

La Raza C3smica? The Construction of Latinx Racial Identities Beyond Panethnicity

Chris Flores*

2 April 2023

Abstract

How do Latines classify their race, and what factors shape this process? While past research has explored the predictors of panethnic identification as Hispanic or Latino, comparatively less work has examined the construction of racial identities such as Afro-Latino, white, multiracial, or indigenous, categories which are recognized across Latin America. In this paper, I argue that incorporation (also known as assimilation or Americanization), phenotype, socioeconomic status, and party ID are key predictors of these labeling choices. I employ survey data from four nationally representative surveys of Latinos to test these hypotheses, and I use a random forest regression to explore other potentially important predictors among hundreds of demographic, political, and psychological variables. My results show that skin tone and incorporation are indeed the most important predictors of these classification choices. Latinos with darker skin tones are much less likely to identify as white and more likely to identify with nonwhite labels like Afro-Latinx. Less incorporated Latinos are also more likely to identify with any labels beyond Latinx. However, more incorporated Latinos are sometimes more likely to identify as white. SES and party ID do not emerge as central predictors, but national origin and age are. These findings are broadly in line with a triracial understanding of U.S. racial structures, in which some Latinos "become" white or Latine.

*PhD Student, Stanford University

Introduction

Latinos¹ are widely recognized as a distinct ethnoracial group in the U.S., the second largest behind white Americans (Cobas, Duany, and Feagin 2015). A wide body of work in the social sciences has now examined the construction of race for different Latino groups and the racialization of Latinos (Itzigsohn and Dore-Cabral 2000; Hitlin, Brown, and Elder 2007; Telles and Ortiz 2008; Mora 2014; Vargas and Stainback 2016; Feliciano 2016; Gómez 2018; Porter and Snipp 2018). Scholars have explored the predictors of panethnic or national origin identities (Masuoka 2006; Sanchez and Masuoka 2010; Fraga et al. 2010; de la Garza and Yang 2020; Martínez and Gonzalez 2021), the assimilation of Latine immigrants into the U.S. racial hierarchy (Portes and Zhou 1993; Alba and Nee 2003; Golash-Boza 2006; Alba and Islam 2009; Frank, Redstone Akresh, and Lu 2010; Schacter 2016), and the implications of panethnic identity for political behavior (Jones-Correa and Leal 1996; García Bedolla 2005; Barreto 2012; Huddy, Mason, and Horwitz 2016; Zepeda-Millán 2017).

Yet Latinos are racially diverse. On one dimension, Latines may be understood as a distinct ethnoracial group, but on a second dimension Latinos can be of any race – of varying skin tones, phenotypes, and identities (Telles 2018). Indeed, scholars of Latin America have examined racial differences in various countries and their implications for inequality and politics (Comas-Díaz, Lykes, and Alarcón 1998; Hooker 2005; Schwartzman 2007; Sue and Golash-Boza 2009; Wade 2010; Telles and Flores 2013; Telles 2014; Monk 2016; Telles and Torche 2019). Comparatively less attention has been paid to racial differences among Latines in the U.S. While past work has explored the predictors of Latinos' racial classification choices on Census-style questions (Tafoya 2005; Irizarry, Monk, and Cobb 2022), the role of racial differences for inequality (Denton and Massey 1989; Telles and Murguia 1990; Massey and Denton 1992), the construction of racial identity for Dominican and Puerto Rican immigrants (Itzigsohn and Dore-Cabral 2000; Roth 2010, 2012, 2016), and the role of ideology in shaping self-reported skin tone (Cutiaia Wilkinson and Earle 2013; Ostfeld and Yadon 2021), much remains to be known about variation in Latinxs' racial classification or identity choices. Despite the variation in racial labels Latin Americans may adopt – including labels such as Afro-Hispanic or Afro-Latino, indigenous, and multiracial labels of various forms – little research has explored the adoption of these racial identities beyond panethnic or national origin identities. Which Latinos identify with racial labels tapping into this second dimension of race, and which do not? And what are the consequences of these choices for racial understandings and processes of

¹I use Latine, Latinx, and Latino interchangeably throughout this paper.

assimilation?

In this paper, I argue that Latinxs have considerable flexibility to identify their race in different ways beyond their ancestry given the historical and present understandings of race across the Americas. Latinos can identify with labels like Afro-Latinx, indigenous, multiracial, or white, or reject these labels and identify primarily as Hispanic or Latino. I contend that this choice depends critically on demographic, political, and psychological predictors – specifically, phenotype and Latinos’ incorporation into U.S. society and politics. I argue that incorporation (also known as assimilation or Americanization) shapes racial self-classification because it determines which ways of understanding race Latinos primarily rely upon. I develop two competing hypotheses for the role of incorporation. On the one hand, less incorporated Latinos may be more likely to hold *mestizaje* ideologies – the belief that all Latin Americans are of one, mixed race – and more incorporated Latines may be more familiar with the rigidity and discreteness of U.S. racial categories. On the other hand, recent movements by Black and Indigenous Latin Americans have brought discrete racial differences to the fore, and past work shows that more incorporated Latinos increasingly see themselves racially as Latinx. I explore the role of phenotype, hypothesizing that nonwhite skin tones or appearances are positively related to nonwhite identification and negatively related to white identification (Telles 2014; Irizarry, Monk, and Cobb 2022). Additionally, I examine socioeconomic status (Schwartzman 2007; Telles 2014; Davenport 2016), and politics (Egan 2020; Ostfeld and Yadon 2021), which have both been identified by past literature as key factors shaping identification choices.

I investigate these factors by employing data from four nationally representative surveys of Latinos that tap into these intra-group racial labels. In particular, I focus on the construction of four classification choices: as Afro-Latinx; indigenous; *mestizo*, *mulato*, or multiracial; and white, labels which are common in both Latin America and the U.S. In order to explore the predictors of these identity choices, I use random forest regression models (Breiman 2001; Montgomery and Olivella 2018). This method allows me to leverage the hundreds of variables across these datasets to determine which ones best explain variance in these outcomes and is an ideal approach for analyzing the complexity of identity choices. I first adopt an exploratory approach using variable importance measures (VIMs) to better understand the contours of these identification choices. Then, I produce partial dependence plots from these models, which explore the marginal association of key predictors with identity choices net of all other correlates to test my core hypotheses.

My results show that incorporation and skin tone are indeed strong predictors of Latines’ racial self-classification choices. As expected, Latinxs with darker skin tones are more likely to identify as

Afro-Latinx and much less likely to identify as white – for the latter, by as much as 20 percentage points, controlling for all other variables available. Incorporation matters, but not in the exact way hypothesized. All things equal, less incorporated Latinxs are more likely to identify with labels beyond panethnicity, but sometimes more incorporated Latinxs are more likely to identify as white. Additionally, I show that SES and politics are not central predictors of these identity choices, and the direction of their association with identity choices is at times counterintuitive – in some cases, Latines who are more Republican or have higher incomes are more likely to identify as Afro-Latinx.

My findings reveal some support for divergent racial understandings as one mechanism by which incorporation relates to Latinos’ racial classification choices, but they also suggest that some Latinos may assimilate – or aspire to assimilate – into whiteness. This identity choice, however, is constrained by phenotype, with assimilation into whiteness being limited for Latinos with darker skin tones or nonwhite appearances. These results speak to broader theories of assimilation, racial positioning, and boundary construction in the U.S. They indicate support for a triracial model of the U.S. racial hierarchy, with some Latinos becoming part of the dominant white racial group and others experiencing downward assimilation into Latino or “collective Black” categories (Portes and Zhou 1993; Bonilla-Silva 2002, 2004; Alba and Nee 2003; Redstone Akresh, and Lu 2010). Further work is necessary to better understand what it means to adopt these racial labels and whether they entail a psychological sense of attachment or group consciousness. Nonetheless, my findings demonstrate that Latinos do classify themselves with racial labels beyond Latinx at high rates, that these choices are not completely determined by ancestry, and that they are intricately related to other social factors. This paper builds upon and contributes to recent scholarship that has examined the fluidity of racial labels and categories in the U.S., despite the historical rigidity of race (Saperstein and Penner 2012; Davenport 2016, 2018, 2020), and I join other scholars in the call to rigorously unpack intra-group differences among Latines (Beltrán 2011; Mora 2014; Alberto 2017).

Race in Comparison: The United States and Latin America

Race is socially constructed and contested by political elites, activist groups, and others. Like the U.S., Latin America has a multiracial history rooted in colonization, enslavement, and immigration, but Latin American countries have adopted different understandings of race over time (Davenport 2020). During the Spanish colonial period, a racial caste system existed based on Spanish ancestry, with white people receiving the greatest legal rights and social opportunities. The Spanish colonies, however, experienced more race mixing and had a greater multiracial population than

English colonies, to which family-based immigration was more common (Telles 2004, 2014). Following independence in the 19th century, Latin American governments – motivated by ideologies of scientific racism – promoted European migration to whiten their largely nonwhite populations. Eventually, elite thinkers turned from scientific racism to celebrating *mestizaje* in the 20th century to foster national unity, continue to whiten their populations, and erase Black and Indigenous people (Helg 1990; Telles 2014; Davenport 2020). As a result, *mestizaje* has become key to the national identities of many Latin American countries, and many Latin Americans identify with multiracial categories (Wade 2010). I refer to the belief that all Latin Americans are of a multiracial background – advanced by prominent thinkers such as José Vasconcelos (Hooker 2014) – as the ideology of *mestizaje*. The title of this paper refers to the famous work by Vasconcelos in which he argues that Latin Americans are forming a “new race” distinct from other existing races – *la raza cósmica*, or the cosmic race – which captures the idea of *mestizaje* well.

Importantly, elites employed *mestizaje* strategically to erase racial distinctions and racism in Latin America. This was rooted in efforts to resist U.S. imperialism, by contrasting the racial exceptionalism of Latin America with the U.S.’ rigid racial hierarchy (Hooker 2014). Racial ideologies also vary significantly by country, with different countries constructing distinct mythologies or narratives around race mixing. Some countries, such as Argentina and Costa Rica, did not embrace *mestizaje* and instead continued to strive toward whiteness (Telles and Flores 2013). These national differences have important consequences for Latinas’ classification choices, on which I elaborate later.

In recent decades, there has been a growing recognition of racial and ethnic categories and inequality in Latin America (Wade 2010; Telles 2014). This is in large part due to movements by Black and Indigenous groups, which have existed for a long time but have increased in prominence since the 1980s. The result has been growing recognition of indigenous rights by governments and affirmative action for Black Brazilians (Wade 2010), though indigenous groups have generally been more successful in their claims to rights based on their distinct cultural identity, which is less clear for Afro-Latinos (Hooker 2005). While *mestizaje* has strongly influenced racial understandings and identity choices in the region, these changes suggest that Latin Americans are increasingly aware of racial distinctions and willing to identify with categories like Afro-Latino or Indigenous (Telles and Torche 2019).

In the U.S., on the other hand, race has been historically understood through the lens of a rigid Black-white binary, and the practice of hypodescent has classified multiracial people as nonwhite (Cornell and Hartmann 2006). However, increases in interracial marriages, the growth

of the multiracial population, and immigration from Asia and Latin America are fundamentally changing the boundaries of race in the U.S. and making racial fluidity more common (Bonilla-Silva 2002, 2004; Davenport 2020). Norms of hypodescent appear to be weakening, as many people with biracial parentage identify as mixed race or white (Davenport 2016). Continued immigration from Latin America and the growth of the racially diverse Latinx population has also been critical to this shift. Despite substantial within-group racial differences, Latinos have come to be recognized as a distinct ethnoracial group in the U.S. (Mora 2014; Cobas, Duany, and Feagin 2015; Telles 2018). Roth (2012) argues that the U.S. is moving toward a “Hispanicized” triracial hierarchy in which Latinos are seen by others – and see themselves – as an intermediate group between Black and white. Relatedly, Bonilla-Silva (2002, 2004) contends that some Latinos will assimilate into whiteness, others into the “collective Black,” and others into an intermediary “honorary white” category.

