A Constructivist Perspective on Empirical Discrimination Research

Evan K. Rose[*]

January 2022

Abstract

Contemporary scholars view race as a constructed social category, not a biological fact. Yet most empirical discrimination research treats race no differently than other individual characteristics typically observed in data. This article considers the implications of adopting a constructivist perspective instead. I develop a simple model where agents use observable characteristics to both interpret membership in racial social categories and make decisions. Discrimination is the result of acting based on perceived social identity. The model highlights the need to measure the racial "first stage"—the social identity contrast between individuals—instead of relying on race as coded in data, and draws a novel distinction between race-based and direct statistical discrimination. I illustrate some implications using data on wages, speech patterns, and skin color and conclude with strategies for future research that build on the constructivist model.

An enormous empirical literature studies the influence of race in labor markets, the criminal justice system, health care, and other settings. Virtually all of this work takes racial categories as facts about people—individuals are Black, white, Asian, and so on as much as they might be immigrants, have blue or brown eyes, or have completed high school. A parallel literature in sociology and anthropology questions this approach, asking, What is meant by a variable B_i that takes a value of one if individual i is "Black"? Does B_i measure melanin levels above a threshold or other phenotypic information? If not determined by physical traits, does B_i refer to i's own ethnoracial identity claims or the category assigned by another actor? Is B_i fixed, or might it change across contexts and as other individual traits evolve? And what, if at all, does it matter for quantitative research on racial discrimination?

For decades, scholars across the humanities and social sciences have regarded race as, to quote Paul Holland, "a socially determined construction with complex biological associations" (Holland, 2008). No surjective map associates individual ancestry or phenotypic and biological characteristics, such as skin tone or hair texture, with a consistent racial taxonomy. Who is considered white, Black, Native American, or Asian in one time or place may not be in another, or even in the same time and place but wearing different clothing. The constructivist argues that race exists not as a natural but a social category forged over hundreds of years of political and historical processes. As a result, while individuals may *observe* others' physical traits, they *interpret* race; race in data and economic models therefore reflects both physical facts about people and the potentially non-neutral mental models people use to digest those facts.

These ideas are not simply the abstract concerns of progressive scholars in the humanities. Substantial quantitative research demonstrates that interpretations of race are both fluid and contextually dependent. In a notable experiment, for example, Freeman et al. (2011) demonstrate that identical faces are more likely to be categorized as white when wearing a suit and tie instead of janitorial overalls. The boundaries of racial categories, as well as who belongs where, are constantly contested in public and political discourse. President Barack Obama is famously the frequent subject of diverging racial perceptions. A 2010 Pew Research poll found that more than half of respondents who identified as Black saw Obama as Black, but less than a quarter of white respondents did the same even though Obama himself has stated "I identify as African-American—that's how I'm treated and that's how I'm viewed. I'm proud of it" (Reynolds, 2007).[1]

Constructivism is also central to a rich literature in sociology and economics that explores the emergence and consequences of social identities (Akerlof and Kranton, 2000; Darity, Mason and Stewart, 2006; Akerlof and Kranton, 2010). A non-essentialist per-

---

[1]During the 2012 presidential election, Fox News contributor Laura Ingraham pondered whether Herman Cain, if elected, would be the "first Black president" because he does not "have a white mother, white father, grandparents" (Somanader, 2011). Cain, for his part, claimed Obama "was raised in Kenya" and was thus more of an "international" (Reeve, 2011).

spective on race is central to this strand of research. The core idea, in fact, is that racial differences emerge *endogenously* as an outcome of stereotypes that, in equilibrium, engender real disparities across groups and reinforce perceptions of racial difference (Coate and Loury, 1993). Acknowledging the plasticity of racial cues is a natural sequitur. When race reflects the social significance of potentially manipulable signals, individuals face incentives to "pass" or "assimilate" across groups if the reputational benefits exceed the costs (Austen-Smith and Fryer, 2005; Eguia, 2017; Kim and Loury, 2019). The fact that passing or partial passing is possible points to the fragility of an essentialist notion of racial identity in the first place.

This paper considers the import of constructivism for empirical research on racial discrimination. Rather than taking race as a fixed characteristic—one of potentially many essential X_i attributable to each individual—I present a simple model where agents use observed physical facts about people and contexts to make a judgment about racial social identities. These judgments reflect the myriad ways physical and contextual cues influence perceived membership in racial social categories that the decision maker inherits from long-standing political, social, and historical processes (Goffman, 1963; Tajfel, 1974; Loury, 2002; Emirbayer and Desmond, 2021). In line with recent research, judgments are not dichotomous; individuals with the same observables may present as "more white" or "more Black" depending on contextual factors. Ostensibly non-racial characteristics, such as incarceration history and attire, may influence how race is perceived as well. In this model, race is therefore neither a simple binary variable nor a composite of individual traits.[2] Instead, it is in the eye of the beholder.

The agent's utility may depend on perceived race directly due to prejudice or because the agent "statistically discriminates" about decision-relevant unobservables on the basis of racial social categories. I call either case discrimination. Utility may also depend on underlying individual characteristics directly or for what they signal about unobservable traits. Testing when differential decisions happen "because of" race thus faces an obvious and immediate identification problem. When the same X_i that determine perceived race enter utility directly, it is not possible to separate direct effects from discrimination. To do so, the empiricist requires either racial instruments—factors that influence perceived race but are excludable from utility—or strategies that compare individuals whose combinations of individual characteristics, absent differences in perceived race, ought to generate the same utility.

Some characteristics may be obviously excludable from utility on normative grounds. It would be unreasonable, for example, to argue that direct preferences over skin color reflect anything other than racial discrimination. Ceteris paribus comparisons of individuals with diverging skin tones is rarely physically possible, however, and the set of other potential racial cues that are obviously excludable quickly becomes controversial. While

---

[2]That is, a "bundle of sticks" as suggested by Sen and Wasow (2016).

some scholars view the differential treatment of trained and matched Black and white testers sent to negotiate over automobile purchase as "decisive" evidence of discrimination, others argue that it is "unlikely that all characteristics that might affect [utility] will be perfectly matched" (Arrow, 1998; Heckman, 1998). Behind this critique lies an implicit definition of discrimination as the ceteris paribus "treatment effect" of race, subjecting discrimination to the same rigor as a pharmaceutical in a randomized controlled trial. In the constructivist perspective, however, there is no hypothetical experiment that could measure the treatment effect of race because race is not an isolable characteristic. Instead, manipulating race as a symbolic category requires varying at least one of the ceteris (Greiner and Rubin, 2011; Sen and Wasow, 2016; Kohler-Hausmann, 2018).[3]

The difficulty of parsing racial from non-racial characteristics in the constructivist framework presents an important measurement error challenge. Consider the long-standing empirical tradition of "kitchen-sink" benchmarking regressions and Oaxaca-Blinder decompositions, which attempt to compare outcomes across racial groups on an equal footing (Blinder, 1973; Corcoran and Duncan, 1979; Oaxaca and Ransom, 1994; Darity Jr, Guilkey and Winfrey, 1996). Race disparities that survive controls for a large set of observable factors are thought to be more reliable measures of discrimination.[4] Without further restrictions, however, such controls may both eliminate potential confounders and attenuate the "first-stage" effect of coded race on perceived racial differences, leaving it unclear how to interpret gaps that diminish as more controls are added. Put simply, the social identity contrast between an Emily and Latisha who are both summa cum laude graduates of the Harvard mathematics department may be significantly weaker than the unconditional contrast. From the constructivist perspective, benchmarking thus runs the risk of overcontrol even in cases where the goal is to measure disparate treatment.[5]

This measurement problem is straightforward to demonstrate using data on wages, speech patterns, and skin color from the 1997 cohort of the National Longitudinal Survey of Youth (NSLY97). Consistent with the results in Neal and Johnson (1996), the large unconditional wage gap between Black and white respondents attenuates substantially after controlling for education, geography, and scores on aptitude tests. These same controls, however, also reduce between-group differences in other important racial cues, including skin shade and how likely one's speech is to be categorized as that of a Black speaker.[6] If the goal is to measure the possible extent of labor market discrimination based on social identity, augmented Mincer (1974)-style controls clearly adjust for wage-

---

[3]Disagreements persist today about whether modern audit and correspondence studies capture discrimination or employers' direct preferences over names (Mullainathan, 2002; Fryer and Levitt, 2004; Gaddis, 2017; Kline, Rose and Walters, 2021).

