy the FDA's requirement to have African Americans represented in it research subject pool, for example, by attracting subjects from the Somali American community in Minneapolis. The researcher would then have the required data on "African Americans." But the data wouldn't tell you anything about the bulk of the African American population, and indeed, to the extent it leads to any medical conclusions applied to the general African American population, could be harmful.

One can point out related problems with the white, Hispanic, and Native American classifications. Each of them takes very internally diverse groups, genetically, sociologically, culturally, and otherwise, and shoehorns them into an arbitrary classification that has no scientific basis, and yet are used by government fiat in scientific research.

Given the points raised above, it's time for OMB to completely overhaul the classifications. In considering what to do, OMB must keep in mind that it isn't simply creating uniform classifications for government data collection, but rather is dictating, among other things, how social science researchers study the American population, who is eligible for affirmative action and other forms of race-conscious preferences, and how scientific research is undertaken.

A new rule incporporating the Hispanic ethnic classification into the existing racial classification and creating a new Middle Eastern and North African (MENA) classification,[17] does not properly or adequately address these problems. Indeed, it may make them worse.

Hispanic is not even a coherent ethnic classification; conflating it with racial classifications further confuses matter. The "Hispanic or Latino" classification includes everyone from individuals whose ancestors were Indigenous Mexicans to children of Spanish movie stars to Chinese Peruvians to Panamanians of African descent, and any combination of these ethnic origins—at least if they move to the US. Hispanic makes even less sense as a racial classification, for the

16. DOROTHY E. ROBERTS, FATAL INVENTION: HOW SCIENCE, POLITICS, AND BIG BUSINESS RECREATE RACE IN THE TWENTY-FIRST CENTURY 159 (2012).

17. Office of Information and Regulatory Affairs, Office of Management and Budget, Executive Office of the President, Notice of Decision, Revisions to OMB's Statistical Policy Directive No. 15: Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity, March 29, 2024, https://www.federalregister.gov/documents/2024/03/29/2024-06469/revisions-to-ombs-statistical-policy-directive-no-15-standards-for-maintaining-collecting-and#:~:text=OMB%20accepts%20the%20recommendation%20to,a%20separate%20MENA%20minimum%20category [https://perma.cc/8DKB-NDVE].

obvious reason that "Hispanic" tells you no more about someone's "racial" origins than "American" does.

Meanwhile, while I support the notion of the government collecting more granular data about ethnic groups, the MENA classification, originally proposed primarily as a way to get data on the Arab American population, will just make matters more confusing. There are approximately 3.5 million Arab Americans in the U.S. About half are descended from early 20th century immigrants, mostly from Lebanon and predominately Christians. The other half are primarily Muslim immigrants who have arrived since the 1970s. Lumping these groups together as "Arab Americans" obscures more than it illuminates. Making matters much worse, the classification also includes approximately 500,000 Israeli immigrants and their descendants, about 300,000 other Jews whose families immigrated to the US from MENA, 600,000 Iranian Americans, and 500,000 Chaldean Americans.

Regardless of official definition, one can also expect many Armenian Americans, Afghan Americans, and others to check the MENA box on forms rather than or in addition to white. One can also expect some percentage of Ashkenazic Jews to check the MENA box, especially given the current vogue of many Jews insisting that they are not white, and DNA evidence showing that "European Jews" have significant Middle Eastern genetic heritage.

In short, MENA is not a coherent classification, any more than is the white classification it has been broken off from. Rather than reducing the arbitrary nature of the current classification scheme, adopting a MENA classification will just add an additional layer of arbitrariness.

Instead of playing around with the existing categories, the classifications need to be fundamentally reconsidered. First, OMB should review every instance in which the government uses, or requires private parties to use, racial classifications. Consistent with Supreme Court precedent on racial classification, government should get out of the racial classification business unless there is a "compelling interest" in using such classifications. The relevant language should be changed from the caution that the classification should not be used for eligibility for government programs to something like "without a detailed justification that goes through the notice and comment process, these classifications MAY NOT be used by government agencies for regulating scientific research or for eligibility for government programs."

Second, to the extent that racial classifications are justified by a compelling government interest, use of the classifications should be narrowly tailored to that interest. For example, as noted previously minority business enterprise preferences were created in the 1970s primarily to assist ADOS who had been excluded from government contracts for generations. Yet today most of these preferences go to post-1965 immigrants their descendants because they fit one of the Directive 15 classifications.

More generally, OMB needs to recognize that these classifications are used for distinct purposes, and tailor the classifications for those purposes. To the extent

the Directive 15 classifications continue to be used, it would makes sense in many instances to break up the broad, crude classifications into more meaningful sub-categories.

One study, for example, shows vast differences in health outcomes for residents of different tribal reservations in just one state, Arizona. Health data that uses the classification of "American Indians" writ large—many of whom are in any event of mixed race and assimilated into the general white population–obscures such findings. Similarly, Appalachian whites and Malaysian Americans have far worse socioeconomic indicia on average than do whites or Asians in general, respectively. But the statistics for the groups overall obscure the problems of the subgroups within them. The Census Bureau has been gradually moving to more granular data, but other government agencies have not. And even the Census Bureau fails to distinguish between the ADOS and the rest of the black population.

One model to look at is the FBI's hate crime statistics. Instead of simply copying the Directive 15 categories, the agency tabulates hate crimes against twenty-nine different identifiable groups, including Mormons and gender-nonconforming people.

For affirmative action purposes, to the extent such programs continue, the focus should logically be on ADOS and residents of Indian reservations, the two groups that have by far suffered the most from state and private violence in the US.

For scientific purposes, Directive 15 racial classifications have no scientific basis should be ignored entirely. Instead, the government should be encouraging the use of scientifically relevant genetic data.

The government should also consider whether the definitions are overbroad, in that they allow people with only distant ancestry, such as one great-grandfather, to claim membership in a group. For most statistical, scientific, or affirmative action purposes, having people with distant minority ancestry list their race by that ancestry acts at cross-purposes with the purposes of collecting that information.

As a concluding thought, we should consider whether having any sort of official racial classification system in the US is something we want to plan to retain in the long run. Many law professors and other academics, particularly those writing from a Critical Race Theory perspective, start with the presumption that racial division is a permanent part of the American landscape, and therefore if we want to achieve justice, we must permanently divide the population by race and make sure each group gets its share.

In fact, though, all sorts of ethnic and religious conflicts once prominent in American life have faded into distant memory. In the 1920s, perhaps the most powerful political movement in the US was the Ku Klux Klan, which focused on hostility to Catholics. As of the 1960 presidential election, it was unclear that Protestant Americans would vote for a Catholic for president. And indeed, many Protestant Democrats voted for Nixon for that reason, and JFK was saved only

because many otherwise Republican Catholics voted for him as a matter of identity politics.

Yet, sixty years later, in January 2021, the US had a Catholic president, a Catholic speaker of the House, six Catholic (and two Jewish) Justices on the Supreme Court., a Jewish Senate majority leader, and a multiracial, black-identified Vice-President with not significant controversy. This would have been unimaginable in 1960. Trends in racial tolerance and intermarriage suggest that a a similar outcome with regard to race, with a common, non-racial or multi-racial American identity coming to the fore, is a real possibility in the future. The government should be careful not to stand in the way.