Theory and Argument

The histories of race in the Americas and their changes over time – *mestizaje* in Latin American thought, the recent prominence of Black and indigenous movements, rigid notions of hypodescent in the U.S., changing racial structures, and recognition of Latinos as a racial group – indicate that Latinx have considerable latitude in identifying or self-categorizing their race. I argue that – within certain constraints – Latinx can choose to identify with a variety of different racial classifications such as Afro-Latinx or Black, indigenous, or multiracial, or reject such classifications and instead identify in panethnic or national origin terms. Critically, this depends on demographic, political, and social psychological factors. I argue that “incorporation” – or the processes by which Latinx become integrated into U.S. society and politics (Abrajan 2010) – is a key predictor of the adoption of racial identities beyond Latinx. Below, I outline two hypotheses that lead us to expect either a positive or negative association between incorporation and the adoption of racial identities, which I will test empirically. I also detail expectations for other central predictors including SES, ancestry and phenotype, party ID, and others.

In this paper, I focus on racial self-classification, or how people identify their race in a close ended survey. Race and racial identity are multidimensional (Saperstein 2012; Roth 2016; Davenport 2016, 2020), and the different dimensions may or may not overlap. For example, how someone answers a racial identity question on a survey may not align with the internal sense of racial identity they describe in open-ended terms. Racial identity often does not overlap with phenotype (physical

characteristics), how others might perceive someone’s race, or how one thinks that others perceive their race (Roth 2016). Racial self-classification also may not reflect the importance that one attaches to that identity – one can identify with a category without feeling a strong sense of attachment to it (Tajfel and Turner 1986; Ellemers, Kortekaas, and Ouwerkerk 1999). As I discuss toward the end, future work would do well in exploring the affective or evaluative components of Latinx racial identities. For my purposes, however, I treat racial self-classification as itself an important outcome and requisite first step toward affective attachment or group consciousness (Miller et al. 1981). Classifying ones’ race on a survey is a common and consequential practice given the history of Census racial classifications (Loveman 2014; Mora 2014) and may influence one’s internal sense of identity. I treat this choice as a valuable signal of identity, though further work should clarify whether behavioral differences can be observed as a consequence of self-classification.

Incorporation

Past literature has used various terms such as acculturation, assimilation, and Americanization to refer to related processes of cultural integration into the United States. I use the term “incorporation” following Abrajano (2010) and Yang and de la Garza (2017), who succinctly describe the concept as indicating “the likelihood that ordinary Latinos are familiar with, knowledgeable about, and engaged with mainstream society, institutions, and social practices” (691). I expect incorporation to be a key predictor of racial identity choice because of the historically distinct ways of conceptualizing race in the U.S. and in various Latin American countries. However, race is not static, and racial understandings have changed over time. This leads to two competing expectations: a negative association with the propensity to racially self-classify as something other than Latinx – e.g., less incorporated Latinxs are more likely to do so – or a positive association with this propensity.

Interviewing Puerto Rican and Dominican migrants in New York, Roth (2012) found that more incorporated migrants – those who have experienced more socialization in the U.S., have social networks with more non-Latinos, or have attended educational institutions in the U.S. – were more likely to identify themselves as Hispanic or Latino and adopt a “triracial schema,” or a cognitive way of understanding race in which White, Black, and Latino people form the major racial groups in the U.S. On the other hand, less incorporated interviewees – those with strongly Latinx networks, who speak predominantly Spanish, or who have lived in the U.S. for only some time – were more likely to identify themselves and others in national origin terms, with “continuum” labels (i.e., the variety of racial labels used in different Latin American countries such as *trigueño* or *jabao*), or express some

uncertainty around the concept of race. Of course, there is significant variation across these groups, and Roth argues that sustained immigration is shaping racial understandings both in the U.S. and in the Caribbean. These findings align with earlier work by Itzigsohn and Dore-Cabral (2000), who find that Dominican immigrants adopt panethnic identities as they become incorporated into the U.S.

This work suggests that incorporation influences the racial identification process because it signals which ways of conceptualizing race people adopt. Less incorporated respondents are more proximate to Latin American understandings, with their focus on *mestizaje*, fluidity, and a multitude of categories; more incorporated respondents are more proximate to U.S. understandings, with its history of racial rigidity and modified Black-white binary. However, I diverge from Roth here in describing two competing conjectures that emerge from these insights. First, we might expect that due to the prevalence of *mestizaje* ideologies, less incorporated respondents would be more likely to reject racial classifications in favor of panethnic or national origin terms. More incorporated respondents would be more likely to classify in racial terms beyond panethnicity because of their experience with racial discrimination or rigidity in the U.S. or because of their incorporation into a society which has historically categorized people discretely. Irizarry, Monk, and Cobb (2022) discern some support for this conjecture, finding that less incorporated Latine youth are more likely to reject any racial classification, whereas more incorporated youth are more likely to identify with more than one racial category. This leads to hypothesis 1a:

H1a: Less incorporated respondents will be less likely to self-classify with racial terms beyond Hispanic or Latinx.

On the other hand, because of the historical usage of a number of racial categories in Latin America, the more recent visibility of Black and Indigenous people, and persistent racial discrimination in the region (Telles 2014), we might expect less incorporated Latinos to instead be more likely to identify with racial labels other than Latinx. Likewise, more incorporated Latinos may instead adopt the triracial perspective advanced by Roth and related to arguments put forth by Bonilla-Silva (2004). This leads to hypothesis 1b:

H1b: Less incorporated respondents will be more likely to self-classify with racial terms beyond Hispanic or Latinx.

Socioeconomic status (SES)

We can expect SES to similarly influence racial classification choices to the extent that it is negatively associated with incorporation. However, independent of incorporation, income and education may relate to classification choices through two distinct channels. First, research on multiracial populations in the U.S. and Latin America finds that higher SES is associated with classification as white due to the racialization of income and wealth (Schwartzman 2007; Davenport 2016). Past work notes that nonwhite Latinos experience more discrimination, earn lower wages, have worse health outcomes, and are more segregated from white people (Telles and Murguía 1990; Massey and Denton 1992; Chavez-Dueñas, Adames, and Organista 2014), so it is possible that these race-class experiences shape the adoption of nonwhite racial labels. This leads to hypothesis 2a:

H2a: Respondents with higher family incomes will be more likely to identify as white and less likely to identify with nonwhite labels.

The alternative to this hypothesis is that we will observe little to no association between income and classification choices. This is plausible given that Telles (2014) and Telles, Flores, and Urrea-Giraldo (2015) find a consistent association between skin tone and income but inconsistent association between race and income because the racial classifications people adopt are intertwined with other factors and often do not align with experiences of discrimination.

Education is closely related to income but may have an independent influence by shaping the way people understand race. Roth (2012) shows that more educated respondents in both the Caribbean and New York tend to distinguish between race and ethnicity, and Davenport (2016) finds that multiracial college students with highly educated parents tended to more conscious of their nonwhite backgrounds. Because Latinxs are racially and phenotypically diverse, I argue that education will be related to the adoption of any classifications beyond Latinx, including white and nonwhite labels. These findings lead to hypothesis 2b:

H2b: More educated respondents will be more likely to self-classify with racial terms beyond Hispanic or Latinx.

Ancestry and Phenotype

Though race is constructed, it is constrained by ancestry and phenotype, particularly given the historically rigid boundaries of race in the U.S. (Cornell and Hartmann 2006; Davenport 2020; Davenport, Jefferson, and Rendelman 2022). This makes it less likely that Latinos who have darker skin

tones or appear phenotypically nonwhite will be socially accepted as white or even Latinx, given the historical erasure of Afro-Latinidad. Latinos with lighter skin tones or phenotypically white features may have less latitude in choosing a Black, Indigenous, or other nonwhite label. Because having a darker skin tone or phenotypically Black appearance is associated with the experience of racial discrimination and a number of negative health and occupational outcomes (Telles and Murguía 1990; Eberhardt et al. 2006; Chavez-Dueñas, Adames, and Organista 2014), such individuals may be more likely to adopt a nonwhite racial label beyond Latinx. Likewise, ancestry has historically constrained race in the U.S. such that individuals with multiracial heritage have been excluded from whiteness.

Along these lines, Irizarry, Monk, and Cobb (2022) find that youth in the Add Health study with a "dark" interviewer-rated skin tone were more likely to shift their racial classification to Black in a later wave of the study, while those with a "medium" skin tone were more likely to shift from white to a nonwhite category. They argue that this represents a "sedimentation of the color line"—youth bring their racial classification in alignment with how others perceive their race as they age and experience racial ascription by others. This leads to the third hypothesis:

H3: Respondents with darker skin tones, nonwhite ancestry, and nonwhite phenotypic features will be more likely to identify with nonwhite racial labels beyond Latinx.

I note, however, that Latinx itself is a racialized category (Cobas, Duany, and Feagin 2015), so no association may be observed between skin tone and self-classification if respondents can still identify as Latinx in some way. That is, if respondents identify their race as Latinx in one question and self-classify with different labels in another, we may observe no relationship. However, if respondents are asked to identify their race where Hispanic or Latino is not an option, the hypothesis should be more likely to hold.

Party Identification

A substantial body of work in political science shows the strong relationship between race and political behavior (Berelson, Lazarsfeld, and McPhee 1954; Dawson 1994; Jardina 2019). Since the partisan realignment beginning in the 1930s, race has become a major cleavage between the Republican and Democratic Parties (Schickler 2016). Today, the two major political parties are sorted along ethnoracial lines, with white people being more likely to identify as Republicans and nonwhite people being more likely to identify as Democrats (Mason 2018); White people who score lower on

measures of racial prejudice are more likely to identify as Democrats (Sides, Tesler, and Vavreck 2019). Importantly, race is endogenous to political behavior. Recent work by Egan (2020) shows that people may shift their racial self-classifications based on partisan and ideological attachments. Ostfeld and Yadon (2022) show that Latinos who overestimate the lightness of their skin tone in a self-report relative to a machine rating are more likely to hold conservative policy positions, and vice-versa. This leads to hypothesis four:

H4: Respondents who are Democratic are more likely to adopt a nonwhite racial classification, while respondents who are Republican are more likely to adopt a white racial classification.

Other Factors: National Origin, Identity Prioritization, and Age

There are a few more factors which I expect to be associated with racial self-classification. First, related to ancestry, national origin should be a significant predictor of racial self-classification. I expect that Latines from countries with a history of whiteness over *mestizaje* ideology should be more likely to identify as white. Indeed, studying Latin America, Telles and Flores (2013) find that people from countries with historical ideologies of whiteness – such as Argentina, Chile, Costa Rica, and Uruguay – are more likely to identify as white, controlling for characteristics such as skin tone and education, among others. On the other hand, respondents from countries with a history of *mestizaje* ideologies – such as Mexico and Peru – are less likely to identify as white. With reference to some of the major Latinx groups, I expect that Mexican and Central American respondents will be more likely to identify as indigenous or multiracial, Dominican and Puerto Rican respondents as Black or Afro-Latinx, and Cuban respondents as white.

Psychological factors should also shape classification choices. Past research has found that the category “American” is implicitly imbued with racial meaning – people associate Americanness with whiteness (Devos and Banaji 2005). Hickel et al. (2021) argue Latinos may prioritize their American identity as a way to conditionally belong to the higher-status white category, and they show that Latinos who prioritize their American identity support more restrictive immigration policies. Thus, I expect people who prioritize their American identity should be more likely to adopt a white racial classification. Finally, I expect age to be correlated with racial self-classification. To the extent that younger people are more exposed to and socialized within multiracial societies – both U.S. and Latin American – I expect them to be more likely to adopt classifications beyond Latinx.

Empirical Strategy

Data

To examine the predictors of Latinx racial identities, I draw upon four nationally representative surveys of Latinos: the Latino National Survey (LNS, 2006), the National Survey of Latinos conducted by Pew Research (NSL, 2014), and the Collaborative Multiracial Postelection Survey (CMPS, 2012 and 2016). These surveys are publicly available and draw from probability samples of Latino adults.² The LNS was conducted face-to-face, the NSL by telephone, and the CMPS online, all in both English and Spanish. The surveys contain a host of demographic, political, and social psychological variables which allow me to descriptively explore the correlates of self-classification choices and quantitatively test expectations derived from the literature. Comparing across multiple different surveys conducted over a period of 10 years makes it possible to see which variables are strong predictors among multiple different samples, increasing our confidence that a result is not specific to one unique survey.

