Perceptions of Race in the Labor Market*

Pedro C. Sant'Anna† Sulin Sardoschau‡ Aiko Schmeißer§

13th November 2026

Abstract

Racial disparities are measured as if race were a fixed attribute though scholars across disciplines view it as fluid, context-dependent, and observer-specific. We study the determinants of racial perceptions in the labor market and its consequences for measured wage disparities, leveraging Brazilian administrative data on 330,000 workers with multiple race measures: employer-recorded and self-identified race, skin tone extracted algorithmically from photographs, and names. Self- and employer perceptions diverge in over a quarter of cases and employers categorize workers with the same skin tone into different racial categories depending on the region, time period, and the worker's social status. University graduates are less likely to be recorded as non-White, conditional on skin tone, self-identified race, distinctly non-White name, and within the same firm and occupation. We compare racial wage disparities across multiple measures of race, showing that accounting for endogenous perceptions changes both magnitude and interpretation of conventional estimates.

Keywords: Race, identity, disparity, wage gap, Brazil
JEL Codes: J15; J50; J71; Z10

*We would like to thank Anjali Adukia, Ellora Derenoncourt, Amy Finkelstein, Michael França, Richard Hornbeck, Peter Hull, Larry Katz, Erzo Luttmer, Ben Olken, Elias Papaioannou, Rohini Pande, Nina Roussille, Evan Rose, Frank Schilbach, as well as participants of the Labor Workshop at the 2025 RIDGE Forum, ifo Junior Labor Economist Workshop, 40th Meeting of the European Economic Association, 5th Arne Ryde Workshop on Culture, Institutions, and Development at Lund University, 1st Workshop on the Political Economy of Identity at RFBerlin, Columbia-Cornell Political Economy of Work Junior Scholars Workshop, MIT's Behavioral Lunch and Algorithms Coffee, Columbia's Labor & Public Finance Colloquium, Research Seminar at the Immigration Policy Lab ETH Zurich, Annual Meeting of the Society of Labor Economists, Monash-Paris-Warwick-Zurich-CEPR Text-as-Data Workshop, Department Seminar of Universidad Carlos III de Madrid, Rockwool Joint Berlin-Kopenhagen Workshop, BFI Chicago AI in Social Science Conference, and BFI Women in Empirical Microeconomics Conference. We also thank the George & Obie Shultz Fund at MIT. The authors declare no conflict on interest.

†MIT Department of Economics and Núcleo de Estudos Raciais (INSPER). E-mail: p_stanna@mit.edu
‡Humboldt Universität zu Berlin, RFBerlin and IZA@LISER. E-mail: sulin.sardoschau@hu-berlin.de
§ROCKWOOL Foundation Berlin (RFBerlin) and IZA@LISER. E-mail: a.schmeisser@rfberlin.com

1 Introduction

Nearly every constitution and international human rights instrument prohibits discrimination on the basis of race. But race is a social category with malleable boundaries: who counts as White or Black has shifted over decades, differs across regions governed by near-identical legal language, and frequently diverges across observers judging the same person. Legal scholars and economists increasingly argue that the operative object is therefore not a fixed trait such as a person's name, ancestry, or even self-identity, but rather the racial category that a decision-maker ascribes at the moment of decision (Greiner, 2008; Lopez, 2006).

Although racial perceptions are central to our understanding of discrimination, empirical evidence on what determines them remains scarce. Race is widely understood as a *bundle of sticks* that includes – beyond phenotype or ancestry – markers of social status such as education, occupation, or income (Sen and Wasow, 2016; Omi et al., 2014). Perceived race is thus endogenous to the very socioeconomic outcomes along which disparities are measured, complicating both measurement and interpretation (Rose, 2023). This raises two questions for research on racial disparities: What shapes perceptions of race? And what do disparity estimates capture when race is itself endogenous?

This paper studies these questions empirically in the Brazilian labor market, bringing together information that is rarely available for the same individual: the racial category held by the decision-maker, outcomes from the same decision environment, and independent measures of the individual's identity, social status, and appearance. Most disparity estimates rely on self-reports, which capture how individuals classify themselves. Enumerator classifications and audit-study treatments capture how researchers or third parties classify them. Yet, none of them captures the perception of the decision-maker under study. We link these data for the same worker and show that race is not fixed: self-perceptions and employers' perceptions frequently diverge, and employers' classifications respond to workers' social status and to the context in which they are formed. We then estimate racial wage disparities across multiple measures of race and find that accounting for employers' endogenous perceptions substantially changes both the magnitude and the interpretation of these disparities.

Our data combine two administrative sources. The *Relação Anual de Informações Sociais* (RAIS) covers the universe of formal jobs in Brazil and records, for each job, the worker's race as reported by the employer, alongside full names, wages, employment histories, and firm characteristics. We link RAIS to records on the universe of political candidates from the *Tribunal Superior Eleitoral* (TSE), which contain candidates' self-declared race and standardized photographs. From these photographs, we extract an algorithmic measure of skin

tone following Adukia et al. (2023) and a machine-learning prediction of race based on all visual cues.[1] The linked sample covers 330,000 workers between 2003 and 2015. Relative to the full RAIS workforce, these workers are more often male and slightly older, but similar in education and average earnings. Labor market attachment remains strong because most candidates run for municipal council positions that require no more than 10 hours per week, and roughly 85% of them do not win office.

We begin by documenting mismatched racial identities in the labor market. Employer-recorded and self-declared race diverge for 28 percent of workers, and mismatches are asymmetric: employers are much more likely to record self-identified non-White workers as White than to classify self-identified Whites as non-White. Moreover, among job movers, the previous and new employers disagree on the same worker's racial category in 23 percent of cases. These mismatches concentrate among individuals with intermediate skin tones, and we provide evidence that measurement error alone is unlikely to explain this mismatch.

To understand the sources of this divergence, we examine the determinants of employer-recorded race. Skin tone predicts how employers classify workers, but the same skin tone maps into different categories in different contexts. At a given skin tone, employers are more likely to record a worker as non-White *i*) in the North than in the South of Brazil, *ii*) in 2015 than in 2003, and *iii*) when the worker has no high school degree rather than a university degree. Skin tone is therefore neither an objective marker nor an essentialist dimension of race. It is one of the features through which individuals are racialized, and how it maps into racial categories depends on the local context, on social norms at the time, and on the worker's social status.

We then estimate the determinants of employer-recorded race jointly, successively adding demographics, region, industry, occupation and, in our most stringent specification, firm-by-occupation fixed effects, which effectively hold the observer fixed. Conditional on skin tone, university graduates are 3.2 percentage points less likely to be recorded as non-White than workers without a high school degree, even among workers in the same firm and occupation. Among workers with multiple jobs, longer tenure and higher wages at the previous employer are similarly associated with White ascription. Because these patterns persist within firm-by-occupation cells, racial perceptions are formed at the job level, in the interaction between worker and employer, and not only through the sorting of workers into firms with different

[1] Ideally, we would observe the race ascribed by the wage-setter at the moment of the pay decision and the racial identity workers present in economic life. Our measures come close but are not perfect: RAIS race is recorded at the establishment level, typically at hiring, rather than by the individual wage-setter, and TSE self-declarations are made under candidacy-specific stakes in a selected sample. We discuss these issues in Sections 3.2, 5, and 6.3 and Appendix C. We also replicate the main results for welfare recipients in the *Cadastro Único* (CadÚnico) linked to RAIS, a poorer and less educated population that declares race privately but has no photographs. Results are provided in a Supplementary Online Appendix.

classification practices.

Since employers observe more than appearance, we next consider workers' self-declared race and indices of how distinctively non-White their given name and surname are (Fryer and Levitt, 2004).[2] Each cue predicts employer-ascribed race, both separately and jointly. Self-declared race and names are strong predictors, but employers systematically depart from them. Conditional on self-declared race and name distinctiveness – as well as skin tone and firm-by-occupation fixed effects – workers without a university degree remain more likely to be recorded as non-White, highlighting the role of social status in racialization.

This has direct consequences for how racial wage disparities are measured: the characteristics typically used to account for productivity differences also shape how workers are racially classified. We examine these consequences empirically, starting with conventional wage gaps based on employer-recorded race and successively adding the same controls and fixed effects used above. Workers recorded as non-White earn 14.8% less on average. The gap shrinks by 86% in the most stringent specification, to 2.1% between workers with the same demographic characteristics in the same firm and occupation. This attenuation reflects two forces: controls absorb differences in productivity and sorting, but they also net out determinants of race itself. What the remaining race coefficient captures therefore depends on how racial categories are formed and which variation in race survives the controls.

A more transparent alternative focuses on a single cue. Following the colorism literature, we estimate wage disparities along the skin tone distribution: unconditionally, a one standard deviation darker skin tone is associated with 2.6% lower wages. Such colorism disparities are informative about earnings differences by skin tone, but they do not directly translate to disparities based on race. Because the same skin tone racializes workers to different degrees across regions, periods, and levels of social status, they combine two margins: how strongly skin tone shifts the employer's racial perception, and how strongly perceived race shifts wages. A small colorism disparity may thus reflect either a small penalty for being perceived as non-White or a large penalty combined with a skin tone that barely moves perceptions. Likewise, differences across groups may reflect differences in racialization rather than in discrimination based on perceived race.

The same ambiguity arises when controls are added. Like the conventional gap, the colorism disparity shrinks by about 80% in the most stringent specification. Importantly however, controls also attenuate the coefficient of skin tone on employer-recorded race by about 40%, and part of the decline in the colorism disparity coincides with this drop in

[2]Studies that use self-declared race as a proxy for outside perceptions approximate them only insofar as the two coincide. Moreover, self-declared race is itself endogenous: it responds to education much like employer-ascribed race.

racialization. Separating smaller disparities from weaker racialization requires measuring how strongly skin tone shifts employers' perceptions in each specification.

Our data allow us to measure both margins. We combine them in a third measure, the perception-normalized disparity (PND), which scales the colorism disparity by the skin tone gradient of employer perceptions. Intuitively, the PND expresses disparities in units of racialization rather than physical appearance, giving social meaning to a cue. The same colorism disparity implies something different in a world where skin tone is used to racialize individuals compared to one where skin tone matters little. The PND makes this explicit by expressing the disparity per unit of change in employer-recorded race. This normalization offers two advantages, which we illustrate empirically.

First, the PND makes cue-based disparities comparable across settings in which the same cue racializes workers to different degrees. For instance, the colorism disparity is similar in the Southeast and the North of Brazil. In the North, however, skin tone is a much weaker racial signal: two otherwise identical workers must differ by a much larger margin in skin tone to be classified as different races. The same wage gap therefore corresponds to a smaller difference in perceived race, and normalizing by perception reveals a substantially larger PND in the North compared to the Southeast.[3]

Second, the PND is less sensitive to benchmarking controls because it accounts for the fact that these controls also weaken the racialization margin. Unconditionally, the PND amounts to 23.5%, compared to 14.8% for the conventional disparity based on employer-recorded race. It declines by about two-thirds in the most stringent specification, a drop roughly 20% smaller than for the conventional and colorism disparities, and remains about four times as large as the conventional wage gap.

Taken together, our results highlight that different approaches to measuring racial disparities target different objects when race is endogenous. Our aim is not to rank these measures – they are all informative in their own right – but to offer a comparative empirical perspective. Conventional estimates capture the gap in the recorded racial category net of its own determinants, cue-based estimates capture the wage gap associated with a single worker characteristic, and the PND expresses this gap per unit of difference in perceived race. By making the perception process explicit and measurable, we connect measured disparities to the notion of discrimination based on racial perceptions.

Overall, our findings suggest that racial identity is not only something individuals bring

[3]This also applies across education groups. Because skin tone is a weaker racial signal for university-educated workers than for workers without a high school degree, the colorism disparity is lower than the PND. Perception normalization also makes disparities comparable across racial cues. A one standard deviation darker skin tone and a more distinctively non-White name are not measured in the same units, but their PNDs are: both express the wage gap per unit of difference in employer-recorded race.

to social and economic life, but also something produced by it. Workers with the same skin tone are classified differently depending on the region, the period, and their education, and no single cue, whether phenotype, name, ancestry, or self-identity, determines how a worker is racialized. Individual characteristics become racial cues only in a social context, where they shape how others perceive a person. This has implications for two of the most studied objects in labor economics: returns to education and discrimination. Returns to education may include not only productivity gains and signals of ability but also returns to being perceived as White. And controlling for education when measuring disparities is not neutral, since education itself shapes how race is perceived. When disparities shrink as controls are added, part of the decline may reflect controls absorbing determinants of race rather than differences in productivity.

Our paper contributes to three literatures. We first add to research on the social construction of race. Sociologists, psychologists, and legal scholars have long argued that racial categories are not fixed attributes but classifications whose boundaries shift over time, differ across contexts, and diverge across observers (Sen and Wasow, 2016; Telles, 2014; Saperstein and Penner, 2012; Greiner and Rubin, 2011). Survey evidence from the United States shows that interviewer-recorded race responds to incarceration, unemployment, and poverty (Saperstein and Penner, 2012; Penner and Saperstein, 2008), and laboratory studies show that status cues such as attire change how the same face is categorized (Freeman et al., 2011). Similar fluidity has been documented for self-reported race, ethnicity, caste, and religion (Antman and Duncan, 2023; Jia and Persson, 2021; Atkin et al., 2021; Antman and Duncan, 2015; Cassan, 2015). Most closely related, Adukia et al. (2026) show that wealth and literacy predict being recorded as White in the 1870 U.S. Census conditional on bank clerks' descriptions of complexion, and that college completion predicts White classification conditional on interviewer-rated skin tone in the NLSY97. Cornwell et al. (2017) show that workers' employer-recorded race in Brazil often changes when they move between firms, and that these changes are associated with wage gains and losses.

We add to this literature in three ways. First, we observe racial classifications by employers, recorded in the same decision environment that produces the outcome we study. Census enumerators, interviewers, clerks, and laboratory subjects, by contrast, classify individuals they observe once or only from a photograph, and their classifications have no bearing on the outcome. Second, because each employer classifies many workers, we can hold the observer fixed by comparing workers in the same firm and occupation. This rules out that status predicts recorded race only because individuals are sorted to observers with different classification practices, a concern that historical and survey data cannot address. Third, we observe workers' self-declared race and phenotype alongside the employer's classification.

This allows us to examine the role of socio-economic status in shaping perceptions while holding constant the most frequently used proxies of race in the literature.

We also contribute to the measurement of racial disparities along a single racial cue. A large literature documents that darker skin is associated with worse labor market outcomes in Latin America and the United States (Woo-Mora, 2026; Kreisman and Rangel, 2015; Telles, 2014; Monk, 2014; Hersch, 2008; Goldsmith et al., 2007, 2006; Hersch, 2006), studies the consequences of distinctively Black names (Kreisman and Smith, 2023; Fryer and Levitt, 2004), and manipulates such cues in audit and correspondence studies to estimate discrimination (Ajzenman et al., 2025; Evsyukova et al., 2025; Kline et al., 2022; Bertrand and Mullainathan, 2004). Recent audit designs pre-test how names are perceived by survey respondents (Crabtree et al., 2023; Gaddis, 2017), but this margin is rarely measured among the decision-makers whose behavior is under study and is examined only in experimental settings. We add to this literature in two ways. First, we measure this margin directly among employers – in a natural setting – and examine how it varies across contexts, worker characteristics, and specifications, which allows us to distinguish differences in racialization from differences in discrimination. Second, we measure skin tone algorithmically from standardized photographs, following Adukia et al. (2023), rather than through observer ratings, which can themselves respond to a person's status (Hannon and DeFina, 2016).

Finally, we contribute to the literature that estimates racial gaps conditional on proxies for productivity (Bayer et al., 2025; Bayer and Charles, 2018; Fryer, 2011; Charles and Guryan, 2008). The literature increasingly uses matched employer-employee data to account for sorting (Sorkin, 2025; Gerard et al., 2021; Card et al., 2018), and proposes frameworks for interpreting what such estimates capture (Bohren et al., 2025a,b; Arnold et al., 2022; Bohren et al., 2019; Arnold et al., 2018; Charles and Guryan, 2011). These frameworks, like the canonical models of discrimination (Aigner and Cain, 1977; Arrow, 1973; Phelps, 1972; Becker, 1957), take group membership as given. Rose (2023) argues that a constructivist view of race questions this premise, since conditioning on status partly conditions on race when status shapes the recorded race itself. We provide the empirical counterpart to this argument by quantifying how these controls capture determinants of racial perceptions and how this affects the measurement and interpretation of multiple disparity estimates.

The remainder of the paper is structured as follows. Section 2 develops a simple conceptual framework to guide the empirical analysis. Section 3 describes the data and background. Section 4 documents race mismatches, and Section 5 examines the determinants of employers' race perceptions. Section 6 analyzes the implications for measuring wage disparities, and Section 7 concludes.

2 Conceptual Framework

We start by introducing a constructivist approach to race in the labor market. Building on the framework proposed by Rose (2023), we distinguish between self-identified race and employer-ascribed race and analyze when and why these measures coincide or diverge. We show that when racial perceptions depend on social cues, the interpretation of conventional disparity measures may change, as variables that directly affect outcomes may also shape the process of racial categorization.

Self-Identification and Employer-Perceptions. Let R^g denote the race category recorded by source $g \in \{S, E\}$, where S indexes self-identification and E indicates the race ascribed by the decision maker – in our case, the employer. We consider a binary coding

$$R^g \in \{0, 1\},$$

where 0 corresponds to White and 1 to non-White.[4]

We model racial classification as the discretization of a latent racialization index. Specifically, observer g maps a set of racial cues $Z^g \in \mathcal{Z}_g$ into a scalar index

$$a^g \;=\; A_g(Z^g) + u^g \in \mathbb{R},$$

where $A_g : \mathcal{Z}_g \to \mathbb{R}$ is the mapping of cues into a latent race index, and where u^g captures noise in the form of administrative inattention or recording errors. The recorded category R^g is then generated by a threshold rule,

$$R^g \;=\; \begin{cases} 1 & \text{if } a^g \geq \tau_g, \\ 0 & \text{if } a^g < \tau_g, \end{cases}$$

with $\tau_g \in \mathbb{R}$ an observer-specific threshold that captures category boundaries in a given social and institutional environment. In this formulation, Z^g may include phenotypic features (e.g., skin tone or hair texture), ancestry (e.g., family origin or surname-based cues), and status signals (e.g., education, income, dress or speech). Moreover, the interpretation and weighting of these cues and their aggregation into racial categories, through $A_g(\cdot)$ and τ_g, may be shaped by contextual factors in which these interactions take place (e.g., location or firm).

This setup makes clear when self- and employer-perceptions can be treated as equivalent.

[4] In principle, the model can be extended to accommodate multiple racial categories. In Section 3, we justify the use of a binary racial categorization in the Brazilian context.

Suppose there exists a common information set Z and a common mapping (A, τ) such that both sources implement the same classification rule,

$$R^S = \mathbf{1}\{A(Z) + u^S \geq \tau\}, \qquad R^E = \mathbf{1}\{A(Z) + u^E \geq \tau\},$$

then R^S and R^E differ only because of observer-specific noise. In the limiting case without such noise, $R^S = R^E$ and the two racial concepts are identical. This equivalence would hold in settings where racial cues were entirely unambiguous, or if individuals internalize repeated external feedback so that self-identification converges to dominant outside perceptions.

In other settings, equivalence need not hold because the underlying objects determining R^g differ across observers. First, information sets can differ ($Z^E \neq Z^S$). Employers typically rely on observable cues in the hiring or wage-setting environment, whereas self-identification may incorporate private information about ancestry, family history, or social affiliation. Second, even when the same cues are available, observers may weight them differently or apply different category boundaries. This corresponds to differences in $A_S(\cdot)$ versus $A_E(\cdot)$ and in thresholds τ_S versus τ_E. In all these cases, disagreement between R^S and R^E is not merely measurement error around a single underlying "true" race. Rather, it reflects heterogeneity in which cues are available and how they are translated into racial categories.

Which measure is most relevant depends on the research question. If the objective is to describe broad labor market inequality associated with racial identity, and one believes that self-identification captures channels such as differential social networks or behavioral responses to anticipated or past discrimination, then R^S may be the more relevant margin. If instead the objective is to study disparate treatment by decision-makers, then employer-ascribed race R^E is the relevant object because it is the category that enters the decision environment. In either case, choosing between R^S and R^E is a substantive modeling choice about the perception process, not a purely technical decision about measurement.

Observational studies that rely on self-declared race to approximate outside perceptions often carry implicit assumptions. They either assume that R^S is included in the set of racial cues Z that enter $A_E(\cdot)$. For instance, workers signal their racial identity to employers who incorporate this information in their own assessment. In this case, R^S can be interpreted as the reduced form effect of a racial cue. Or, they assume that R^S is an approximation of R^E given that $A_S(\cdot)$ and $A_E(\cdot)$ draw on a similar information set, apply similar weights, and hold similar classification thresholds. While we will document substantial mismatch rates between R^S and R^E, we also show that both employer perceptions and self-perceptions are endogenous to socio-economic status. As a result, the implications for measured disparities discussed next are equally applicable to both concepts.

Conventional Disparity Measures. Conventional measures of disparity compare mean outcomes across racial categories, either unconditionally or conditional on other individual characteristics. If the goal is to describe differences in outcomes across recorded racial categories then these quantities are well-defined descriptive objects. If the goal is, however, to identify *disparate impact* (capturing both structural/institutional and direct forms of discrimination) or *disparate treatment* (direct discrimination at a specific decision-node), then one must take into account the process of racial categorization by decision makers.

More specifically, researchers attempting to estimate the "causal effect of race" face a longstanding disagreement over what it means to estimate such an effect (e.g., Hu and Kohler-Hausmann, 2025; Rose, 2023). There are at least two reasons to hesitate to interpret recorded race itself as a cause. First, race is fluid, observer-specific, and context-dependent. Second, the notion of a counterfactual race is conceptually difficult. For these reasons, researchers often prefer to interpret the causal effect of race as the causal effect of *racial perceptions*, particularly when studying direct discrimination. This approach asks what would happen if a relevant decision-maker perceived the same individual differently.

In this section, we focus on employers as the relevant decision-makers in our labor market setting. Let $Y \in \mathbb{R}$ denote a labor market outcome of interest, such as (log) wages or an indicator for employment. X denotes productivity-relevant characteristics of the worker. Then, the conditional racial disparity based on employer perceptions is defined as

$$\Delta(x) \;=\; \mathbb{E}[Y \mid R^E = 1, X = x] \;-\; \mathbb{E}[Y \mid R^E = 0, X = x].$$

The constructivist approach clarifies that racial disparity measures depend on how R^E is generated, because R^E is endogenous to observable cues and contextual information. Specifically, the key empirical complication is that the cue set Z^E can contain characteristics that are also productivity-relevant, i.e., $Z^E \cap X \neq \varnothing$. This creates an identification problem in conventional disparity estimates. If the target is the effect of employer-ascribed race holding productivity fixed, omitting elements of Z^E that affect outcomes and are correlated with R^E confounds perceived race with omitted productivity-relevant variation.

If, instead, the researcher observes and controls for these elements, the omitted-variable concern is reduced, but the interpretation of this conditional comparison changes: the comparison is conducted while holding fixed the inputs that generate the perceived racial category in the first place. When R^E is highly predictable from X, conditioning on X leaves little identifying variation in R^E. As controls absorb inputs into racial classification, the comparison increasingly holds fixed the cues through which employer-ascribed race is generated, and the interpretation would depend on what remaining variation in racial perceptions exists.

Cue-Based Disparity Measures. An alternative approach that sidesteps direct race measures and is more explicit about the variation exploited, focuses on a single salient cue $z \in Z$. Colorism and audit studies, for example, study disparities by skin tone and names, respectively.[5] Let us simplify to a case where $z \in \{z^0, z^1\}$ denotes two values of a cue. Cue-based disparities compare outcomes across values of z, often conditional on X:

$$\Delta_z(x) \;=\; \mathbb{E}\big[Y \mid z = z^1, X = x\big] \;-\; \mathbb{E}\big[Y \mid z = z^0, X = x\big]\,.$$

The interpretive challenge is that $\Delta_z(x)$ is not, by itself, a disparity in perceived race. If z is as-good-as random conditional on X, $\Delta_z(x)$ can be interpreted as an intent-to-treat (ITT) effect of the cue. Thus, it is informative about perception-based disparities only to the extent that the cue shifts employer-ascribed race. The sensitivity of racial ascription to the cue can be summarized by the first-stage:

$$\pi_z(x) \;=\; \mathbb{E}\big[\mathbf{1}\{R^E = 1\} \mid z = z^1, X = x\big] \;-\; \mathbb{E}\big[\mathbf{1}\{R^E = 1\} \mid z = z^0, X = x\big]\,.$$

When $\pi_z(x)$ is small, the cue rarely changes perceived race in that group, so even substantial disparate treatment based on R^E can generate only a small reduced-form disparity in z. Conversely, when $\pi_z(x)$ is large, the same cue difference induces larger changes in perceived identity, making cue-based disparities more informative about perception-based disparities. For these reasons, cue-based estimates are best interpreted jointly with evidence on how the cue maps into perceived race.