[4]In Rose (2020), for example, I examine the sensitivity of racial disparities in technical probation violations to criminal history, demographic, and standardized test score controls.

[5]Ayres (2010) argues that "included-variable" bias is a concern for statistical tests of disparate impact but not disparate treatment.

[6]The speech data were created and studied originally in Grogger (2019).

relevant skills and adjust the implicit social contrasts across workers. Indeed, if distinctive speech is taken as the relevant determinant of social identity, then the reduction in the Black-white wage gap measured in Neal and Johnson (1996) would be 40% smaller.

One way out of the benchmarking morass is to take an explicit stand on what *is* decision relevant, thereby reducing the dimension of the problem. In Becker (1957)'s original analysis of labor market discrimination, for example, workers' marginal product fully characterizes their value to firms. Discrimination occurs when an employer acts as if the equally productive Black worker commands a higher wage. The researcher can therefore test for discrimination by finding sets of characteristics that yield diverging perceptions of race but hold utility fixed. Defining the set of decision-relevant factors, however, typically requires normative claims on what *ought* to matter to decision makers absent discrimination. Making these claims can be controversial. Moreover, decision makers do not typically directly observe latent traits such as productivity. They must instead make guesses about them using the characteristics they do see. Comparing two individuals with equal *expected* productivity to the agent poses further challenges, as I discuss below.

Finally, the constructivist model highlights an inherent tension in the classic distinction between "taste-based" and statistical motivations for discrimination. Statistical reasoning about decision-relevant unobservables based on perceived race necessarily reflects a coarsening of underlying characteristics—many people are inferentially grouped together as "Black" or "white" despite their underlying differences (Mullainathan, 2002; Fryer and Jackson, 2008; Bordalo et al., 2016). Yet these groupings are not necessarily neutrally or exogenously constructed. To the extent that who is seen as Black or white reinforces in part the social meaning of race, group-based reasoning at the expense of within-group heterogeneity is difficult to cast as a purely statistical exercise and may no longer be the efficient solution to a decision problem under uncertainty.

Studying discrimination empirically therefore requires institutional or normative restrictions on what information affects preferences and racial perceptions and how. Many disagreements about whether and when discrimination has been reliably identified empirically, the nature of its motivations, and what to do about it stem from disagreements about what such restrictions are reasonable. Since many battles over discrimination occur in courtrooms, one might expect the law to offer sharper guidance on how to define and measure race and discrimination. As I detail below, however, the same difficulties that challenge empirical discrimination research reappear in legal contexts. Though decades of legal contests have produced multiple theories of discrimination and evidentiary standards, in many cases what it means to say legally that an action was illegally taken "because of" race remains unclear.

These arguments should not be construed as claims that race and racial discrimination are not pressing and real social problems. The fact that social identity is constructed

does not make its impacts on individuals' and society's well-being any less sharply felt. And while one can imagine a universe in which the social meaning of race were radically different, we live, for better or (more likely) for worse, in this one, where historical and social processes have shaped social identities in particular ways. Nor is the objective here to argue that individuals' internal notions of identity are unimportant or wrong in any sense. Rather, my objective here is to encourage empirical researchers to think critically about what their data tell us about how members of different racial groups are treated and also how membership in these groups is constructed and perceived.

In what follows, I begin with a review of theoretical and empirical research on racial identity. The goal here is not to be comprehensive. Instead, I focus on a subset of key ideas and results that underscore the quantitative importance of constructivist ideas. I then discuss how legal notions of discrimination, from which recent empirical research draws much inspiration, grapple with the meaning of race. Finally, I sketch the model of discrimination introduced above and discuss its implications for applied research. I conclude with some suggested solutions and directions for future work.

1   What is race?

Two broad theories have dominated scholarship on race, each with its own important implications for empirical research on discrimination.[7] The first, often dubbed "essentialism," views race as a primarily biological taxonomy of people: one's race is a fixed set of natural characteristics determined primarily by ancestry. These views have strong ties to the history of slavery and colonialism, underlie many arguments for racial supremacy, and were of deep interest to eugenicists. One can see clear traces of essentialism in American racial classifications, such as a Louisiana law passed in 1970 that decreed anyone with at least one thirty-second or more "Negro blood" was legally "Black."[8]

Yet Louisiana's law also highlights an important challenge to essentialist ideas: borders between seemingly "natural" racial categories are fundamentally arbitrary. Why draw the line at one-thirty second? What, exactly, is "Negro blood"? The second prominent theory—constructivism—argues that such confusion arises because no natural and essential racial taxonomy exists. Racial classifications such as "Black" and "white" in

---

[7]I provide an extremely brief overview of these complex topics here. For more thorough introductions, see, for example, the discussions in Zuberi (2001), Bonilla-Silva (1999), Sen and Wasow (2016), Emirbayer and Desmond (2021), and Kohler-Hausmann (2018). Sewell (2021) provides an overview of empirical issues in discrimination research in sociology.

[8]Desmond and Emirbayer (2009) relate this story about Louisiana's law: "In 1982, Susie Guillory Phipps sued Louisiana for the right to be White. She lost. The state genealogist discovered that Phipps was the great-great-great-great-grandchild of a White Alabama plantation owner and his Black mistress and, therefore—although all of Phipps's other ancestors were White—she was to be considered 'Black.'" *United States* v. *Bhagat Singh Thind*, 261 U.S. 204 (1923) provides an earlier example of legal ambiguity over the definition of race. Thind sued for the right to claim naturalized citizenship, which was restricted to "free white persons," under the claim the he was "Aryan." The court rejected his claim.

the U.S. represent social classifications forged by long-standing social and political processes intimately tied to the history of chattel slavery, the legacy of Jim Crow, the War on Drugs, the rise of mass incarceration, and persistent socio-economic divides. Through these processes, certain physical and contextual markers—particularly, but not solely, the color of one's skin—became markers of social difference and social identity. Race, in the constructivist view, refers to these social categories and their attendant social meanings, not to one's genetics or ancestry directly.

While constructivist views are the product of relatively modern qualitative and theoretical scholarship, growing quantitative empirical evidence calls for more nuance than allowed by treating race as an unambiguous demographic category. For one, simply assigning basic racial labels to people is non-trivial. For example, Liebler et al. (2017) find that 6% of people changed race and Hispanic-origin responses between the 2000 and 2010 Decennial censuses, including 6% of people reported as non-Hispanic Black in 2000. Innumerable other examples illuminate how cleanly defining where the boundaries of racial categories fall and who belongs where is not a simple exercise (Kennedy, 2012; Davenport, 2016). Between 2000 and 2020, for example, the number of people who identified as Native American nearly doubled, according to census figures.

Some may object that much purported racial fluidity is concentrated among people from multi-racial families or with Hispanic origins who appear racially "ambiguous." Yet racial labels appear to be not only flexible generally but also responsive to simple social and contextual factors, pointing to a deeper fluidity in how race is interpreted and perceived. Studying the 1979 National Longitudinal Survey of Youth cohort, Saperstein and Penner (2012) report that changes in both self-reported and interviewer-assigned race are associated with job loss, marriage, incarceration, and other major life events.[9] Charles and Guryan (2011) report similar fluidity and sensitivity to earnings and education in the Current Population Survey, while Cornwell, Rivera and Schmutte (2017) show that changes in racial classifications in Brazil are associated with wage changes. In yet another context, Blouin and Mukand (2019) find that the salience of ethnic identity is sensitive to government propaganda.

Thus even in cases where demographic racial labels remained unchanged, social cues likely affect perceptions of how *cleanly* an individual belongs to particular social categories. Gaddis (2017) makes this point clearly when studying perceptions of names more commonly given to children of Black mothers. Children named Bria and Tamika are equally likely to be recorded as Black on birth records (roughly 80%), yet mothers of Bria are nearly four times more likely to have some college education. Less than 30% of Gaddis' survey respondents associated being Black with the name Bria, however, while

---

[9]Alba, Insolera and Lindeman (2016) argue that much of the racial fluidity documented in Saperstein and Penner (2012) is concentrated among respondents with mixed racial family backgrounds. Regardless, shifting racial identify seems to be a real feature of the NSLY data.

close to 90% did so for Tamika.[10] The results strongly suggest that "Blackness" for many respondents is more closely associated with social class than straightforward demographic measures, a pattern noted in many analyses of the intertwined roles race and class play in contemporary inter-group relations (Jaynes, 2004).