Important for my purposes, each of these surveys contains different items measuring racial self-classification beyond Latine which make up my key dependent variables. In 2016, the CMPS asked if respondents identify as Afro-Latino, and in 2012 they asked respondents to identify with a variety of Spanish-language racial labels in separate questions. The 2014 NSL asked respondents if they identify as Afro-Latinx, indigenous, or multiracial in three separate questions. Finally, the 2006 LNS asked a standard Census-style race question, where respondents are asked to choose between white, some other race, and several other categories in which Hispanic or Latino was not an option.³

²The datasets and more details about survey methodology can be accessed at the links below:

Latino National Survey, 2006: <https://www.icpsr.umich.edu/web/RCMD/studies/20862>

National Survey of Latinos, 2014:

<https://www.pewresearch.org/hispanic/dataset/2014-national-survey-of-latinos/>

Collaborative Multiracial Postelection Survey, 2012 and 2016: <https://cmgsurvey.org/>

³The exact question wording and response choices are listed below.

CMPS, 2016: “Do you consider any part of your ancestry or family origin to be Afro-Latino, that is to include both African and Latin American ancestry?” The response options were: Yes, No, Don’t know.

CMPS, 2012: “Would you use any of the below terms to describe yourself racially?” Respondents could mark yes or no to any of the following: Indio, Moreno, Mestizo, Negro, Blanco, Mulatto, Trigueño, and None of these. In total, these make up 8 separate questions.

NSL, 2014: a) “Do you consider yourself to be mixed race, that is belonging to more than one racial group, such as Mestizo, Mulatto or some other mixed race, or not?” Response options were Yes, some mixed race; No, not mixed race; Don’t know.

b) “Do you consider yourself to be Afro-[Hispanic/Latino], Afro-Caribbean, or Afro-[National Origin] or not?” Respondents were shown Afro-Hispanic or Afro-Latino based on their stated preferences earlier in the survey. Respondents were also shown a term corresponding to their national origin group, such as Afro-Cuban. The response options were: Yes, No, Don’t know.

c) “Do you consider yourself to be indigenous or Native American, such as (Purepecha, Mixteco, Zapoteco, Nahua, Maya, Tzotzil/ Taino, Arawak/ Maya, Quiche, Aymara, Quechua, Kwichua, Mapuche, Guarani/ Maya, Nahua, Taino, Quiche, Aymara, Quechua) or some other indigenous or Native American origin, or not?” Respondents were shown certain indigenous groups based on their stated national origin. For example, if respondents were Central or South

To construct my dependent variables, I dichotomize the response options as follows. Questions from the 2016 CMPS and 2014 NSL were all Yes or No questions with “Don’t know” response options. I code these variables – called “Afro-Latino,” “Native,” and “Multiracial” – by assigning a value of 1 to Yes and 0 to No or “Don’t know.” For the 2014 NSL, I code an additional variable (“Nonwhite”) which assigns a value of 1 to respondents who said yes to at least one of the three classification choices (Native, Afro-Latino, or multiracial) and 0 to respondents who selected “No” or “Don’t know” for all three.⁴ In the 2014 NSL, very few respondents selected the “Don’t know” option for any of the three categories I examine – between 1.8 to 3.7 percent of respondents, depending on the question. For the 2016 CMPS, however, over 24 percent of respondents select “Don’t know” when asked if they consider their ancestry to be Afro-Latino or not. Why the distributions of these DVs are so different across surveys is not immediately clear, but I identify two possible explanations. First, the mode of the survey is different – the NSL being over the phone and the CMPS online – so perhaps the presence of an interviewer made respondents more likely to select Yes or No in the NSL. Second, the exact question wording varies (see footnote 3). The 2016 CMPS asked respondents if they identify *any part of their ancestry or family origin to be Afro-Latino*, while the NSL question is more akin to self-classification. The much higher rate of “Don’t knows” could therefore represent respondents who are unsure about their ancestry or family tree.

How to code such responses for my purposes is not obvious, but I group the “Don’t knows” with people who respond No both for ease of interpretability and because my interests lie in identity selection and self-classification. I would argue that “Don’t knows” are substantively most similar to respondents who select No because they are not making an affirmative, expressive identity choice. Another approach would be to conduct an ordinal analysis where “Don’t know” is an intermediary category between No and Yes, but this overlooks the possibility that “Don’t know” is simply outside of – and not between – the other options. A third approach is to treat it as a nominal categorical variable, which I do not pursue because it complicates interpretability. Either way, comparing results from multiple surveys should decrease the likelihood that the idiosyncrasies of any one survey drive the findings.

American, they were shown “Maya, Nahua, Taino, Quiche, Aymara, Quechua.” The response options were: Yes, indigenous or Native American origin; No, not indigenous or Native American origin; Don’t know.

LNS, 2006: “What is your race? Are you White, Black, American Indian, Asian, Native Hawaiian/Pacific Islander, some other race or more than one?” The response categories are: White, Black or African American, American Indian or Alaska Native, Asian Indian, Native Hawaiian or Pacific Islander, Some other race (specify). The question was open-ended, and interviewers were instructed to probe for one of the listed categories if respondents said something not listed.

⁴I note that not answering affirmatively to any of these three questions does not imply respondents choose a white classification, but I label this variable as such for ease of discussion.

Turning to the 2012 CMPS, I construct two dichotomous variables given the large number of categories. First, I code a variable called “Labels” where a value of 1 indicates that a respondent chose any one of the eight Spanish-language racial categories and a value of 0 indicates that a respondent selected none of these labels. Second, I code a variable called “White,” where I subset to respondents who did select one of the eight categories and assign a value of 1 to those who selected only white and 0 to those who selected any other combination of categories. This procedure allows me to address two important questions: first, what predicts the adoption of any racial category beyond Latinx, and second, what predicts adopting a white only label versus a nonwhite label or combination of them.⁵

Finally, turning to the 2006 LNS, I subset to respondents who select “White” or “Some other race,” dropping those who are coded as Black or African American, American Indian or Alaska Native, Asian Indian, or Native Hawaiian or Pacific Islander, and those who select more than one race (most commonly “White” and “Some other race”). This removes 849 respondents, dropping the sample size by around nine percent, from 8,634 to 7,785. I then create a “White” variable by assigning a value of 1 to those who select White and 0 to those who select Some other race. While this last survey does not contain distinct racial categories beyond those normally asked in public opinion surveys, it contains a very large sample size that enables me to describe the individual-level characteristics of White versus Some other race classifiers and examine similarities or differences with the other labeling choices analyzed.

Turning to independent variables, I operationalize incorporation through immigrant generation and interview language. Immigrant generation is coded as a numerical variable ranging from 0 to 1, with 1 representing the furthest generations from the immigration experience. Interview language is dichotomous, either English or Spanish. Some surveys also ask for self-reported English and Spanish language ability, which I examine in the appendix. I code income as a numerical variable ranging from 0-1 and education as a categorical factor with values ranging from “less than high school” to “advanced degree.” I measure phenotype through self-reported skin tone (asked in all but the 2014 NSL dataset) and reflected race (only in the 2016 CMPS) – or how respondents think others perceive their race (Roth 2016). I measure ancestry for the 2014 NSL, which asks respondents if they have Afro-Latino, indigenous, or multiracial parents or grandparents. Finally, party ID is coded numerically from 0 to 1, age from 1 to 100, and national origin is a categorical variable.

Notably, I exclude “multiracial” respondents – respondents who report having parentage that

⁵Respondents who selected white and something else were assigned a 0.

Survey	Year	Number of Predictors	Original N	Final N	DVs
Latino National Survey (NSL)	2006	115	8634	7785	White (vs. some other race)
Collaborative Multiracial Post-election Survey (CMPS)	2012	134	934	934	Choosing racial labels (vs. none) White (vs. nowwhite labels)
CMPS	2016	240	3002	2761	Afro-Latino
National Survey of Latinos (NSL)	2014	54	1520	1298	Afro-Latino Indigenous Multiracial / mestizo / mulato Nonwhite

Table 1: This table displays information on the surveys employed and corresponding DVs.

is both Latinx and non-Latinx or who identify with a non-Latinx racial category – from all the analyses. For example, respondents who say that one of their parents is not Latinx or who identify as both African American and Latinx are excluded. I do this to be sure that my results are not driven by these multiracial respondents. For example, an African American-Latinx respondent may report more closeness to Black people because of their African American background. This removes a small percent of respondents – no more than eight percent.⁶ Table 1 summarizes the final sample sizes of the different surveys and corresponding dependent variables.

Table 1 also includes the number of predictors I examine for each survey. Each survey asks a number of questions only to subsets of respondents, so I exclude variables which are not asked of all respondents for which there is no easy fix. For example, in every survey only respondents who report being registered to vote were asked if they voted in the last election. In this case, there is an easy fix: impute “not voted” for those who report not being registered to vote. In other cases, however, it is less clear. For example, only respondents who report being employed are asked about their job satisfaction. I exclude such variables in order to avoid imputing large amounts of data and in order to avoid listwise deletion, which removes all observations for which there is not complete information. There is still some missingness for the remaining questions asked of all respondents, particularly for sensitive topics like income, so I impute the remaining missingness using random forest.

⁶I also exclude 16 respondents from the 2016 CMPS for quality purposes. While coding a variable for percent of life lived in the U.S., I came across respondents who said that they first moved to the U.S. in a year before they were born. As this is impossible, I exclude these 16 respondents as potentially inattentive.

Methods

I estimate separate random forest regression models for each dependent variable. Random forests is a machine-learning ensemble method originally proposed by Breiman (2001). It is a nonparametric method (i.e., does not impose a functional form on the data) that builds a “random forest” of trees broadly based on the Classification and Regression Trees (CART) method (Montgomery and Olivella 2018). CART works by searching for splits in independent variables that minimize prediction error in the dependent variable. It builds a “tree” by determining the optimal split on one variable and proceeding downward – splitting on other variables – until some stopping criterion is reached. A random forest aggregates from these individual trees to a forest in two ways. First, it relies on bootstrapping (known as “bagging,” short for bootstrap aggregation). This involves sampling with replacement from the original dataset, which ensures independence between the different trees. Second, the model randomly selects a constant number of independent variables for each tree, which further ensures independence and enables a more accurate description of the underlying data generating process.

There are several benefits of using random forests over traditional regression models or other regularization techniques, particularly for the questions I address. First, since the method does not impose a functional form, it accounts for potential nonlinearities and complex interactions in the data. This is particularly useful when studying the construction of identities, which are multifaceted and potentially related to unknown interactions of other variables. Second, it allows for feature selection, through which we can identify the most important predictors of the dependent variables. This is ideal for the kinds of data analyzed here, where there are up to hundreds of variables per observation which may or may not predict the dependent variable well. Rather than the researcher selecting a number of variables to analyze, the model draws upon all available variables and returns the most important predictors. This is valuable for exploratory analyses where we want to describe the nuance behind identity selection. Finally, random forests avoid overfitting, which can lead a model to make predictions based on irregular variables when there is a large number of predictors (Montgomery and Olivella 2018). These benefits are born out in the data: when examining the predictive performance of a random forest model compared to OLS, the random forest leads to substantially better predictions measured by mean squared error (MSE). It performs marginally better than other regularization techniques like LASSO, but these methods also perform well compared to OLS.

To conduct feature selection, I rely upon permutation variable importance measures (VIMs, Wei, Lu, and Song 2015). The idea behind a permutation VIM is the following. A random forest model is run, and one-by-one, the values of each predictor are permuted, or randomly shuffled, and a model is refit on the data. Then, the decrease in MSE of the model is calculated when the values of the given variable are permuted versus when they are at their true values. Important predictors will then decrease the MSE of the model substantially when their true values are used, and less important predictors will not decrease the MSE much or even increase it if they are poor predictors. Because this procedure is subject to some variation due to the randomized shuffling of values, I re-estimate each variable importance measure 100 times and average over the runs to guard against the idiosyncrasies of any particular permutation.

Results

Turning to the findings, I proceed by first examining the distributions of the dependent variables and their relation to ancestry for the 2014 NSL. Then, I run several random forest models and use the permutation VIM discussed above to explore which features emerge as the top 5 predictors of self-classification choices. Finally, I produce partial effects plots derived from these models for the key independent variables, which display the marginal associations with identity choices net of all other influences.