Importantly, $\pi_z(x)$ may not be a constant function. The same racial cue can carry a different weight depending on the regional context or other worker characteristics, such as education or income. For example, a darker skin tone may shift employer classification toward non-White more strongly among low-educated workers than among highly educated workers. If the estimated reduced-form effect of skin tone on wages is close to zero among highly educated workers, this may reflect genuinely smaller disparities conditional on being perceived as non-White. But it may also reflect that skin tone is less likely to change perceived race in the first place among the highly educated.

Perception Normalization. This suggests a simple normalization that puts cue-based gradients on a common scale. Let us define the *Perception-Normalized Disparity* (PND) as

$$\mathrm{PND}_z(x) \;\equiv\; \frac{\Delta_z(x)}{\pi_z(x)} \;=\; \frac{\mathbb{E}[Y \mid z = z^1, X = x] - \mathbb{E}[Y \mid z = z^0, X = x]}{\mathbb{E}[\mathbf{1}\{R^E = 1\} \mid z = z^1, X = x] - \mathbb{E}[\mathbf{1}\{R^E = 1\} \mid z = z^0, X = x]}.$$

[5] These studies rarely equate racial categories with the cues themselves; rather, they treat (implicitly or not) cues as a shifter of perceived race.

Intuitively, $\text{PND}_z(x)$ rescales the outcome gradient in the cue by the extent to which the cue changes the likelihood of being perceived as non-White. When $\pi_z(x)$ is close to zero, the cue does not move perceived race in that group, so cue-based disparities are not informative about perception-based disparities along that margin, and the normalization makes this lack of a racialization margin explicit. Compared to the reduced form, the $\text{PND}_z(x)$ expresses disparities in units of perception rather than units of the cue.

$\text{PND}_z(x)$ should be interpreted as a scaling exercise rather than the causal effect of perceived race unless one can argue that the cue affects outcomes only through employer racial ascription. Under such an exclusion restriction, $\text{PND}_z(x)$ coincides with an instrumental-variables estimand and can be interpreted as a local average treatment effect (LATE) of being perceived as non-White for observations whose perceived race changes when the cue shifts. The estimand changes depending on the specific racial cue used and does not need to be the same, such that, for instance, $\text{PND}_{z=\text{skin tone}}(x) \neq \text{PND}_{z=\text{name}}(x)$.

In sum, our conceptual framework suggests that conventional disparity measures based on R^g are inherently sensitive to control choices whenever cues that enter the racial classification rule are also productivity-relevant ($Z \cap X \neq \varnothing$). Cue-based disparities $\Delta_z(x)$ are, in turn, sensitive to the strength with which a given cue shifts perceived race, which may vary systematically across worker characteristics and contexts. The PND's scaling approach accounts for this variation, thereby making cue-based disparity estimates more comparable across settings. Importantly, each of these measures is well-defined for a distinct estimand and rests on different identifying assumptions. Our empirical analysis does not seek to rank these measures, but to quantify the magnitude of the resulting measurement differences in a real-world setting and to make explicit how the formation of racial perceptions affects the interpretation of disparity estimates.

3 Background and Data

3.1 Race in Brazil

Race in Brazil is shaped by the country's colonial history and institutional legacy. Brazil's racial system originated during the colonial era, when Portuguese colonizers enslaved millions of Africans and forced Indigenous peoples into labor. The country was the last in the Americas to abolish slavery in 1888. Even though segregation policies after abolition never matched those of the United States, the absence of formal segregation did not eliminate racial hierarchy. While scholars such as Freyre (1933) advanced the idea of "racial democracy," suggesting that widespread miscegenation fostered more harmonious race relations than in

the United States, critics like Silva and Paixão (2014), Schwarcz (2013), Fernandes (1965) and Holanda (1936) highlight persistent racial gaps in various socioeconomic outcomes.

In official statistics, the Brazilian Institute of Geography and Statistics (IBGE) recognizes five racial categories: *branco* (White), *preto* (Black), *pardo* (Brown), *amarelo* (Asian), and *indígena* (Indigenous). However, racial boundaries in Brazil typically follow a White versus non-White divide (pooling *preto* and *pardo*). Asian and Indigenous individuals make up only about one percent of the Brazilian population (IBGE, 2022). Sociological and ethnographic work emphasizes that discrimination is organized around the boundary of whiteness, whereas distinctions within Afro-descendant categories are comparatively narrow and context-dependent (Schwarcz, 2013; Telles, 2004). This aggregation is also conventional in the literature quantifying racial disparities in Brazil (e.g., Miller and Schmutte, 2023; Gerard et al., 2021; Firpo et al., 2021). Policy design mirrors this practice: affirmative-action rules treat *pretos* and *pardos* jointly as beneficiaries, rather than targeting them separately (Francis and Tannuri-Pianto, 2013). Anti-discrimination laws (including the Constitution) also protect non-Whites in general. Moreover, the White versus non-White boundary is reflected in measured inequalities. *Preto* and *pardo* workers are shown to exhibit relatively similar levels of education, employment, and wages (e.g., Miller and Schmutte, 2023; Gerard et al., 2021; Silva, 2000). At the same time, both groups experience substantially worse socio-economic outcomes than White individuals: non-Whites earn 35% less, are more likely to be unemployed, are underrepresented in leadership positions, and have fewer years of schooling (França and Portella, 2023; IBGE, 2022).

Nevertheless, an influential strand of the sociological literature in Brazil emphasizes the role of skin color in shaping racial relations and discrimination (Telles, 2014; Bailey et al., 2014, 2013). Importantly, the focus on skin tone does not amount to a naturalization of race, since the social significance of phenotypic differences is itself the outcome of a historical process (Abel, 2026; Wade, 1993). Conceptual critiques nevertheless caution that approaches centered on skin tone risk reducing racialization to a single phenotypic dimension (Silva, 2020) and highlight that skin tone is itself interpreted, along other cues, to form racial perceptions (Roth et al., 2022; Guimarães, 2003). Related concerns appear in Brazilian antiracist thought. For example, Carneiro (2004) argues that color distinctions can undermine a shared Black identity by excluding lighter-skinned Black people from it.

There is also mounting evidence that race in Brazil is not static. Individuals and observers classify or perceive the same person differently depending on economic conditions or institutional settings. In the early 2000's, many Brazilians changed their self-identification towards *preto* and *pardo* (Miranda, 2015), possibly in response to increased opportunities and lower stigma (De Micheli, 2021) or policy incentives (Francis and Tannuri-Pianto, 2013). Improve-

ments in socio-economic status are associated with changes in self- and external-identification towards White (Cornwell et al., 2017; Schwartzman, 2007; Maggie and Fry, 2004).

The phenomenon is not specific to Brazil, however. Racial boundaries shift in the United States, where individuals passed as White even under the one-drop rule and changes in racial identification remain common across censuses (Adukia et al., 2026; Noghanibehambari and Fletcher, 2025; Liebler et al., 2017; Antman and Duncan, 2015). Similar fluidity appears for ethnicity in China (Jia and Persson, 2021) and Romania (Mitrut et al., 2025), for caste in India (Cassan, 2015), and for religious identity (Atkin et al., 2021), and ties into the larger question of endogenous social identity.

3.2 Data

We combine information from large-scale administrative data sources that allow us to observe various racial identity measures for the same individual. First, we draw on the *Relação Anual de Informações Sociais* (RAIS) that provides matched employer-employee data on all formal jobs in Brazil and, importantly, records workers' race *as reported by the employer*. Second, we merge RAIS with data on the universe of political candidates in Brazil, for whom we observe their *self-declared race* as well as algorithmic measures of *phenotype* derived from standardized electoral photographs. Finally, we explore the racial content of individuals' names. This section describes the data sources, the recording or construction of the racial identity measures, and the steps used to obtain our estimation sample.

Employer-Ascribed Race. We use RAIS records covering all formal sector employment relations in Brazil between 2003 and 2015. The data is collected annually by Brazil's Ministry of Labor and Employment for the administration of various social security programs. All employers are mandated to submit information on their employees via an online platform provided by the Ministry (Appendix Figure C.1). Compliance with reporting requirements is high because incomplete submissions can trigger substantial penalties (MTE, 2024).

Each record in the data corresponds to a worker-establishment contract in a given year and contains detailed information on average monthly earnings, contracted hours, occupation, contract type, the start and end date of the contract, along with workers' sociodemographic characteristics (gender, age, education) and the firm's industry and location. Workers are identified by their unique tax identifier *Código de Pessoa Física* (CPF) and their names, and can be tracked across employers.

Crucially, RAIS collects data on employees' race since 1999. The race question asked in RAIS and the available racial categories are the standard ones in Brazilian statistics, as discussed in the previous section. There is no specific guidance on how to report a worker's

race, either in the official RAIS Manual or in the reporting interface, and the reported race information is not subject to any systematic auditing (MTE, 2024). Most employers do not seem to consult workers and instead classify race based on their own perceptions at the time of hiring (Silveira, 2022; Osório, 2003). Appendix C.1 provides more details on how employers collect and report worker information, including results from qualitative interviews we conducted with employers. Reporting of RAIS is done at the establishment level, by a dedicated HR or administrative employee, or by the owner or manager themselves in the case of smaller establishments (Silveira, 2022; Osório, 2003). Thus, the race measure should be interpreted as an establishment-level report of worker race, rather than necessarily reflecting of the specific decision-maker responsible for wage setting within the firm.

There are no direct economic incentives for reporting a more or less racially diverse workforce (e.g., no tax incentives or affirmative action laws in the private labor market). Hence, we do not expect strategic reporting in the sense of changes in employees' ascribed race to achieve a desired racial composition. At the same time, the presence of a race question in RAIS is not unusual for respondents, as Brazilians have long been accustomed to being asked about their race and that of others in official statistics. Taken together, these features suggest that employers take reporting of race in RAIS seriously and that employer-reported race is a good proxy for how employers perceive the racial identity of their workers. Nevertheless, we will subject our main results to a series of robustness checks designed to explicitly account for potential measurement error in employer-ascribed race.

Self-Declared Race. To enrich RAIS with independent measures of racial identity, we link it to administrative records from Brazil's *Tribunal Superior Eleitoral* (TSE), which contain information on the universe of individuals who run for political office. Beginning in 2014, the TSE registry includes each candidate's self-declared race using the five standard categories. Importantly, only about 15% of all candidates are elected, and about 90% run for municipal council positions. These are typically part-time roles – on average, municipal councilors are required to attend the council only four days per month (Ferraz and Finan, 2011) – during which councilors maintain their private-sector jobs. The strong labor market attachment makes the candidate sample well-suited for comparing self-declared and employer-ascribed racial identities. Candidates in the TSE records are identified by their tax identifier (CPF), enabling a deterministic linkage to RAIS and avoiding any potential mismatch in racial identities that might arise under probabilistic record-linkage procedures.

Candidate registration is an administrative declaration to the TSE, publicly available on the TSE website and electoral apps. Self-declared race in TSE records is therefore a relatively public signal, though candidates' race is not displayed on the ballot, and prior work finds

limited electoral returns to self-declaring race in TSE registration (Janusz, 2023). To the extent that the signal is public, employers may plausibly observe it. Self-identification is also not necessarily fixed, and racial self-classification can change over time (Janusz, 2021). In the empirical analysis, we examine whether strategic considerations or changes in self-identification help account for mismatches in racial identity.

Skin Tone. The candidate registry also provides standardized photographs used in electronic voting machines for nearly all candidates between 2004 and 2024, yielding close to three million photos. We use these head-shots to construct an algorithmic measure of skin tone that is derived directly from the color value of each pixel in a photo. In contrast to existing measures in the literature, which typically rely on enumerator-reported skin tone (Adukia et al., 2026; Bueno and Dunning, 2017; Telles, 2014), this approach delivers an objective and reproducible measurement of skin tone.[6] This circumvents the issue that characteristics and experiences of the enumerator or respondents' other racial cues affect how racial cues are reported in the first place.

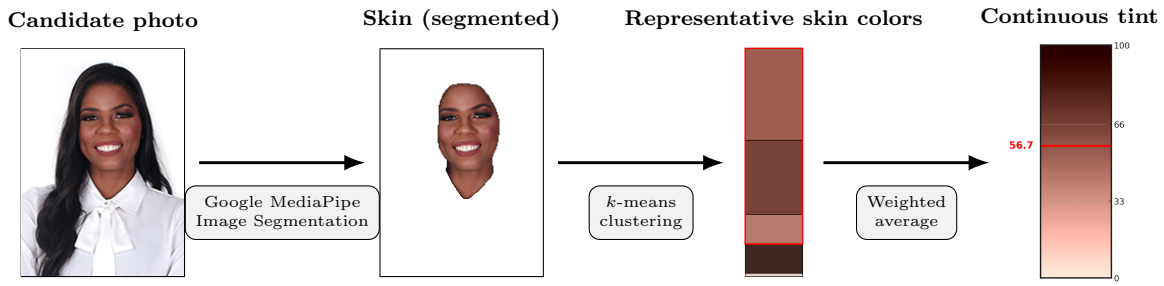
The algorithmic pipeline is illustrated in Figure 1. We first isolate the skin region in each photo using Google MediaPipe's Image Segmentation tool, which segments the skin part from other elements of the image (background, hair, clothing, accessories). We then summarize the color of the skin region by closely following the algorithm of Adukia et al. (2023). We use a k-means clustering approach to group pixels into $k = 5$ clusters and exclude the smallest two clusters that often refer to shadows or other non-skin facial features. For each remaining cluster, we compute its characteristic color and express these values in the $L^*a^*b^*$ color space, ranging from zero (for the lightest shades) to one hundred. A weighted average of these cluster colors produces a single representative skin color for each candidate. Appendix C.2 provides more information on the characteristics of electoral photos, examples across the skin tone distribution, details on the measurement procedure, and a validation exercise that compares our measure to human assessments of candidates' skin tone.[7]

Figure 2 depicts the skin tone distribution in our data, separately by self-declared White and non-White candidates (Appendix Figure A.1 divides the latter between *pardos* and *pretos*). Skin tone has a mean of 38.6 and a standard deviation of 10.8, with self-declared White

[6]There are few other papers using algorithmic procedures to extract some racial identity measure from TSE's photos (Baerlocher and Schneider, 2026; de Lucena Coelho et al., 2024). All these papers use off-the-shelf "race-detection" algorithms to predict broad racial categories, with the goal of either estimating the effects of electing non-White candidates or of using candidates' race as an outcome. In contrast, we extract specific phenotypic traits from photos to understand how racial perceptions form.

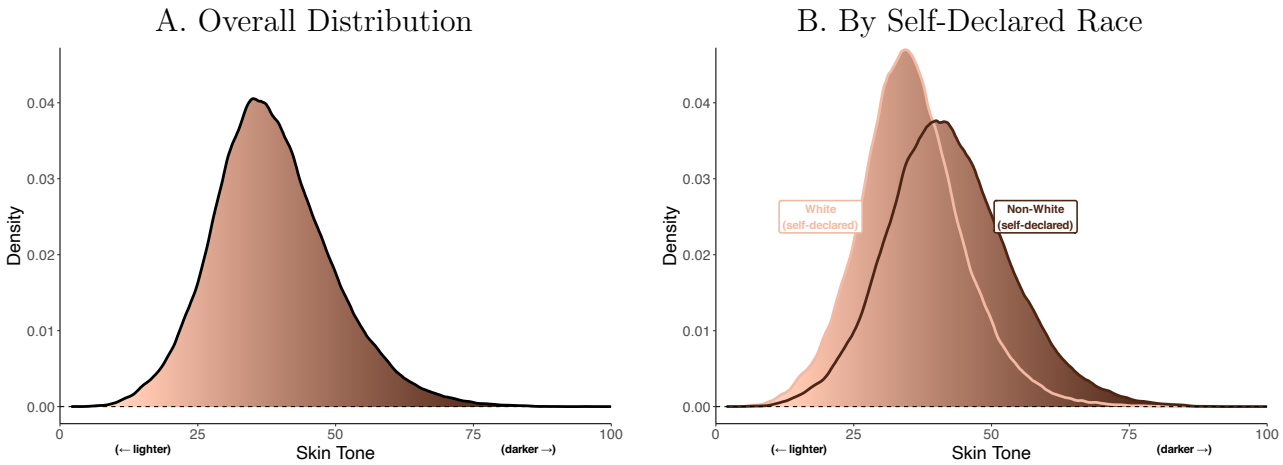
[7]About half of all candidate photos in our estimation sample are in grayscale. In Appendix C.2, we also assess the measurement quality across grayscale and colored photos, providing evidence that motivates the pooling of both types of photos in our main analysis.

Figure 1: Skin Tone Extraction Pipeline



Notes: The figure illustrates how we extract skin tone from electoral photos. The example shown corresponds roughly to the 90th percentile of the skin tone distribution. Appendix Figure C.2 provides additional examples across the distribution.

Figure 2: Skin Tone Distribution



Notes: The figure shows the distribution of skin tone as measured from the electoral photos of candidates. As will be done in all figures in paper, the skin tone distribution is visualized by varying the L^* value of a skin tone with $a^* = 17.2$ and $b^* = 19.1$, which correspond to the sample averages of a^* and b^* across all our colored photos. Panel A depicts the skin tone distribution among all candidates. Panel B splits candidates by their self-declared race, distinguishing White and non-White (*pretos* and *pardos*) candidates.

and non-White candidates differing by an average of 0.62 standard deviations. Throughout the analysis, we standardize skin tone and report estimates in standard deviation units.

Names. RAIS records workers' full names, which allows us to measure the racial content of a name directly from how employers classify the workers who carry it. Following Fryer and Levitt (2004), we compute for each name n a non-White name distinctiveness index, $P(n \mid \text{NW})/[P(n \mid \text{NW}) + P(n \mid \text{W})]$, where the probabilities are the frequencies of the name among workers recorded as non-White and as White by their employers in the full RAIS workforce. The index ranges from zero, for a name observed only among White workers, to one, for a name observed only among non-White workers, and equals 0.5 for a name that is equally common in both groups. We compute it separately for the first given name and the last surname. The underlying sample is at the worker level, using the modal name and

modal employer-ascribed race across a worker's yearly records, and we leave the worker's own record out of the frequency calculation so that a worker's own racialization does not enter their score. We present descriptive results on the racial content of Brazilian names and discuss the literature on the topic in Appendix C.4.

Main Estimation Sample. Our estimation sample consists of approximately 330,000 unique individuals who appear as political candidates at least once between 2014 and 2024 and who can be linked to at least one formal work contract in RAIS. For these individuals, we construct a yearly panel capturing all of their jobs between 2003 and 2015. We focus on private-sector, full-time ($\geq$ 30 hours/week), open-ended contracts held by individuals aged 20 to 54. For each year, we retain the primary job, defined as the contract with the highest annual earnings.

Our analysis compares employer-ascribed race in these jobs to the candidate's self-declared race and skin tone measured at the time of candidacy.[8] If an individual has run for a political office multiple times, we assign their modal self-declared race (across 2014-2024) and their mean skin tone (across 2004-2024). This approach implicitly treats self-identity and skin tone as time-invariant characteristics, which we compare to potentially divergent racial perceptions of different employers. In robustness checks, we relax this assumption and restrict comparisons to periods in which the racial identity measures overlap in time.

Appendix Table B.1 presents summary statistics for our estimation sample of political candidates and compares them with the full RAIS workforce. Relative to other workers, candidates are more likely to be male (by 13 percentage points) and slightly older (by 0.84 years), but they are similar in education and average earnings. Moreover, as shown in Appendix Table B.2, the wage gap between White and non-White workers (based on employer-ascribed race) is 10 percentage points smaller among candidates than among other workers, suggesting that our estimated disparities likely represent lower bounds for population-level gaps. Although political candidates are not representative of the Brazilian workforce, this sample is uniquely suited for our purposes because it allows us to compare various race measures for the same worker: race as perceived by the employer, self-declared racial identity, skin tone as a phenotypic marker, and name-based racial cues.

Extension to Welfare Recipients. Our main analysis relies on the candidate sample because the photographs allow us to measure skin tone, often viewed as the central racial cue in Brazil. We complement it with data on welfare recipients from the *Cadastro Único*

[8] We exclude observations where either the employer-ascribed or self-declared race is Asian or Indigenous, following standard practice in the literature (e.g., Gerard et al., 2021).

(CadÚnico), the federal registry through which low-income households apply for Brazil's main social programs. CadÚnico records self-declared race in the same five categories. Linked to RAIS through the CPF, it yields 117 million worker-years for registrants between 2012 and 2024, about 40% of the RAIS workforce. The sample covers the lower part of the income distribution, and race is declared privately to a social security office, with no electoral or material stake attached. In the absence of photographs, we replicate all exercises except those involving skin tone, and the patterns are similar throughout.[9]

4 Mismatched Racial Identities

In this section, we document how employer-ascribed race differs from self-declared race and how the same worker is classified across employers. Throughout, we focus on the binary distinction between White and non-White. Appendix Table B.3 also reports results when disaggregating the non-White category into *pretos* and *pardos*.

4.1 Self-Declared versus Employer-Ascribed Race

Panel A of Table 1 cross-tabulates self-declared and employer-ascribed race. Roughly 72% of workers fall on the diagonal, that is, they self-identify and are classified by their employer as belonging to the same racial group (White or non-White). The remaining 28% exhibit a mismatch between self-identification and employer-ascription. The mismatch is asymmetric. Among self-declared Whites, about 18% are recorded as non-White by their employer. Among self-declared non-Whites, about 40% are reported as White. In other words, conditional on how workers describe themselves, employers are much more likely to "whiten" non-White workers than to "darken" self-identified Whites.

Skin tone helps to interpret these discrepancies. Within each cell of Panel A, we report the mean standardized skin tone measure. Workers who self-identify as non-White and are also classified as non-White have, on average, the darkest skin tones. Self-identified Whites who are classified as White have the lightest skin tones. The two off-diagonal cells, where self-identification and employer-ascription differ, are characterized by intermediate skin tone levels. Panel (a) of Appendix Figure A.2 shows the full skin tone distribution, confirming that mismatches concentrate among workers with mid-range skin tones rather than at the extremes of the distribution. Notably, the average tendency of employers to "whiten" rather than "darken" their workers does not appear to be driven by skin tone differences: self-identified non-Whites classified as White by their employer have skin tones that are 0.30 SD

[9]The results are available in the Supplementary Online Appendix.

Table 1: Mismatch in Racial Identities

[A] Mismatch between Self-Declared and Employer-Ascribed Race

	Employer-Ascribed Race			
	White		Non-White	
Self-Declared Race	%	Avg. Skin Tone	%	Avg. Skin Tone
White	45.20%	-0.32	9.71%	-0.11
Non-White	17.84%	0.19	27.25%	0.45

[B] Mismatch across Subsequent Employers

	Employer-Ascribed Race, Current			
	White		Non-White	
Employer-Ascribed Race, Previous	%	Avg. Skin Tone	%	Avg. Skin Tone
White	49.82%	-0.22	12.18%	0.15
Non-White	10.88%	0.15	27.12%	0.40

Notes: The table displays the (mis)match shares for different race measures in our sample, as well as the average skin tone for each cell. Skin tone is standardized to have a mean of zero and a standard deviation of one across the whole sample (higher values represent darker shades, while lower values represent lighter shades). Panel A presents the distribution of workers by self-declared race (rows) and employer-ascribed race (columns). It considers all worker-year observations ($N = 1,510,833$). Panel B presents the distribution of ascribed race for the same worker across different employers in consecutive years. It considers all workers who changed jobs ($N = 153,366$) and presents the distribution of workers by their ascribed race in the previous employer (rows) and current employer (columns).

darker than those of self-identified Whites classified as non-White.

4.2 Mismatch across Employers

We next examine how the same individual is classified across employers. Panel B of Table 1 restricts the sample to workers who change jobs and compares employer-ascribed race in the previous and current establishment across consecutive years. About 77% of job switchers are assigned the same racial category by both employers. The remaining 23% change racial classification when they move. As in Panel A, the off-diagonal cells are sizable on both sides. Roughly 30% of workers who were previously classified as non-White are recorded as White by their new employer, while close to 20% of previously White workers are coded as non-White in their subsequent job. Panel (b) of Appendix Figure A.2 shows that mismatches across employers are again concentrated among individuals with intermediate skin tones.[10] These patterns indicate that employer-ascribed race is not a fixed attribute of the worker

[10]While Table 1 focuses on the matched sample of political candidates, Appendix Table B.4 shows that the mismatch rates across employers are very similar within the full sample of workers in RAIS, suggesting that the candidate sample is not atypical in terms of racial ambiguity.

but the outcome of an interaction between the worker's appearance and the classification practices of each firm.