Related work shows that physical markers—and skin tone in particular—play an important role in determining social identity among those widely viewed as members of the same race (Monk Jr, 2021). Among those who identify as Black, dark skin is more closely associated with low social status and negative stereotypes about ability and criminality, a phenomenon known as colorism. A pointed example comes from popular media. When *Time Magazine* reported on the "American Tragedy" of O.J. Simpson's murder trial, editors darkened Simpson's skin in the mug shot splashed across the magazine's cover. Rarely is the feedback loop from social stigma to perceived race so plainly documented.[11]

The phenomenon of "passing" provides yet another instructive window into the constructed nature of racial categories. Passing is the act of signaling with physical and contextual cues—by, for example, whistling Vivaldi while walking the streets at night—that an outside observer should not apply the naive social classification they might otherwise. Passing recognizes that race is, in part, a performance: "You are not Black because you are (in essence) Black; you are Black... because of how you act... because of how you juggle and combine many differently racialized and class(ed) actions (walking, talking, laughing, watching a movie, standing, emoting, partying) in an everyday matrix of performative possibilities" (Jackson Jr, 2010). Naturally, when social categorization is manipulable, members of stigmatized groups face incentives to present otherwise when the benefits outweigh the costs (Kim and Loury, 2019). Doing so is only possible because racial social categorization is not strictly tied to innate physical characteristics.

Of course, individuals also have their own notions of racial identity. When census surveys ask a respondent to declare themselves as white, Black or African American, American Indian or Alaska Native, Chinese, Filipino, Asian Indian, Vietnamese, Korean, Japanese, Native Hawaiian, Samoan, Chamorro, other Asian, or other Pacific Islander, or some other race,[12] they likely use a variety of facts about themselves and their experiences to do so. One's own racial identity is not typically directly observed, however. And even if it were, what matters for discrimination is not how a person views themselves but how they are viewed by others.[13]

---

[10]Similar comparisons are possible with other distinctively Black names and for men, such as Kimani versus Latoya or Reginald versus Tyrone.

[11]The photo editor who produced the cover denied any racial motivations and claimed he darkened the photo "much like a stage director would lower the lights on a somber scene" (Angeletti and Oliva, 2010).

[12]These are the categories available on the 2020 Census form.

[13]In some cases, these two notions of identity may correspond closely. Regardless of the degree of congruence, however, measuring the contributions of discrimination to the disparities in outcomes of those who view themselves as members of one racial category or another is still a well-defined exercise.

Many quantitative and physical scientists continue to view race primarily in biological terms (Morning, 2007). Much research in economics implicitly adopts this view, treating race as a fixed demographic characteristic readily measured by categorical variables. Yet economic theory on the nature of identity has long recognized the constructed and manipulable nature of race as a social category. And for decades, quantitative researchers have argued the extent to which talking about the "effect" of race is logically coherent in an essentialist paradigm (Greiner and Rubin, 2011).

Putting aside philosophical issues of causality, however, the constructivist challenge to empirical discrimination research is this: Do we intend to ask whether an agent treats an individual differently because of his skin color, as if manipulation of that single trait alone would lead to different outcomes? Or do we intend to ask whether an agent treats an individual differently because a constellation of physical and contextual features strongly suggest they belong in a social category that has particular meaning to the decision maker? I argue it is the latter. This article explores the import of taking this idea seriously.

2 What is race and discrimination under the law?

Since many battles over race and discrimination are fought in courtrooms, one would think the law offers more clear-cut guidance for empiricists. This section argues that, unfortunately, it does not. The Constitution's Fourteenth Amendment, the basis for cornerstone racial civil rights cases, including *Brown* v. *Board of Education*, *Loving* v. *Virginia*, and *Shelley* v. *Kraemer*, makes no mention of race, color, or ancestry and their meaning.[14] The nation's most comprehensive civil rights legislation, the 1964 Civil Rights Act, prohibits discrimination in public facilities and programs, public education, and employment by restricting differential treatment "on the basis of," "because of," "on the ground of," "on account of," or "by reason of" race, color, religion, sex, or national origin. No definition is given of race, aside from an implicit differentiation from color, nor is the meaning of terms like "because of" elaborated.

In light of this ambiguity, substantial case law has developed interpreting when exactly behavior falls afoul of the legal protections against discrimination. Two distinct doctrines have emerged. The first—disparate treatment—covers *intentional* discrimination and is intimately tied to the Fourteenth Amendment's promise of "equal protection." The most obvious disparate treatment cases occur when documentary evidence makes race's role as a motivating factor explicit. In the infamous case of *Foster* v. *Chatman*, for

---

[14] "All persons born or naturalized in the United States, and subject to the jurisdiction thereof, are citizens of the United States and of the State wherein they reside. No State shall make or enforce any law which shall abridge the privileges or immunities of citizens of the United States; nor shall any State deprive any person of life, liberty, or property, without due process of law; nor deny to any person within its jurisdiction the equal protection of the laws."

example, the Supreme Court ruled that Georgia prosecutors had unconstitutionally struck all prospective Black jurors in Timothy Foster's murder trial. Key to the case were several things that the prosecutors did: wrote notes highlighting which jurors were Black; circled juror's races on their questionnaires; labeled jurors as "B#1," "B#2," etc.; and wrote an annotation stating "No Black Church" near the name of one Black juror's place of worship.

There is little to debate in such cases.[15] Race clearly played an illegal role in the decision regardless of whether it referred to an essential or constructed trait of the prospective jurors. Yet the law also recognizes that discrimination can be proved without "smoking gun" evidence of the relevant party's state of mind. Circumstantial evidence is frequently used to establish disparate treatment (*Teamsters* v. *United States*, 431 U.S. 324, 1977), typically using frameworks that leave discrimination as the residual explanation after eliminating plausible alternatives (*McDonnell Douglas Corp.* v. *Green*, 411 U.S. 792, 1973).[16] In fact, statistical disparities alone can establish evidence of disparate treatment that must be the result of discrimination unless otherwise explained (*Hazelwood School Dist.* v. *United States*, 433 U.S. 299, 1977).[17]

These frameworks illustrate how the law has essentially dodged the definitional challenges. Discrimination has occurred unless the employer can explain why *something else* motivated an adverse outcome for a protected worker. Whether a plaintiff can actually claim membership in a protected category is rarely contested. Nor have employers argued that their behavior cannot be discriminatory because they did not view the plaintiff as a member of the class.[18] The heart of many cases thus lies in what characteristics are and are not job relevant and which are and are not racial characteristics. Naturally, there is no bright line rule that can be universally applied, and as such, legal definitions of discrimination face many of the same empirical challenges faced by the benchmarking exercises I highlight below.

The second core legal doctrine—disparate impact—covers cases of "unintentional" discrimination, or situations when polices and behaviors that are "fair in form, but dis-

---

[15]Much of the argument in the Foster case centered on whether the prosecutors actually assigned to the case had made the notes.

[16]The court held that Green, a Black mechanic passed over for new positions despite his qualifications, had established a presumption of intentional discrimination by demonstrating that a) he belongs to a protected group, b) he was qualified for the position but rejected, and c) the employer continued to seek applicants with similar qualifications afterwards. The burden then fell on McDonnell Douglas to articulate "some legitimate, nondiscriminatory reason for the employee's rejection."

[17]The Hazelwood case centered on whether the district's share of Black teachers was suspiciously low. The majority held that "where gross statistical disparities can be shown, they alone may, in a proper case, constitute prima facie proof." Quoting an earlier case (*Castenda* v. *Partida*), the opinion also defined approximate standards for statistical significance: "'as a general rule for such large samples, if the difference between the expected value and the observed number is greater than two or three standard deviations,' then the hypothesis that teachers were hired without regard to race would be suspect."

[18]Onwuachi-Willig and Barnes (2005) argue that an employer who discriminates against racially distinctive names such as Latisha or Jamal violates the Civil Rights Act regardless of the self-identified race of the name's owner.

criminatory in operation" (*Griggs* v. *Duke Power Co.*, 401 U.S. 424, 1971). Title VII's disparate impact law prohibits facially race-neutral policies, such as Duke Power's requirement that employees have a high school degree and minimum scores on aptitude tests to be eligible for promotion, if they have a "disproportionate" impact on a protected group and are shown not to be job related. However, the disparate impact doctrine typically sidesteps complicated issues of the meaning of race as well. In fact, because differences in treatment need not happen "because of" race at all, a claim simply requires pointing to a specific policy that causes a significant disparate impact based on race (as the plaintiff's identify). Even if the policy serves some business purpose, the plaintiff can still prevail by demonstrating that the employer refuses to adopt an alternative policy with less disparate impacts that meets the same needs.[19]

It is worth noting that without explicit evidence of intent, the distinction between disparate impact and disparate treatment can be tricky to make. Advertising job vacancies with the admonishment that "dark-skinned workers need not apply" would be difficult to justify as "fair in form" regardless of the business necessity and would likely constitute disparate treatment. A blanket ban on candidates from historical Black colleges and universities (HBCUs), on the other hand, might support a disparate impact claim since some HBCU graduates identify and are viewed as white. Then again, not all Black individuals have dark skin, nor do all white individuals have light skin. And a policy against hiring HBCU graduates may be ultimately motivated as either a practical means to intentionally discriminate, by race-neutral beliefs about the quality of *all* candidates from HBCUs, or both.