While previous qualitative work has argued that Latinos primarily identify their race in panethnic or national origin terms, substantial numbers of Latines choose to identify with a label beyond Hispanic or Latinx. Figure 1 displays the frequency distributions of the dependent variables across the four surveys. Most Latinos choose not to identify with terms like white, multiracial or mixed race, Afro-Latino, indigenous, or various other terms. However, in the 2014 NSL, most Latinxs identified either as indigenous, multiracial, or Afro-Latinx, and less than half (around 45 percent) rejected all of these classifications. In the 2012 CMPS, over 40 percent of respondents identified as Blanco or with another Spanish-language term such as Indio or mestizo. Thus, many Latinos do identify beyond panethnic labels when asked.

Before turning to the predictors of these various classification choices, I examine the extent to which ancestry constrains identification. The 2014 NSL is the only survey I analyze which asked respondents if they have parents or grandparents who are Afro-Latino, multiracial or mixed race, or Indigenous. Figure 2 displays the percent of respondents who say they have family or grandparents

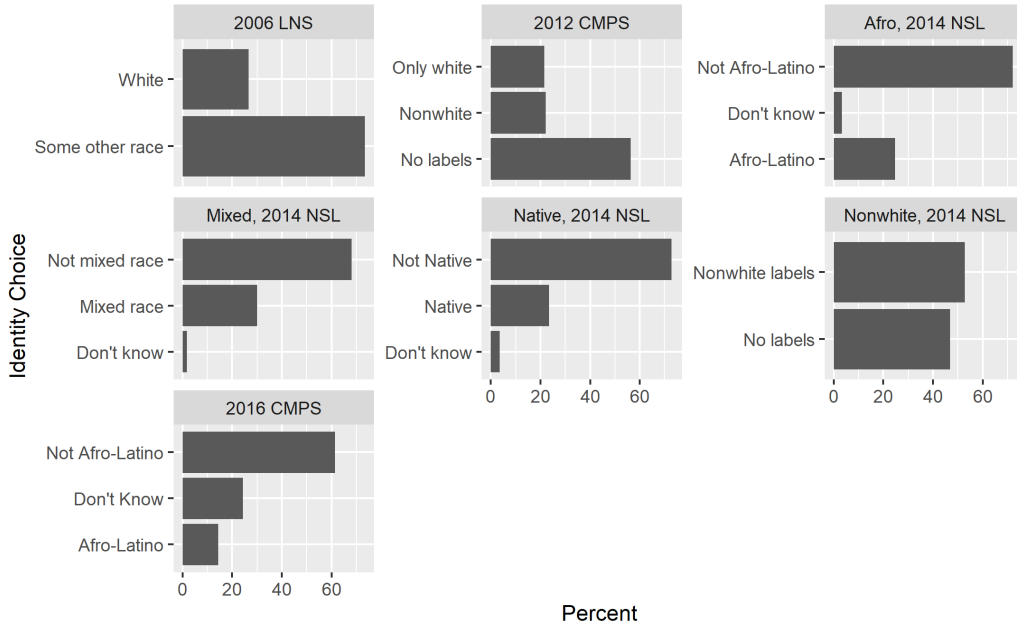


Figure 1: Distributions of DVs. This figure displays the distribution of each dependent variable. Each plot is a distribution for a different DV. The y-axis plots the values each DV takes and the x-axis plots the percent of respondents selecting that label.

who belong to a certain racial category split by whether they themselves identify with that category. We can see that very few respondents who do not identify with one of these labels say they have family belonging to one of these groups. On the other hand, while the majority of respondents who identify as Afro-Latinx, multiracial, or indigenous say they also have family sharing this identity, almost one third of respondents who do so say they do *not* have family that also identify like them. This suggests that many Latinos identify in ways not constrained by their ancestry, and these identification choices can be shaped by a number of other factors.

I turn next to examining these factors with the variable importance measures (VIMs) derived from random forest models for each dependent variable, presented in Figure 3. For brevity, I examine the top five predictors for each dependent variable, but I present the top ten in the appendix. I also exclude ancestry from the 2014 dependent variable graphs in order to compare better across years because it is highly predictive of the outcomes. However, the appendix graphs include ancestry, and we can see that ancestry explains substantially more variation in the outcomes than any of the other predictors. The x-axis of these plots measure the decrease in MSE resulting from the permutation VIM discussed in the methods section, similar to an R^2 measure in OLS. For example, in the bottom right panel – where Afro-Latinx identity in the 2016 CMPS is the dependent variable – national origin decreases the predictive error of the model by around 1 percent of the full scale, accounting

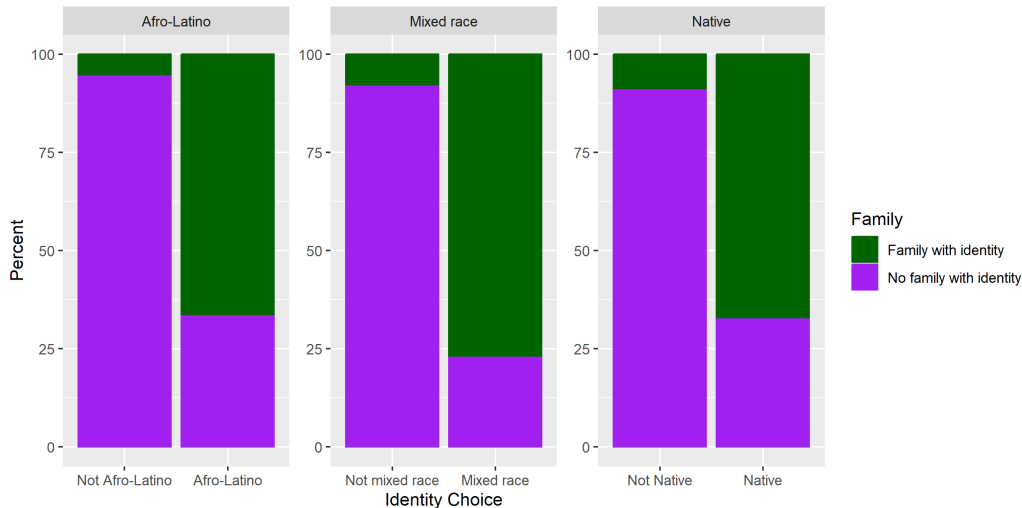


Figure 2: The role of ancestry in classification choices. This figure displays the percent of respondents who say they have family belonging to a given category by whether they identified with that category or not. Each plot represents one labeling choice from the 2014 NSL. The x-axis plots whether respondents identified with that label or not, and the y-axis displays the percent of respondents who say they have parents or grandparents who belong to that racial category.

for all other variables.

Across the board, incorporation, phenotype, and national origin emerge as top variables explaining racial self-classification choices. Markers of incorporation – including language of interview, generation, and language ability – emerge as important predictors of identification as white over some other race (White, 2006 LNS), of selecting a Spanish-language racial label (Labels, 2012 CMPS), and of identifying as Afro-Latino. For the first of these – choosing white over some other race – language of interview (“In Spanish”) is a more important predictor than skin tone. Notably, however, incorporation is absent from the top ten predictors of indigenous identification in the 2014 NSL (Generation is the 6th most important predictor of Afro-Latino identity in the 2016 CMPS, and English ability is the 9th most important predictor of multiracial identity, which are in the appendix graph).

Skin tone, however, is also an important predictor, particularly for the models where the dependent variable is identification as white. This is instructive for the 2012 dataset. When we look at the predictors of the adoption of any racial labels beyond Latinx including white (Labels, 2012 CMPS), skin tone is not an important predictor, but markers of incorporation are. Looking at the predictors of choosing a white label over others (White, 2012 CMPS), however, skin tone emerges as the top predictor and is magnitudes more predictive than the next predictor, national origin. This suggests that incorporation shapes the racial understandings people adopt, but skin tone specifically

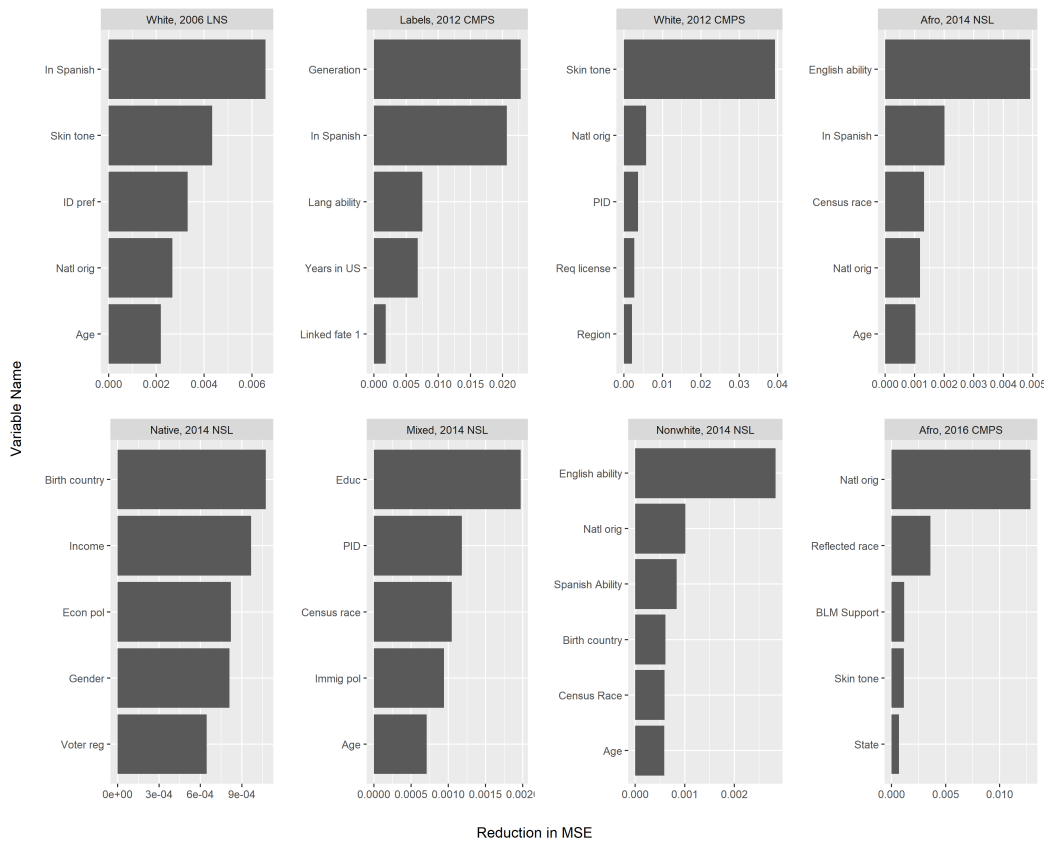


Figure 3: Top 5 predictors by variable importance for each DV. This figure displays the top 5 predictors of each dependent variable using permutation variable importance. The y-axis displays variable names, and the x-axis displays the reduction in mean-squared error (MSE) when the variable is randomly shuffled versus when it is held at its true values. The scale of MSE runs from 0 to 1, so a reduction of 0.01 represents a reduction in MSE of 1 percent of the full scale. More concretely, if the classification error of the model is 15 percent, then a variable with a value of 0.01 reduces the error of the model from 16 to 15 percent. Ancestry variables are excluded from the 2014 NSL DVs but are present in the appendix figure.

shapes the adoption of a nonwhite classification.

Though income and education appear to be important predictors of multiracial and native identification, SES does not appear to explain much of the variance for other racial labels, lending limited support to H2a and H2b. There is less support for H4 on party ID, though it emerges as a primary predictor of multiracial identity in the 2014 NSL and white identification in the 2012 CMPS. National origin is well supported as a valuable predictor – it is the top predictor of Afro-Latinx identity in the 2016 CMPS. Age emerges as a less central but important predictor of multiracial, Afro-Latine, and white identification. Finally, there is limited support for identity prioritization as a primary predictor, though it does explain variance in white identification in the 2006 LNS.

Figure 3 reveals that many of the factors identified in past literature indeed emerge as central predictors, even taking into account the hundreds of other potential variables in these datasets. There are also some unexpected predictors that appear in Figure 3. Region and state of residence appear to reduce MSE substantially for white classification in the 2012 CMPS and Afro-Latinx identity in the 2016 CMPS. Examining partial dependence plots, we can see that respondents from New York and Florida are more likely to identify as Afro-Latinx, and respondents from California and Texas are less likely to. Respondents from the East Coast are more likely to identify as white, while respondents from the Pacific and Mountain states are least likely to do so.