4.3 Interpretation and Robustness

Measurement Error in Employer-Ascribed Race. A natural concern is that the observed mismatches may simply reflect noisy reporting rather than differences in latent race concepts – for instance, if employers pay little attention to reporting in RAIS. Several pieces of evidence indicate that measurement error alone is unlikely to account for the observed mismatch. First, as noted above, the mismatch is concentrated among intermediate skin tones, suggesting that self- and employer-ascribed racial categories diverge when phenotypic features allow for ambiguity. This is inconsistent with a scenario whereby employers record race with noise uniformly across the skin tone distribution. Instead, it suggests systematic divergence in ascribed race in cases where visual cues are ambiguous.

Second, we show that the mismatch patterns are similar across sub-samples that should exhibit higher reporting quality. Appendix Table B.4 repeats the comparison between self-declared and employer-ascribed race under a series of restrictions. We begin by excluding firms that uniformly assign the same race to their entire workforce. Because these firms might plausibly be less attentive in their reporting, we re-estimate mismatch rates in the subset of establishments where we observe workers assigned to different racial groups (only about 12% of workers in our sample are employed in fully-segregated establishments). We then restrict the sample to firms with dedicated human-resources staff to address concerns that variation reflects firms' capacity to collect information on workers. As an additional benchmark, the same table reports mismatches in gender, showing that disagreement between self-declared and employer-reported gender and changes in reported gender across employers are negligible. If employers filled out workers' demographic information in RAIS with large inattention, we would expect to observe similar mismatches for gender, which we do not.

Third, we exploit the presence of multiple employer observations for a sub-sample of workers and aggregate employer-ascribed race across jobs. Appendix Table B.5 compares several measures of race perception – including the modal employer classification and the share of employers that code the worker as non-White – to self-declared race, following the standard approach in studies with RAIS data (e.g., Gerard et al., 2021). To address measurement error in these aggregated measures, we estimate split-sample IV regressions in which the measure constructed from one random half of a worker's yearly observations is instrumented with the measure constructed from the other half. Even after averaging across employers and applying measurement error corrections, we find that the association between

employer-ascribed and self-declared race remains far from one-to-one.

Incentives in Self-Declaration. Our previous results suggest that measurement error is unlikely to be the sole explanation for the large mismatch rates we document – 28% between self-declared and employer-ascribed race and 23% in ascribed race across subsequent employers. Most of the previous research has documented the fluidity of race by studying changes in self-identification over time (e.g., Muniz et al., 2024; Janusz, 2021; Liebler et al., 2017; Miranda, 2015; Saperstein and Penner, 2012). In our sample of political candidates, among those who ran for office multiple times between 2014 and 2024, self-declared race changes between White and non-White across consecutive elections in 16.8% of cases. Thus, the divergences between self- and hetero-identification, and across different observers, are substantially more common than changes in self-identification over time.

To probe the possibility that the large mismatch is an artifact of the TSE sample or related to incentives in electoral competition, we run several checks in Appendix Table B.4. First, we limit the sample to worker-year observations in which both workers and employers report race in the same period, ruling out that discrepancies arise from changes in self-declared race over time. Second, we focus on candidates for local councilor positions, among whom potential strategic incentives to report particular self-identified race are arguably weaker than for higher political offices. Across both sub-samples, both the magnitude and the asymmetry of mismatches remain virtually unchanged.[11]

5 Determinants of Employer-Ascribed Race

The mismatch patterns in the previous section suggest that race can be ambiguous and that employers' classifications depend on how they perceive observable worker characteristics. This section examines the determinants of employer-ascribed race. Observable characteristics are not universal markers of race but features through which individuals are racialized, in ways that depend on context. Skin tone is a natural starting point in Brazil. A prominent strand of the sociological literature describes Brazil, and Latin America more broadly, as a *Pigmentocracy*, in which skin tone is a central determinant of racial perceptions and discrimination (Telles, 2014, 2004). Yet the meaning of color in Brazil has long been debated: color terms have always drawn on other physical traits and on social status, so that color never meant skin color alone (Guimarães, 2012). In this section, we add nuance to interpretations

[11] In addition, we find similarly high mismatch rates between employer-ascribed and self-declared race in the CadÚnico sample, in which welfare recipients are not subject to electoral incentives when declaring their race in private to the social security office.

of *Pigmentocracy* that center on skin tone: even conditional on skin tone, racial perceptions vary systematically over time, across contexts, and with status signals. Other cues, such as distinctively non-White names, also matter. Single cues do not map neatly into racial categories; the social context shapes how these cues are interpreted (Silva, 2020).

5.1 Cues in Context: The Case of Skin Tone

Figure 3 shows that skin tone is one input into employers' racial ascription, but it is far from deterministic. The mapping from skin tone into racial categories varies systematically over time, with workers' education, and across regions. For each dimension, we distinguish the *level* of the curve, which determines who is considered White or non-White at a given skin tone, from its *slope*, which captures how strongly employers translate incremental differences in skin tone into racial categories.

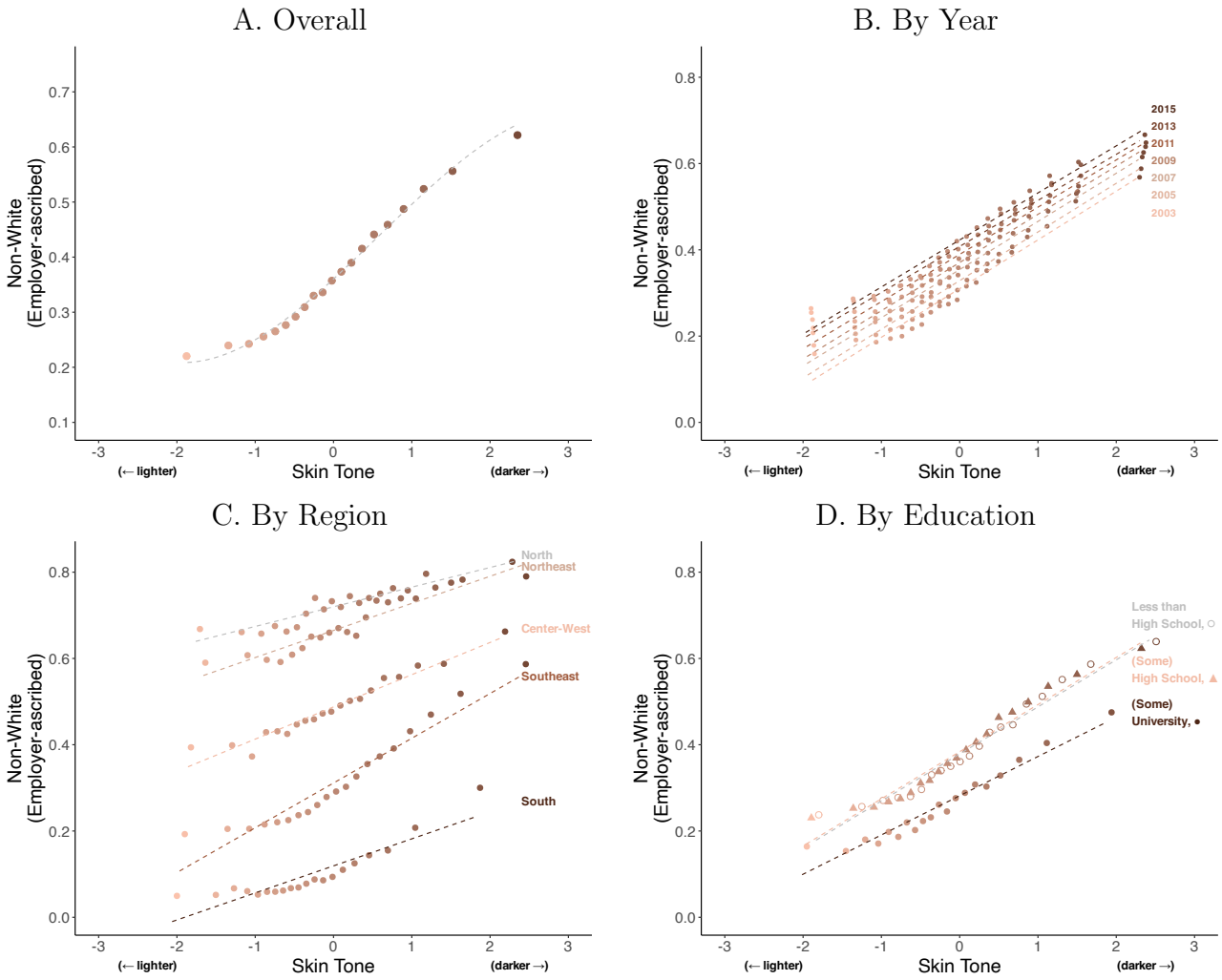
Panel A plots the probability that a worker is coded as non-White against skin tone. The relationship is steep and monotone: moving from the 10th to the 90th percentile of the skin tone distribution raises the probability of non-White ascription by 32 percentage points. However, about 20% of the lightest-skinned workers are classified as non-White, and more than 30% of the darkest-skinned workers are classified as White, which suggests that factors other than skin tone shape racial ascription.[12]

Panel B splits the sample by the year in which race is recorded. The curves shift upward monotonically between 2003 and 2015: at a given skin tone, workers are more likely to be classified as non-White in later years, by roughly 10 percentage points over the period. The slopes are nearly parallel, so the change operates through the threshold for Whiteness rather than through the weight placed on skin tone. This pattern coincides with a broader shift in racial classification in Brazil linked to declining stigma and the expansion of race-conscious narratives (De Micheli, 2021; Francis and Tannuri-Pianto, 2013).

Panel C shows that employers in the North classify workers with the same skin tone as non-White at substantially higher rates than employers in the South. "Whiteness" thus includes individuals with much darker skin tones in the South. In the North, where the Afro-descendant population is largest, the racial norm is non-White and only very light-skinned workers stand out as White. In the South, shaped by European immigration, the racial norm is White, and only very dark-skinned workers are classified as non-White. The curves also differ in slopes: the Southeast exhibits a steeper gradient, suggesting that small tone differences carry more weight in racial ascription. Again, this is in line with the local racial

[12]Measurement error in our skin tone algorithm could explain part of this residual variation. We discuss measurement error in skin tone in Section 6.2.

Figure 3: Relationship between Employer-Ascribed Race and Skin Tone



Notes: The figures show binned scatterplots of employer-ascribed race (an indicator equal to one if the worker is ascribed non-White) and workers' skin tone (standardized), in the full sample (Panel A), by year of recorded race (Panel B), by region of the establishment (Panel C), and by workers' educational attainment (Panel D). The plots are constructed following the procedure in Cattaneo et al. (2024). For each sample, we divide the skin tone distribution in twenty equal-sized bins and, for each bin, plot the average of employer-ascribed race. We fit third-degree polynomials based on the underlying data of each sample. For the relationship by year (Panel B), we show only odd years to reduce visual clutter, but the relationship is monotonic across all years.

context in the Southeast, characterized by a highly ethnically diverse population and thus sensitivity to racial categories.

Panel D splits the sample by education. Workers with a university degree are less likely to be classified as non-White than workers with the same skin tone and less schooling, across the entire skin tone distribution. The curve for university-educated workers is also somewhat flatter. Education thus shifts the threshold for Whiteness and attenuates the elasticity of recorded race with respect to skin tone.

Together, these patterns show that employers' racial ascription is not a fixed function of phenotype: the same skin tone maps into different racial categories depending on the year of

classification, the local context, and worker's education. Differences across these dimensions may reflect shifting racial boundaries, for example driven by changing race consciousness or local social norms, but they may also reflect differences in other correlated characteristics that influence racialization. In the next subsection, we examine how social status shapes racialization once we account for a rich set of worker and job characteristics.

5.2 The Many Cues Shaping Racial Perceptions

Skin Tone and Socio-economic Status. To quantify how employers translate multiple worker attributes into racial classifications, Table 2 reports coefficients of a linear probability model relating employer-ascribed race to a broad set of worker and labor market characteristics. Our goal is twofold: to isolate the direct effects of these characteristics and to examine how the coefficients of skin tone and education evolve as richer controls and fixed effects are added.

We begin with the direct effects. Column 1 reports the unconditional association between skin tone and employer-ascribed race: a one-standard-deviation increase in skin tone raises the probability of being classified as non-White by 11 percentage points. Notably, skin tone explains less than 6% of the variation in employer ascriptions, highlighting the strong potential for other characteristics to determine how employers perceive the racial identity of their workers. Column 2 introduces demographic characteristics, that is, education, gender, and a squared age term. Conditional on skin tone, university graduates are 9.6 pp less likely to be coded as non-White than those with less than a high school education. Column 3 adds regional fixed effects, which explains a substantial share of the variation in employer-ascribed race, consistent with the patterns documented above. Relative to the North and Northeast, employers in other regions are more likely to report workers as White at a given skin tone.

Columns 4 to 6 introduce progressively richer labor-market fixed effects, which allow us to compare workers who face similar task requirements, pay scales, and employer environments but differ in phenotypic and demographic attributes. Column 4 includes industry fixed effects; column 5 adds three-digit occupation by industry fixed effects.[13] Column 6 includes establishment-by-three-digit-occupation fixed effects, exploiting variation among workers that hold the same job and work at the same firm. Across all these specifications, the qualitative pattern remains: education decreases the likelihood of being classified as non-White, conditional on skin tone, even within narrowly defined labor-market segments.

[13]We consider 9 industry categories following the Brazilian national classification of economic activities: agriculture, extractive industries, manufacturing, construction, retail, transport & communication, finance & real estate, education & health, and other services. Three-digit occupations correspond to the 192 subcategories in the Brazilian Occupation Classification (2002 CBO).

Table 2: Determinants of Employer-Ascribed Race

	(1)	(2)	(3)	(4)	(5)	(6)
Skin Tone	0.110***	0.107***	0.082***	0.081***	0.078***	0.068***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Demographics						
Male		0.002	-0.004**	-0.011***	-0.013***	-0.014***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Age		-0.025***	-0.014***	-0.016***	-0.013***	-0.012***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age^2		-0.007***	-0.003***	-0.003***	-0.002***	-0.002**
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
(Some) High School		-0.009***	-0.034***	-0.027***	-0.012***	-0.007**
		(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
(Some) University		-0.096***	-0.115***	-0.110***	-0.058***	-0.032***
		(0.003)	(0.003)	(0.003)	(0.003)	(0.006)
Region						
Northeast			-0.052***	-0.053***	-0.055***	
			(0.004)	(0.004)	(0.004)	
Southeast			-0.400***	-0.398***	-0.399***	
			(0.003)	(0.003)	(0.003)	
South			-0.584***	-0.578***	-0.578***	
			(0.003)	(0.003)	(0.003)	
Center-West			-0.226***	-0.224***	-0.223***	
			(0.004)	(0.004)	(0.004)	
N	1,510,833	1,510,833	1,510,833	1,510,510	1,510,445	1,249,942
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	No	No	Yes	No	No
Industry × Occupation FE	No	No	No	No	Yes	No
Firm × Occupation FE	No	No	No	No	No	Yes
R-squared	0.057	0.065	0.209	0.213	0.223	0.779
Adjusted R-squared	0.057	0.065	0.209	0.213	0.222	0.719

Notes: The table reports OLS estimates of the relationship between employer-ascribed race (an indicator equal to one for non-White) and various covariates. The first column displays the relationship with skin tone, when only controlling for year fixed effects. The subsequent columns progressively add demographic covariates as well as region, one-digit industry, three-digit occupation, and firm fixed effects. Standard errors clustered at the individual level are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

To substantiate this claim further, we leverage information on other potential status signals for a sub-sample of workers who hold multiple jobs over our observation period. Appendix Table B.6 adds tenure within the firm (conditional on age) and the employee's final wage at their previous job to the model. Both seniority and higher wages at the previous firm reduce the likelihood of being classified as non-White by the employer.

Turning again to the effect of skin tone in Table 2, we find that the magnitude of the skin tone coefficient decreases by approximately 40% in column 6 compared to column 1. The skin tone elasticity attenuates as individuals with varying skin tones differ in worker and job

characteristics that also matter for employers' race perceptions. Once we compare a dark-skinned and a light-skinned individual with the same characteristics, the perceived identity contrast between them is far less pronounced than the unconditional contrast between two individuals drawn at random from the population.

The advantage of our preferred specification in column 6 is that we can hold the observer constant. This means that we capture all components that are common to workers in the same occupation and firm, accounting for observer-specific race thresholds, the available information set (including other workplace characteristics), and the weights associated with each cue. This allows us to mitigate a confounding mechanism that attributes differences in worker-characteristics to differences in classification practices across observers. These results also emphasize that racial perceptions are formed at the granular employer-worker pair level.

Other Cues. Skin tone is only one of many cues employers can draw on. A distinctive feature of our data is that we observe multiple potential racial cues for the same worker. This allows us to assess how much of the insights of Table 2 are specific to skin tone, and how much can be generalized to other cues. First, we examine workers' names as racial cues using the non-White name distinctiveness index described in Section 3.2, computed separately for the first given name and the last surname. Column 1 of Table 3 replicates our most stringent specification with skin tone, while column 2 presents coefficients for the name indices. We find that more racialized names significantly predict the current employer's ascription within the same establishment and occupation. The surname index is the stronger predictor: moving from a surname found only among White workers to one found only among non-White workers raises the probability of a non-White ascription by 29.8 percentage points, compared to 6.7 percentage points for the same change in the given name index. Surnames are inherited and therefore carry information on ancestry and, in Brazil, on the national origin of European immigrant families. For instance, the most frequent surnames classified as White are "Becker" and "Schneider" (see Appendix C.4). In contrast, given names are chosen by parents and, conditional on the surname, retain less independent racial content.

Self-declared race is the measure most observational studies rely on. As discussed in Section 2, it is the appropriate measure when the object of interest is disparity by a person's own identity. However, when it stands in for how others perceive the person, it is implicitly treated as a racial cue that decision-makers adopt, either because both read the same underlying characteristics or because workers convey their identity in ways employers incorporate. Column 3 measures how far this holds. Within the same establishment and occupation, workers who declare themselves non-White are 25.9 percentage points more likely to be coded non-White by their employer than workers who declare themselves White.

Thus, self-declaration is a strong predictor for employer-ascribed race but the relation is far from one to one, such that the mismatch recorded previously remains within fine-grained establishment-occupation cells.

Table 3: Determinants of Employer-Ascribed Race: Other Racial Cues

	(1)	(2)	(3)	(4)
Skin Tone	0.068***			0.040***
	(0.001)			(0.001)
Non-White Index Given Name		0.067***		0.023**
		(0.012)		(0.011)
Non-White Index Surname		0.298***		0.147***
		(0.008)		(0.008)
Non-White Self-Declared			0.259***	0.230***
			(0.003)	(0.003)
N	1,208,444	1,208,444	1,208,444	1,208,444
Time FE	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes
Firm × Occupation FE	Yes	Yes	Yes	Yes
R-squared	0.781	0.778	0.792	0.795
Adjusted R-squared	0.721	0.717	0.736	0.739

Notes: The table reports OLS estimates of the relationship between employer-ascribed race (an indicator equal to one for non-White) and various racial cues. All columns include year dummies, demographic covariates, and firm × three-digit occupation fixed effects, mimicking column (6) of Table 2. Standard errors clustered at the individual level are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Column 4 examines the joint role of skin tone, names, and self-identity in determining employer-ascribed race. The attenuation of skin tone, from 0.068 in column 1 to 0.040 in column 4 suggests that only about 40 percent of the skin tone gradient in the most demanding specification is shared with names and self-declared race, and the remainder is specific to skin tone as captured in the photograph.[14] In addition, the coefficients on the name indices and self-ID remain statistically significant, and each cue retains a substantial share of its stand-alone magnitude: 30 percent for the given name index, 50 percent for the surname index, and 90 percent for self-declared race. This suggests that no cue is subsumed by the other, either because all employers use several cues simultaneously to infer race or because different employers rely on different cues.

Importantly, two central conclusions from our skin tone results extend to other cues. First, the coefficient on the racial cue attenuates once other worker and job characteristics are controlled for. Appendix Table B.7, Panel A, replicates the full sequence of specifications

[14]This is primarily driven by the inclusion of self-ID but not names, suggesting only a moderate correlation between skin tone and name distinctiveness, once we account for sorting across firms and occupation.

in Table 2 with the name indices in place of skin tone: the surname coefficient falls by 58 percent between the first and the last column and the given-name coefficient by 88 percent.[15] Panel B of Appendix Table B.7 shows all specifications with self-declared race: similar to skin tone, the coefficient on self-ID falls by 40 percent between the first and the last column.[16]

Second, education continues to predict employers' racial ascription. Panel C of Appendix Table B.7 shows that education matters, even conditional on the racial distinctiveness of names, self-identification, and skin tone. Employers therefore depart from workers' self-identity, phenotypic traits and markers of ancestry, when they are of higher social status. This is, of course, a conservative test of the role of socio-economic status, since all of these racial cues correlate with social status, including family background or social networks.

We emphasize that this association need not capture the causal effect of an exogenous increase in education. Rather, education may proxy for a broader bundle of socio-economic characteristics that vary with education and shape how employers racialize workers, such as income, dress, or speech patterns. The substantive implication is unchanged: racial ascription responds to social-status cues beyond phenotype. In the next subsection, we discuss additional tests that support this interpretation.

5.3 Interpretation and Robustness

Education and Other Phenotypic Cues. One potential concern is that additional phenotypic features correlate with education and simultaneously determine employers' racial ascription. Among workers with the same measured skin tone, other phenotypical features such as facial morphology and hair texture could still be correlated with education. If this is true, the education coefficient would partly capture phenotype that our skin tone measure does not hold constant. To address this concern, we replace skin tone with a machine-learning prediction of employers' racial ascription from the full photograph (see Appendix C.3 for details). The prediction summarizes all visual information associated with employer ascription, including facial features and hair, but also photo quality and lighting.

The visual-cue measure explains about 17% of the variation in ascribed race, roughly three times the explanatory power of skin tone alone. Conditioning on it is a demanding test, because it absorbs not only phenotype but also manipulable features of presentation

[15] Names – in particular given names – differ from skin tone in one important respect: they may signal socio-economic status directly, not only through racial ascription. The correlation between social background and names can explain why we see the strong attenuation when adding controls.

[16] As mentioned before, self-declared race differs from skin tone in that it is itself the output of a racialization process and responds to the same socio-economic characteristics that shape employer ascription. Appendix Table B.8 estimates this relationship directly: education and local context shift self-ID as much as they shift employer ascription, though with different weights. Self-declared race is therefore subject to the same endogeneity concerns when measuring disparities, and is an imperfect proxy for outside perceptions.

that may themselves reflect social status. Appendix Table B.9 shows that education remains significantly associated with ascribed race in every specification: in the most demanding specification, university graduates are 1.8 percentage points less likely to be recorded as non-White, compared to 3.2 percentage points when conditioning on skin tone (column 6 of Table 2). These results assuage the concern that additional phenotypic features correlated with social status drive the education gradient.

Manipulation of Skin Tone in Photographs. A second concern is that candidates choose the photograph they submit and may present their skin tone strategically. The photographs are highly standardized: the TSE requires a frontal bust portrait in attire appropriate for an official document, and the image is not used in campaigning but displayed on the electronic voting machine, next to the candidate's name, number, and party, for the voter to check before confirming the vote (Fujiwara, 2015).

Candidates may nonetheless have incentives to present a particular way, and this would confound the education coefficient if more educated candidates presented a darker skin tone than they would otherwise. To examine the potential for skin tone manipulation, we consider changes in skin tone within the sample of candidates who run more than once. We test whether candidates who change their self-declared race between elections – and may therefore have particularly strong incentives to manipulate their appearance, for example, in response to changes in their electorate (Janusz, 2021) – also exhibit changes in measured skin tone. We regress the change in skin tone between consecutive candidacies on the change in self-declared race, separately for candidates moving from White to non-White and from non-White to White. Appendix Table C.2 shows that changes in self-declared race are not associated with changes in measured skin tone in either direction.

6 Racial Wage Disparities

Our results so far are consistent with a constructivist view of race: racial identities diverge between workers and employers and across employers, and racial ascription is a function of socio-economic cues. In the following, we examine the implications for measured disparities. In a first step, we present estimates of racial wage disparities using employer-ascribed race, skin tone, and a perception-normalization approach. In a second step, we illustrate how accounting for race perceptions improves the comparability of cue-based disparities by examining *i*) conditional versus unconditional contrasts *ii*) heterogeneity across workers and contexts and *iii*) different racial cues. Lastly, we discuss the interpretation and limitations of our results.

6.1 Estimates of Racial Disparity

Conventional Disparity. We start by quantifying racial wage gaps that are commonly estimated using RAIS data. Panel A of Table 4 shows differences in log hourly wages between workers who are ascribed White vs. non-White by their employer. In column 1, conditioning only on year fixed effects, workers classified as non-White earn 14.8% less than workers classified as White.[17] This unconditional gap captures the disparate impact of race which entails both direct and systemic forms of discrimination (Bohren et al., 2025b).