Finally, though I have discussed "the law" generally in this section, it is important to note that the legal system does not treat discrimination in all domains equally. Employment discrimination is especially developed due to the special provisions in Title VII and the Civil Rights Act's creation of the Equal Employment Opportunity Commission. For example, Title VI, the part of the Civil Rights Act that prohibits discrimination in federally funded programs, has no disparate impact provisions, which were only added to Title VII in 1991.[20] In all contexts, however, it seems that the law offers relatively little guidance to empirical researchers about how to operationalize modern views of race.

---

[19]A simple example comes from *Bradley* v. *Pizzaco of Nebraska, Inc.*, 7 F.3d 797, 799 (8th Cir. 1993), where the court ruled that Domino's Pizza's no-exceptions policy against beards discriminated against Black men, who are disproportionately affected by pseudofolliculitis barbae, a painful skin condition that makes shaving difficult, and served no legitimate business purpose.

[20]Writing for the majority in *Alexander* v. *Sandoval*, 532 U.S. 275 (2001), Justice Scalia argued that it is "beyond dispute—and no party disagrees—that [Title VI] prohibits only intentional discrimination," and citing *Regents of Univ. of Cal.* V. *Bakke*, 438 U. S. 265 (1978), he noted that it "proscribe[s] only those racial classifications that would violate the Equal Protection Clause." Scalia noted, however, that federal agencies are free to promulgate and enforce their own disparate impact regulations, and many do.

3   A constructivist model of discrimination

This section formalizes the constructivist perspective on race in an explicit but simple model. The model recognizes that race is not an independent characteristic but rather is a function of how those characteristics are perceived in particular contexts and by particular people. I then discuss the model's implications for identifying discrimination in common empirical settings.

3.1   The race function

To model the process of racial social categorization, I first introduce a *race function* that maps individual characteristics into perceived membership in racial social identities. Specifically, a decision maker observes a set of characteristics X. For example, a bail judge deciding whether to grant pretrial release may observe a defendant's skin color, criminal history, defense attorney's arguments, attire, hair texture, etc. Not all characteristics are necessarily observed by the researcher: X can be partitioned into $[X_1 \ X_2]'$ to reflect observed and unobserved factors.

The race function $r(X) : \mathbb{R}^p \to [0,1]$ captures how these characteristics affect perceived membership in racial social categories. For simplicity, social categories are modeled as binary continuum—i.e., from the most "white" to the most "Black"—but in principle could instead be represented as a simplex of arbitrary dimension. As discussed above, r reflects social classification, not demographic classification or individuals' internal identities. That is, it reflects "a process of bringing together social objects or events in groups which are equivalent with regard to an individual's actions, intentions, attitudes and systems of beliefs" (Tajfel, 1974). Classification need not be stable over time, consistent across individuals, intentionally constructed by agent, or even explicitly known to them. And while in some cases classification may serve as a useful heuristic for making judgments under uncertainty (Fryer and Jackson, 2008), even under full information the agent still perceives and interprets social identity.

After observing individual characteristics and perceiving race, the agent takes an action $a \in A$. The bail judge, for example, can decide to release the defendant on their own recognizance, set monetary bail, or deny pretrial release all together. Their utility from action a depends on the action, perceived race, and individual characteristics: $U(a, r(X), X)$. The agent maximizes utility by selecting the best action given r and X.

Allowing utility to depend on X captures the direct effects of individual characteristics on the decision. The judge, for example, may face guidelines specifying that defendants arrested for violent crimes should be denied release by default. She may also be swayed by the defense attorney's attestations that the defendant presents no risk to the community, observe a recommendation produced by an algorithm, factor in the defendant's poverty when setting a bail amount, etc. The judge may also value individual factors for the

signal value about an ultimate, latent objective, like the potential for pretrial misconduct. Utility from detaining individuals with a prior history of pretrial misconduct, for example, may be higher because the judge believes these individuals are more likely to misbehave again.

Utility also depends directly on perceived race. It may do so first because of explicit or implicit prejudice, however motivated. For example, if the judge views themselves as white, they may consciously or subconsciously seek to treat members of the racial "outgroup" more harshly (Luttmer, 2001; Chen and Li, 2009; Feigenberg and Miller, 2021). The direct effects of perceived race thus reflect preferences in the sense of Becker (1957). Responding to someone's race directly captures a "taste" for discrimination— utility from a given action is simply higher or lower *because of* perceived race. Allowing for this effect captures the most uncontroversial form of discrimination.

Perceived race may also affect utility because the agent uses racial categories to make inferences about relevant unobservables. The judge may believe, for example, that white defendants are more likely to be involved in drug than violent crime and thus feel more comfortable releasing them on their own recognizance. These beliefs may be correct, on average, or reflect distorted views of between-group differences (Bordalo et al., 2016). When reasoning based on social categories, what matters is whether the agent views an individual as Black or white and their beliefs about the behaviors and traits of both groups; the specific characteristics that produced the racial inference are glossed over. That is, it is a between-group inference rather than within. I return to the distinction between these two types of statistical inferences below.

The goal in discrimination research is to determine whether and how decisions depend on perceived race. Translating the model into the language of potential outcomes, one might let $Y \in A$ reflect the utility-maximizing action taken by the decision maker. Potential outcomes $Y(r, X)$ depend on individual characteristics and perceived race. The goal is to test whether $Y(r, X) \neq Y(r', X)$ *for some* $r \neq r'$.

At this most general level, the identification problem is clear. Without additional restrictions on the functional form of U or r, it will not be possible to tell whether differences in decisions across individuals reflect differences in their characteristics or how their race is perceived, because r is an arbitrary function of X. Not only are potentially decision-relevant elements of X potentially *correlated* with race due to the covariance structure of observables (e.g., if the distribution of crime types is not evenly distributed across individuals with particular phenotypic traits), but decision-relevant elements of X may also *directly* affect perceived race (Freeman et al., 2011; Saperstein and Penner, 2012; Guryan and Charles, 2013).

The fundamental challenge to studying discrimination from the constructivist perspective is to overcome this identification problem. I discuss potential solutions further below. First, however, I highlight the basic implications of this model for common tests

and theories of discrimination.

3.2   Benchmarking and the measurement problem

Empirical research typically confronts the challenge of decision-relevant observables and unobservables that correlate with perceived race by controlling for as many factors as possible when estimating the "effects" of race (Ayres, 2010). So-called benchmarking regressions are often thought to be more reliable when a particularly rich set of factors are accounted for, making it more likely that any residual differences in treatment are attributable to race and not some other non-racial factor. In cases where *all* factors observed by the decision maker are also observed by the agent, sufficiently flexible controls are often argued to identify the effect of race (and therefore discrimination) (Hangartner, Kopp and Siegenthaler, 2021).

Yet in most empirical benchmarking exercises, the data simply record categorical variables for race. Often these variables reflect the answer to a question asked of the individual in the data, as in census records. In other settings, racial data may reflect the categorization assigned by another observer. In addition to cases where reported race is intentionally manipulated to avoid scrutiny (Luh, 2019), coded race may also simply not correspond to race as perceived by the decision maker. As Guryan and Charles (2013) write:

> A person calling herself Black, who is so noted by the analyst, may not be regarded as Black by the market actors with whom she interacts, and vice versa. Because discrimination research aims to determine whether people are treated differently by various market actors because of their race, knowing the race that observers ascribe to an individual would seem an important precondition for calling differential treatment discrimination.