Though there is not strong support for party ID as influential in the construction of Latinx identities, other political variables do emerge in this analysis. Respondents who believe the Democratic Party does a better job at handling immigration policy are more likely to identify as multiracial in the 2014 NSL, and respondents who believe the Democratic Party does a better job at handling economic policy are more likely to identify as native. Respondents who believe that people should be required to have a license to vote are more likely to identify as white in the 2012 CMPS. Afro-Latino identifiers in the 2016 CMPS are more likely to support Black Lives Matter. “Census” race in the 2014 NSL – how respondents identify in a Census style question – is predictive of identification choices, suggesting that Census classification questions without a Hispanic or Latino option do capture racial classification choices beyond Latinx to a certain extent. Finally, gender is predictive of native identification in the 2014 NSL, with men being more likely than women to identify as such.

While the previous graphs indicate which variables are explaining the most variation in the outcome, they do not tell us the direction or magnitude of the association between a predictor and the dependent variable. To examine my hypotheses more closely, I turn to partial dependence plots, which display the conditional probability of choosing a given classification, net of all the other variables in the model. In figures 4, 5, 6, and 7, I display partial dependence plots for immigrant generation, income, party ID, and skin tone, respectively. The remainder I discuss but are displayed in the appendix.

Starting with immigrant generation, hypothesis 1b is well-supported and 1a somewhat supported. Those of a later immigrant generation are less likely to identify with nonwhite labels and more likely to identify with white labels, suggesting that some incorporated Latines may assimilate into whiteness. The magnitudes of the difference are small, except for choosing one of the racial self-classifications in 2012. In this dataset, the likelihood of choosing a racial label beyond Latinx –

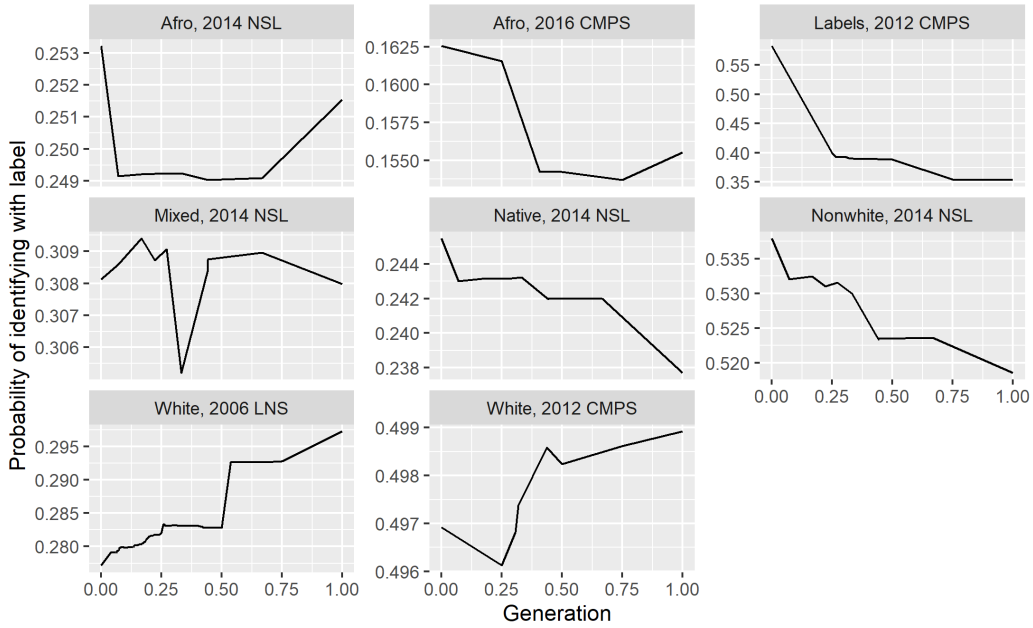


Figure 4: Partial dependence plot of immigrant generation. This figure plots values of immigrant generation on the x-axis and corresponding predicted probabilities of identifying with each DV – derived from random forest regressions – on the y-axis. Higher values of immigrant generation indicate being further from the immigration experience.

net of all other variables – decreases over 20 percentage points going from the lowest to the highest levels of immigrant generation, which demonstrates support for hypothesis 1b.

Turning to other markers of incorporation – language of interview and self-reported language ability, results for which are presented in the appendix – we see similar support for hypothesis 1b. Looking at language of interview, we can see that respondents who completed the 2006 LNS in Spanish (bottom left panel) are around 7 percentage points more likely to identify as white. Likewise, respondents who completed the 2012 CMPS in Spanish are 20 percentage points more likely to identify with racial labels beyond Latinx. Respondents who rated *both* their English and Spanish language ability as higher are less likely to identify with almost all racial labels. This latter finding is somewhat puzzling, but since the models already control for language of interview these variables could be capturing some aspect of socioeconomic status or levels of education. Overall, there is support for both 1a and 1b. Less incorporated respondents are on the whole more likely to identify with these racial labels, though more incorporated respondents are sometimes more likely to identify as white.

Turning to the second hypothesis, there is limited support for H2a and H2b. In figure 5, we can see that income is negatively associated with racial self-classification for some nonwhite

labels, particularly for native identification. However, it actually has a positive association with Afro-Latino identity in 2016 and multiracial identity in 2014. As with generation, the magnitudes are small, being largest for native identification, where going from the lowest to highest income is associated with a 4-percentage point drop in the probability of selecting a native label. Earlier, we saw that education was only an important predictor for multiracial identification in the 2014 NSL. The appendix figure demonstrates that H2b does not hold. There is an inconsistent relationship between education levels and the predicted probability of adopting a racial classification. For some DVs, like Afro-Latinx identity in the 2014 NSL, there is a negative association, with more formally educated respondents being slightly less likely to adopt this label. For multiracial identity, those with college degrees or higher are more likely to adopt this label choice.

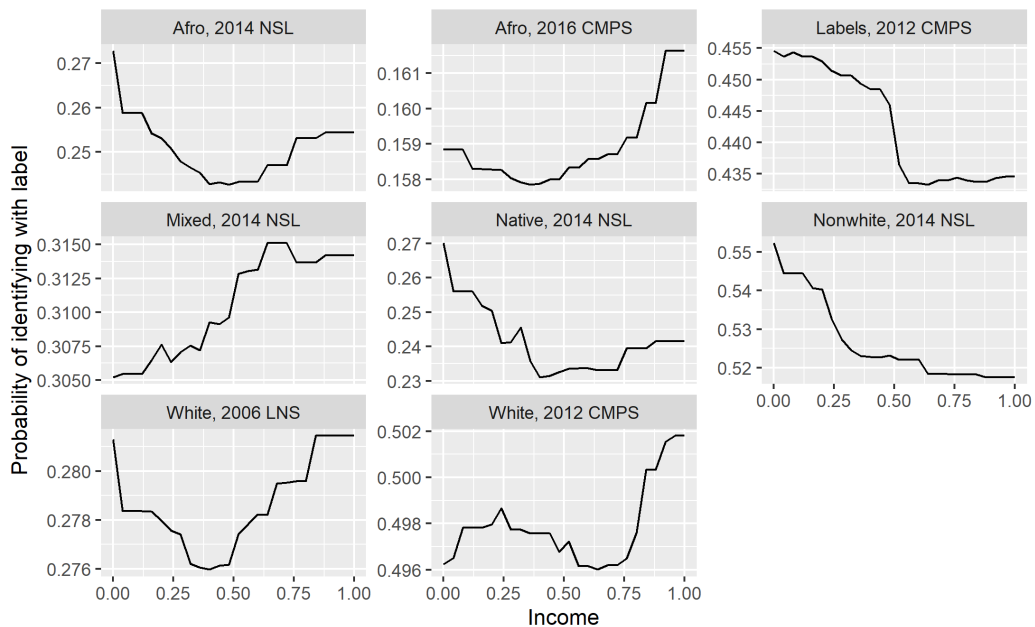


Figure 5: Partial dependence plot of income. This figure plots values of income on the x-axis and corresponding predicted probabilities of identifying with each DV – derived from random forest regressions – on the y-axis. Higher values represent higher self-reported family incomes.

There are similarly unexpected results for party ID (Figure 6). Both white and nonwhite labels are positively associated with greater Republican party identification. These findings are not simply the result of accounting for many different variables. While Afro-Latinos in the 2016 CMPS are significantly more likely to identify as Democrats, in the 2014 NSL Afro-Latino identifiers are a few percentage points more likely to be *Republican*, even before controlling for many other factors. These differences are quite unexpected, and I return to them in the discussion.

Turning to skin tone, we find strong support for H3: darker skin tones are associated with

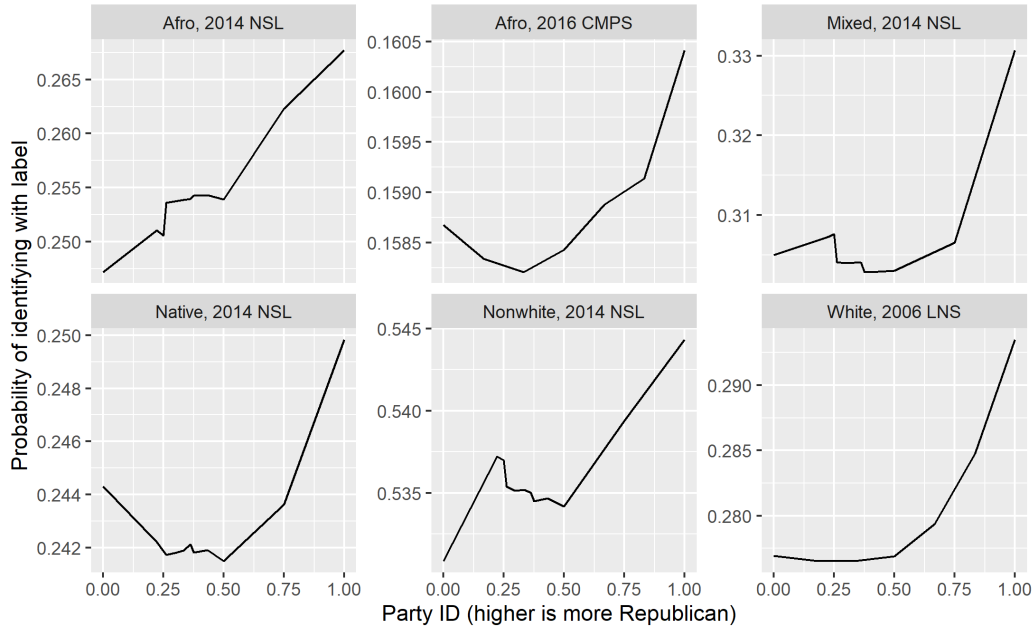


Figure 6: Partial dependence plot of party ID. This figure plots values of party ID on the x-axis and corresponding predicted probabilities of identifying with each DV – derived from random forest regressions – on the y-axis. Higher values represent more Republican party ID. The 2012 CMPS is excluded because they only asked a 3-point version of party ID.

greater likelihood of adopting nonwhite labels and lower likelihood of adopting white labels. For 2016, those with the lightest skin tones are around 5 percentage points less likely to identify as Afro-Latine than those with the darkest skin tones, net of all other influences. The maximum difference is around 10 percentage points in the 2006 LNS and 30 percentage points in the 2012 CMPS. However, there are clear nonlinearities for both of these. Respondents with self-reported skin tones above the midpoint do not have a substantially different probability of identifying as white. This is possibly because very few respondents select the darkest skin tones in each survey, but it indicates that there are not very large self-classification differences between those selecting medium-dark and dark skin tones. The 2016 CMPS also asks reflected race – what race respondents think others think they are, another rough proxy for phenotype. Results for this variable are presented in the appendix, and we can see that respondents who think they are perceived by others as Black are around 15 percentage points more likely to identify as Afro-Latino.