Researchers often run "benchmarking regressions" to account for productivity differences among workers. They seek to quantify direct discrimination (both taste-based and statistical), understood as an employer's disparate treatment based on race. Columns 2 to 6 implement a standard benchmarking sequence, adding the same covariates and fixed effects used in our race-function analysis (Table 2). The coefficient attenuates substantially when we control for sociodemographics and region fixed effects in columns 2 and 3, and shrinks further once we hold fixed sorting across industries and occupations in columns 4 and 5. The most stringent specification in column 6 compares workers within the same firm and three-digit occupation; even within these narrow cells, being classified as non-White is associated with a 2.1% lower wage.

The large attenuation across columns is not, by itself, evidence that "discrimination disappears" once we benchmark. Rather, controls also partial out determinants of racial classification itself. Therefore, it may not be obvious what the residual variation in race perceptions entails and how this changes the interpretation of the conventional disparity estimates. We return to this issue in Section 6.2.

Colorism Disparity. An alternative approach that is more explicit about the exploited variation focuses on one specific racial cue that affects employers' racial perceptions. We focus on our algorithmic measure of skin tone. Panel B of Table 4 quantifies the direct colorism disparity. In column 1, a one standard deviation darker skin tone is associated with 2.6% lower wages.[18] This estimate is essentially unchanged when adding demographics, region

[17] Appendix Table B.10 expresses wage gaps in terms of workers' self-declared race. The unconditional wage gap for self-declared race lies at 16.3% and is thus modestly larger than for employer-ascribed race.

[18] Existing estimates for Latin America rely on interviewer-rated skin tone palettes and broader economic outcomes, which limits direct comparison. Using the 11-point PERLA palette in the AmericasBarometer, Woo-Mora (2026) finds that a one-step darker skin tone is associated with about 4% lower household income per capita, conditional on location-by-interviewer fixed effects and demographics. For Brazil, Monk (2016) finds that a one-step darker tone, roughly one standard deviation of the palette, lowers the odds of higher educational attainment by 26% and is associated with lower occupational status conditional on education. Our estimates differ in three respects: the outcome is hourly wages in formal employment rather than household income or attainment, skin tone is measured algorithmically rather than by an interviewer, and our most stringent specifications compare workers within the same firm and occupation.

Table 4: Wage Disparities

	(1)	(2)	(3)	(4)	(5)	(6)
[A] OLS: Conventional Disparity						
Non-White Employer-Ascribed	-0.148***	-0.103***	-0.054***	-0.064***	-0.038***	-0.021***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
[B] Reduced Form: Colorism Disparity						
Skin Tone	-0.026***	-0.032***	-0.026***	-0.028***	-0.015***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
[C] First Stage: Racialization						
Skin Tone	0.110***	0.107***	0.082***	0.081***	0.078***	0.068***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
[D] 2SLS: Perception-Normalized Disparity						
Non-White Employer-Ascribed	-0.235***	-0.301***	-0.314***	-0.348***	-0.193***	-0.080***
	(0.011)	(0.010)	(0.014)	(0.014)	(0.012)	(0.013)
KP F-stat	17,195.0	15,429.4	10,082.1	9,713.9	9,082.4	2,820.5
N	1,510,833	1,510,833	1,510,833	1,510,510	1,510,445	1,249,942
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	No	Yes	Yes	Yes	Yes	Yes
Region FE	No	No	Yes	Yes	Yes	No
Industry FE	No	No	No	Yes	No	No
Industry × Occupation FE	No	No	No	No	Yes	No
Firm × Occupation FE	No	No	No	No	No	Yes

Notes: The table reports estimates of racial wage gaps based on employer-ascribed race. All regressions include year fixed-effects. The first column always shows the unconditional relationship (apart from year fixed-effects), and the remaining columns progressively add controls. Panel A shows coefficient estimates of OLS regressions of log hourly wage on employer-ascribed race. Panel B shows reduced-form regressions of log hourly wage on skin tone. Panel C presents the first stage coefficient from regressing employer-ascribed race on skin tone. Finally, Panel D shows 2SLS estimates of regressing log hourly wage on employer-ascribed race, instrumented by skin tone. Standard errors clustered at the individual level are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

fixed effects, and industry fixed effects, but shrinks substantially when adding occupation by industry or occupation by firm fixed effects. In the most stringent specification in column 6, a one standard deviation darker skin tone is associated with 0.5% lower wages.[19]

Perception-Normalized Disparity. The colorism disparity measures wage gaps along the skin tone distribution, but skin tone racializes workers to different degrees depending on the context. Because we observe both the colorism disparity and how strongly skin tone

[19]Notably, declines in the reduced form track declines in the first stage across columns. For instance, adding region fixed effects lowers both by about 20 percent. We return to this point in the following subsection.

shifts employer-ascribed race, we can scale the former by the latter. Panel D reports the resulting PND. Column 1 shows a large unconditional perception-normalized disparity of 23.5%. Adding demographics in column 2 and then region fixed effects in column 3 leaves the PND large and, if anything, increases it in magnitude. The estimate becomes largest in column 4 when we additionally include industry fixed effects, consistent with the idea that once we compare workers within broad sectors the remaining wage differences associated with skin-tone based perception are even larger. Introducing industry-by-occupation fixed effects in column 5 reduces the estimate considerably, indicating that a meaningful share of the wage gap operates through sorting into occupations. In the most stringent specification, which compares workers with the same demographics who work in the same job (column 6), we estimate a PND of 8.0%.

The PND is primarily a scaling exercise that expresses cue-based disparities in units of racial perception rather than units of the cue itself. This normalization facilitates comparisons across specifications, across settings in which the same cue shifts racial perceptions to different degrees, and across racial cues. Section 6.2 illustrates this by showing how the racialization margin affects measured disparities *i*) in conditional versus unconditional contrasts, *ii*) across education groups and regions, and *iii*) across skin tone, distinctively non-White names, and self-declared race. In our most stringent specification, the PND is about four times as large as the wage gap based directly on employer-ascribed race. This difference could be attributed to several factors: the endogeneity of employer perceptions, measurement error in perceptions, and the greater weight the PND places on worker-employer pairs whose perceptions respond to skin tone. Section 6.3 clarifies the interpretation of the PND by *i*) discussing the assumptions required for a causal interpretation, *ii*) assessing the role of measurement error, and *iii*) characterizing the complier population.

6.2 Accounting for Differences in Racialization

Comparability Across Unconditional and Conditional Contrasts. Racial disparity estimates based on employer-ascribed race and skin tone decline substantially as we add controls, whereas the PND falls much less. Panel A of Table 4 shows that the employer-ascribed race wage gap shrinks by 86% in the within-firm specification relative to the unconditional gap. Part of this decline reflects differences in productivity-related characteristics. But, as discussed in Section 2, part of it reflects netting out the racial identity contrast itself: the same covariates used to proxy for productivity also enter the racial perception process. Interpreting employer-ascribed race disparities therefore requires making explicit that the

estimand depends on the residual variation in recorded race after conditioning on controls.

Panel B of Table 4 shows an analogous pattern for skin tone. The reduced-form colorism estimates shrink by about 81% when moving from column 1 to column 6. Under the constructivist framework, this decline can reflect two mechanisms simultaneously. Conditioning can make workers more comparable in productivity-relevant dimensions, but it can also make skin tone less predictive of employer-ascribed race. As shown in Table 2 (replicated in Panel C of Table 4), the relationship between skin tone and employer-ascribed race falls by about 40% when moving from the unconditional specification to one with demographics and detailed firm-by-occupation fixed effects. The elasticity declines as skin tone is correlated with other worker and job characteristics that enter the employer's classification rule.

The PND approach explicitly takes into account the attenuation of the first stage. After scaling the reduced form by the corresponding change in perceptions, the PND declines by about 66% between columns 1 and 6 in Panel D – a decline that is about 20% smaller than that for the employer-ascribed or colorism disparity. These results highlight how conventional conditional disparity estimates can overstate the role of productivity differences (and other differences in worker and job characteristics) in explaining observed wage gaps. Moreover, they emphasize that measuring and incorporating the first stage of race perceptions is crucial to allow for differences in the racialization margin under conditional contrasts.

Comparability Across Workers and Regions. Heterogeneity in cue-based disparity is not automatically evidence of heterogeneous discrimination. This is because the same racial cue need not enter the employer's classification rule with the same weight for all workers and contexts. Appendix Figures A.3 and A.4 illustrate this point for heterogeneity by education and region, respectively. Each figure reports group-specific estimates of wage disparities using the relevant race measure.

Considering first heterogeneity by education, we see that skin tone wage disparities are 23–64% larger among university-educated workers than among those without a high school degree in specifications with at least demographic controls. However, given that the elasticity of employers' race perceptions with respect to skin tone is smaller among university-educated workers than among low-educated workers, the PND — which takes into account these differences in skin tone elasticity — yields even larger gaps among university-educated workers (30-90% larger). Regional heterogeneity reinforces this point. For instance, skin tone disparities do not differ significantly in the North and the Southeast. However, the skin tone elasticity of employer-ascribed race is more than twice as high in the Southeast than in the North, a pattern that gets even stronger within stringent fixed-effect cells. Taking into account these differences in racialization, the PND estimates show that perception-normalized

disparities are in fact significantly smaller in the Southeast than in the North.

Taken together, these results underscore that heterogeneity in reduced-form disparities need not reflect heterogeneity in discrimination. Differences in how strongly skin tone maps into perceived race can mask or exaggerate group-level comparisons unless the first stage is explicitly accounted for.

Comparability Across Racial Cues. Section 5 showed that names and self-declared race shape employers' racial perceptions alongside skin tone. Table 5 compares disparities across these cues in the most demanding specification: the conventional disparity based on employer-ascribed race, the reduced-form disparities for skin tone, self-declared race, and the two name indices in columns 2 to 4. Within the same firm and occupation, a one standard deviation darker skin tone is associated with 0.5% lower wages, and workers who self-declare as non-White earn 2.0% less, close to the conventional disparity of 2.1%. For names, the penalty operates through surnames: a worker whose surname is carried exclusively by non-White workers earns 5.0% less than one whose surname is carried exclusively by White workers, whereas given names have no detectable association with wages. Skin tone, surnames, and self-declared race remain significant when entered jointly (Appendix Table B.14, Panel B). While existing studies typically isolate a single cue, these results show that several racial cues carry wage penalties among workers in the same job.

Table 5: Wage Disparities: Other Racial Cues

	OLS: Conventional	Reduced Form: Cue-Based			2SLS: Perception-Normalized		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Racial cue		Skin tone	Self-decl.	Names	Skin tone	Self-decl.	Names
Employer-Ascribed	-0.021***				-0.079***	-0.076***	-0.164***
	(0.002)				(0.013)	(0.008)	(0.024)
Skin Tone		-0.005***					
		(0.001)					
Self-Declared			-0.020***				
			(0.002)				
Index Surname				-0.050***			
				(0.007)			
Index Given Name				-0.005			
				(0.009)			
KP F-stat					2,684.5	7,529.8	687.5
N	1,208,444	1,208,444	1,208,444	1,208,444	1,208,444	1,208,444	1,208,444
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm × Occupation FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table reports different estimates of racial wage gaps based on employer-ascribed race and various racial cues. Column (1) shows coefficient estimates of OLS regressions of log hourly wage on employer-ascribed race. Columns (2)-(4) show reduced-form regressions of log hourly wage on the respective racial cue. Columns (5)-(7) present 2SLS estimates of regressing log hourly wage on employer-ascribed race, instrumented by each racial cue. All columns include year dummies, demographic covariates, and firm × three-digit occupation fixed effects, mimicking column (6) of Table 4. Standard errors clustered at the individual level are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Because the cues are measured in different units, however, the reduced forms cannot be compared directly. The corresponding PNDs, each instrumenting employer-ascribed race with one cue, are reported in columns 5 to 7.[20] Expressed in units of racial perception, the skin tone and self-declared race PNDs are nearly identical, at 7.9% and 7.6%, while the name-based PND is about twice as large, at 16.4%. All three exceed the conventional within-job disparity by a wide margin. Section 6.3 examines whether these differences reflect who is racialized through each cue.

6.3 Interpretation and Robustness

Scaling vs. Causal Interpretation. The PND scales a reduced-form disparity by how strongly the racial cue shapes employers' racial perceptions. A stronger reading interprets the PND as the causal effect of being perceived as non-White, as illustrated by the following stylized wage equation:

$$Y \;=\; \beta R^E \;+\; \delta z \;+\; \nu \;+\; X'\theta \;+\; \varepsilon,$$

where R^E is employer-ascribed race, z is the racial cue, and ν is productivity unobserved by the researcher, which may be correlated with the cue. The parameter β is the object of interest: how the same worker would be paid if the employer classified them differently. The parameter δ captures any direct effect of the cue on wages that does not run through the employer's racial category. Comparing two values of the cue, z_0 and z_1, and dividing by the first stage $\pi_z(x)$, defined in Section 2, yields

$$\text{PND}_z(x) \;=\; \beta \;+\; \frac{\delta(z_1 - z_0) + \lambda_z(x)}{\pi_z(x)},$$

where $\lambda_z(x)$ is the difference in unobserved productivity across values of the cue. The PND therefore equals β under two conditions: the cue is unrelated to unobserved productivity conditional on controls ($\lambda_z = 0$), and it affects wages only through employer-ascribed race ($\delta = 0$). A non-zero λ_z is a conventional omitted-variable problem: the PND would attribute a return to productivity to perceived race. A non-zero δ instead reflects employers' preferences over, or inferences from, the cue itself, for instance valuing lighter skin, independent of the racial category they assign. If darker skin is associated with lower unobserved productiv-

[20] The first stages are shown in Table 3. The sample slightly differs from Table 4 because of the sample restriction required for the name indices. Appendix Tables B.12 and B.13 replicate the attenuation pattern across specifications: moving from the unconditional specification to firm-by-occupation cells, the name-based PND declines by 61% and the self-declared-race PND by 81%, compared to 85% and 88% for the respective cue-based disparities.

ity or penalized directly by employers, the PND provides an upper bound on the magnitude of the wage penalty from being perceived as non-White.

We examine the plausibility of these assumptions empirically for the PND based on skin tone. The first condition fails if skin tone is correlated with unobserved productivity, for instance, through family or socio-economic background ($\lambda_z \neq 0$). Names and self-declared race are strong correlates of family background and identity, so removing the part of the skin tone variation correlated with them is a demanding test.[21] Appendix Table B.14 re-estimates the skin tone PND controlling for the name indices, self-declared race, and both. If the correlation of skin tone with background inflated the PND, removing the background-correlated part of the skin tone variation should shrink the estimate: the reduced form would fall by more than the first stage. Instead, both fall by the same 40%, and the PND is essentially unchanged: 7.9% unconditionally, 7.4% conditional on names, 8.1% conditional on self-declared race, and 7.7% conditional on both.

The second condition fails if employers reward lighter skin directly, or infer productivity from it, independent of the racial category they assign ($\delta \neq 0$).[22] Empirically, a direct skin tone effect has a testable implication: it produces a wage gradient in skin tone even where skin tone does not move racial classification. We have shown in section 5 that skin tone moves employer-ascribed race more than twofold between the Southeast and the North, and more for less- compared to more-educated workers. We exploit the fact that the first stage varies substantially across groups and examine whether the reduced-form effect of skin tone on wages varies proportionately with its effect on racial perceptions.

Appendix Figure A.5 plots the estimated first stage against the reduced form across state-education cells, weighting cells by the number of wage observations and using the full set of fixed effects and controls. As expected, there is substantive heterogeneity in the elasticity of skin tone in employer-recorded race. The fitted line has slope -0.051 and intercept -0.002 (the dashed line indicates the PND for the full sample). Assuming that the direct effect is approximately constant across those first stage cells, the intercept of the fitted line yields an estimate of direct skin-tone preferences. The intercept is statistically indistinguishable from

[21] Employers' beliefs about workers' social status are part of β when they operate through racial ascription (statistical discrimination based on perceived race). Status inferred from skin tone itself falls into δ, and names may carry status signals independent of race.

[22] A long tradition in social psychology and economics holds that individuals interpret, reason, and act through the lens of social categories (e.g., Charness and Chen, 2020; Shayo, 2020; Tajfel, 1981; Taylor et al., 1978; Tajfel et al., 1971). This suggests that skin tone and other characteristics determine the ascribed racial category and decision-makers act upon that ascribed category. Colorism studies document skin tone disparities among individuals within the same racial categories (e.g., Woo-Mora, 2026; Monk, 2016; Kreisman and Rangel, 2015; Monk, 2014; Goldsmith et al., 2006), but such conditional analyses also yield biased estimates if racial categories are endogenous objects (Adukia et al., 2026). For the same reason, we do not use the within-category tone gradient in our own data.

zero, with bootstrapped standard error of 0.007, as predicted if skin tone affects wages only through employer-assigned race.

As an additional check, Appendix Figure A.6 examines the first stage and reduced form at the tails of the skin tone distribution, using our most stringent specification. Among very light- or dark-skinned workers, differences in skin tone may be insufficient to shift employers' racial classifications. Indeed, skin tone does not significantly predict employer-ascribed race among the most dark-skinned workers, and only noisily in the lighter tail. In the absence of a first stage, any remaining reduced-form effect of skin tone would suggest direct skin tone effects rather than an effect mediated by racialization. We find no negative reduced-form effect of skin tone on wages in these parts of the distribution that do not produce a significant first stage, consistent with the effect operating through employers' racial perceptions rather than through a direct effect of skin tone.[23]

Finally, the race category recorded by the employer is an observable proxy for a broader process of racialization, not itself the object that determines wages. For instance, a purely clerical change to this field, holding people's beliefs and behavior fixed, would not change a worker's pay. It therefore serves as an approximation of racial perceptions in the wage-setting environment. In addition, skin tone may also shape how coworkers or customers perceive the worker. In this case, the PND is best understood as a wage disparity based on workplace perceptions more broadly, but anchored in employer-recorded race.

Compliers. The conventional disparity averages over all workers whose employers record them as non-White. The PND, by contrast, captures disparities for compliers, that is, workers whose racial classification responds to the cue. Differences between the two estimates, and between PNDs based on different cues, may therefore partly reflect who is racialized rather than how strongly racial perceptions are penalized. Appendix Table B.15 characterizes compliers for each cue following Hull (2025), alongside the sample mean, in the firm-by-occupation specification.

Compliers differ from the average worker in two respects common to all three cues. First, they work in much larger and higher-paying establishments: mean log establishment size is between 6.2 and 6.3 among compliers against 4.3 in the full sample, and compliers are concentrated in manufacturing and finance rather than retail. Second, they are concentrated in the Southeast (64–72% against 51%) and underrepresented in the North and Northeast. The PND thus speaks to a population that differs systematically from the one underlying the conventional disparity.

[23] Among the 5% most light-skinned workers, skin tone does predict employer-ascribed race but the estimate is comparatively noisy. If anything, the reduced form in the lighter tail is slightly positive, which would bias the PND toward zero rather than inflate it.

Across cues, complier groups differ in education and earnings, in the direction consistent with the differences in the PNDs documented in Section 6.2. Skin tone compliers are less educated than the average worker: 4.0% hold a university degree against 8.9% in the full sample, and they are over-represented among workers with elementary schooling. Name compliers are the mirror image: 10.3% hold a university degree, their hourly wage is 19 log points above the sample mean and 16 log points above that of skin tone compliers, and they work in establishments with higher average wages. Self-declared-race compliers lie in between but closer to skin tone compliers (5.6% university graduates), consistent with their similar PND. Since racial wage gaps are larger among the highly educated (Appendix Figure A.3), part of the larger name-based PND reflects who is racialized through names rather than a stronger penalty per se. Name compliers are also the only group over-represented in the South (26% against 20%), the region with the highest share of distinctively European surnames that strongly signal White racial ascription.

Measurement Error. Perhaps most obviously, racial wage gaps may be attenuated when race variables or racial cues are measured with noise. To quantify the role of measurement error in skin tone, we implement a split-sample IV design that uses individuals with multiple photographs from a subsample who ran for office at least twice. For each individual, we randomly split their photos into two halves, compute average skin tone in each half, and instrument the average from one half with the average from the other. Appendix Table B.11 reveals a substantial role of measurement error in attenuating skin tone differences. The unconditional split-sample IV results indicate that a one-standard-deviation darker skin tone reduces wages by 11.4% (versus 5.1% in OLS) and increases the probability of being reported non-White by the employer by 34.6 percentage points (versus 15.2 percentage points in OLS). We find the same pattern across all specifications with controls: the split-sample IV coefficients are always more than twice as high as the corresponding OLS coefficients.

A useful feature of the PND is its invariance to classical measurement error in skin tone. Noise in the skin tone measure attenuates the reduced-form and first-stage by the same factor, thus leaving their ratio unchanged. Consistent with this, the PND implied by the OLS and split-sample IV estimates in Appendix Table B.11 is nearly identical, despite large differences in the reduced-form and first-stage levels.[24]

The PND approach may also help bound the bias from measurement error in employer-ascribed race. On the one hand, measurement error may attenuate conventional OLS estimates in which ascribed race is the binary explanatory variable. On the other hand, PND estimates of disparities may be inflated because they scale reduced-form effects of skin tone

[24]They are not exactly equal because of finite-sample behavior of the SSIV.

by the first-stage effect of skin tone on employer ascription, and this first-stage relationship is itself attenuated by noise in the binary outcome (see, e.g., Bingley and Martinello (2017) on bias in IV estimators with a mis-measured binary endogenous regressor). Thus, if the gap between OLS and PND estimates were driven solely by measurement error in ascribed race, the true perception-based disparity is expected to lie between the two.

7 Conclusion

Anti-discrimination law protects race as it is perceived by the decision-maker, yet empirical research on racial disparities typically treats race as a fixed attribute of the individual. This paper takes the perception of the decision-maker seriously. Using linked administrative data from Brazil, we show that self-declared and employer-recorded race diverge for more than a quarter of workers, that employers disagree on the classification of the same worker, and that social status shapes employers' classification conditional on skin tone, self-declared race, and name, even within the same firm and occupation. Racial classification thus responds to the very characteristics along which disparities are usually benchmarked.

This has three implications for the measurement of racial disparities. First, which classification a study uses matters, because self-perceptions and decision-makers' perceptions can diverge. Second, controls for productivity also absorb determinants of racial classification. When the perception process is ignored, disparities are therefore easily attributed to productivity differences that partly reflect how workers are classified. Third, cue-based estimates combine the strength of racialization with the penalty for perceived race, so that differences across groups or settings need not reflect differences in racial discrimination based on the cue. The perception-normalized disparity separates these two margins. We view it primarily as a way to scale cue-based disparities by racial perception and to make them comparable across settings, groups, and cues.

Our results point to several directions for future research. Experimental designs could manipulate the cues that shape racial perception, while eliciting perceptions themselves, to study how decision-makers map observable characteristics into categories. Our evidence that classification norms vary across employers, regions, and time also suggests studying how these norms form and change, and how they interact with policies that rely on racial classification, such as affirmative action. Finally, the same logic applies to other settings in which decisions rely on socially perceived categories, such as judicial sentencing, school admissions, or credit markets, and to other categories, including religion, caste, and migrant status. Measuring inequality along these dimensions requires treating them similarly, namely as categories produced by the perceptions of those who make decisions.

Declarations. We gratefully acknowledge funding from the George & Obie Shultz Fund at MIT and declare no conflict of interest.

References

Abel, Sarah, "Objectivity and objectification. On the ethics and epistemologies of skin colour measurements in the social sciences," *Studies in History and Philosophy of Science*, 2026, *115*, 102112.

Adukia, Anjali, Alex Eble, Emileigh Harrison, Hakizumwami Birali Runesha, and Teodora Szasz, "What We Teach About Race and Gender: Representation in Images and Text of Children's Books," *The Quarterly Journal of Economics*, 2023, *138* (4), 2225–2285.

_ , Richard Hornbeck, Daniel Keniston, and Benjamin Lualdi, "The Social and Statistical Construction of Race," *NBER Working Paper 33502*, 2026. Updated May 2026.

Aigner, Dennis J. and Glen G. Cain, "Statistical Theories of Discrimination in Labor Markets," *Industrial and Labor Relations Review*, 1977, *30* (2), 175–187.

Ajzenman, Nicolás, Bruno Ferman, and Pedro C Sant'Anna, "Discrimination in the formation of academic networks: A field experiment on #EconTwitter," *American Economic Review: Insights*, 2025, *7* (3), 357–375.

Antman, Francisca and Brian Duncan, "Incentives to identify: Racial identity in the age of affirmative action," *Review of Economics and Statistics*, 2015, *97* (3), 710–713.

_ and _ , "American Indian Casinos and Native American Self-Identification," *Journal of the European Economic Association*, 2023, *21* (6), 2547–2585.

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

_ , _ , and Peter Hull, "Measuring Racial Discrimination in Bail Decisions," *American Economic Review*, 2022, *112* (9), 2992–3038.

Arrow, Kenneth J., "The Theory of Discrimination," in Orley Ashenfelter and Albert Rees, eds., *Discrimination in Labor Markets*, Princeton, NJ: Princeton University Press, 1973, pp. 3–33.