This measurement problem interacts with the challenges of benchmarking in an important way. While conditioning on X may absorb differences in decision-relevant characteristics across individuals, it may also change the extent to which the decision maker views them as racially different. Imagine, for example, comparing bail decisions for two individuals who live in similar homes in the same neighborhood, went to the same high school and got the same grades, and have the same sets of tattoos, the same name, the same accent and patterns of speech, the same courtroom demeanor, the same jewelry and attire, and the same criminal conduct but differ in the color of their skin. The judge's decisions for these two individuals may be more similar than for two random individuals with different skin color. Yet the judge is also very likely to view the differences in social identity for the matched pair differently as well.

Controlling for successively more comprehensive sets of characteristics may therefore ensure that coded race is uncorrelated with potential confounds but also may attenuate

the unmeasured first-stage effect of coded race on perceived racial differences.[21] If the goal is to detect whether race as a social category affects decisions, as I argue it should be, such attenuation is an important concern. An observed racial disparity that disappears after conditioning on a large set of controls may reflect either decisive evidence that race is not a factor in decisions or the seriousness of the measurement problems at hand.

I illustrate these issues using data from the 1997 NLSY cohort. In 2011 and 2012, respondents were recorded speaking in both formal and informal settings. Grogger (2019) recruited individuals to listen to the recorded speech and to classify each speaker's sex, race, and region of origin. Up to six separate listeners classified each participant's speech. Members of the NLSY97 cohort also had their skin color recorded by interviewers in several previous rounds using color cards that ranked shades from 1 (the lightest) to 10 (the darkest). The detailed data in the NSLY on education, employment, and scores on aptitude tests administered by the Department of Defense allow me to assess not only how racial gaps in wages are affected by detailed controls but also these other important racial cues.

Panel a of Table 1 first regresses log wages in 2011, when cohort members were in their late 20s and early 30s, on an indicator for coded race and increasingly detailed controls. I only include respondents who are coded as either Black or white, are not currently enrolled in school, and have a valid log wage for their primary job. Column 1 shows that the unconditional wage gap is large—roughly 25%. Including demographic controls, which capture sex, birth year, and census region of birth, reduce the gap to roughly 20%. Adding controls for highest grade achieved further attenuates the gap by roughly half. Finally, adding aptitude test controls reduces the gap by roughly 5%.

These results suggest that much of the unconditional Black-white wage gap may reflect differences in skills rewarded by employers. The rest of Table 1, however, shows that these same controls also meaningfully attenuate between-group differences in other important racial cues. Panel b shows that while Black NLSY97 respondents are naturally coded as having darker skin, adding the full suite of controls reduces between-group skin shade differences by about 5%. Panel c uses the share of listeners who classified each respondent's speech as Black as the outcome. While Black members of the NLSY97 cohort are more likely to speak in distinctive ways overall, the conditional difference is meaningfully smaller. Controlling for a large set of covariates thus changes not only the distribution of job-relevant skills between groups but also the conditional differences in how each group member looks and sounds.

One can go further by taking seriously the idea that regressions with race dummies on the right-hand side should be viewed as reduced forms. Indeed, if panels b or c reflect the

---

[21]To continue the instrumental variables analogy, if the researcher observed perceived race, she could potentially correct for such attenuation by re-scaling the reduced-form effects of coded race on outcomes. Though perceived race is rarely measured, this is an interesting avenue for future research.

typically unmeasured first stage—implying that perceived race is captured entirely by skin shade or speech—the reduction in the racial wage gap due to Neal and Johnson (1996)-style proxies for pre-market skills would be 5% or 40% smaller, respectively, than what is suggested in panel a.[22] Of course, in reality, coded race, skin shade, and speech may all be imperfect proxies for how race is perceived in the labor market. More complicated models that formalize this idea in the style of Bound (1991) are an interesting area for future research.

If including *all* individual characteristics is not a foolproof strategy, the benchmarker faces the challenge of deciding when to stop. Which characteristics are decision-relevant confounds and which ones are simply racial cues that should only affect utility through race itself? While there are some characteristics that are uncontroversially non-racial (perhaps) and others that are not, there is a vast gray area and no uncontroversial answer as to where to draw the line. Saperstein and Penner (2012), for example, suggest that even criminal history, the most common control included in benchmarking regressions in research on the justice system, is not necessarily non-racial. Faced with this ambiguity, the empirical tendency has been to control for as much as possible. The constructivist perspective suggests this approach may be misguided, or at the very least highly conservative.

3.3 Within- and between-group statistical discrimination

Putting aside measurement issues, another key lesson from the constructivist perspective is the distinction between direct and indirect—or race-based—statistical discrimination. Making statistical inferences about other people based on how they look and behave is an innate part of human cognition. We frequently do so automatically and uncontroversially. For example, a recruiting manager may assume that a candidate from an Ivy League school may be more productive than one who did not go to college. A bail judge may predict that defendants with histories of violent conduct are more likely to be violent again. A police officer may ask for a breathalyzer test after a stopped motorist slurs their speech.

All of these are examples of statistical discrimination based on observed characteristics. But what does it mean to reason about others' unobserved traits on the basis of race? If race reflects not physical characteristics but how individuals are socially categorized, then race-based reasoning reflects statistical discrimination based on social identity. The agent observes traits that make her more likely to view someone as white, and she makes an inference about that person based on the characteristics and behaviors of white people more generally. Because the social category necessarily groups together a large swath

---

[22]These calculations compare the reduction in panel a (–0.248 to –0.047) to the reduction in the implied instrumental variable estimates (e.g., $\frac{-0.248}{4.289}$ to $\frac{-0.047}{4.095}$).

Table 1: Wage, Skin Shade, and Speech Gaps in the NLSY97

	(1)	(2)	(3)	(4)
	(a) Log wages			
Black	-0.248***	-0.202***	-0.115***	-0.0471
	(0.0283)	(0.0303)	(0.0281)	(0.0304)
Constant	2.822***	2.974***	3.132***	3.036***
	(0.0166)	(0.0440)	(0.356)	(0.359)
	(b) Skin shade			
Black	4.289***	4.209***	4.167***	4.095***
	(0.0879)	(0.0946)	(0.0975)	(0.104)
Constant	1.812***	1.551***	1.038*	1.140*
	(0.0275)	(0.0908)	(0.479)	(0.484)
	(c) Distinctive speech			
Black	0.284***	0.164***	0.143***	0.111***
	(0.0154)	(0.0153)	(0.0152)	(0.0160)
Constant	0.169***	0.0630***	0.174	0.220
	(0.00748)	(0.0162)	(0.122)	(0.122)
N	1,905	1,905	1,905	1,905
Demographics		Yes	Yes	Yes
Education			Yes	Yes
ASVAB				Yes

Notes: This table reports regressions of log wages, skin shade, and racially distinctive speech patterns on an indicator for coded race and demographic, educational, and aptitude test controls. The outcome in panel a is the log wage. In panel b it is interviewer-coded skin shade on a 1–10 scale. In panel c it is the share of listeners who classified the respondent's recorded speech as Black. The sample includes all NLSY97 respondents surveyed in 2011 with a valid log wage, skin shade measure, and speech measure whose indicated race is either Black or white. Respondents currently enrolled in school are dropped. Demographic controls include year of birth, sex, and census region at age 12 dummies. Educational controls include indicators for highest year of education. ASVAB includes a linear effect of scores on the Armed Services Vocational Aptitude Battery test. Robust standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

of individuals with varying characteristics, personal histories, and behaviors, race-based reasoning necessarily glosses over potentially meaningful within-group heterogeneity.

Superficially, reasoning on the basis of social identity may appear no different than reasoning about impending weather from cloud formations. From the perspective of the decision maker, social categories exist. She knows generally who fits into which group and how members of the group typically behave. Using group identity to make judgments about people is therefore just a simple application of Bayes' rule. Unlike

verifiable characteristics like years of schooling or behaviors like slurring one's speech, however, social categories are not facts about people. They are facts about how people are viewed and constructed by political and social processes. While social categories may contain decision-relevant information, they are just one of many potential ways people can be grouped and categorized. And in the case of race, research shows that racial categories are not neutrally constructed.

To be more concrete, take the finding from the research cited earlier that class and income are strongly associated with race so that individuals in janitorial overalls with a history of incarceration are less likely to be categorized as white. The Black social category lumps together richer individuals with no criminal history but darker skin into the same bucket as individuals with lighter skin and an extensive criminal history. If one reasons on the basis of race and at the expense of within-group heterogeneity, those viewed as Black may be artificially viewed as more criminal than if one reasoned on the basis of individual characteristics alone.