Finally, I turn to the additional factors: age, national origin, and identity prioritization, for which results can be found in the appendix. National origin has a strong relationship with outcome variables. I display only results for major U.S. Latinx groups from the 2016 CMPS and the 2006 LNS because their sample sizes are sufficiently large. We can see that past literature holds true: those

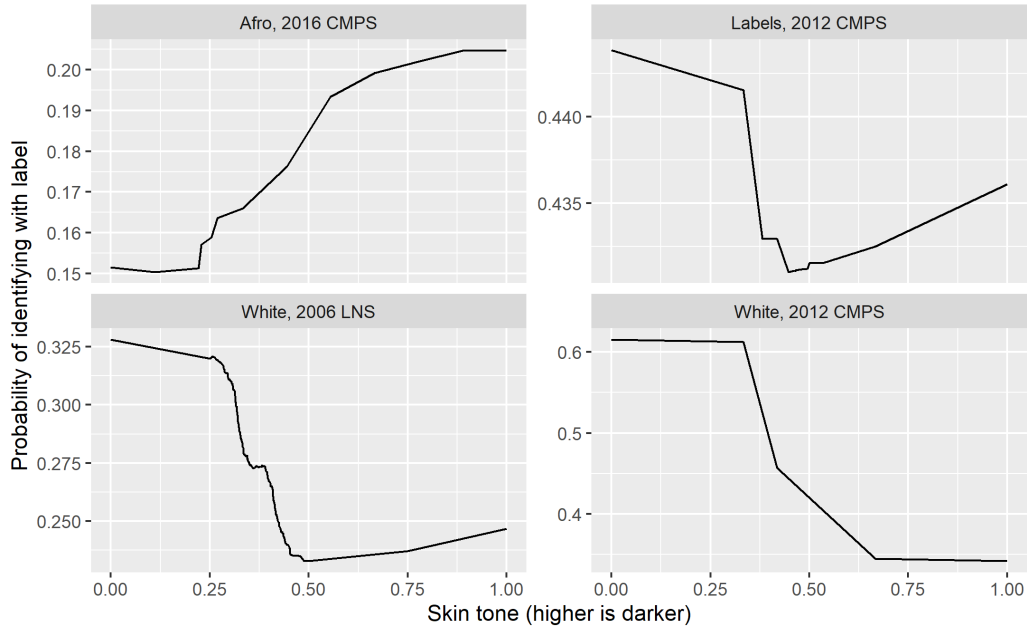


Figure 7: Partial dependence plot of skin tone. This figure plots values of self-reported skin tone on the x-axis and corresponding predicted probabilities of identifying with each DV – derived from random forest regressions – on the y-axis. Higher values represent darker self-reported skin tones. The 2014 NSL DVs are missing because skin tone was not asked on this survey.

from countries with a history of *mestizaje* – such as Mexico and El Salvador – as opposed to whiteness ideology are less likely to identify as white. Cuban Americans are more likely to identify as white, and Dominican and Puerto Rican respondents are most likely to identify as Afro-Latinx. Looking at age, younger respondents are more likely to identify with nonwhite labels, net of all other influences. The youngest respondents are about 5 percentage points more likely than the oldest respondents to identify as Afro-Latinx. For white identification, the opposite is true. Younger respondents are less likely to identify as white. In the 2006 LNS, the youngest respondents are around 10 percentage points less likely to identify as white (over some other race). There is also some support for the role of identity prioritization – those who prioritize their American identity are slightly more likely to identify with white labels, and those who prioritize their national origin identity are more likely to identify with nonwhite labels.

Comparing the results from the partial dependence analysis to variable importance, we come to slightly different conclusions between *prediction* goals and *marginal relationship* goals. The random forest models demonstrate that incorporation is a key predictor of identification choices – indeed, the top predictor for white identification in the 2006 LNS – though the exact differences are substantively not large. This tells us that it is an important variable for explaining variation in identity

choices, though the magnitude of its association is small. Skin tone, on the other hand, has a much stronger relationship with the dependent variables while also being an importance predictor. Overall, compared to ancestry, the demographic, political, and psychology predictors I examine explain a small portion of variation. People’s identities and identity choices are complicated – further work is needed to understand the complex interaction of factors that go into understanding identity.

Discussion

Overall, my results show strong support for the role of ancestry and skin tone constraining ethnoracial identification, in line with prior work (Irizarry, Monk, and Cobb 2022), I find that the adoption of ethnoracial labels beyond Latinx – such as white, Afro-Latino, indigenous, or multiracial – is closely related to self-reported skin tone. Using the 2014 NSL data, I also show that ancestry is closely related to the choice of identity labels, though respondents show substantial latitude in choosing an identity label beyond that of their family.

Incorporation emerges as a primary predictor but not in the exact way that I had hypothesized. While less incorporated respondents are more likely to identify with nonwhite labels, they are sometimes more likely to identify as white. I take these findings to suggest that less incorporated Latinos do identify with a variety of nonwhite racial labels common in Latin America and adopt racial understandings recognizing this diversity, in line with H1b. However, the results also imply that some Latinos assimilate into whiteness net of all other differences, in line with H1a. As the findings on skin tone and ancestry demonstrate, some Latinos are more constrained than others in doing so, but this result suggests that more Latinos will “become” white over generations. In addition to indicating the importance of incorporation for racial ideologies and understandings, the results point to the intricate connection between racial classification and the process of incorporation.

More broadly, these findings contribute to theories on the processes of assimilation and race in the U.S. Older theories of assimilation primarily based upon the experiences of European Americans viewed assimilation as a linear process in which later generations shed their cultural ties and became part of the white majority (Alba and Nee 2003). Later theorists, however, posited that immigrants from Asian and Latin American countries did not straightforwardly fit this account, with their racialization preventing them from joining the white majority and with many retaining cultural ties over generations (Portes and Zhou 1993; Schacter 2016). Yet other scholars, such as Bonilla-Silva (2002, 2004) and Frank, Redstone Akresh, and Lu (2010) posit that some light-skinned Latinos

may push to expand the boundaries of whiteness to include them. My findings here align with this account, with some more incorporated Latinos being more likely to view themselves as white. Other Latinos – particularly those with darker skin tones and nonwhite ancestry – have less latitude in identifying as white and instead “become” Latinx (Golash-Boza 2006, Roth 2012) or choose a classification within “honorary white” and “collective Black” categories.

Nonetheless, incorporation and assimilation are not one-way processes, and other white people may not view Latinos who adopt a white racial-self classification as part of their group (Schacter 2016; Abascal 2020). Since the data on skin tone here is self-reported, I cannot assess to what extent external classification shapes racial self-classification, but past work suggests they align to a certain extent (Irizarry, Monk, and Cobb 2022). Skin tone self-reports may also be capturing a politicized identity rather than accurately capturing phenotype (Yadon 2020; Ostfeld and Yadon 2021), but I argue that they are nonetheless important markers of experiences with racialization. Future work might examine to what extent Latinos who select nonwhite labels such as Afro-Latinx or indigenous are accepted as members of these categories by other Black and Indigenous people.

I note that incorporation was an important predictor of all identity choices except native or indigenous classification; future work might explore the predictors of Indigenous Latinidad in particular. Gender emerged as an important predictor of indigenous self-classification but not any of the other categories, with men being markedly more likely to identify as indigenous than women. Prior work has found that gender shapes racial classification because men are more likely to be racialized as nonwhite and experience racial discrimination distinctly (Ho et al. 2011; Penner and Saperstein 2013; Davenport 2016). It is possible that some Latinx men are more likely to be classified by others as indigenous and their self-classification choice reflects this experience. However, further work is necessary to uncover why this key social variable did not emerge for other racial labels, and much work is needed to center indigeneity among the study of Latines (Alberto 2017).

My findings showed limited support for the second hypothesis on SES – neither income nor education appear to be important drivers of racial classification choices except for multiracial and indigenous identification. Results on income are in the expected direction except for Afro-Latine identity, for which there is a positive relationship. These are surprising results, given the association of racial discrimination with socioeconomic outcomes among Latines (Telles and Murguia 1990; Massey and Denton 1992; Chavez-Dueñas, Adames, and Organista 2014). However, they fit within the arguments advanced by Telles (2014) and Telles, Flores, and Urrea-Giraldo (2015) that income may be less consistently related to racial classification choices than it is to measures of phenotype

or skin tone. Prior work shows that racial classification choices often do not align with how others classify one's race, particularly for Latinas (Roth 2016), so future work may examine misalignments between these two dimensions of race for the categories studied here.

I observe similarly surprising results for party ID, where greater Republican identification is positively associated with all identity choices – even Afro-Latinx identity. The results are not simply due to the modeling strategy either, as Afro-Latino identifiers in the 2014 NSL are on average more Republican even before controlling for other variables. These are puzzling findings, particularly in light of the strong association between race and party identification (Mason 2018; Westwood and Peterson 2021). They suggest that Republican Latinas may take on distinct racial understandings than Democrats – perhaps Democratic Latinas are more likely to see themselves as one distinct, racialized group. How Latinas with different partisan attachments perceive racial inequality among Latinas and between Latinas and other racial groups – and how this relates to racial identity – is an interesting area for further inquiry.

Besides the key factors I examine, national origin, age, and region surface as valuable predictors. National origin differences are in the expected directions based on Latin American countries' histories of *mestizaje* (Telles and Flores 2013). They suggest that people may make accurate inferences when they substitute national origin for race among Latinas (Roth 2012), though substantial differences remain even after controlling for national origin. Age, on the other hand, is positively associated with white identification and negatively associated with nonwhite identification. This is an interesting factor to consider in light of the findings on incorporation. Do Latinas shift their racial classification as they age, are younger Latinas understanding race differently, or both? Lastly, geography appears to play a role in racial self-classification. For example, independent of national origin, Latinas from the West – which is heavily Mexican American – are less likely to identify as white. This suggests that notions of race are not constant across the U.S. and shaped by local or regional context.

Conclusion and Future Directions

While Latinas are widely recognized as one of the fastest growing ethnoracial groups in the U.S., significant racial differences remain within the broader group that have received insufficient scholarly attention. In this paper, I have endeavored to add to our knowledge of intra-group difference among Latinas by examining one aspect of race – self-classification. Using four nationally representative surveys, I have shown that substantial numbers of Latinas classify their race as something other than

Latinx when asked – potentially as large a number as a majority. These classification choices are closely related to other meaningful social factors such as incorporation, phenotype, national origin, and age. These factors emerge as important predictors among the hundreds of variables available through VIMs derived from random forest regression models. Crucially, my findings indicate that less incorporated Latines bring with them a Latin American racial classification system with many distinct categories and labels. As they become incorporated, some Latinos strive to expand the boundary of whiteness, but others may not have this route available to them and instead retain an identification as Latine.

Critical to my argument here is the role of ideologies of *mestizaje*; however, the surveys employed here do not ask questions that capture this construct well. The 2006 LNS does ask respondents if they believe Latinos make up a distinct racial group in the U.S., and this variable is the 7th most important predictor of white identification, with respondents who answer affirmatively to this statement being less likely to identify as white. However, future work would do well in probing *mestizaje* beliefs more directly, or designing survey items which capture this idea. One possibility is to ask respondents whether they believe all Latin Americans are multiracial or mixed race and explore how this relates to other key demographic and political variables.

The analysis here focuses on racial self-classification on a survey, which is itself a consequential choice and signal of identity. However, questions remain about the scope and strength of racial identities beyond panethnicity for Latinxs. On top of classification, do some Latinos identify with the racial labels studied here in open-ended terms, and do similar factors predict this open-ended identification? There is substantial room for further qualitative and quantitative survey work in this vein. In-depth interviews could shed light on what it means to classify oneself as Indigenous, Afro-Latino, white, or multiracial. It is promising to develop better survey items to capture the centrality or strength of these identities and discern whether some Latines hold a sense of racial group consciousness (Miller et al. 1981; Tajfel and Turner 1986; Ellemers, Kortekaas, and Ouwerkerk 1999). Do some Latines prioritize these racial identities over their panethnic, national origin, or even non-racial identities such as gender or religion, and why?

The findings here speak to debates about the proper measurement of racial identity or classifications of Latines on surveys. This paper is distinct from many others on racial identity in analyzing dependent variables that do not force respondents to choose between different options. However, another promising area of investigation is the development of survey items that better assess Latinxs' classification choices beyond Census-style items. Many Latinos feel poorly represented by Census

race questions (Roth 2016). It is possible that many more Latinos would choose categories like Black or American Indian if they were worded differently, like the questions on the surveys analyzed here. If the answer choices resemble more common ways of expressing Indigenous or Black Latinidad – e.g., not limiting an Indigenous category to those who consider themselves “American Indian” in particular – then the number of Latinos selecting “some other race” on these kinds of questions may be lowered. At the same time, many Latinos view their race through the lens of their panethnicity (Hitlin, Brown, and Elder 2007; Porter and Snipp 2018), so there is no easy solution to the challenges faced by the Census. As Latinos make up an increasing share of the U.S. population, race and racial structures will continue to shift and defy classification.