Atkin, David, Eve Colson-Sihra, and Moses Shayo, "How do we choose our identity? A revealed preference approach using food consumption," *Journal of Political Economy*, 2021, *129* (4), 1193–1251.

Baerlocher, Diogo and Rodrigo Schneider, "Racial self-classification, group consciousness, and public employment representation," *Journal of Development Economics*, 2026, *181*, 103755.

Bailey, Stanley, Aliya Saperstein, and Andrew Penner, "Race, Color, and Income Inequality Across the Americas," *Demographic Research*, 2014, *31*, 735–756.

Bailey, Stanley R., Mara Loveman, and Jeronimo O. Muniz, "Measures of "Race" and the analysis of racial inequality in Brazil," *Social Science Research*, 2013, *42* (1),

106–119.

Bayer, Patrick and Kerwin Kofi Charles, "Divergent Paths: A New Perspective on Earnings Differences Between Black and White Men Since 1940," *The Quarterly Journal of Economics*, 2018, *133* (3), 1459–1501.

_ , _ , and Ellora Derenoncourt, "Racial Inequality in the Labor Market," in Christian Dustmann and Thomas Lemieux, eds., *Handbook of Labor Economics*, Vol. 6, Elsevier, 2025, pp. 159–228.

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

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

Bingley, Paul and Alessandro Martinello, "Measurement Error in Income and Schooling, and the Bias for Linear Estimation," *Journal of Labor Economics*, 2017, *35* (4), 1117–1148.

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

_ , Kareem Haggag, Alex Imas, and Devin G. Pope, "Inaccurate Statistical Discrimination: An Identification Problem," *Review of Economics and Statistics*, 2025, *107* (3), 605–620.

_ , Peter Hull, and Alex Imas, "Systemic Discrimination: Theory and Measurement," *The Quarterly Journal of Economics*, 2025, *140* (3), 1743–1799.

Bueno, Natálie S and Thad Dunning, "Race, Resources, and Representation: Evidence from Brazilian Politicians," *World Politics*, 2017, *69* (2), 327–365.

Card, David, Ana Rute Cardoso, Jörg Heining, and Patrick Kline, "Firms and Labor Market Inequality: Evidence and Some Theory," *Journal of Labor Economics*, 2018, *36* (S1), S13–S70.

Carneiro, Sueli, "Negros de pele clara," *Correio Braziliense*, 2004, p. 21. 29 May. Available at: https://www.geledes.org.br/negros-de-pele-clara-por-sueli-carneiro/. Accessed November 08, 2026.

Cassan, Guilhem, "Identity-based policies and identity manipulation: Evidence from colonial Punjab," *American Economic Journal: Economic Policy*, 2015, *7* (4), 103–131.

Cattaneo, Matias D, Richard K Crump, Max H Farrell, and Yingjie Feng, "On binscatter," *American Economic Review*, 2024, *114* (5), 1488–1514.

Charles, Kerwin Kofi and Jonathan Guryan, "Prejudice and Wages: An Empirical Assessment of Becker's *The Economics of Discrimination*," *Journal of Political Economy*, 2008, *116* (5), 773–809.

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

Charness, Gary and Yan Chen, "Social Identity, Group Behavior, and Teams," *Annual*

Review of Economics, 2020, *12*, 691–713.

Cornwell, Christopher, Jason Rivera, and Ian M. Schmutte, "Wage Discrimination When Identity Is Subjective: Evidence from Changes in Employer-Reported Race," *Journal of Human Resources*, 2017, *52* (3), 719–755.

Crabtree, Charles, Jae Yeon Kim, S. Michael Gaddis, John B. Holbein, Chelsea Guage, and Wesley W. Marx, "Validated Names for Experimental Studies on Race and Ethnicity," *Scientific Data*, 2023, *10* (130).

de Holanda, Sérgio Buarque, *Raízes do Brasil*, Rio de Janeiro: Livraria José Olympio Editora, 1936.

de Lucena Coelho, Thiago, Fernanda Estevan, Marcos Nakaguma, and Alexandre Rabelo, "Do Black Politicians Matter? Political Leadership and Racial Composition in Top Public Sector Positions," *Political Leadership and Racial Composition in Top Public Sector Positions*, 2024.

do Valle Silva, Nelson, "A research note on the cost of not being white in Brazil," *Studies in Comparative International Development*, 2000, *35*, 18–27.

Evsyukova, Yulia, Felix Rusche, and Wladislaw Mill, "LinkedOut? A field experiment on discrimination in job network formation," *The Quarterly Journal of Economics*, 2025, *140* (1), 283–334.

Fernandes, Florestan, *A Integração do Negro na Sociedade de Classes*, São Paulo: Dominus Editora, 1965.

Ferraz, Claudio and Frederico Finan, "Motivating Politicians: The Impacts of Monetary Incentives on Quality and Performance," *Unpublished Manuscript*, 2011.

Firpo, Sergio, Michael França, and Alysson Portella, "Racial Inequality in the Brazilian Labor Market and the Role of Education," *Available at SSRN 3967828*, 2021.

França, Michael and Alysson Portella, *Números da discriminação racial: Desenvolvimento humano, equidade e políticas públicas*, Editora Jandaíra, 2023.

Francis, Andrew M and Maria Tannuri-Pianto, "Endogenous Race in Brazil : Affirmative Action and the Construction of Racial Identity among Young Adults," *Economic Development and Cultural Change*, 2013, *61*, 731–753.

Freeman, Jonathan B, Andrew M Penner, Aliya Saperstein, Matthias Scheutz, and Nalini Ambady, "Looking the part: Social status cues shape race perception," *PloS one*, 2011, *6* (9), e25107.

Freyre, Gilberto, *Casa-Grande & Senzala*, São Paulo: Companhia das Letras, 1933.

Fryer, Roland G., "Racial Inequality in the 21st Century: The Declining Significance of Discrimination," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 4B, Elsevier, 2011, pp. 855–971.

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

Fujiwara, Thomas, "Voting Technology, Political Responsiveness, and Infant Health: Evidence From Brazil," *Econometrica*, 2015, *83* (2), 423–464.

Gaddis, S Michael, "How black are Lakisha and Jamal? Racial perceptions from names used in correspondence audit studies," *Sociological Science*, 2017, *4*, 469.

Gerard, François, Lorenzo Lagos, Edson Severnini, and David Card, "Assortative matching or exclusionary hiring? The impact of employment and pay policies on racial wage differences in Brazil," *American Economic Review*, 2021, *111* (10), 3418–3457.

Goldsmith, Arthur H., Darrick Hamilton, and Jr. Darity William A., "From Dark to Light: Skin Color and Wages Among African-Americans," *Journal of Human Resources*, 2007, *42* (4), 701–738.

_ , _ , and William Darity Jr., "Shades of Discrimination: Skin Tone and Wages," *American Economic Review*, 2006, *96* (2), 242–245.

Greiner, D James, "Causal inference in civil rights litigation," *Harvard Law Review*, 2008, *122*, 533.

Greiner, D. James and Donald B. Rubin, "Causal Effects of Perceived Immutable Characteristics," *Review of Economics and Statistics*, 2011, *93* (3), 775–785.

Guimarães, Antonio Sérgio Alfredo, "Como trabalhar com" raça" em sociologia," *Educação e pesquisa*, 2003, *29* (1), 93–107.

_ , "The Brazilian System of Racial Classification," *Ethnic and Racial Studies*, 2012, *35* (7), 1157–1162.

Hannon, Lance and Robert DeFina, "Reliability Concerns in Measuring Respondent Skin Tone by Interviewer Observation," *Public Opinion Quarterly*, 01 2016, *80* (2), 534–541.

Hersch, Joni, "Skin-Tone Effects Among African Americans: Perceptions and Reality," *American Economic Review*, 2006, *96* (2), 251–255.

_ , "Profiling the New Immigrant Worker: The Effects of Skin Color and Height," *Journal of Labor Economics*, 2008, *26* (2), 345–386.

Hu, Lily and Issa Kohler-Hausmann, "What is Perceived When Race is Perceived and Why It Matters for Causal Inference and Discrimination Studies," *Law & Society Review*, 2025, *59* (2), 239–264.

Hull, Peter, "One Weird Trick to Characterize Effective Populations in Design-Based Specifications," Technical Report, Metrics Note 2025.

IBGE, "Desigualdades Sociais por Cor ou Raça no Brasil," *Estudos e Pesquisas -Informação Demográfica e Socioeconômica*, 2022, *48* (2nd edition), 1–16.

Janusz, Andrew, "Electoral incentives and elite racial identification: Why Brazilian politicians change their race," *Electoral Studies*, 2021, *72*, 102340.

_ , "The electoral consequences of racial fluidity," *Electoral Studies*, 2023, *82*, 102597.

Jia, Ruixue and Torsten Persson, "Choosing ethnicity: The interplay between individual and social motives," *Journal of the European Economic Association*, 2021, *19* (2), 1203–1248.

Kline, Patrick, Evan K Rose, and Christopher R Walters, "Systemic discrimination among large US employers," *The Quarterly Journal of Economics*, 2022, *137* (4), 1963–

2036.

Kreisman, Daniel and Jonathan Smith, "Distinctively Black Names and Educational Outcomes," *Journal of Political Economy*, 2023, *131* (4), 877–897.

_ **and Marcos A. Rangel**, "On the Blurring of the Color Line: Wages and Employment for Black Males of Different Skin Tones," *Review of Economics and Statistics*, 2015, *97* (1), 1–13.

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

Lopez, Ian Haney, "White by law 10th anniversary edition: The legal construction of race," in "White by Law 10th Anniversary Edition," New York University Press, 2006.

Maggie, Yvonne and Peter Fry, "Qual Cor, Qual Raça?," *Estudos Avançados*, 2004, *18* (50), 57–67.

Micheli, David De, "Racial reclassification and political identity formation," *World Politics*, 2021, *73* (1), 1–51.

Miller, Conrad and Ian M Schmutte, "The Dynamic Effects of Co-Racial Hiring," *Unpublished Manuscript*, 2023.

Miranda, Vítor, "A resurgence of black identity in Brazil? Evidence from an analysis of recent censuses," *Demographic Research*, 2015, *32*, 1603–1630.

Mitrut, Andreea, Gabriel Kreindler, Margareta Matache, Andrei Munteanu, and Cristian Pop-Eleches, "Education and Selection into Ethnic Identification: Evidence from Roma People in Romania," *NBER Working Paper 34383*, 2025.

Monk, Ellis P., "Skin Tone Stratification among Black Americans, 2001–2003," *Social Forces*, 2014, *92* (4), 1313–1337.

_ , "The Consequences of "Race and Color" in Brazil," *Social Problems*, 2016, *63*, 413–430.

MTE, *Manual de Orientação da RAIS*, Brasília: Ministério do Trabalho e Emprego, 2024.

Muniz, Jerônimo, Aliya Saperstein, and Bernardo L. Queiroz, "Racial classification as a multistate process," *Demographic Research*, 2024, *50* (17), 457–472.

Noghanibehambari, Hamid and Jason Fletcher, "Passing as White: Racial Identity and Old-Age Longevity," *NBER Working Paper 33394*, 2025.

Omi, Michael, Howard Winant et al., *Racial formation in the United States*, Vol. 3, Routledge New York, 2014.

Osório, Rafael Guerreiro, "O sistema classificatório de cor ou raça do IBGE," Technical Report, Instituto de Pesquisa Econômica Aplicada (IPEA) 2003.

Penner, Andrew M and Aliya Saperstein, "How social status shapes race," *Proceedings of the National Academy of Sciences*, 2008, *105* (50), 19628–19630.

Phelps, Edmund S., "The Statistical Theory of Racism and Sexism," *American Economic Review*, 1972, *62* (4), 659–661.

Rose, Evan K, "A constructivist perspective on empirical discrimination research," *Journal*

of Economic Literature, 2023, *61* (3), 906–923.

Roth, Wendy D, Patricio Solís, and Christina A Sue, "Beyond money whitening: racialized hierarchies and socioeconomic escalators in Mexico," *American Sociological Review*, 2022, *87* (5), 827–859.

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

Schwarcz, Lilia Moritz, *Nem preto nem branco, muito pelo contrário: cor e raça na sociabilidade brasileira*, Editora Companhia das Letras, 2013.

Schwartzman, Luisa Farah, "Does money whiten? Intergenerational changes in racial classification in Brazil," *American Sociological Review*, 2007, *72* (6), 940–963.

Sen, Maya and Omar Wasow, "Race as a bundle of sticks: Designs that estimate effects of seemingly immutable characteristics," *Annual Review of Political Science*, 2016, *19*, 499–522.

Shayo, Moses, "Social identity and economic policy," *Annual Review of Economics*, 2020, *12*, 355–389.

Silva, Graziella Moraes, "Race and Racisms: Why and How to Compare?," in John Solomos, ed., *Routledge International Handbook of Contemporary Racisms*, New York: Routledge, 2020, pp. 67–77.

_ **and Marcelo Paixão**, "Mixed and Unequal: New Perspectives on Brazilian Ethnoracial Relations," in Edward Telles, ed., *Pigmentocracies: Ethnicity, Race, and Color in Latin America*, Chapel Hill, NC: University of North Carolina Press, 2014, pp. 172–217.

Silveira, Leonardo, "Imputação da informação de raça/cor na Rais para o setor público brasileiro," Technical Report, Instituto de Pesquisa Econômica Aplicada (IPEA) 2022.

Sorkin, Isaac, "Quantifying racial disparities using consecutive employment spells," *NBER Working Paper 33946*, 2025.

Tajfel, Henri, *Human groups and social categories: Studies in social psychology*, Cambridge University Press, 1981.

_ **, Michael G. Billig, Robert P. Bundy, and Claude Flament**, "Social categorization and intergroup behaviour," *European Journal of Social Psychology*, 1971, *1* (2), 149–178.

Taylor, Shelley E., Susan T. Fiske, Nancy L. Etcoff, and Audrey J. Ruderman, "Categorical and contextual bases of person memory and stereotyping," *Journal of Personality and Social Psychology*, 1978, *36* (7), 778–793.

Telles, Edward E., *Race in Another America: The Significance of Skin Color in Brazil*, Princeton: Princeton University Press, 2004.

_ , *Pigmentocracies: Ethnicity, race, and color in Latin America*, UNC Press Books, 2014.

Wade, Peter, "'Race', Nature and Culture," *Man*, 1993, *28* (1), 17–34.

Woo-Mora, L. Guillermo, "Unveiling the Cosmic Race: Skin tone and intergenerational economic disparities in Latin America and the Caribbean," *Journal of Development Economics*, 2026, *179*, 103594.

Online Appendix to
"Perceptions of Race in the Labor Market"

Pedro C. Sant'Anna Sulin Sardoschau Aiko Schmeißer

13th November 2026

A Additional Figures **1**

B Additional Tables **5**

C Details on Data **18**

C.1 Employer-Ascribed Race 18

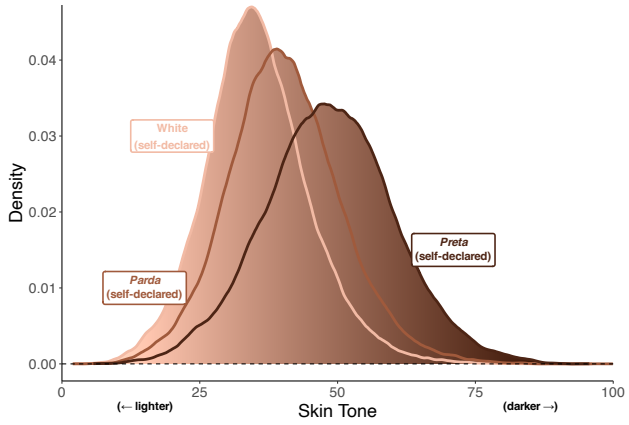
C.2 Skin Tone 20

C.3 Machine-Learning Prediction of Visual Cues 25

C.4 Name Index 26

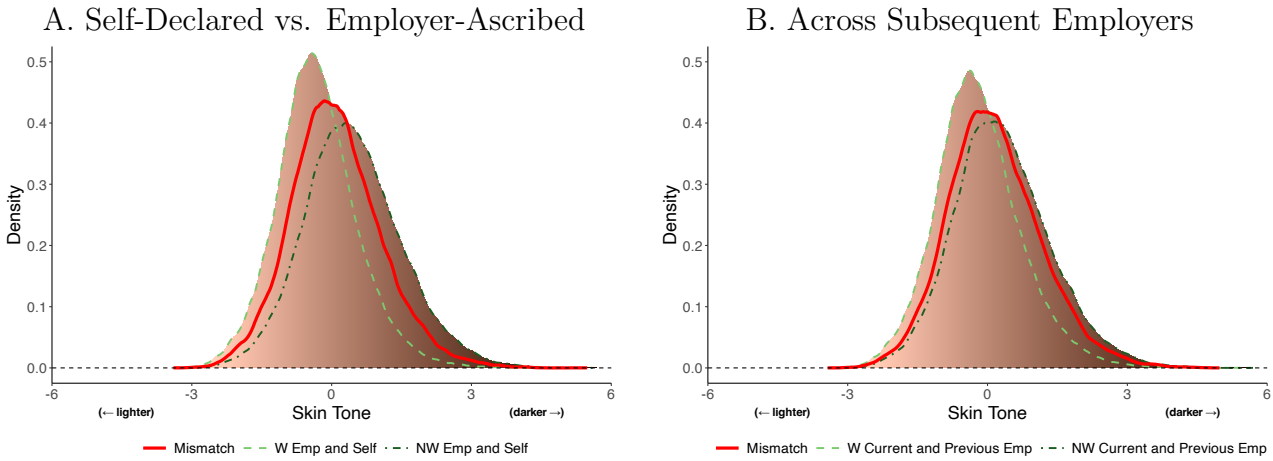
A Additional Figures

Figure A.1: Skin Tone Distribution: Detailed Racial Categories



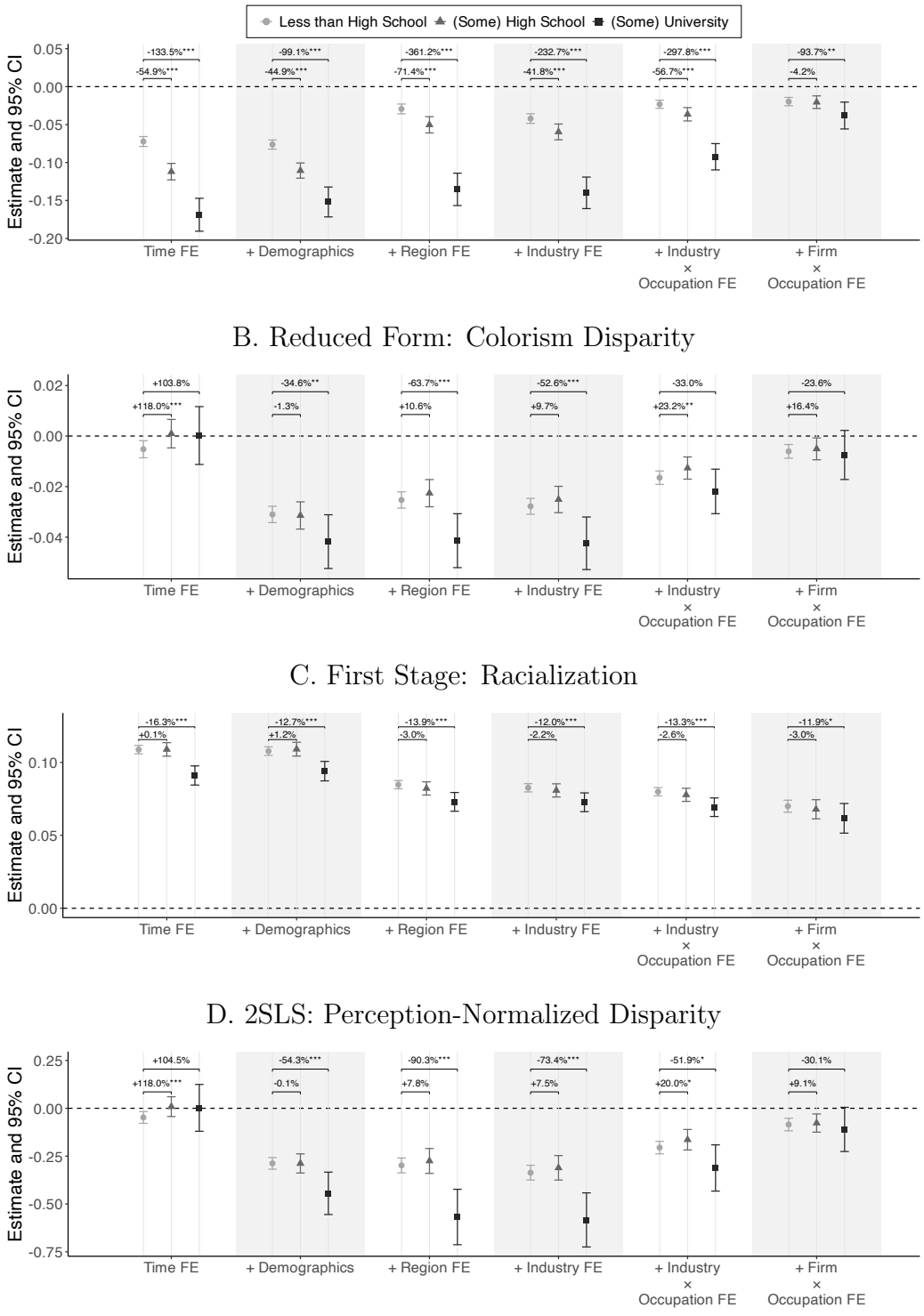
Notes: The figure shows the distribution of skin tone as measured from the electoral photos of candidates, splitting candidates by their self-declared race, distinguishing White, *Pardo* (Brown), and *Preto* (Black) candidates.

Figure A.2: Skin Tone Distribution by Mismatch Category



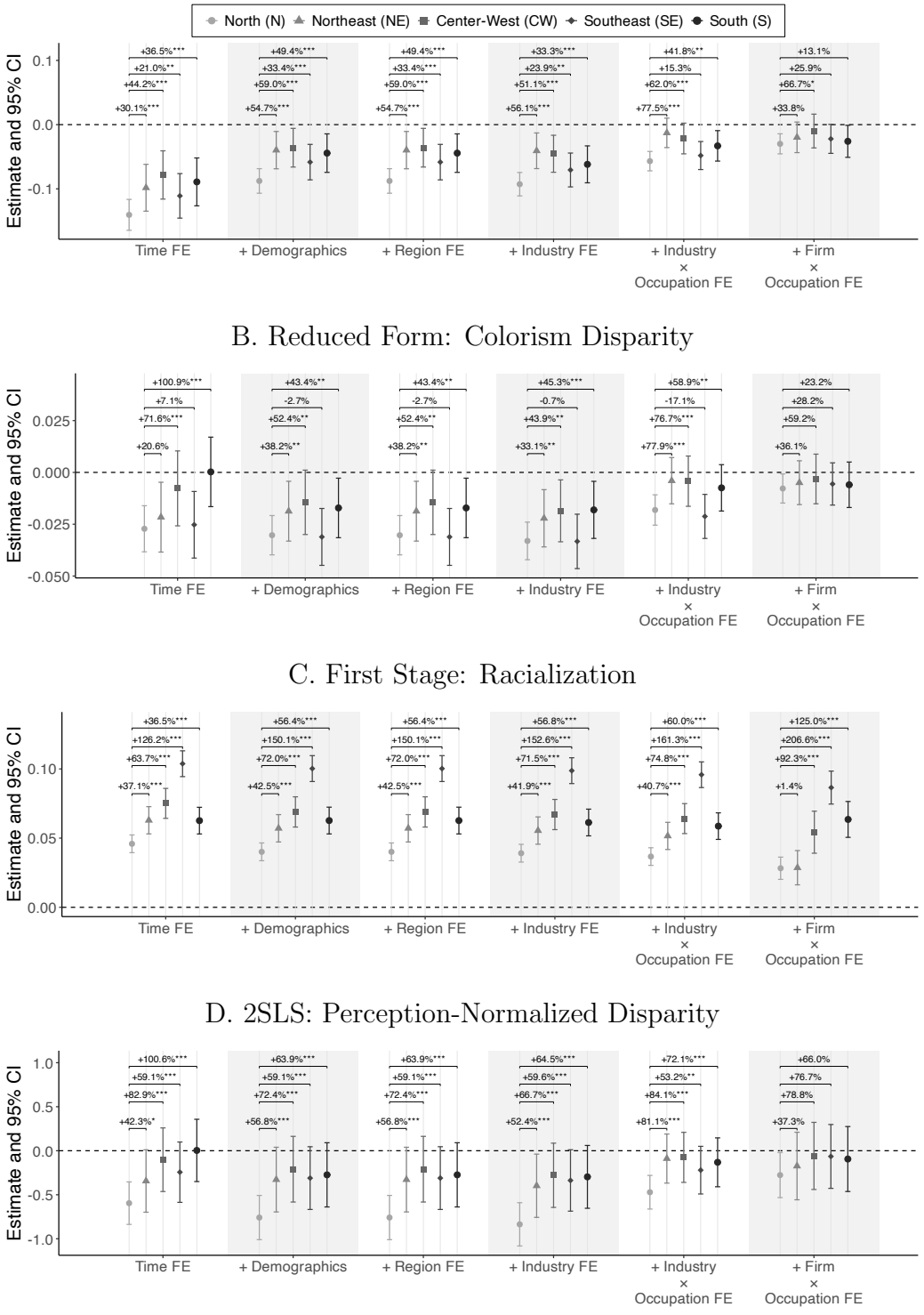
Notes: The figures show histograms of skin tone by categories of racial mismatch. The samples are the same as in Table 1. Panel A shows the skin tone distribution by cells of self-declared and employer-ascribed race, while Panel B shows the skin tone distribution by cells of ascribed race by the current and previous employer. The skin tone measure is standardized to have a mean of zero and a standard deviation of one within our sample.