Importantly, the decision maker's reasoning is not statistically flawed, at least superficially. Given any population, it is always possible to group people into different categories and to form accurate beliefs about the behaviors of each category. But doing so may also not be the "efficient" solution to a decision problem under uncertainty, as statistical discrimination is typically understood. In other words, "accurate" statistical discrimination is only accurate if one conditions on social categories and social meaning. The decision maker could do weakly better if she used all information available to her and ignored social identities. Doing so may be hard, however, and many models of cognitive inattention could rationalize the use of coarse categories in decision-making (Mullainathan, 2002; Fryer and Jackson, 2008).[23]

Of course, the decision maker may also hold inaccurate beliefs. She may, for example, overestimate the prevalence of some group traits that are relatively more common (Bordalo et al., 2016). Non-neutral construction of social categories may exacerbate mistaken beliefs if groupings respond to inaccurate beliefs as well, leading to a hardening of perceived social differences. If the decision maker, for example, mistakenly believes that *all* Asian students are mathematically talented, then she may also be more likely to view a strong mathematics student as Asian, reinforcing the belief. Indeed, if racial social identity is flexible enough, initially inaccurate beliefs about group behavior may ultimately be confirmed as the decision maker reshapes group membership to better reflect her beliefs.

The distinction between direct and race-based statistical discrimination presents some interesting and somewhat uncomfortable challenges. In principle, it is possible to reason statistically on the basis of an important racial cue without reference to social identity— that is, not through its effect on perceived race. Is this discrimination? Although the

---

[23]Loury (2002) describes the problem as the difference between applying a given specification, which is easy, and uncovering a specification error, which is harder.

agent is not using race-based reasoning, their utility still depends on a characteristic that would surely enter the race function as well. Given the obvious identification challenges, normative claims are needed. Most people would recognize that preferences over skin color are wrong, however motivated.[24] Yet many other physical characteristics, such as height and build or the friendliness of a smile, routinely and uncontroversially affect decisions. Discrimination is thus defined only through a normative restriction that one *ought* not to have preferences over skin color. I return to this important point further below.

Economic research on statistical discrimination has historically elided any difference between direct and race-based statistical inferences, both from a modeling and welfare perspective. Canonical frameworks beginning with Phelps (1972) and Aigner and Cain (1977) assume there are typically two groups (indexed by an observable $b_i \in \{0, 1\}$). The agent observes group membership and updates their beliefs according to the population distributions of behavior. Learning models and their empirical applications (e.g., Altonji and Pierret, 2001; Bohren, Imas and Rosenberg, 2019) typically make similar assumptions.[25]

In contrast, the constructivist perspective does not imply statistical discrimination is poorly defined or uninteresting. Instead, it offers new directions for research that seeks to better understand how race is used in decision-making and to distinguish between race-based and direct inference. Because the former fundamentally relies on between- versus within-group comparisons, there are important empirical implications that can be explored. While substantial work has explored the motivations and applications of stereo- typic thinking (Hilton and Von Hippel, 1996), the formation and maintenance of groups over which stereotypes are formed has received less attention, especially in economics.

4 Solutions

There are many ways to study discrimination that acknowledge the socially constructed nature of racial social identities. Though not originally discussed in such terms, many classic projects, such as Bertrand and Mullainathan (2004)'s correspondence study of employment discrimination, have a simple interpretation under the constructivist model. Viewing race as a social category, however, does raise new challenges for several common strategies, and I discuss these below.

---

[24]As is suggested by the fact that color demarcates a specifically protected class under the law.

[25]Tests for learning and the accuracy of beliefs face the same challenges as benchmarking discussed above. As the decision maker accumulates more experience and information, their perceptions of social identity may fluctuate as well.

4.1   Exclusion

By now, one potential strategy for studying discrimination should be clear: racial instruments. If the researcher has access to variables that shift the race function but not utility, then it is straightforward to study the reduced-form effect of these variables on outcomes. Continuing with the potential outcomes notation introduced above, one requires the following:

Definition 1 (Racial instruments) *The variable Z is a valid racial instrument if $r(Z, X) \neq r(Z', X)$ and $Y(r, Z) = Y(r, Z')$.*

Ideally, one would use these variables to instrument for a measurement of perceived race, but in many cases the empirical goal is to simply test whether discrimination exists at all, not to quantify its magnitude. In such cases the reduced form alone is sufficient.

Perhaps the most famous example of racial instruments comes from correspondence studies (Bertrand and Mullainathan, 2004; Bertrand and Duflo, 2017; Kline, Rose and Walters, 2021). These studies test whether employers respond differently to resumes with distinctively Black names, such as Latisha and Jamal, than those with distinctively white names, such as Emily or Greg, and typically find that Black names are contacted as much as 30% less often. Exclusion is central to the audit study's claim to measure racial discrimination. Names must affect employers' perceptions of race but not other decision-relevant factors; that is, they must enter the race function but not utility directly.

Much ink has been spilled on testing this assumption, including the question of whether employers respond to information about social class encoded in names. This critique argues that employers would be no more likely to contact a "Cleatus" than a Jamal (Fryer and Levitt, 2004; Gaddis, 2017).[26] One might be tempted to compare audit study responses to distinctively Black and white names typically given to people from similar socio-economic backgrounds. But as in the benchmarking example above, doing so may both eliminate a potential confound and attenuate names' perceived racial differences (Gaddis, 2017). These tests therefore either require an analogous exclusion restriction—that socio-economic measures associated with names do not enter the race function—or strategies to correct for attenuation. Actually measuring the first-stage effects on perceptions is a promising route forward.

There are myriad other examples of racial instruments, mostly focused on experimental settings where manipulating explicit racial cues is possible.[27] As in cases where researchers use instruments to identify causal effects, the plausibility of the instrument is context and design specific. The researcher must take a stand on how and why the

---

[26]I use this name here in homage to Cletus Spuckler, the stereotypical "yokel" portrayed on *The Simpsons*.

[27]A particularly interesting example comes from Doleac and Stein (2013), who study responses to classified advertisements for iPods where the product was photographed in with light- or dark-skinned hands.

instrument affects utility and (carefully) test those assumptions against plausible alternatives.

4.2 Observed dimension reduction

Rather than taking a stand on how experimentally manipulated racial cues affect utility, researchers can alternatively take a stand on everything else. This strategy requires arguing that all decision-relevant factors can be observed or isolated by the researcher. The classic example is Becker (1957)'s analysis of labor market discrimination, where workers' value to firms is characterized by a single dimension of heterogeneity: productivity. If the researcher observes productivity, she can simply examine how factors that affect racial categorization impact treatment (e.g., wages) conditional on productivity. The virtue of this approach is that many observed and unobserved characteristics may contribute to qualification as long as the research design isolates qualification itself.

Definition 2 (Observed dimension reduction) *There exists an index Z such that no discrimination implies $Y(r(X), Z, X) = Y(r(X'), Z, X') \; \forall \; X, X'$.*

Goncalves and Mello (2021) present a very simple contemporary example. Studying the ticketing behavior of Florida Highway Patrol officers, they argue minority motorists are less likely to receive a reduced penalty conditional on being caught driving at the same speed as a white motorist. Here, decision-relevant factors can be collapsed to a single number: driving speed. Since a core part of the officers' job is arguably to catch speeders, this seems fairly uncontroversial. Yet even in this simple case, the implicit dimension reduction does not come for free. Goncalves and Mello (2021)'s approach requires that absent discrimination, officers would be equally likely to ticket Black and white motorists when caught driving at the same speed. This means factors such as the vehicle's features, the driver's attitude, or even the duration the motorist sustained the driving speed are either irrelevant or uncorrelated with perceived race.

Since one can likely always invent potential omitted decision-relevant factors, dimension reduction should be viewed as a normative exercise. The argument is that Black and white motorists *ought* to face the same penalties if caught speeding to the same degree regardless of what else the officer may observe about them. The normative aspects of dimension reduction are even clearer in the famous example of Goldin and Rouse (2000). By studying gender differences in musicians' evaluations when performing behind a screen that obscured their appearance, all that *ought* to matter is the quality of the performance. Because dimension reduction hinges on normative restrictions on what factors should be decision relevant in the absence of discrimination, it faces distinct challenges relative to testing for discrimination using racial instruments.