References

- Abascal, Maria. "Contraction as a Response to Group Threat: Demographic Decline and Whites' Classification of People Who Are Ambiguously White". *American Sociological Review* 85, no. 2 (Apr. 2020): 298–322. ISSN: 0003-1224, 1939-8271, visited on 03/06/2023. <https://doi.org/10.1177/0003122420905127>. <http://journals.sagepub.com/doi/10.1177/0003122420905127>.
- Abrajano, Marisa. *Campaigning to the new American electorate: advertising to Latino voters*. Stanford, Calif: Stanford University Press, 2010. ISBN: 978-0-8047-6895-5 978-0-8047-6896-2.
- Alba, Richard D., and Tariqul Islam. "The Case of the Disappearing Mexican Americans: An Ethnic-Identity Mystery". *Population Research Policy Review* 28 (2009): 109–121.
- Alba, Richard D., and Victor Nee. *Remaking the American Mainstream*. Cambridge: Harvard University Press, 2003.
- Alberto, Lourdes. "Coming Out as Indian: On being an Indigenous Latina in the US". *Latino Studies* 15 (2017): 247–253.
- Banaji, Mahzarin R., and Thierry Devos. "American = White?" *Journal of Personality and Social Psychology* 88 (3): 447–466.
- Barreto, Matt A. *Ethnic cues: the role of shared ethnicity in Latino political participation*. OCLC: 8318410721 [OCLC: 831841072. Ann Arbor: University of Michigan Press, 2012. ISBN: 978-0-472-03495-6.
- Beltrán, Cristina. *The trouble with unity: Latino politics and the creation of identity*. Oxford ; New York: Oxford University Press, 2010. ISBN: 978-0-19-537590-9 978-0-19-537591-6.
- Berelson, Bernard, Paul F. Lazarsfeld, and William N. McPhee. *Voting: a study of opinion formation in a presidential campaign*. Midway reprint ed. Midway reprint. Chicago: University of Chicago Press, 1954. ISBN: 978-0-226-04350-0.
- Bonilla-Silva, Eduardo. "From bi-racial to tri-racial: Towards a new system of racial stratification in the USA". *Ethnic and Racial Studies* 27, no. 6 (Nov. 2004): 931–950. ISSN: 0141-9870, 1466-4356, visited on 05/07/2021. <https://doi.org/10.1080/0141987042000268530>. <http://www.tandfonline.com/doi/abs/10.1080/0141987042000268530>.
- . "We are all Americans!: the Latin Americanization of racial stratification in the USA". *Race and Society* 5, no. 1 (Jan. 2002): 3–16. ISSN: 10909524, visited on 05/04/2021. <https://doi.org/10.1016/j.racsoc.2003.12.008>. <https://linkinghub.elsevier.com/retrieve/pii/S1090952403000275>.

- Breiman, Leo. "Random Forests". *Machine Learning* 45, no. 1 (2001): 5–32.
- Chavez-Dueñas, Nayeli Y., Hector Y. Adames, and Kurt C. Organista. "Skin-Color Prejudice and Within-Group Racial Discrimination: Historical and Current Impact on Latino/a Populations". *Hispanic Journal of Behavioral Sciences* 36, no. 1 (Feb. 2014): 3–26. ISSN: 0739-9863, 1552-6364, visited on 05/07/2021. <https://doi.org/10.1177/0739986313511306>. <http://journals.sagepub.com/doi/10.1177/0739986313511306>.
- Cobas, José A, Jorge Duany, and Joe R Feagin. *How the United States racializes Latinos: white hegemony and its consequences*. OCLC: 931534660 [OCLC: 931534660. 2016. ISBN: 978-1-317-25802-5 978-1-315-63410-4 978-1-317-25803-2 978-1-317-25801-8, visited on 05/05/2021. <https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1106206>.
- Comas-Díaz, Lilian, M. Brinton Lykes, and Renato D. Alarcón. "Ethnic Conflict and the Psychology of Liberation in Guatemala, Peru, and Puerto Rico". *American Psychologist* 53, no. 7 (July 1998): 778–792.
- Cornell, Stephen, and Douglas Hartmann. *Ethnicity and race: Making identities in a changing world*. Sage Publications, 2006.
- Cutaia Wilkinson, Betina, and Emily Earle. "Taking a New Perspective to Latino Racial Attitudes: Examining the Impact of Skin Tone on Latino Perceptions of Commonality with Whites and Blacks". *American Politics Research* 41, no. 5 (Sept. 2013): 783–818. ISSN: 1532-673X, 1552-3373, visited on 05/07/2021. <https://doi.org/10.1177/1532673X12464546>. <http://journals.sagepub.com/doi/10.1177/1532673X12464546>.
- Davenport, Lauren. "The Fluidity of Racial Classifications". *Annual Review of Political Science* 23, no. 1 (May 2020): 221–240. ISSN: 1094-2939, 1545-1577, visited on 05/07/2021. <https://doi.org/10.1146/annurev-polisci-060418-042801>. <https://www.annualreviews.org/doi/10.1146/annurev-polisci-060418-042801>.
- Davenport, Lauren, Hakeem Jefferson, and Hunter Rendleman. "Deconstructing Race". July 2022.
- Davenport, Lauren D. *Politics beyond Black and White: Biracial Identity and Attitudes in America*. 1st ed. Cambridge University Press, Mar. 2018. ISBN: 978-1-108-69460-5 978-1-108-42598-8 978-1-108-44433-0, visited on 01/15/2023. <https://doi.org/10.1017/9781108694605>. <https://www.cambridge.org/core/product/identifier/9781108694605/type/book>.
- . "The Role of Gender, Class, and Religion in Biracial Americans' Racial Labeling Decisions". *American Sociological Review* 81, no. 1 (Feb. 2016): 57–84. ISSN: 0003-1224, 1939-8271, visited on

- 06/10/2022. <https://doi.org/10.1177/0003122415623286>. <http://journals.sagepub.com/doi/10.1177/0003122415623286>.
- Dawson, Michael C. *Behind the mule: race and class in African-American politics*. Princeton, N.J: Princeton University Press, 1994. ISBN: 978-0-691-08770-2.
- De la Garza, Rodolfo O., and Alan S. Yang. *Americanizing Latino politics, latinoizing American politics*. First edition. New York: Routledge, 2020. ISBN: 978-1-351-05464-5 978-1-351-05466-9 978-1-351-05465-2 978-1-351-05463-8.
- Denton, Nancy A., and Douglas A. Massey. "Racial Identity among Caribbean Hispanics: The Effect of Double Minority Status on Segregation". *American Sociological Review* 54, no. 5 (Oct. 1989): 790–808.
- Eberhardt, Jennifer L., et al. "Looking Deathworthy: Perceived Stereotypicality of Black Defendants Predicts Capital-Sentencing Outcomes". *Psychological Science* 17, no. 5 (May 2006): 383–386. ISSN: 0956-7976, 1467-9280, visited on 05/04/2021. <https://doi.org/10.1111/j.1467-9280.2006.01716.x>. <http://journals.sagepub.com/doi/10.1111/j.1467-9280.2006.01716.x>.
- Egan, Patrick J. "Identity as Dependent Variable: How Americans Shift Their Identities to Align with Their Politics". *American Journal of Political Science* 64, no. 3 (July 2020): 699–716. ISSN: 0092-5853, 1540-5907, visited on 01/15/2023. <https://doi.org/10.1111/ajps.12496>. <https://onlinelibrary.wiley.com/doi/10.1111/ajps.12496>.
- Ellemers, Naomi, Paulien Kortekaas, and Jaap Ouwerkerk. "Self-categorisation, commitment to the group and group self-esteem as related but distinct aspects of social identity". *European Journal of Social Psychology* 29 (1999): 371–389. [https://onlinelibrary.wiley.com/doi/pdf/10.1002/\(SICI\)1099-0992\(199903/05\)29:2/3%3C371::AID-EJSP932%3E3.0.CO;2-U](https://onlinelibrary.wiley.com/doi/pdf/10.1002/(SICI)1099-0992(199903/05)29:2/3%3C371::AID-EJSP932%3E3.0.CO;2-U).
- Feliciano, Cynthia. "Shades of Race: How Phenotype and Observer Characteristics Shape Racial Classification". *American Behavioral Scientist* 60, no. 4 (2016): 390–419.
- Fraga, Luis, et al. *Latino Lives in America: Making it Home*. Philadelphia: Temple University Press, 2010.
- Frank, Reanne, Ilana Redstone Akresh, and Bo Lu. "Latino Immigrants and the U.S. Racial Order: How and Where Do They Fit In?" *American Sociological Review* 75, no. 3 (2010): 378–401.
- García-Bedolla, Lisa. "The Identity Paradox: Latino Language, Politics and Selective Dissociation". *Latino Studies* 1, no. 2 (July 2003): 264–283. ISSN: 1476-3435, 1476-3443, visited on 03/14/2023. <https://doi.org/10.1057/palgrave.lst.8600038>. <http://link.springer.com/10.1057/palgrave.lst.8600038>.

- Golash-Boza, T. “Dropping the Hyphen? Becoming Latino(a)-American through Racialized Assimilation”. *Social Forces* 85, no. 1 (Sept. 2006): 27–55. ISSN: 0037-7732, 1534-7605, visited on 04/12/2021. <https://doi.org/10.1353/sof.2006.0124>. <https://academic.oup.com/sf/article-lookup/doi/10.1353/sof.2006.0124>.
- Gómez, Laura E. *Manifest Destinies: The Making of the Mexican American Race*. New York: New York University Press, 2018.
- Helg, A. “Race in Argentina and Cuba, 1880–1930: Theory, policies, and popular reaction”. In *The idea of race in Latin America, 1870–1940*, 37–69. Austin: University of Texas Press.
- Hickel, Flavio R, et al. “The Role of Identity Prioritization”. *Public Opinion Quarterly* 84, no. 4 (June 2021): 860–891. ISSN: 0033-362X, 1537-5331, visited on 03/14/2023. <https://doi.org/10.1093/poq/nfaa048>. <https://academic.oup.com/poq/article/84/4/860/6205964>.
- Hitlin, S., J. S. Brown, and G. H. Elder. “Measuring Latinos: Racial vs. Ethnic Classification and Self-Understandings”. *Social Forces* 86, no. 2 (Dec. 2007): 587–611. ISSN: 0037-7732, 1534-7605, visited on 08/19/2022. <https://doi.org/10.1093/sf/86.2.587>. <https://academic.oup.com/sf/article-lookup/doi/10.1093/sf/86.2.587>.
- Ho, Arnold K., et al. “Evidence for hypodescent and racial hierarchy in the categorization and perception of biracial individuals”. *Journal of Personality and Social Psychology* 100 (3): 492–506.
- Hooker, Juliet. “Hybrid subjectivities, Latin American mestizaje, and Latino political thought on race”. *Politics, Groups, and Identities* 2, no. 2 (Apr. 2014): 188–201. ISSN: 2156-5503, 2156-5511, visited on 05/07/2021. <https://doi.org/10.1080/21565503.2014.904798>. <http://www.tandfonline.com/doi/abs/10.1080/21565503.2014.904798>.
- . “Indigenous Inclusion/Black Exclusion: Race, Ethnicity and Multicultural Citizenship in Latin America”. *Journal of Latin American Studies* 37, no. 2 (May 2005): 285–310. ISSN: 0022-216X, 1469-767X, visited on 08/19/2022. <https://doi.org/10.1017/S0022216X05009016>. https://www.cambridge.org/core/product/identifier/S0022216X05009016/type/journal_article.
- Huddy, Leonie, Lilliana Mason, and S. Nechama Horwitz. “Political Identity Convergence: On Being Latino, Becoming a Democrat, and Getting Active”. *RSF: The Russell Sage Foundation Journal of the Social Sciences* 2, no. 3 (2016): 205. ISSN: 23778253, visited on 06/10/2022. <https://doi.org/10.7758/rsf.2016.2.3.11>. <https://www.jstor.org/stable/10.7758/rsf.2016.2.3.11>.
- Irizarry, Yasmiyn, Ellis P. Monk Jr., and Ryon J. Cobb. “Race-shifting in the United States: Latinxs, Skin Tone, and Ethnoracial Alignments”. *Sociology of Race and Ethnicity* (2022): 1–19.