Figure A.3: Wage Disparities: Heterogeneity by Education



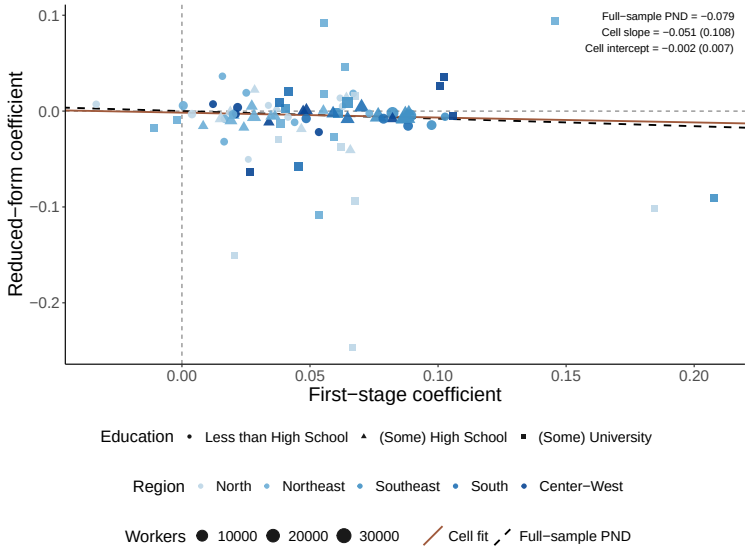
Notes: The figures display different estimates of racial wage gaps for different educational levels. We use the same specifications as columns (1) to (6) of Table 4, with interaction terms for the three educational levels. The plots present point estimates and 95% confidence intervals computed using clustered standard errors at the individual level. The brackets above pairs of point estimates report the percent difference between the estimates (relative to the absolute value of the reference category), with stars indicating whether the difference is statistically significant. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Figure A.4: Wage Disparities: Heterogeneity by Region



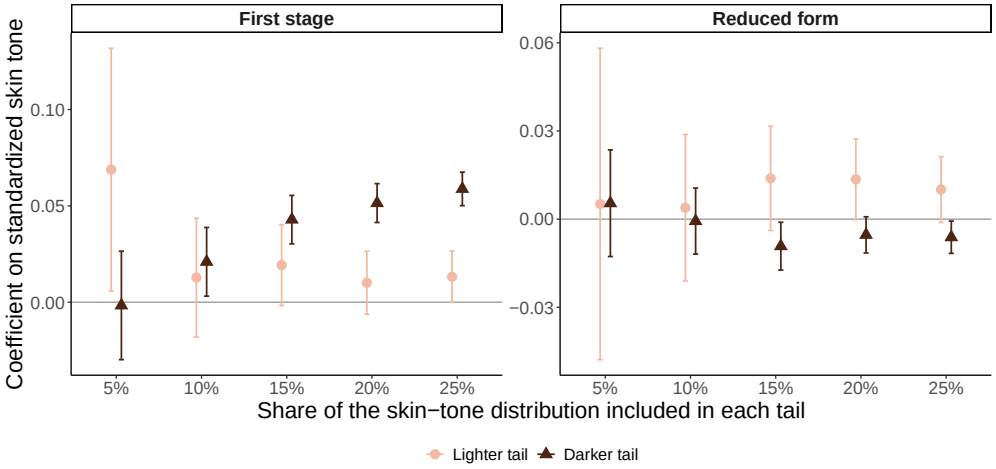
Notes: The figures display different estimates of racial wage gaps for the five different regions of the country. We use the same specifications as columns (1) to (6) of Table 4, with interaction terms for each region. The plots report point estimates and 95% confidence intervals, computed using clustered standard errors at the individual level. The brackets above pairs of point estimates report the percent difference between the estimates (relative to the absolute value of the reference category), with stars indicating whether the difference is statistically significant. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Figure A.5: First-Stage and Reduced-Form Estimates across State–Education Cells



Notes: The figure plots first-stage and reduced-form estimates across workplace-state-by-education cells. Each point represents one cell. The horizontal axis reports the coefficient from a regression of employer-ascribed Non-White on skin tone (the "first stage"), while the vertical axis reports the coefficient from a regression of log hourly wages on skin tone (the 'reduced form"). Horizontal and vertical bars report 95% confidence intervals. Point size indicates the number of workers in the cell; colors identify regions and symbols identify education groups. The dashed line passes through the origin and has slope equal to the full-sample Perception Normalized Disparity (PND). The solid line is the linear fit across cells, with an unrestricted intercept and weighted by cell size. Regressions follow the most strict specification of Table 4, i.e., they control for gender and a quadratic in age and include year and establishment-by-three-digit-occupation fixed effects. Standard errors in parentheses are based on 500 bootstrap replications. In each replication, workers are sampled with replacement, and the cell-level first stages, reduced forms, and fitted line are re-estimated.

Figure A.6: First-Stage and Reduced-Form Estimates at the Tails of the Skin-Tone Distribution



Notes: The figure reports first-stage and reduced-form estimates separately within the lighter and darker tails of the skin-tone distribution. The horizontal axis indicates the share included in each tail, ranging from 5% to 25%; consequently, samples are nested across cutoffs. Skin tone is standardized using the full analysis sample. The first-stage estimates measure the relationship between skin tone and employer-ascribed Non-White status. The reduced-form estimates measure the relationship between skin tone and log hourly wages. Vertical bars report 95% confidence intervals. Regressions follow column (6) of Table 4, i.e., they control for gender, a quadratic in age, and education, and include year and establishment-by-three-digit-occupation fixed effects. Standard errors are clustered at the worker level. If skin tone affects wages only through employers' racial perceptions, the reduced form should approach zero in ranges of the distribution where the first stage is negligible.

B Additional Tables

Table B.1: Summary Statistics of Candidate Sample and Full RAIS Sample

	(1) Candidates Mean	(2) Non-candidates Mean	(3) Difference (1) - (2)
Worker Characteristics			
Non-White (Employer-Ascribed)	0.37	0.37	-0.00**
Male	0.75	0.62	0.13***
Age (years)	34.48	33.65	0.84***
Log Hourly Wage (2018 BRL)	2.18	2.20	-0.02***
Highest Degree Attained			
Less than Elementary	0.03	0.04	-0.01***
Elementary School	0.14	0.15	-0.01***
Middle School	0.23	0.23	0.00***
High School	0.52	0.49	0.03***
Undergraduate or More	0.09	0.10	-0.01***
Establishment Characteristics			
Firm Size	536.54	580.17	-43.62***
Average Log Hourly Wage	2.15	2.21	-0.06***
Sector			
Agriculture	0.03	0.02	0.01***
Extractive Industry	0.01	0.01	0.00***
Manufacturing	0.24	0.25	-0.01***
Construction	0.06	0.07	-0.00***
Retail	0.25	0.25	-0.01***
Transport & Communication	0.13	0.13	-0.00***
Finance & Real Estate	0.15	0.17	-0.02***
Education & Health	0.09	0.07	0.01***
Other Services	0.05	0.03	0.02***
Region			
North	0.05	0.04	0.01***
Northeast	0.17	0.16	0.01***
Southeast	0.51	0.55	-0.04***
South	0.19	0.18	0.02***
Center-West	0.08	0.07	0.01***
No. of worker-year obs.	1,510,833	310,051,787	
No. of workers	331,531	61,075,388	
No. of establishments	358,899	5,616,123	

Notes: The table shows mean characteristics of workers who have run for a political office between 2014 and 2024 (our estimation sample, column (1)) and all other workers in RAIS (column (2)). Column (3) shows the difference in means between both samples. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.2: Racial Gaps in Candidate Sample and Full RAIS Sample

	(1)	(2)	(3)	(4)	(5)	(6)
	Candidates			Non-candidates		
	White Mean	Non-White Mean	Racial gap (1)-(2)	White Mean	Non-White Mean	Racial gap (4)-(5)
Worker Characteristics						
Male	0.73	0.77	-0.04***	0.60	0.66	-0.06***
Age (years)	34.72	34.09	0.63***	33.87	33.25	0.62***
Log Hourly Wage (2018 BRL)	2.22	2.10	0.13***	2.28	2.06	0.23***
Highest Degree Attained						
Less than Elementary	0.02	0.04	-0.02***	0.03	0.06	-0.03***
Elementary School	0.13	0.15	-0.02***	0.14	0.16	-0.03***
Middle School	0.23	0.23	0.00	0.23	0.24	-0.01***
High School	0.51	0.52	-0.01***	0.49	0.49	-0.01***
Undergraduate or More	0.11	0.06	0.05***	0.12	0.05	0.07***
Establishment Characteristics						
Firm Size	441.86	698.04	-256.17***	501.76	707.93	-206.17***
Average Log Hourly Wage	2.18	2.09	0.09***	2.27	2.10	0.17***
Sector						
Agriculture	0.02	0.03	-0.01***	0.01	0.02	-0.01***
Extractive Industry	0.01	0.01	-0.00***	0.01	0.01	-0.00***
Manufacturing	0.26	0.21	0.05***	0.27	0.22	0.05***
Construction	0.05	0.09	-0.05***	0.05	0.10	-0.05***
Retail	0.26	0.23	0.03***	0.26	0.25	0.01***
Transport & Communication	0.13	0.13	-0.00***	0.13	0.13	0.00
Finance & Real Estate	0.13	0.17	-0.04***	0.16	0.18	-0.02***
Education & Health	0.09	0.08	0.01***	0.08	0.07	0.01***
Other Services	0.05	0.05	0.01***	0.03	0.03	0.00***
Region						
North	0.02	0.11	-0.08***	0.02	0.09	-0.07***
Northeast	0.09	0.31	-0.22***	0.08	0.30	-0.22***
Southeast	0.55	0.44	0.12***	0.60	0.48	0.12***
South	0.28	0.05	0.23***	0.25	0.04	0.21***
Center-West	0.06	0.10	-0.04***	0.06	0.09	-0.04***
N	952,450	558,383		192,996,780	113,674,338	

Notes: The table shows racial gaps (by employer-ascribed race) in mean characteristics of workers who have run for a political office between 2014 and 2024 (our estimation sample, columns (1)-(3)) and all other workers in RAIS (columns (4)-(6)). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.3: Mismatch in Ascribed Race: Detailed Racial Categories

[A] Mismatch between Self-Declared and Employer-Ascribed Race

	Employer-Ascribed Race					
	White		*Parda*		*Preta*	
Self-Declared Race	%	Skin Tone	%	Skin Tone	%	Skin Tone
White	45.20%	-0.32	9.22%	-0.13	0.49%	0.14
Parda	14.86%	0.07	17.82%	0.21	1.90%	0.55
Preta	2.99%	0.77	4.16%	0.80	3.35%	1.19

[B] Mismatch across Subsequent Employers

	Employer-Ascribed Race, Current					
	White		*Parda*		*Preta*	
Employer-Ascribed Race, Previous	%	Skin Tone	%	Skin Tone	%	Skin Tone
White	49.82%	-0.22	10.79%	0.08	1.38%	0.71
Parda	9.57%	0.07	20.71%	0.24	1.73%	0.81
Preta	1.31%	0.72	1.83%	0.75	2.87%	1.11

Notes: The table displays the (mis)match shares for different race measures in our sample, as well as the average skin tone for each cell. This table considered the detailed racial categories available in Brazilian statistics, White, *Parda* (brown), and *Preta* (black), excluding indigenous and Asian individuals. Skin tone is standardized to have a mean of zero and a standard deviation of one across the whole sample (higher values represent darker shades, while lower values represent lighter shades). Panel A presents the distribution of workers by self-declared race (rows) and employer-ascribed race (columns). It considers all worker-year observations ($N = 1,510,833$). Panel B presents the distribution of ascribed race for the same worker across different employers in consecutive years. It considers all workers who changed jobs ($N = 153,366$) and presents the distribution of workers by their ascribed race in the previous employer (rows) and current employer (columns).

Table B.4: Mismatch in Ascribed Race: Robustness

[A] Mismatch between Self-Declared and Employer-Ascribed Race

Employer-ascribed / Self-declared	White / White	Non-White / Non-White	White / Non-White	Non-White / White	N
Baseline sample	45.20%	27.25%	17.84%	9.71%	1,510,833
Firms with variation in ascribed race	42.30%	29.24%	18.08%	10.38%	1,328,249
Firms with HR employees	43.48%	28.88%	17.24%	10.41%	910,506
Self-declared race in same period	43.28%	29.23%	16.04%	11.45%	48,316
Councilor candidates	44.97%	27.49%	17.99%	9.55%	1,412,050

Employer-ascribed / Self-declared	Male / Male	Female / Female	Male / Female	Female / Male	N
Baseline sample	73.75%	24.26%	1.06%	0.93%	1,510,833

[B] Mismatch across Subsequent Employers

Employer-ascribed, current / Employer-ascribed, previous	White / White	Non-White / Non-White	White / Non-White	Non-White / White	N
Baseline sample	49.82%	27.12%	10.88%	12.18%	153,366
Firms with variation in ascribed race	45.68%	29.96%	11.51%	12.85%	128,142
Firms with HR employees	47.74%	30.64%	10.22%	11.40%	71,501
All workers in RAIS	50.44%	26.70%	10.78%	12.08%	33,898,971

Employer-ascribed, current / Employer-ascribed, previous	Male / Male	Female / Female	Male / Female	Female / Male	N
Baseline sample	76.58%	20.76%	1.32%	1.34%	153,366

Notes: The table shows mismatch rates between self-declared and employer-ascribed race (Panel A) and in employer-ascribed race across different employers in subsequent years (Panel B) in alternative subsamples. Specifically, we restrict the sample to firms that do not ascribe the same racial category to all their workers, firms that have at least one employee in an HR occupation, jobs in 2015 and candidates running in 2014 or 2016 (such that self-declared and employer-ascribed race are report at roughly the same time), as well as candidates that run for local councilor positions. Panel B also considers job changes among all workers in RAIS (not limited to the candidate sample). Both panels also report mismatch rates across reported gender instead of race.

Table B.5: Mismatch in Ascribed Race: Aggregating Across Employers

	(1) OLS	(2) OLS	(3) SSIV	(4) OLS	(5) SSIV
Non-White Employer	0.454***				
	(0.002)				
Non-White Employer Mode		0.472***	0.613***		
		(0.002)	(0.003)		
Non-White Employer Mean				0.597***	0.659***
				(0.002)	(0.002)
Constant	0.283***	0.286***	0.240***	0.230***	0.207***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N	1,510,833	1,510,833	1,439,433	1,510,833	1,439,433

Notes: The table reports results from regressions of self-declared race (indicator equal to one if the worker identifies as non-White) on employer-ascribed race (indicator equal to one if the worker is ascribed non-White). While column (1) uses race as ascribed by the current employer, columns (2) to (5) aggregate across employers by considering the mode or mean across all yearly observations of the worker. Columns (3) and (5) implement a split-sample IV (SSIV) approach, in which the mode or mean constructed from one randomly selected half of a worker's yearly observations is instrumented with the mode or mean constructed from the other half of observations. Standard errors clustered at the individual level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.6: Determinants of Employer-Ascribed Race: Employment History Controls

	(1)	(2)	(3)	(4)	(5)	(6)
Skin Tone	0.110***	0.104***	0.082***	0.081***	0.079***	0.068***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Demographics						
Male		0.037***	0.016***	0.009**	0.001	-0.000
		(0.004)	(0.004)	(0.004)	(0.004)	(0.009)
Age		0.007***	0.003*	0.001	-0.000	-0.002
		(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Age^2		-0.010***	-0.005***	-0.004***	-0.003**	-0.001
		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
(Some) High School		-0.001	-0.028***	-0.019***	-0.010**	-0.008
		(0.004)	(0.004)	(0.004)	(0.004)	(0.007)
(Some) University		-0.044***	-0.087***	-0.078***	-0.040***	-0.014
		(0.006)	(0.005)	(0.006)	(0.006)	(0.017)
Work History						
Tenure		-0.055***	-0.043***	-0.036***	-0.035***	-0.048***
		(0.003)	(0.002)	(0.002)	(0.002)	(0.004)
$Tenure^2$		0.005***	0.005***	0.003**	0.003*	0.002**
		(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Final Wage in Previous Job		-0.046***	-0.030***	-0.033***	-0.028***	-0.015***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.005)
Region						
Northeast			-0.057***	-0.060***	-0.062***	
			(0.008)	(0.008)	(0.008)	
Southeast			-0.374***	-0.370***	-0.372***	
			(0.007)	(0.007)	(0.007)	
South			-0.569***	-0.560***	-0.561***	
			(0.007)	(0.007)	(0.007)	
Center-West			-0.191***	-0.188***	-0.188***	
			(0.009)	(0.009)	(0.009)	
N	1,510,833	374,036	374,036	374,006	373,905	298,044
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	No	No	Yes	No	No
Industry × Occupation FE	No	No	No	No	Yes	No
Firm × Occupation FE	No	No	No	No	No	Yes
R-squared	0.057	0.065	0.209	0.213	0.223	0.779
Adjusted R-squared	0.057	0.065	0.209	0.213	0.222	0.719

Notes: The table reports OLS estimates of the relationship between employer-ascribed race (an indicator equal to one for non-White) and various covariates, enriching the analysis of Table 2 with work history variables. The sample in columns 2-6 only includes workers whom we observe in multiple establishments. Standard errors clustered at the individual level are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table B.7: Determinants of Employer-Ascribed Race: Other Racial Cues

	(1)	(2)	(3)	(4)	(5)	(6)
[A] Name Cues						
Non-White Index Given Name	0.552***	0.540***	0.177***	0.170***	0.160***	0.067***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.012)
Non-White Index Surname	0.717***	0.693***	0.385***	0.377***	0.359***	0.298***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.008)
(Some) High School		-0.011***	-0.037***	-0.029***	-0.013***	-0.008***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
(Some) University		-0.080***	-0.114***	-0.110***	-0.055***	-0.034***
		(0.003)	(0.003)	(0.003)	(0.003)	(0.006)
N	1,460,574	1,460,574	1,460,574	1,460,263	1,460,199	1,208,444
R-squared	0.093	0.097	0.197	0.202	0.212	0.778
Adjusted R-squared	0.093	0.097	0.197	0.202	0.211	0.717
[B] Self-Declared Race						
Non-White Self-Declared	0.426***	0.419***	0.318***	0.314***	0.308***	0.257***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
(Some) High School		0.000	-0.022***	-0.016***	-0.006***	-0.003
		(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
(Some) University		-0.044***	-0.073***	-0.069***	-0.033***	-0.018***
		(0.003)	(0.002)	(0.002)	(0.003)	(0.006)
N	1,510,833	1,510,833	1,510,833	1,510,510	1,510,445	1,249,942
R-squared	0.198	0.200	0.272	0.275	0.283	0.791
Adjusted R-squared	0.198	0.200	0.272	0.275	0.282	0.734
[C] All Racial Cues						
Skin Tone	0.042***	0.044***	0.044***	0.043***	0.042***	0.040***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Non-White Index Given Name	0.309***	0.329***	0.092***	0.088***	0.083***	0.023**
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.011)
Non-White Index Surname	0.362***	0.355***	0.184***	0.180***	0.173***	0.147***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.008)
Non-White Self-Declared	0.348***	0.345***	0.280***	0.278***	0.274***	0.230***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
(Some) High School		0.006***	-0.015***	-0.010***	-0.002	-0.000
		(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
(Some) University		-0.019***	-0.054***	-0.050***	-0.021***	-0.011*
		(0.003)	(0.002)	(0.002)	(0.003)	(0.006)
N	1,460,574	1,460,574	1,460,574	1,460,263	1,460,199	1,208,444
R-squared	0.198	0.200	0.272	0.275	0.283	0.791
Adjusted R-squared	0.198	0.200	0.272	0.275	0.282	0.734
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	No	No	Yes	No	No
Industry × Occupation FE	No	No	No	No	Yes	No
Firm × Occupation FE	No	No	No	No	No	Yes

Notes: The table reports OLS estimates of the relationship between employer-ascribed race (an indicator equal to one for non-White) and various racial cues. Panel A includes name cues, Panel B considers self-declared race, and Panel C adds them jointly with skin tone. The first column displays the relationship with racial cues, when only controlling for year fixed effects. The subsequent columns progressively add demographic covariates as well as region, one-digit industry, three-digit occupation, and firm fixed effects, as in Table 2. Standard errors clustered at the individual level are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table B.8: Determinants of Self-Declared Race

	(1)	(2)	(3)	(4)	(5)	(6)
Skin Tone	0.154***	0.149***	0.126***	0.125***	0.122***	0.115***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Demographics						
Male		0.007***	0.002	-0.001	-0.005**	-0.006
		(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
Age		-0.025***	-0.017***	-0.019***	-0.014***	-0.009***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Age^2		-0.011***	-0.008***	-0.007***	-0.007***	-0.006***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
(Some) High School		-0.033***	-0.054***	-0.050***	-0.031***	-0.025***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
(Some) University		-0.159***	-0.173***	-0.174***	-0.107***	-0.077***
		(0.003)	(0.003)	(0.003)	(0.004)	(0.008)
Region						
Northeast			-0.071***	-0.072***	-0.072***	
			(0.004)	(0.004)	(0.004)	
Southeast			-0.344***	-0.342***	-0.343***	
			(0.004)	(0.004)	(0.004)	
South			-0.535***	-0.529***	-0.531***	
			(0.004)	(0.004)	(0.004)	
Center-West			-0.180***	-0.179***	-0.177***	
			(0.005)	(0.005)	(0.005)	
N	1,510,833	1,510,833	1,510,833	1,510,510	1,510,445	1,249,942
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	No	No	Yes	No	No
Industry × Occupation FE	No	No	No	No	Yes	No
Firm × Occupation FE	No	No	No	No	No	Yes
R-squared	0.097	0.110	0.214	0.217	0.226	0.772
Adjusted R-squared	0.097	0.110	0.214	0.217	0.225	0.711

Notes: The table reports OLS estimates of the relationship between self-declared race (an indicator equal to one for non-White) and various covariates. The first column displays the relationship with skin tone, when only controlling for year fixed effects. The subsequent columns progressively add demographic covariates as well as region, one-digit industry, three-digit occupation, and firm fixed effects, as in Table 2. Standard errors clustered at the individual level are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table B.9: Determinants of Employer-Ascribed Race: ML Prediction of Visual Cues

	(1)	(2)	(3)	(4)	(5)	(6)
Visual Cues ML Prediction	1.030***	1.014***	0.828***	0.820***	0.806***	0.676***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.007)
(Some) High School		0.003*	-0.021***	-0.015***	-0.007***	-0.004
		(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
(Some) University		-0.044***	-0.070***	-0.064***	-0.034***	-0.018***
		(0.003)	(0.002)	(0.003)	(0.003)	(0.006)
N	1,409,846	1,409,846	1,409,846	1,409,546	1,409,476	1,164,572
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	No	No	Yes	No	No
Industry × Occupation FE	No	No	No	No	Yes	No
Firm × Occupation FE	No	No	No	No	No	Yes
R-squared	0.167	0.168	0.276	0.279	0.286	0.796
Adjusted R-squared	0.167	0.168	0.276	0.279	0.286	0.741

Notes: The table reports OLS estimates of the relationship between employer-ascribed race (an indicator equal to one for non-White) and various covariates, including a machine-learning prediction of employer non-White ascriptions based on candidate photos. See Appendix C.3 for details on the prediction algorithm. The first column displays the coefficient of the predicted probability to be non-White, when only controlling for year fixed effects. The subsequent columns progressively add demographic covariates as well as region, one-digit industry, three-digit occupation, and firm fixed effects, as in Table 2. Standard errors clustered at the individual level are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table B.10: Wage Disparities by Self-Declared Race