4.3 Unobserved dimension reduction

The decision makers in the preceding examples observe the relevant dimension—the motorists' speed or the musicians' performance—directly. The more common case involves situations where even if utility can be inarguably restricted to depend on a single or set of latent factors, it is not observed by the decision maker herself. In this case, the decision maker acts on her best guess of the relevant factor based on other characteristics. The relevant definition of non-discrimination is the following:

Definition 3 (Unobserved dimension reduction) *There exists a mapping $g(X)$: $\mathbb{R}^p \to \mathbb{R}^k$ such that no discrimination implies*

$$Y(r(X), g(X), X) = Y(r(X'), g(X'), X') \ \forall \ X, X' \ such \ that \ g(X) = g(X').$$

Now the researcher must identify not only the relevant latent factor (e.g., productivity) but also the appropriate information set the decision maker uses to make inferences about the latent factor. Doing so requires another type of normative claim. One might posit, for example, that $g(\cdot)$ captures the expected net benefits of taking a particular action given all information observed by the decision maker. In other words, the decision maker forms rational expectations based on all data available. But what information she uses and how is not always uncontroversial. Even if she forms rational expectations, she may use extra information unobserved by the researcher or rely on only a subset of the information the researcher observes.

An important example comes from pretrial detention decisions in criminal courts (Arnold, Dobbie and Yang, 2018; Arnold, Dobbie and Hull, 2020; Marx, 2021). The dimension reduction in this context argues that all that matters is the defendant's likelihood of pretrial misconduct. Judges detain defendants whose expected costs of misconduct exceed the costs of detention. They discriminate if they act as if detention costs are lower for Black versus white defendants, either because they place less value on Black defendants' freedom or because they systematically overestimate their probability of misconduct.[28] Many factors can contribute to expected misconduct, including defendant characteristics unobserved to the researcher.

One natural response to the problem posed by unobserved dimension reduction is to collect information on decision makers' subjective beliefs about latent factors directly, as suggested by Manski (2004). It would then be straightforward to test whether expectations are systematically biased for some groups (relative to $g(\cdot)$, assuming it can be

---

[28]This definition is not uncontroversial. Canay, Mogstad and Mountjoy (2020) argue that the testable implications for discrimination depend on whether and how other "non-racial" characteristics affect judges' expectations of misconduct and costs of detention. Putting aside the somewhat murky distinction between racial and non-racial characteristics discussed above, the critique highlights the important sense in which dimension reduction arguments embody normative claims about how agents ought to behave in the absence of discrimination.

estimated), whether choices appear to be consistent with the model sketched above, and whether there are racial gaps in detention rates conditional on beliefs. An even better approach would be to simultaneously collect data on perceived race so that the researcher can also tease out how beliefs about misconduct affect racial perceptions conditional on other observed factors.

Various quasi-experimental approaches to measuring discrimination in this setting have also been developed. The definitional and econometric issues involved in doing so are beside the point of this paper.[29] The solution to the challenge posed by a constructivist model of race remains the same: restrict utility such that comparisons across groups can be made while holding fixed factors that ought to solely explain behavior in the absence of discrimination. Just as in the observed case, the validity of this approach hinges on the plausibility of the restrictions.

5   Conclusion

Constructivist theories of race—the idea that race refers to socially constructed identity groups and not innate biological factors—can seem loosely defined and difficult to grapple with empirically. A wealth of evidence suggests, however, that constructivist perspectives offer a better description of how race is interpreted and deployed in the real world. How one perceives both one's own and others' race can change over time, depend on contextual factors, and be infected by various "non-racial" characteristics. Even when someone might be relatively unambiguously categorized *demographically*, for example, by a census enumerator, the *social meaning* of race depends on other complex cues, as shown by the diversity of peoples' views on how "Black" President Obama is.

This paper offers a framework for operationalizing constructivist ideas in empirical research on racial discrimination. Rather than treating race as a fixed characteristic of people, one easily measured by demographic variables in data, I model race as a process of social categorization based on observable characteristics and contextual factors. Discrimination is the act of treating someone differently based on perceived social identity. The model offers new perspectives on whether and how traditional measures, such as the Oaxaca-Blinder benchmarking exercise, capture discrimination, and draws a novel distinction between race-based and direct statistical discrimination. Two avenues for testing for discrimination are offered: one relying on instruments that shift social identity but are not decision relevant themselves and another that requires inherently normative

---

[29]Dimension reduction in the justice system is particularly challenging because of a lack of a clear objective function. Firms may relatively safely be assumed to maximize profits, in which case the relevant dimension of worker heterogeneity is their productivity. But qualification in the justice system is more complex, and there are myriad examples where a decision maker may justifiably treat equally "qualified" (e.g., guilty, likely to reoffend, or repentant) individuals differently to serve other socially desirable purposes.

claims about the isolable set of decision-relevant factors.

Does acknowledging the constructed nature of race invalidate a vast body of work in the social sciences that has largely treated race as coded in data as no different than age or place of birth? Absolutely not. Many research projects have interesting and useful interpretations under a constructivist lens, and measures of inter-group disparities based on self-reported or census-provided racial categorizations are still well defined and useful. The challenge for future research on discrimination, however, is to develop new tools that directly measure and study the influence of race as a social category, both to understand the impacts of discrimination and an end in itself.

References

Aigner, Dennis J, and Glen G Cain. 1977. "Statistical theories of discrimination in labor markets." *Ilr Review*, 30(2): 175–187.

Akerlof, George A., and Rachel E. Kranton. 2000. "Economics and Identity*." *The Quarterly Journal of Economics*, 115(3): 715–753.

Akerlof, George A, and Rachel E Kranton. 2010. *Identity economics.* Princeton University Press.

Alba, Richard D., Noura E. Insolera, and Scarlett Lindeman. 2016. "Is Race Really So Fluid? Revisiting Saperstein and Penner's Empirical Claims." *American Journal of Sociology*, 122(1): 247–262.

Altonji, Joseph G, and Charles R Pierret. 2001. "Employer learning and statistical discrimination." *The quarterly journal of economics*, 116(1): 313–350.

Angeletti, Norberto, and Alberto Oliva. 2010. *Time: The illustrated history of the world's most influential magazine.* Rizzoli.

Arnold, David, Will Dobbie, and Crystal S Yang. 2018. "Racial Bias in Bail Decisions." *The Quarterly Journal of Economics*, 133(4): 1885–1932.

Arnold, David, Will S Dobbie, and Peter Hull. 2020. "Measuring racial discrimination in bail decisions." National Bureau of Economic Research.

Arrow, Kenneth J. 1998. "What has economics to say about racial discrimination?" *Journal of economic perspectives*, 12(2): 91–100.

Austen-Smith, D., and Roland Fryer. 2005. "An economic analysis of "Acting White"." *Quarterly Journal of Economics*, 120(2): 551–583. Cited By :173.

Ayres, Ian. 2010. "Testing for Discrimination and the Problem of "Included Variable Bias"." Working Paper.

Becker, Gary S. 1957. *The Economics of Discrimination.* University of Chicago Press.

Bertrand, Marianne, and Esther Duflo. 2017. "Field experiments on discrimination." In *Handbook of Field Experiments.* Vol. 1, , ed. Esther Duflo and Abhijit Banerjee. Elsevier.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination." *American Economic Review*, 94(4): 991–1013.

Blinder, Alan S. 1973. "Wage discrimination: reduced form and structural estimates." *Journal of Human resources*, 436–455.

Blouin, Arthur, and Sharun W. Mukand. 2019. "Erasing Ethnicity? Propaganda, Nation Building, and Identity in Rwanda." *Journal of Political Economy*, 127(3): 1008–1062.

Bohren, J. Aislinn, Alex Imas, and Michael Rosenberg. 2019. "The Dynamics of Discrimination: Theory and Evidence." *American Economic Review*, 109(10): 3395–3436.

Bonilla-Silva, Eduardo. 1999. "The essential social fact of race." *American Sociological Review*, 64(6): 899–906.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. 2016. "Stereotypes." *The Quarterly Journal of Economics*, 131(4): 1753–1794.

Bound, John. 1991. "Self-Reported Versus Objective Measures of Health in Retirement Models." *Journal of Human Resources*, 26(1): 106–138.

Canay, Ivan A, Magne Mogstad, and Jack Mountjoy. 2020. "On the use of outcome tests for detecting bias in decision making." National Bureau of Economic Research.

Charles, Kerwin Kofi, and Jonathan Guryan. 2011. "Studying Discrimination: Fundamental Challenges and Recent Progress." *Annual Review of Economics*, 3(1): 479–511.

Chen, Yan, and Sherry Xin Li. 2009. "Group Identity and Social Preferences." *American Economic Review*, 99(1): 431–57.

Coate, Stephen, and Glenn C. Loury. 1993. "Will Affirmative-Action Policies Eliminate Negative Stereotypes?" *The American Economic Review*, 83(5): 1220–1240.