	(1)	(2)	(3)	(4)	(5)	(6)
[A] OLS: Conventional Disparity						
Non-White Self-Declared	-0.163***	-0.107***	-0.070***	-0.077***	-0.043***	-0.019***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
[B] Reduced Form: Colorism Disparity						
Skin Tone	-0.026***	-0.032***	-0.026***	-0.028***	-0.015***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
[C] First Stage: Racialization						
Skin Tone	0.154***	0.149***	0.126***	0.125***	0.122***	0.115***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
[D] 2SLS: Perception-Normalized Disparity						
Non-White Self-Declared	-0.168***	-0.217***	-0.204***	-0.225***	-0.123***	-0.048***
	(0.008)	(0.008)	(0.009)	(0.009)	(0.008)	(0.008)
KP F-stat	26,503.1	23,066.8	17,089.7	16,775.0	15,857.0	5,460.5
N	1,510,833	1,510,833	1,510,833	1,510,510	1,510,445	1,249,942
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	No	Yes	Yes	Yes	Yes	Yes
Region FE	No	No	Yes	Yes	Yes	No
Industry FE	No	No	No	Yes	No	No
Industry × Occupation FE	No	No	No	No	Yes	No
Firm × Occupation FE	No	No	No	No	No	Yes

Notes: The table reports racial wage gaps based on self-declared race. All columns include year fixed-effects. The first column always shows the unconditional relationship (apart from year fixed-effects), and the remaining columns progressively add controls. Panel A shows coefficient estimates of OLS regressions of log hourly wage on self-declared race. Panel B shows reduced form regressions of log hourly wage on skin tone. Panel C presents the first-stage coefficient from regressing self-declared race on skin tone. Finally, Panel D shows 2SLS estimates of regressing log hourly wage on self-declared race, instrumented by skin tone. Standard errors clustered at the individual level are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table B.11: Skin Tone Gaps for Repeated Candidates

	(1)	(2)	(3)	(4)	(5)	(6)
[A] Hourly Wage (log), OLS						
Skin Tone	-0.051***	-0.044***	-0.037***	-0.040***	-0.020***	-0.010***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
[B] Hourly Wage (log), Split-Sample-IV						
Skin Tone	-0.114***	-0.112***	-0.099***	-0.109***	-0.054***	-0.029***
	(0.007)	(0.007)	(0.008)	(0.007)	(0.006)	(0.008)
[C] Non-White Employer-Ascribed, OLS						
Skin Tone	0.152***	0.152***	0.119***	0.117***	0.113***	0.099***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.003)
[D] Non-White Employer-Ascribed, Split-Sample-IV						
Skin Tone	0.346***	0.378***	0.315***	0.311***	0.307***	0.303***
	(0.005)	(0.007)	(0.006)	(0.006)	(0.006)	(0.014)
N	652,140	652,140	652,140	651,990	651,902	536,491
KP F-stat	6,651	5,147	4,290	4,224	4,020	959
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	No	Yes	Yes	Yes	Yes	Yes
Region FE	No	No	Yes	Yes	Yes	No
Industry FE	No	No	No	Yes	No	No
Industry × Occupation FE	No	No	No	No	Yes	No
Firm × Occupation FE	No	No	No	No	No	Yes

Notes: The table reports results from regressions of log hourly wages (Panels A and B) and employer-ascribed race (Panels C and D) and on skin tone. The sample consists of all individuals who ran for political office multiple times and for whom multiple electoral photos are available. Panels A and C present OLS estimates, in which skin tone is measured by averaging across all photos of an individual. Panels B and D implement a split-sample IV approach, where the average skin tone from one randomly selected half of an individual's photos is instrumented with the average from the other half of photos. The first column only controls for year fixed-effects, and the remaining columns progressively add controls. Standard errors clustered at the individual level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.12: Wage Disparities by Employer-Ascribed Race: Name Cues

	(1)	(2)	(3)	(4)	(5)	(6)
[A] OLS: Conventional Disparity						
Non-White Employer-Ascribed	-0.146***	-0.102***	-0.054***	-0.063***	-0.038***	-0.021***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
[B] Reduced Form: Cue-Based Disparity						
Non-White Index Given Name	0.047***	-0.193***	-0.065***	-0.076***	-0.036***	-0.005
	(0.012)	(0.011)	(0.011)	(0.011)	(0.009)	(0.009)
Non-White Index Surname	-0.429***	-0.236***	-0.167***	-0.174***	-0.101***	-0.050***
	(0.009)	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)
[C] First Stage: Racialization						
Non-White Index Given Name	0.552***	0.540***	0.177***	0.170***	0.160***	0.067***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.012)
Non-White Index Surname	0.717***	0.693***	0.385***	0.377***	0.359***	0.298***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.008)
[D] 2SLS: Perception-Normalized Disparity						
Non-White Employer-Ascribed	-0.425***	-0.344***	-0.426***	-0.461***	-0.274***	-0.164***
	(0.010)	(0.009)	(0.019)	(0.019)	(0.017)	(0.024)
KP F-stat	24,943.0	22,299.0	4,835.6	4,613.2	4,216.0	687.5
N	1,460,574	1,460,574	1,460,574	1,460,263	1,460,199	1,208,444
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	No	Yes	Yes	Yes	Yes	Yes
Region FE	No	No	Yes	Yes	Yes	No
Industry FE	No	No	No	Yes	No	No
Industry × Occupation FE	No	No	No	No	Yes	No
Firm × Occupation FE	No	No	No	No	No	Yes

Notes: The table reports racial wage gaps based on employer-ascribed race when using non-White indices of given names and surnames as racial cues. All columns include year fixed-effects. The first column always shows the unconditional relationship (apart from year fixed-effects), and the remaining columns progressively add controls. Panel A shows coefficient estimates of OLS regressions of log hourly wage on employer-ascribed race. Panel B shows reduced form regressions of log hourly wage on the non-White name indices. Panel C presents the first-stage coefficient from regressing employer-ascribed race on the non-White name indices. Finally, Panel D shows 2SLS estimates of regressing log hourly wage on employer-ascribed race, instrumented by the non-White name indices. Standard errors clustered at the individual level are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table B.13: Wage Disparities by Employer-Ascribed Race: Self-Declared Race as Racial Cue

	(1)	(2)	(3)	(4)	(5)	(6)
[A] OLS: Conventional Disparity						
Non-White Employer-Ascribed	-0.148***	-0.103***	-0.054***	-0.064***	-0.038***	-0.021***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
[B] Reduced Form: Cue-Based Disparity						
Non-White Self-Declared	-0.163***	-0.107***	-0.070***	-0.077***	-0.043***	-0.019***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
[C] First Stage: Racialization						
Non-White Self-Declared	0.426***	0.419***	0.318***	0.314***	0.308***	0.257***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
[D] 2SLS: Perception-Normalized Disparity						
Non-White Employer-Ascribed	-0.382***	-0.255***	-0.219***	-0.245***	-0.139***	-0.074***
	(0.006)	(0.005)	(0.008)	(0.007)	(0.006)	(0.008)
KP F-stat	73,663.0	68,642.6	32,953.6	31,985.3	30,761.7	7,822.6
N	1,510,833	1,510,833	1,510,833	1,510,510	1,510,445	1,249,942
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	No	Yes	Yes	Yes	Yes	Yes
Region FE	No	No	Yes	Yes	Yes	No
Industry FE	No	No	No	Yes	No	No
Industry × Occupation FE	No	No	No	No	Yes	No
Firm × Occupation FE	No	No	No	No	No	Yes

Notes: The table reports racial wage gaps based on employer-ascribed race when using self-declared race as racial cue. All columns include year fixed-effects. The first column always shows the unconditional relationship (apart from year fixed-effects), and the remaining columns progressively add controls. Panel A shows coefficient estimates of OLS regressions of log hourly wage on employer-ascribed race. Panel B shows reduced form regressions of log hourly wage on self-declared race. Panel C presents the first-stage coefficient from regressing employer-ascribed race on self-declared race. Finally, Panel D shows 2SLS estimates of regressing log hourly wage on employer-ascribed race, instrumented by self-declared race. Standard errors clustered at the individual level are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table B.14: Wage Disparities by Employer-Ascribed Race: Conditional PND of Skin Tone

	(1)	(2)	(3)	(4)
[A] OLS: Conventional Disparity				
Non-White Employer-Ascribed	-0.021***	-0.020***	-0.017***	-0.016***
	(0.002)	(0.002)	(0.002)	(0.002)
[B] Reduced Form: Colorism Disparity				
Skin Tone	-0.005***	-0.005***	-0.003***	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
Non-White Index Given Name		-0.005		-0.002
		(0.009)		(0.009)
Non-White Index Surname		-0.047***		-0.039***
		(0.007)		(0.007)
Non-White Self-Declared			-0.018***	-0.016***
			(0.002)	(0.002)
[C] First Stage: Racialization				
Skin Tone	0.068***	0.065***	0.041***	0.040***
	(0.001)	(0.001)	(0.001)	(0.001)
Non-White Index Given Name		0.062***		0.023**
		(0.012)		(0.011)
Non-White Index Surname		0.259***		0.147***
		(0.008)		(0.008)
Non-White Self-Declared			0.237***	0.230***
			(0.003)	(0.003)
[D] 2SLS: Skin Tone PND				
Non-White Employer-Ascribed	-0.079***	-0.074***	-0.081***	-0.077***
	(0.013)	(0.014)	(0.023)	(0.023)
KP F-stat	2,684.5	2,472.0	1,048.3	1,006.1
N	1,208,444	1,208,444	1,208,444	1,208,444
Time FE	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes
Firm × Occupation FE	Yes	Yes	Yes	Yes

Notes: The table reports racial wage gaps based on employer-ascribed race when using skin tone as racial cue and using other racial cues as controls. Column (1) also includes skin tone as racial cue. Columns (2) to (4) add controls for non-White name indices of given names and surnames and for self-declared race. All columns include year dummies, demographic covariates, and firm × three-digit occupation fixed effects. Panel A shows coefficient estimates of OLS regressions of log hourly wage on employer-ascribed race. Panel B shows reduced form regressions of log hourly wage on skin tone. Panel C presents the first-stage coefficient from regressing employer-ascribed race on skin tone. Finally, Panel D shows 2SLS estimates of regressing log hourly wage on employer-ascribed race, instrumented by skin tone. Standard errors clustered at the individual level are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table B.15: Characteristics of Compliers in Perception-Normalized Disparities Analysis

	(1) Population	(2) Skin Tone Compliers	(3) Name Indices Compliers	(4) Self-Declared Race Compliers
Worker Characteristics				
Male	0.756	0.791	0.754	0.777
	(0.001)	(0.009)	(0.014)	(0.005)
Age (years)	35.013	34.704	35.635	34.904
	(0.019)	(0.196)	(0.299)	(0.113)
Log Hourly Wage (2018 BRL)	2.214	2.244	2.405	2.295
	(0.002)	(0.013)	(0.020)	(0.008)
Highest Degree Attained				
Less than Elementary	0.027	0.033	0.027	0.028
	(0.000)	(0.004)	(0.007)	(0.002)
Elementary School	0.139	0.176	0.116	0.156
	(0.001)	(0.009)	(0.013)	(0.005)
Middle School	0.232	0.258	0.227	0.234
	(0.001)	(0.011)	(0.016)	(0.006)
High School	0.512	0.489	0.520	0.523
	(0.001)	(0.012)	(0.018)	(0.007)
Undergraduate or More	0.089	0.040	0.103	0.056
	(0.001)	(0.005)	(0.009)	(0.003)
Establishment Characteristics				
Log Firm Size	4.310	6.240	6.330	6.294
	(0.005)	(0.034)	(0.047)	(0.021)
Average Log Hourly Wage	2.177	2.309	2.424	2.356
	(0.001)	(0.011)	(0.017)	(0.007)
Sector				
Agriculture	0.026	0.035	0.031	0.031
	(0.000)	(0.003)	(0.005)	(0.002)
Extractive Industry	0.010	0.011	0.009	0.013
	(0.000)	(0.002)	(0.003)	(0.001)
Manufacturing	0.257	0.347	0.383	0.363
	(0.001)	(0.010)	(0.014)	(0.006)
Construction	0.057	0.056	0.059	0.054
	(0.000)	(0.004)	(0.005)	(0.002)
Retail	0.228	0.067	0.065	0.067
	(0.001)	(0.004)	(0.005)	(0.002)
Transport & Communication	0.128	0.146	0.122	0.147
	(0.001)	(0.007)	(0.010)	(0.004)
Finance & Real Estate	0.153	0.212	0.198	0.199
	(0.001)	(0.008)	(0.012)	(0.005)
Education & Health	0.089	0.085	0.092	0.091
	(0.001)	(0.005)	(0.008)	(0.003)
Other Services	0.053	0.040	0.041	0.035
	(0.000)	(0.003)	(0.004)	(0.002)
Region				
North	0.051	0.022	0.012	0.018
	(0.000)	(0.003)	(0.004)	(0.002)
Northeast	0.166	0.069	0.033	0.077
	(0.001)	(0.006)	(0.006)	(0.004)
Southeast	0.514	0.717	0.636	0.681
	(0.001)	(0.010)	(0.014)	(0.006)
South	0.196	0.140	0.255	0.159
	(0.001)	(0.007)	(0.012)	(0.005)
Center-West	0.073	0.052	0.061	0.065
	(0.001)	(0.005)	(0.008)	(0.003)
N	1,208,444	1,208,444	1,208,444	1,208,444

Notes: The table reports estimates of the average characteristics of compliers in the PND analysis of Table 5. To characterize compliers, we follow Hull (2025) and estimate the following IV model: $R^E_{it} \times \mathrm{X}_{it} = \alpha + \beta R^E_{it} + X_{it} + \epsilon_{it}$, where R^E_{it} is the employer-ascribed race of worker i at time t, which we instrument with a racial cue, and X_{it} is the value of the characteristic of interest. All regressions also include include year dummies, demographic covariates, and firm × three-digit occupation fixed effects. The coefficient β identifies the mean of X_{it} among compliers, i.e., among individuals for whom employer-ascribed race changes with their racial cue. Column (1) reports mean characteristics of all individuals in our sample. Columns (2), (3), and (4) show complier mean estimates when using skin tone, non-White name indices of given names and surnames, or self-declared race as racial instruments. Each mean reported in these columns is from a separate IV regression. Standard errors clustered at the individual level are in parentheses.

C Details on Data

C.1 Employer-Ascribed Race

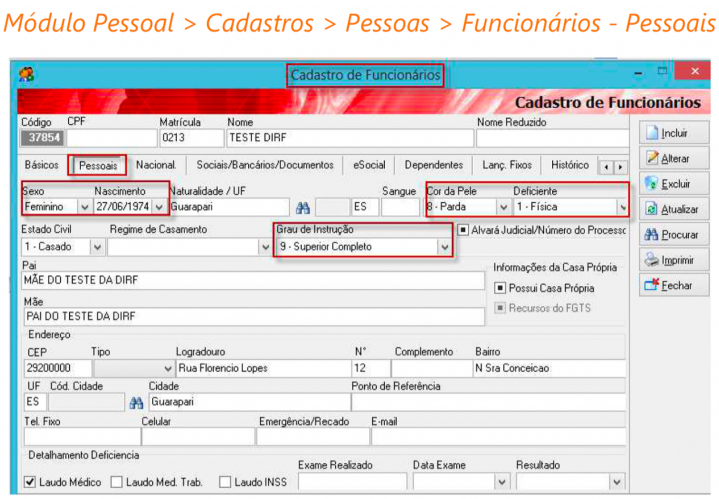
Information on employer-ascribed race, as well as many other employee and firm characteristics used in this paper, comes from the *Relação Anual de Informações Sociais* (RAIS). RAIS is an annual data collection instrument administered by Brazil's Ministry of Labor and Employment. Its objective is to "meet the needs of control, statistics, and information of government entities in the social and labor sectors" (Brasil, 2021). All formal firms (i.e., those registered with a tax identifier) and individuals who hire formal employees (excluding domestic employees) are required to report their employees' contract information to RAIS annually. Since 2003, one of the requested fields in RAIS is the employee's "raça ou cor" (literally, "race or color").

In RAIS, employers are responsible for reporting information on behalf of their employees. In larger firms, these records are typically submitted by a manager, HR department, or accountant, whereas in smaller firms the owner often completes the reporting. Figure C.1 shows the online interface, maintained by Brazil's Ministry of Labor and Employment, through which employers enter the information. Until 2018, RAIS reporting was conducted exclusively by the employer through this online system. The official RAIS completion manual from 2003 to 2018 provides no specific guidance on completing the race field, and there was no standardized training for those who entered the data.

Understanding how race is measured in RAIS thus requires examining how employers obtain information about workers' characteristics. At the time of hiring, workers must present various official documents, including the *Carteira de Trabalho e Previdência Social* (CTPS). The CTPS contains basic identifying information – including name, date of birth, gender, and residence – but does not include a field for race. Workers must also provide a photograph and proof of required education. Employers then record worker information in the *Livro de Registro dos Empregados* (LRE), an internal employee record maintained at the establishment level. Employers must enter each new worker into the LRE, and the information recorded there is used to comply with various reporting requirements, including RAIS. The required information includes date of birth, hire date, identification number, and various job characteristics. There is also space for a photograph of the worker. Unlike the CTPS, the LRE commonly includes a field for race, although completing this field is not mandatory (Cornwell et al., 2017).

To assess the extent to which race recorded in RAIS during our sample period reflects employers' racial perceptions, , we conducted qualitative interviews with employers and

Figure C.1: RAIS Interface



Notes: RAIS interface for reporting employee's information (2017 version).

Human Resources (HR) managers with experience reporting RAIS data. We found that firms rarely asked employees directly for their self-identified race. We conducted eight interviews between October 2025 and March 2026: two with small-business owners employing fewer than 50 workers, and six with HR managers experienced in larger firms.[1] The two small-business owners reported that, until 2018, they filled out the RAIS form themselves and imputed employees' race based on their own perceptions, without asking employees to self-identify.

Practices varied more among larger firms. The two most common procedures described in the interviews, each mentioned roughly equally often, were to report race based on employee photos collected at hiring or to ask direct managers to provide employee information to the firm's central HR team. The HR managers we interviewed both reported using these practices in their own firms and described them as common in other firms during the period we studied. In a single case, a health-sector firm with approximately 10,000 employees, the interviewee reported that the firm collected self-declared race at hiring and used this information to complete the RAIS form. Even in this case, however, the interviewee explained that not all newly hired employees completed the form, and that HR staff would then complete the profile using employee photos.

Overall, the interviews suggest that race in RAIS during this period is better interpreted as reflecting employers' racial perceptions than employees' self-identification. This interpretation is consistent with the language used in technical reports issued by the Brazilian government. For instance, a 2003 report by the *Instituto de Pesquisa Econômica Aplicada*

[1]The two small firms are in the retail sector. The larger firms are in the health, education, retail, and finance sectors, and range from 300 to 20,000 employees.

(IPEA, Institute for Applied Economic Research) states that, in RAIS, "it is usually not the employees who report their characteristics but the company" (Osório, 2003). Similarly, a more recent IPEA report states that, in RAIS, "the classification format is hetero-classification; that is, racial declaration is made by the employer or by an administrative employee responsible for it" (Silveira, 2022).

C.2 Skin Tone

Electoral Photos. We assemble a dataset of nearly three million official photographs of candidates running for political office in Brazil between 2004 and 2024. These photographs are used during elections in Brazil's electronic voting machines, and they typically differ from campaign materials used in advertisements. Electronic voting was introduced by the *Tribunal Superior Eleitoral* (TSE) in 1996 with the aim to reduce the time and costs of vote counting. Voters cast their ballots by typing a candidate's identification number into the machine's keypad. Once the full number is entered, the interface displays the candidate's name, party label, and photograph, after which the voter may either confirm the vote or restart the process. The photograph is thus visible only at the moment of verification, used as a mechanisms to help low-literacy voters cast their ballot and reduce the incidence of misrecorded votes (Fujiwara, 2015).

The submission of a valid photograph is mandatory for a candidacy to be accepted by the TSE. Technical and stylistic guidelines for these photographs are specified in electoral resolutions and have evolved over time. Most recently, Resolution No. 23,609 (2019), Article 27 II, details the following requirements:

- Dimensions: 161 x 225 pixels (W x H); without frame,
- Color depth: 24bpp,
- Background: uniform color,
- Characteristics: frontal (bust) view, with attire suitable for official photography, ensuring the use of ethnic or religious clothing and body paint, as well as accessories necessary for the person with a disability; the use of scenic elements and other adornments is prohibited, especially those that have connotations of electoral propaganda or that induce or hinder the recognition of the candidate by the electorate.

Before that, Resolution No. 23,405 (2014), Article 27 III, imposed similar requirements but specified an 8bpp grayscale format with a uniform, preferably white, background.

Skin Tone Measurement. Next, we provide details on our skin tone measurement algorithm. We begin by detecting the skin part in each photo. For that, we use Google MediaPipe's Image Segmentation tool, which relies on a vision transformer neural network to locate different objects in an image and allows us to segment the skin from non-skin areas of the photo. Specifically, we apply the *multi-class selfie segmentation model*, which classifies pixels as *background*, *hair*, *clothing*, *accessories*, *face-skin*, and *body-skin*.[2] We combine the latter two categories to form the skin mask of each photo. Next, we identify the representative color of the skin area, closely following the algorithm of Adukia et al. (2023). In particular, we use k-means clustering to partition the pixel colors into distinct groups based on their RGB color values. We group the pixels into $k = 5$ clusters but only use the largest three clusters for the skin tone measure, as the smaller clusters often correspond to shadows and non-skin parts such as eyes or lips. For each retained cluster, we compute its centroid, which summarizes its characteristic color, and then convert it from the RGB color space to the $L^*a^*b^*$ color space.[3] To obtain a single representative skin color for each photo, we average the clusters' L^*, a^*, and b^* values, weighting each cluster by its assigned share of the skin-area pixels. In order to reduce the dimensionality of skin color, we focus on the darkness or lightness of the color by retrieving only the L^* value from each photo's representative skin color.[4] Finally, we invert the L^* value, such that its scale ranges from 0 (lightest) and 100 (darkest). Figure C.2 shows examples of candidates with measured skin tones at the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentile of the distribution. For our analysis, we standardize the skin tone measure to have mean zero and unit standard deviation.

Comparison to Human-Assessed Skin Tone. As a validation exercise, we can compare our algorithmic skin tone measure to assessments made by a representative sample of human raters. For that, we draw on data collected by Bueno and Dunning (2017). They conducted an online survey of about 1,100 Brazilian respondents who were shown electoral photos of candidates running in the 2008 and 2010 elections and were asked to rate candidates' skin color on a scale from 1 (lightest) to 10 (darkest). We successfully match 4,826 candidates, each of whom was assessed by an average of three respondents. Figure C.3 presents a binned scatterplot for the relation between our algorithmic skin tone measure and the mean survey

[2] https://ai.google.dev/edge/mediapipe/solutions/vision/image_segmenter.

[3] The $L^*a^*b^*$ color space – also referred to as CIELAB – is a device-independent system defining colors with three values: L^* (lightness), a^* (red-green axis), and b^* (yellow-blue axis). Created by the International Commission on Illumination, the color space is perceptually uniform, meaning equal distances between colors represent similar perceived differences.

[4] In the figures in the paper that depict color differences across the skin tone distribution, we always vary the L^* value for a skin tone with $a^* = 17.2$ and $b^* = 19.1$, which correspond to the sample averages of a^* and b^* across all our colored photos.

Figure C.2: Skin Tone Examples

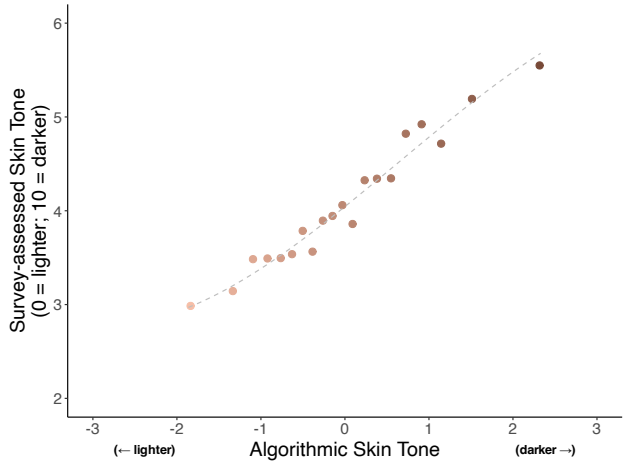


Percentile	10	25	50	75	90
Skin tone	23.3	29.9	37.0	46.1	56.7
Race (self)	*Parda* Brown	*Branco* White	*Pardo* Brown	*Preta* Black	*Parda* Brown

Notes: The figure presents five candidate examples, showing their electoral photographs, the extracted skin tone with its corresponding position in the skin-tone distribution, and the candidates' self-reported race.

rating for the same photo, showing a strong positive relation across the distribution. The overall correlation between both measures is 0.39. This can be compared to how much human raters diverge in their evaluation of the same photo. The average inter-rater correlation is 0.37. Hence, our algorithm's agreement with the average human perception is on par with the level of agreement observed among humans themselves.

Figure C.3: Algorithmic vs. Survey-assessed Skin Tone



Notes: The figure shows a binned scatterplot of survey-assessed skin tone and our algorithmic skin tone measure. The former is obtained from Bueno and Dunning (2017) and captures the average rating across Brazilian human coders who assess the skin tone of a given political candidate on a scale from 1 (lightest) to 10 (darkest). $N = 4,826$ candidates. The plot is constructed following the procedure in Cattaneo et al. (2024). For each sample, we divide the algorithmic skin tone distribution in twenty equal-sized bins and, for each bin, plot the average of survey-assessed skin tone.

Colored vs. Gray-scale Photos. About half of all candidate photos in our estimation sample (47.1%) are in color, with the other half – especially photos from candidacies in earlier years – being in grayscale.[5] Note that the algorithm is suitable for extracting skin tint from both types of images (following Adukia et al. (2023)). To assess potential differences in measurement quality, in Table C.1 we examine correlations with employer-ascribed race separately for skin tone measured from colored and grayscale photos. In column (1), we replicate our main result from Section 5 when using all available photos of each candidate, while in columns (2) and (3), we only use colored or grayscale photos to measure the skin tone of each candidate. For 63.2% (62.2%) of all candidates, we have at least one colored (grayscale) photo. We find that the predictive power of skin tone for racial ascription is only modestly higher when using colored photos rather than grayscale photos. At the same time, comparing columns (1) and (2), we see that the predictive power of colored photos (column (2)) is almost the same as when using all photos. This is because aggregating over more photos to measure the skin tone of each individual improves the measurement, which is highlighted in columns (4) to (6). Here, we focus on candidates who run multiple times and for whom we observe at least one colored and one grayscale photo. As can be seen, averaging across colored and grayscale photos yields a substantially larger correlation with employer-ascribed race compared to using colored or grayscale photos alone. Overall, Table C.1 shows that incorporating both colored and grayscale photos in our skin tone measurement increases the sample of candidates that we can analyze without compromising measurement quality.

Table C.1: The First Stage of Skin Tone: Colored versus Grayscale Photos

	All Candidates			Candidates with Colored and Grayscale Photos		
	(1)	(2)	(3)	(4)	(5)	(6)
Skin Tone - All Photos	0.107***			0.164***		
	(0.001)			(0.002)		
Skin Tone - Colored Photos		0.111***			0.114***	
		(0.001)			(0.002)	
Skin Tone - Grayscale Photos			0.090***			0.094***
			(0.001)			(0.002)
N	1,480,624	939,770	922,032	381,178	381,178	381,178

Notes: The table reports results from regressions of employer-ascribed race (indicator equal to one if the worker is ascribed non-White) on skin tone measured from all candidate photos vs. only colored photos vs. only grayscale photos. All skin tone measures are standardized using the mean and standard deviation from the distribution of skin tone based on all photos. Columns (1) to (3) use the full sample of candidates, while columns (4) to (6) restrict the sample to candidates who run for office multiple times and for whom we observe at least one colored and at least one grayscale photo. All regressions include year fixed effects. Standard errors clustered at the individual level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[5]Specifically, all photos of candidates until 2016 were in grayscale, in 2018 and 2020 they started to be published in color (depending on location and party), and since then almost all photos are colored.

Potential Manipulation of Skin Tone. Although candidate photographs are used on voting machines only for verification and are subject to strict standardization requirements, candidates may still have incentives to influence their appearance, including their skin tone (for instance, through lighting or makeup). We assess this possibility by exploiting repeated candidacies, which allow us to compare photographs of the same candidate across elections. In particular, we ask whether changes in self-declared race across candidacies – which is easy to change and may reflect electoral incentives to alter racial appearance – coincide with changes in measured skin tone. We estimate regressions relating changes in skin tone across consecutive candidacies to changes in self-declared race, considering separately transitions from White to non-White and from non-White to White. As shown in Table C.2, neither type of change in self-declared race predicts a corresponding change in measured skin tone: the estimated coefficients are small and statistically indistinguishable from zero.

Table C.2: Relationship between Changes in Skin Tone and Changes in Self-Declared Race

	(1)	(2)	(3)
[A] Initially White			
White → Non-White	0.037***	0.014	0.014
	(0.011)	(0.011)	(0.011)
N	65,458	65,458	65,458
R-squared	0.000	0.028	0.030
Adjusted R-squared	0.000	0.028	0.030
[B] Initially non-White			
Non-White → White	0.010	-0.016	-0.015
	(0.012)	(0.012)	(0.012)
N	61,213	61,213	61,213
R-squared	0.000	0.022	0.024
Adjusted R-squared	0.000	0.022	0.024
Election cycle FE	No	Yes	Yes
Grayscale dummy	No	No	Yes

Notes: The table reports results from regressions of changes in skin tone on changes in self-declared race. The sample consists of pairs of consecutive elections for political candidates whom we observe running for office in both elections. Panel A considers candidates who self-declared as White in the initial election. The explanatory variable is an indicator equal to one if the candidate self-declares as non-White in the subsequent election, and zero if the candidate remains White. Panel B considers candidates who self-declared as non-White in the initial election. The explanatory variable is an indicator equal to one if the candidate self-declares as White in the subsequent election, and zero if the candidate remains non-White. Column (2) adds election-year fixed effects, and column (3) additionally controls for whether the photo is grayscale. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C.3 Machine-Learning Prediction of Visual Cues

In addition to skin tone, we construct a flexible summary of the racial cues visible in the candidates' photos. For that, we use a machine-learning algorithm to predict employers' racial ascription based on the photos. Following recent computer vision papers on ethnicity inference from face images (Soudbakhsh et al., 2026; Fournier-Montgieux et al., 2025a), we apply a transfer learning approach that integrates a pretrained face recognition encoder with a non-linear classification head that we train on our data. Several studies have documented benefits of transfer learning in face analysis tasks relative to full end-to-end model training (Fournier-Montgieux et al., 2025b; Narayan et al., 2025; Sun et al., 2024; Amato et al., 2019).

For the pretrained face recognition model, we use the *Facenet-512d* encoder (Schroff et al., 2015), which is the best-performing encoder in face recognition benchmarks among those wrapped in the *DeepFace* Python package developed by Serengil and Ozpinar (2024). The encoder extracts a 512-dimensional vector of facial embeddings from each candidate photo. Before applying the face recognition encoder, we segment faces from the photos, relying on the *OpenCV* face detector that is also available in the *DeepFace* package.

We then use the facial embeddings, together with our skin tone measure, to train an RBF-kernel support vector machine (SVM) that predicts whether employers ascribe an individual as non-White.[6] Specifically, the outcome is the mode of employers' racial ascription across all years in which the individual appears in RAIS. We estimate the SVM in R using the *kernlab* implementation through the *tidymodels* interface. We use a random training dataset of 50,000 photos, with the remaining 510,849 photos used as hold-out data. Hyperparameters are tuned with a grid search, using 50% of the training sample to fit candidate models and using 50% as validation sample. After selecting the parameter combination with the highest validation ROC-AUC, we refit the SVM on all 50,000 training photos. The selected parameters are RBF kernel sigma = 0.0001 and regularization cost = 10. The model yields predicted probabilities of being ascribed non-White by the employer. In the hold-out data, we obtain a ROC-AUC of 0.736, indicating that the model has a 73.6% probability to correctly to assign a higher prediction to a worker ascribed as non-White by the employer than to a worker ascribed as White. The R^2 of the predictions in the hold-out data is 0.147. We only use the predictions for hold-out photos in the estimation in Table B.9.[7]

[6] We also experimented with a random forest model, which had notably worse prediction performance.

[7] For individuals with multiple photos, i.e. those running for office multiple times, we average the predicted probabilities across photos.

C.4 Name Index

Names can convey information about perceived racial or ethnic identity. This is especially well-documented in the United States, where racialized names are often used in observational and audit studies to manipulate perceived race (Kreisman and Smith, 2023; Bertrand and Mullainathan, 2004; Fryer and Levitt, 2004). Names have also been used as signals of racial identity in other settings, including Turkey (Assouad, 2025), India (Hanna and Linden, 2012; Banerjee et al., 2009), and Europe more generally (Ghekiere et al., 2025). We examine the extent to which names signal racial classification in Brazil.

Following Fryer and Levitt (2004), we construct a non-White name distinctiveness index. For each name n, this is defined as:

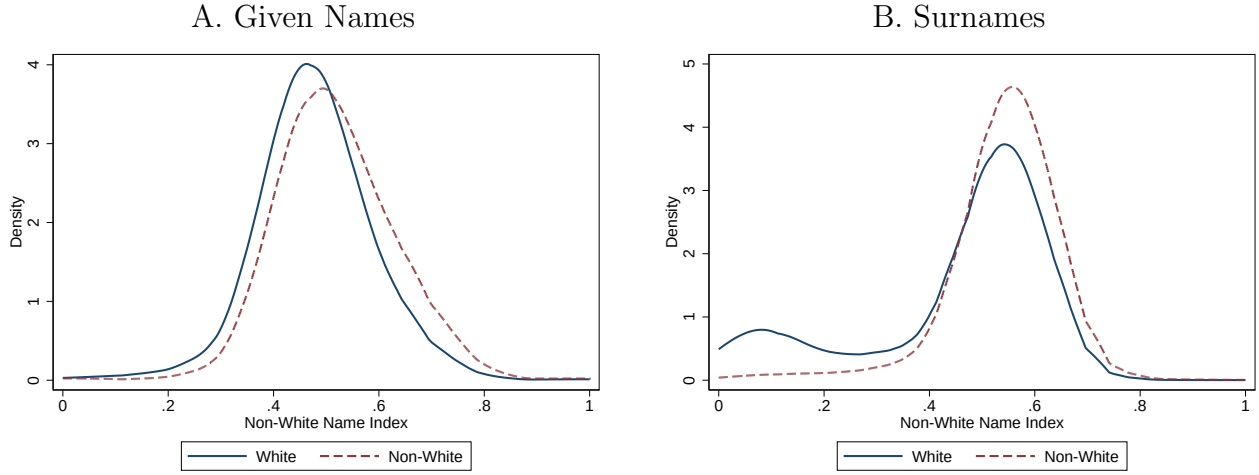
$$I_n = \frac{P(n \mid NW)}{P(n \mid NW) + P(n \mid W)}, \tag{1}$$

where NW denotes individuals classified as non-White by the employer in RAIS and W denotes individuals classified as White. The index ranges from zero, for a name observed only among White individuals, to one, for a name observed only among non-White individuals. It equals 0.5 for a name that is equally common across both groups. The index is computed separately for first and last names and is based on all workers in the full RAIS data. The underlying sample is at the person level, using the modal name and modal racial classification across the person's yearly records. To avoid mechanical predictive power, we construct leave-one-out measures that exclude each worker's own racial classification when calculating their index value. Figures C.4 displays kernel density estimates of the name indices for individuals currently classified as White and non-White.

Brazil's long history of racial mixing and the widespread use of Portuguese and Catholic naming conventions across racial groups provide reasons to expect weaker racial differentiation in given names than in the United States. During slavery, Catholic baptism frequently resulted in enslaved Africans being assigned or recorded under Christian, generally Portuguese, given names (Silva, 2018; Hébrard, 2003). Together with strong racial mixing and common naming norms, this historical process contributed to a repertoire of given names broadly shared across racial groups. Consistent with this, Panel A of Figure C.4 shows that the vast majority of given names have a Non-White Name Index between 0.4 and 0.6, with substantial overlap between White and non-White individuals. Hence, most first names in Brazil provide relatively little information about racial classification.

Surnames, by contrast, contain more information about racial classification in Brazil, particularly as signals of whiteness. Panels B of Figure C.4 shows a substantial mass at non-White name-index values below 0.2, but considerably less above 0.8. This asymmetry

Figure C.4: Distribution of First Name and Surname Indices in RAIS



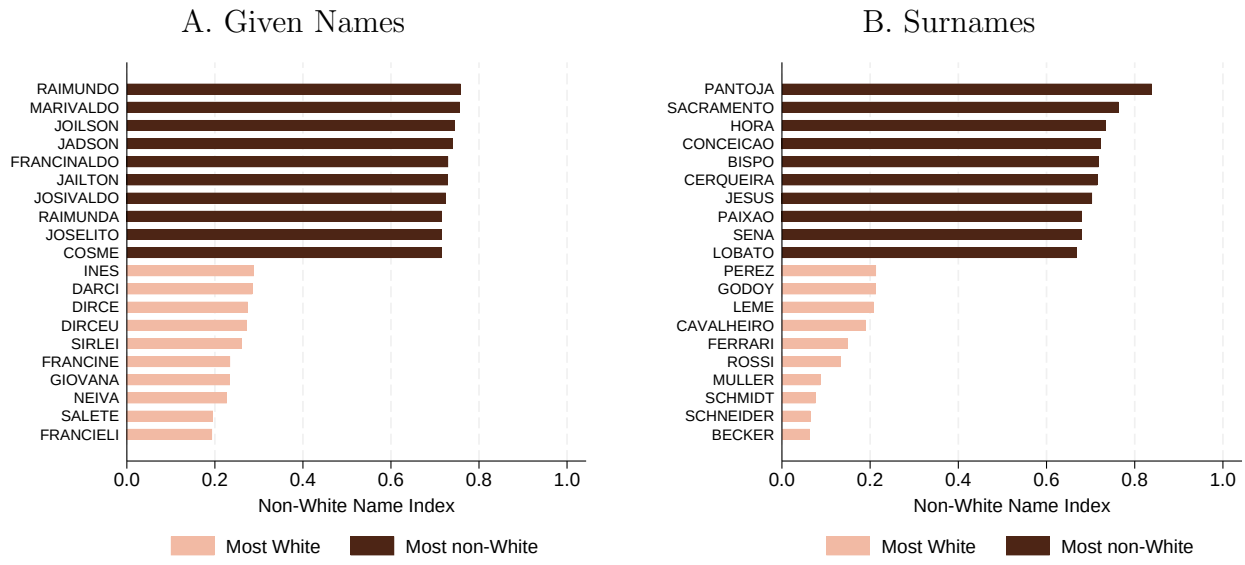
Notes: The figures display the distribution of the non-White name index in RAIS, separately for workers classified as White and non-White. Distributions are smoothed out using an Epanechnikov Kernel with a 0.05 bandwidth. Panel A reports the index based on the worker's first given name, while Panel B reports the index based on the worker's last surname. The unit of observation is a worker-year, and racial classification corresponds to the employer-reported race in that year.

is consistent with nineteenth- and early-twentieth-century European immigration, which introduced relatively distinctive European surnames into Brazil (Monasterio, 2017).

Figure C.5 depicts the given names and surnames with the lowest and highest non-White name indices among names observed for at least 10,000 distinct individuals in each dataset. Several of the surnames with the lowest indices are of European, especially German, origin. At the opposite end of the distribution, several surnames with the highest non-White indices have Catholic or other religious origins, including "Jesus," "Conceição," and "Bispo." These findings are consistent with earlier Brazilian evidence that surnames predict race more strongly than given names, that distinctively White surnames tend to be of non-Portuguese origin, and that Catholic devotional surnames are more strongly associated with Afro-Brazilian and lower-income individuals (Scottini, 2011).

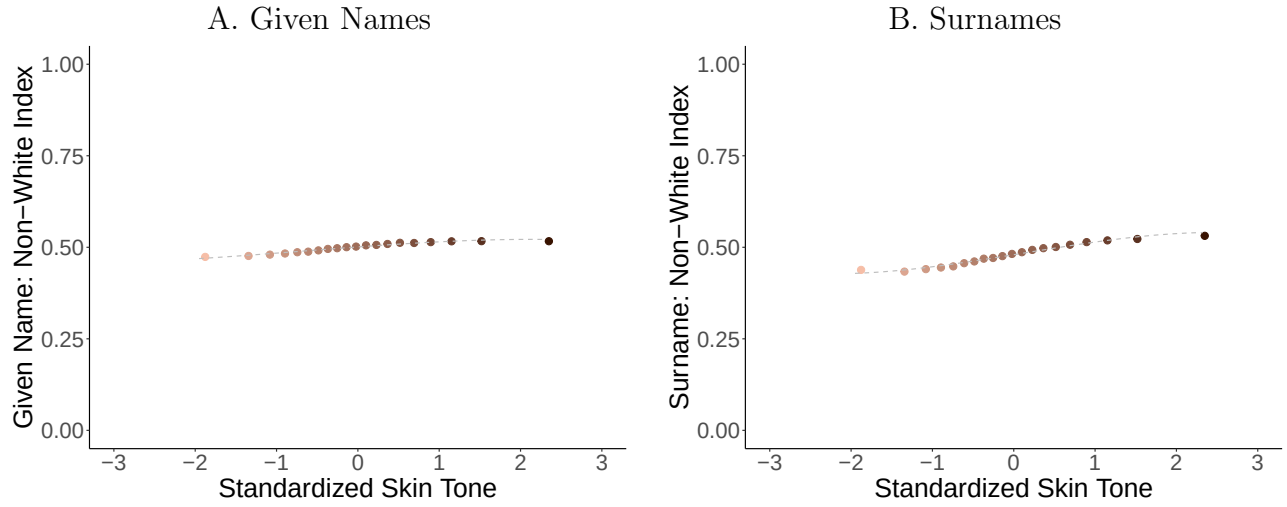
Finally, the name indices are only weakly correlated with standardized skin tone in the politician sample, as shown in Figure C.6. This suggests that names and phenotype capture distinct, or at most weakly related, dimensions of racialization in Brazil. This interpretation is strengthened by our evidence on compliers in the first stage (see Table B.15): relative to skin tone compliers, name compliers are more highly educated and receive higher wages, reflecting exactly the observation that the most racialized names in Brazil are surnames that reveal European descent.

Figure C.5: Names Most Associated with White and Non-White Individuals in RAIS



Notes: The figures display the ten names with the lowest and highest non-White name indices among names observed for at least 10,000 distinct individuals in the full RAIS sample. Panel A uses individuals' first given names, while Panel B uses their last surnames. For each name n, the index is defined as $P(n \mid NW)/[P(n \mid NW) + P(n \mid W)]$, where NW denotes non-White and W denotes White. Racial classification is based on each individual's modal employer-reported race across years. The index is calculated using leave-one-out racial-group and name counts and then averaged within each name for the ranking. Higher values indicate names more strongly associated with non-White individuals. Light bars report the ten names with the lowest index values, and dark bars report the ten names with the highest index values.

Figure C.6: Relationship between Name Indices and Skin Tone



Notes: The figures show binned scatterplots of the non-White name indices and workers' standardized skin tone. Panel A uses the index based on the worker's first given name, while Panel B uses the index based on the worker's last surname. Skin tone is standardized within the politician analysis sample to have mean zero and standard deviation one, with higher values indicating darker skin tone. The name indices are constructed using the full RAIS sample. For each name n, the index is defined as $P(n \mid NW)/[P(n \mid NW)+P(n \mid W)]$, where NW denotes non-White and W denotes White. The plots are constructed following the procedure in Cattaneo et al. (2024). We divide the sample into 20 equal-sized bins and, for each bin, plot the mean skin tone and mean name index within a bin. We fit a third-degree polynomial based on the underlying data.

Appendix References

Adukia, Anjali, Alex Eble, Emileigh Harrison, Hakizumwami Birali Runesha, and Teodora Szasz, "What We Teach About Race and Gender: Representation in Images and Text of Children's Books," *The Quarterly Journal of Economics*, 2023, *138* (4), 2225–2285.

Amato, Giuseppe, Fabrizio Falchi, Claudio Gennaro, Fabio Valerio Massoli, Nikolaos Passalis, Anastasios Tefas, Alessandro Trivilini, and Claudio Vairo, "Face verification and recognition for digital forensics and information security," in "2019 7th International Symposium on Digital Forensics and Security" IEEE 2019, pp. 1–6.

Assouad, Lydia, "Leadership and nation-building," *Unpublished Manuscript*, 2025.

Banerjee, Abhijit, Marianne Bertrand, Saugato Datta, and Sendhil Mullainathan, "Labor market discrimination in Delhi: Evidence from a field experiment," *Journal of comparative Economics*, 2009, *37* (1), 14–27.

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

Brasil, "Decreto nº 10.854, de 10 de novembro de 2021," November 10 2021. Regulamenta disposições relativas à legislação trabalhista e institui o Programa Permanente de Consolidação, Simplificação e Desburocratização de Normas Trabalhistas Infralegais e o Prêmio Nacional Trabalhista, alterando o Decreto nº 9.580/2018. Diário Oficial da União, Seção 1, 11 nov. 2021.

Bueno, Natálie S and Thad Dunning, "Race, Resources, and Representation: Evidence from Brazilian Politicians," *World Politics*, 2017, *69* (2), 327–365.

Cattaneo, Matias D, Richard K Crump, Max H Farrell, and Yingjie Feng, "On binscatter," *American Economic Review*, 2024, *114* (5), 1488–1514.

Cornwell, Christopher, Jason Rivera, and Ian M. Schmutte, "Wage Discrimination When Identity Is Subjective: Evidence from Changes in Employer-Reported Race," *Journal of Human Resources*, 2017, *52* (3), 719–755.

Fournier-Montgieux, Alexandre, Hervé Le Borgne, Adrian Popescu, and Bertrand Luvison, "Reliable and Reproducible Demographic Inference for Fairness in Face Analysis," *arxiv preprint arXiv:2510.20482*, 2025.

_ , Michael Soumm, Adrian Popescu, Bertrand Luvison, and Hervé Le Borgne, "Fairer analysis and demographically balanced face generation for fairer face verification," in "2025 IEEE/CVF Winter Conference on Applications of Computer Vision" IEEE 2025, pp. 2788–2798.

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

Fujiwara, Thomas, "Voting Technology, Political Responsiveness, and Infant Health: Evidence From Brazil," *Econometrica*, 2015, *83* (2), 423–464.

Ghekiere, Abel, Billie Martiniello, Daniel Capistrano, Jeremy Kuhnle, Stefanie

Sprong, Pelin Atay, Héctor Cebolla Boado, Mathew Creighton, Valentina Di Stasio, Mariña Fernández-Reino et al., "The perception of names in experimental studies on ethnic origin: a cross-national validation in Europe," *Scientific data*, 2025, *12* (1), 1883.

Hanna, Rema N and Leigh L Linden, "Discrimination in grading," *American Economic Journal: Economic Policy*, 2012, *4* (4), 146–168.

Hébrard, Jean, "Esclavage et dénomination: Imposition et appropriation d'un nom chez les esclaves de la Bahia au XIXe siècle," *Cahiers du Brésil Contemporain*, 2003, (53–54), 31–92.

Hull, Peter, "One Weird Trick to Characterize Effective Populations in Design-Based Specifications," Technical Report, Metrics Note 2025.

Kreisman, Daniel and Jonathan Smith, "Distinctively Black Names and Educational Outcomes," *Journal of Political Economy*, 2023, *131* (4), 877–897.

Monasterio, Leonardo, "Surnames and Ancestry in Brazil," *PLOS ONE*, 2017, *12* (5), e0176890.

Narayan, Kartik, Vibashan VS, and Vishal M. Patel, "FaceXBench: Evaluating multimodal LLMs on face understanding," *arxiv preprint arXiv:2501.10360*, 2025.

Osório, Rafael Guerreiro, "O sistema classificatório de cor ou raça do IBGE," Technical Report, Instituto de Pesquisa Econômica Aplicada (IPEA) 2003.

Schroff, Florian, Dmitry Kalenichenko, and James Philbin, "FaceNet: A unified embedding for face recognition and clustering," in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition" 2015, pp. 815–823.

Scottini, Lucas Costa, "O que o nome nos ensina? Padrões sociais e raciais de nomes e sobrenomes e performance escolar em São Paulo," Dissertação de mestrado, Universidade de São Paulo, São Paulo 2011.

Serengil, Sefik and Alper Ozpinar, "A Benchmark of Facial Recognition Pipelines and Co-Usability Performances of Modules," *Journal of Information Technologies*, 2024, *17* (2), 95–107.

Silva, Gian Carlo de Melo, "Os filhos da escravidão e o primeiro sacramento: Batismo, compadrio e sociedade escravista na freguesia de Santo Antônio do Recife, Capitania de Pernambuco, no fim do século XVIII," *História (São Paulo)*, 2018, *37*, e2018024.

Silveira, Leonardo, "Imputação da informação de raça/cor na Rais para o setor público brasileiro," Technical Report, Instituto de Pesquisa Econômica Aplicada (IPEA) 2022.

Soudbakhsh, Hasti, Sonjoy Ranjan Das, Bilal Hassan, and Muhammad Farooq Wasiq, "Transfer Learning-Based Ethnicity Recognition Using Arbitrary Images Captured Through Diverse Imaging Sensors," *Sensors*, 2026, *26* (3), 886.

Sun, Haomiao, Mingjie He, Tianheng Lian, Hu Han, and Shiguang Shan, "Face-MLLM: A large face perception model," *arxiv preprint arXiv:2410.20717*, 2024.