Corcoran, Mary, and Greg J Duncan. 1979. "Work history, labor force attachment, and earnings differences between the races and sexes." *Journal of Human Resources*, 3–20.

Cornwell, Christopher, Jason Rivera, and Ian M Schmutte. 2017. "Wage discrimination when identity is subjective evidence from changes in employer-reported race." *Journal of Human Resources*, 52(3): 719–755.

Darity Jr, William, David K Guilkey, and William Winfrey. 1996. "Explaining differences in economic performance among racial and ethnic groups in the USA: the data examined." *American Journal of Economics and Sociology*, 55(4): 411–425.

Darity, William A., Patrick L. Mason, and James B. Stewart. 2006. "The economics of identity: The origin and persistence of racial identity norms." *Journal of Economic Behavior Organization*, 60(3): 283–305.

Davenport, Lauren D. 2016. "The Role of Gender, Class, and Religion in Biracial Americans' Racial Labeling Decisions." *American Sociological Review*, 81(1): 57–84.

Desmond, Matthew, and Mustafa Emirbayer. 2009. "What is racial domination?" *Du Bois Review: Social Science Research on Race*, 6(2): 335–355.

Doleac, Jennifer L, and Luke CD Stein. 2013. "The visible hand: Race and online market outcomes." *The Economic Journal*, 123(572): 469–492.

Eguia, Jon X. 2017. "Discrimination and assimilation at school." *Journal of Public Economics*, 156: 48–58.

Emirbayer, Mustafa, and Matthew Desmond. 2021. *The racial order.* University of Chicago Press.

Feigenberg, Benjamin, and Conrad Miller. 2021. "Racial Divisions and Criminal Justice: Evidence from Southern State Courts." *American Economic Journal: Economic Policy*, 13(2): 207–40.

Freeman, Jonathan B., Andrew M. Penner, Aliya Saperstein, Matthias Scheutz, and Nalini Ambady. 2011. "Looking the Part: Social Status Cues Shape Race Perception." *PLOS ONE*, 6(9): 1–10.

Fryer, Roland, and Matthew O. Jackson. 2008. "A Categorical Model of Cognition and Biased Decision Making." *The B.E. Journal of Theoretical Economics*, 8(1).

Fryer, Roland, and Steven Levitt. 2004. "The causes and consequences of distinctively black names." *The Quarterly Journal of Economics*, 119(3): 767–805.

Gaddis, S. Michael. 2017. "How Black Are Lakisha and Jamal? Racial Perceptions from Names Used in Correspondence Audit Studies." *Sociological Science*, 4(19): 469–489.

Goffman, Erving. 1963. *Stigma: Notes on the Management of Spoiled Identity.* Simon and Schuster.

Goldin, Claudia, and Cecilia Rouse. 2000. "Orchestrating impartiality: The impact of" blind" auditions on female musicians." *American economic review*, 90(4): 715–741.

Goncalves, Felipe, and Steven Mello. 2021. "A few bad apples? Racial bias in policing." *American Economic Review*, 111(5): 1406–41.

Greiner, D James, and Donald B Rubin. 2011. "Causal effects of perceived immutable characteristics." *Review of Economics and Statistics*, 93(3): 775–785.

Grogger, Jeffrey. 2019. "Speech and Wages." *Journal of Human Resources*, 54(4): 926–952.

Guryan, Jonathan, and Kerwin Kofi Charles. 2013. "Taste-based or statistical discrimination: the economics of discrimination returns to its roots." *The Economic Journal*, 123(572): F417–F432.

Hangartner, Dominik, Daniel Kopp, and Michael Siegenthaler. 2021. "Monitoring hiring discrimination through online recruitment platforms." *Nature*, 589(7843): 572–576.

Heckman, James J. 1998. "Detecting Discrimination." *The Journal of Economic Perspectives*, 12(2): 101–116.

Hilton, James L, and William Von Hippel. 1996. "Stereotypes." *Annual review of psychology*, 47(1): 237–271.

Holland, Paul W. 2008. "Causation and race." *White logic, white methods: Racism and methodology*, 93–109.

Jackson Jr, John L. 2010. *Harlemworld.* University of Chicago Press.

Jaynes, Gerald. 2004. "Immigration and the social construction of otherness: Underclass stigma and intergroup relations." *Not just black and white: Historical and contemporary perspectives on immigration, race, and ethnicity in the United States*, 100–116.

Kennedy, Randall. 2012. *Interracial intimacies: Sex, marriage, identity, and adoption.* Vintage.

Kim, Young-Chul, and Glenn C. Loury. 2019. "To be, or not to be: Stereotypes, identity choice and group inequality." *Journal of Public Economics*, 174: 36–52.

Kline, Patrick, Evan K. Rose, and Christopher Walters. 2021. "Systemic Discrimination Among Large U.S. Employers."

Kohler-Hausmann, Issa. 2018. "Eddie Murphy and the dangers of counterfactual causal thinking about detecting racial discrimination." *Nw. UL Rev.*, 113: 1163.

Liebler, Carolyn A., Sonya R. Porter, Leticia E. Fernandez, James M. Noon, and Sharon R. Ennis. 2017. "America's Churning Races: Race and Ethnicity Response Changes Between Census 2000 and the 2010 Census." *Demography*, 54(1): 259–284.

Loury, Glenn C. 2002. *The Anatomy of Racial Inequality.* Harvard University Press.

Luh, Elizabeth. 2019. "Not So Black and White: Uncovering Racial Bias from Systematically Misreported Trooper Reports." *Available at SSRN 3357063.*

Luttmer, Erzo F. P. 2001. "Group Loyalty and the Taste for Redistribution." *Journal of Political Economy*, 109(3): 500–528.

Manski, Charles F. 2004. "Measuring expectations." *Econometrica*, 72(5): 1329–1376.

Marx, Philip. 2021. "An Absolute Test of Racial Prejudice." *The Journal of Law, Economics, and Organization.* ewab002.

Mincer, Jacob. 1974. "Schooling, Experience, and Earnings. Human Behavior & Social Institutions No. 2."

Monk Jr, Ellis P. 2021. "The unceasing significance of colorism: skin tone stratification in the United States." *Daedalus*, 150(2): 76–90.

Morning, Ann. 2007. ""Everyone knows it's a social construct": Contemporary science and the nature of race." *Sociological focus*, 40(4): 436–454.

Mullainathan, Sendhil. 2002. "Thinking through Categories." *Mimeo, Harvard Unviersity.*

Neal, Derek A., and William R. Johnson. 1996. "The Role of Premarket Factors in Black-White Wage Differences." *Journal of Political Economy*, 104(5): 869–895.

Oaxaca, Ronald L, and Michael R Ransom. 1994. "On discrimination and the decomposition of wage differentials." *Journal of econometrics*, 61(1): 5–21.

Onwuachi-Willig, Angela, and Mario L. Barnes. 2005. "By Any Other Name?: On Being "Regarded as" Black, and Why Title VII Should Apply Even If Lakisha and Jamal Are White." *Wisconsin Law Review*, 1283.

Phelps, Edmund S. 1972. "The statistical theory of racism and sexism." *The american economic review*, 62(4): 659–661.

Reeve, Elspeth. 2011. "Herman Cain: Obama 'Was Raised in Kenya'." *The Atlantic.*

Reynolds, Dean. 2007. "Obama's Racial Identity Still An Issue." *CBS.*

Rose, Evan K. 2020. "Who gets a second chance? Effectiveness and equity in supervision of criminal offenders." *The Quarterly Journal of Economics*, 136(2): 1199–1253.

Saperstein, Aliya, and Andrew M. Penner. 2012. "Racial Fluidity and Inequality in the United States." *American Journal of Sociology*, 118(3): 676–727.

Sen, Maya, and Omar Wasow. 2016. "Race as a Bundle of Sticks: Designs that Estimate Effects of Seemingly Immutable Characteristics." *Annual Review of Political Science*, 19(1): 499–522.

Sewell, Alyasah Ali. 2021. "Workshop on Data Equity Issues in Quantifying Racism, Race, and Ethnicity."

Somanader, Tanya. 2011. "Ingraham Suggests Cain Would Be The Real First Black President Because Obama Has White Relatives." *Think Progress*.

Tajfel, Henri. 1974. "Social identity and intergroup behaviour." *Social Science Information*, 13(2): 65–93.

Zuberi, Tukufu. 2001. *Thicker than blood: How racial statistics lie.* U of Minnesota Press.