- Itzigsohn, Jose, and Carlos Dore-Cabral. "Competing Identities? Race, Ethnicity and Panethnicity among Dominicans in the United States". *Sociological Forum* 15, no. 2 (June 2000): 225–247.
- Jardina, Ashley. *White Identity Politics*. 1st ed. Cambridge University Press, Feb. 2019. ISBN: 978-1-108-64515-7 978-1-108-47552-5 978-1-108-46860-2, visited on 05/07/2021. <https://doi.org/10.1017/9781108645157>. <https://www.cambridge.org/core/product/identifier/9781108645157/type/book>.
- Jones-Correa, Michael, and David L. Leal. "Becoming "Hispanic": Secondary Panethnic Identification among Latin American-Origin Populations in the United States". *Hispanic Journal of Behavioral Sciences* 18, no. 2 (May 1996): 214–254. ISSN: 0739-9863, 1552-6364, visited on 06/10/2022. <https://doi.org/10.1177/07399863960182008>. <http://journals.sagepub.com/doi/10.1177/07399863960182008>.
- Loveman, Mara. *National Colors: Racial Classification and the State in Latin America*. Oxford University Press, July 2014. ISBN: 978-0-19-933735-4, visited on 05/07/2021. <https://doi.org/10.1093/acprof:oso/9780199337354.001.0001>. <https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199337354.001.0001/acprof-9780199337354>.
- Martínez, Daniel E., and Kelsey E. Gonzalez. "Panethnicity as a reactive identity: primary panethnic identification among Latino-Hispanics in the United States". *Ethnic and Racial Studies* 44, no. 4 (Mar. 2021): 595–617. ISSN: 0141-9870, 1466-4356, visited on 04/12/2021. <https://doi.org/10.1080/01419870.2020.1752392>. <https://www.tandfonline.com/doi/full/10.1080/01419870.2020.1752392>.
- Mason, Lilliana. *Uncivil agreement: how politics became our identity*. Chicago, Illinois ; London: The University of Chicago Press, 2018. ISBN: 978-0-226-52440-5 978-0-226-52454-2.
- Massey, Douglas A., and Nancy A. Denton. "Racial Identity and the Spatial Assimilation of Mexicans in the United States". *Social Science Research* 21 (1992): 235–260.
- Masuoka, Natalie. "Together They Become One: Examining the Predictors of Panethnic Group Consciousness Among Asian Americans and Latinos". *Social Science Quarterly* 87, no. s1 (Dec. 2006): 993–1011. ISSN: 0038-4941, 1540-6237, visited on 03/17/2021. <https://doi.org/10.1111/j.1540-6237.2006.00412.x>. <http://doi.wiley.com/10.1111/j.1540-6237.2006.00412.x>.
- Miller, Arthur H., et al. "Group Consciousness and Political Participation". *American Journal of Political Science* 25, no. 3 (1981): 494–511.
- Monk, Ellis P., Jr. "The Consequences of "Race and Color" in Brazil". *Social Problems* 63 (2016): 413–430.

- Montgomery, Jacob M., and Santiago Olivella. “Tree-Based Models for Political Science Data”. *American Journal of Political Science* 62, no. 3 (July 2018): 729–744.
- Mora, G. Cristina. *Making Hispanics: How Activists, Bureaucrats, and Media Constructed a New American*. OCLC: 1233041873 OCLC: 1233041873. 2021. ISBN: 978-0-226-03397-6, visited on 04/12/2021. <https://www.degruyter.com/isbn/9780226033976>.
- Ostfeld, Mara C, and Nicole D Yadon. “¿Mejorando La Raza?: The Political Undertones of Latinos’ Skin Color in the United States”. *Social Forces* 100, no. 4 (Apr. 2022): 1806–1832. ISSN: 0037-7732, 1534-7605, visited on 06/10/2022. <https://doi.org/10.1093/sf/soab060>. <https://academic.oup.com/sf/article/100/4/1806/6290180>.
- Penner, Andrew M., and Aliya Saperstein. “Engendering Racial Perceptions: An Intersectional Analysis of How Social Status Shapes Race”. *Gender & Society* 27, no. 3 (June 2013): 319–344.
- Porter, Sonya R., and C. Matthew Snipp. “Measuring Hispanic Origin: Reflections on Hispanic Race Reporting”. *The Annals of the American Academy of Political and Social Science* 677 (2018): 140–152.
- Portes, Alejandro, and Min Zhou. “The New Second Generation: Segmented Assimilation and its Variants”. *The ANNALS of the American Academy of Political and Social Science* 530, no. 1 (Nov. 1993): 74–96. ISSN: 0002-7162, 1552-3349, visited on 12/18/2022. <https://doi.org/10.1177/0002716293530001006>. <http://journals.sagepub.com/doi/10.1177/0002716293530001006>.
- Roth, Wendy. *Race Migrations: Latinos and the Cultural Transformation of Race*. OCLC: 1198931214 OCLC: 1198931214. 2012. ISBN: 978-0-8047-8253-1, visited on 05/05/2021. <https://doi.org/10.1515/9780804782531>.
- Roth, Wendy D. “Racial Mismatch: The Divergence Between Form and Function in Data for Monitoring Racial Discrimination of Hispanics*: Racial Mismatch”. *Social Science Quarterly* 91, no. 5 (Dec. 2010): 1288–1311. ISSN: 00384941, visited on 08/19/2022. <https://doi.org/10.1111/j.1540-6237.2010.00732.x>. <https://onlinelibrary.wiley.com/doi/10.1111/j.1540-6237.2010.00732.x>.
- . “The multiple dimensions of race”. *Ethnic and Racial Studies* 39, no. 8 (June 2016): 1310–1338. ISSN: 0141-9870, 1466-4356, visited on 08/19/2022. <https://doi.org/10.1080/01419870.2016.1140793>. <https://www.tandfonline.com/doi/full/10.1080/01419870.2016.1140793>.
- Sanchez, Gabriel R., and Natalie Masuoka. “Brown-Utility Heuristic? The Presence and Contributing Factors of Latino Linked Fate”. *Hispanic Journal of Behavioral Sciences* 32, no. 4 (Nov. 2010): 519–531. ISSN: 0739-9863, 1552-6364, visited on 04/12/2021. <https://doi.org/10.1177/0739986310383129>. <http://journals.sagepub.com/doi/10.1177/0739986310383129>.

- Saperstein, Aliya, and Andrew M. Penner. "Racial Fluidity and Inequality in the United States". *American Journal of Sociology* 118 (3): 676–727.
- Schacter, Ariela. "From "Different" to "Similar": An Experimental Approach to Understanding Assimilation". *American Sociological Review* 81, no. 5 (2016): 981–1013.
- Schickler, Eric. *Racial realignment: the transformation of American Liberalism, 1932-1965*. OCLC: 1032363540 [OCLC: 1032363540. 2018. ISBN: 978-1-4008-8097-3, visited on 06/07/2022. <https://doi.org/10.23943/princeton/9780691153872.001.0001>.
- Schwartzman, Luisa Farah. "Does Money Whiten? Intergenerational Changes in Racial Classification in Brazil". *American Sociological Review* 72, no. 6 (2007): 940–963.
- Sides, John, Michael Tesler, and Lynn Vavreck. *Identity crisis: the 2016 presidential campaign and the battle for the meaning of America*. First paperback edition. OCLC: on1104602105 [OCLC: on1104602105. Princeton: Princeton University Press, 2019. ISBN: 978-0-691-19643-5.
- Sue, Christina A., and Tanya Golash-Boza. "Blackness in Mestizo America: The Cases of Mexico and Peru". *Latino(a) Research Review* 7, no. 1 (2008): 30–58.
- Tafoya, Sonya. *Shades of Belonging*. Tech. rep. Pew Hispanic Center, Dec. 2004.
- Tajfel, Henri, and John C. Turner. "The Social Identity Theory of Intergroup Behavior". In *Political Psychology*, 0th ed., ed. by John T. Jost and Jim Sidanius, 276–293. Psychology Press, 1986. ISBN: 978-0-203-50598-4, visited on 08/19/2022. <https://doi.org/10.4324/9780203505984-16>. <https://www.taylorfrancis.com/books/9781135151355/chapters/10.4324/9780203505984-16>.
- Telles, Edward. "Latinos, Race, and the U.S. Census". *The ANNALS of the American Academy of Political and Social Science* 677, no. 1 (May 2018): 153–164. ISSN: 0002-7162, 1552-3349, visited on 05/07/2021. <https://doi.org/10.1177/0002716218766463>. <http://journals.sagepub.com/doi/10.1177/0002716218766463>.
- Telles, Edward, and René Flores. "Not Just Color: Whiteness, Nation, and Status in Latin America". *Hispanic American Historical Review* 93, no. 3 (Aug. 2013): 411–449. ISSN: 0018-2168, 1527-1900, visited on 05/07/2021. <https://doi.org/10.1215/00182168-2210858>. <https://read.dukeupress.edu/hahr/article/93/3/411/10799/Not-Just-Color-Whiteness-Nation-and-Status-in>.
- Telles, Edward, René D. Flores, and Fernando Urrea-Giraldo. "Pigmentocracies: Educational inequality, skin color and census ethnoracial identification in eight Latin American countries". *Research in Social Stratification and Mobility* 40 (June 2015): 39–58. ISSN: 02765624, visited on

- 05/04/2021. <https://doi.org/10.1016/j.rssm.2015.02.002>. <https://linkinghub.elsevier.com/retrieve/pii/S0276562415000177>.
- Telles, Edward, and Edward Murguía. “Phenotypic Discrimination and Income Differences among Mexican Americans”. *Social Science Quarterly* 71, no. 4 (1990): 682–696. <https://www.proquest.com/openview/9281d28fa2ad2d6aba954573a9b73e5f/1?pq-origsite=gscholar&cbl=1816420>.
- Telles, Edward E. *Race in Another America: The Significance of Skin Color in Brazil*. Princeton, N.J.: Princeton University Press, 2004.
- Telles, Edward E., and Vilma Ortiz. *Generations of exclusion: Mexican Americans, Assimilation, and Race*. New York: Russell Sage Foundation, 2008.
- Telles, Edward E., and Florencia Torche. “Varieties of Indigeneity in the Americas”. *Social Forces* 97, no. 4 (June 2019): 1543–1570.
- Telles, Edward Eric. *Pigmentocracies: ethnicity, race, and color in Latin America*. 1 Edition. Chapel Hill, NC: University of North Carolina Press, 2014. ISBN: 978-1-4696-1783-1.
- Vargas, Nicholas, and Kevin Stainback. “Documenting Contested Racial Identities Among Self-Identified Latina/os, Asians, Blacks, and Whites”. *American Behavioral Scientist* 60, no. 4 (2016): 442–464.
- Vasconcelos, José, and Didier Tisdell Jaén. *The cosmic race: a bilingual edition*. Johns Hopkins paperbacks ed. Race in the Americas. Baltimore, Md: Johns Hopkins University Press, 1997. ISBN: 978-0-8018-5655-6.
- Wade, Peter. *Race and ethnicity in Latin America*. 2nd ed. Anthropology, culture and society. OCLC: ocn5024155781[]OCLC: ocn502415578. London ; New York : New York: Pluto Press ; Distributed in the United States of America exclusively by Palgrave Macmillan, 2010. ISBN: 978-0-7453-2948-2 978-0-7453-2947-5.
- Wei, Pengfei, Zhenzhou Lu, and Jingwen Song. “Variable importance analysis: A comprehensive review”. *Reliability Engineering and System Safety* 142 (2015): 399–432.
- Westwood, Sean J., and Erik Peterson. “The Inseparability of Race and Partisanship in the United States”. *Political Behavior* (Oct. 2020). ISSN: 0190-9320, 1573-6687, visited on 12/07/2021. <https://doi.org/10.1007/s11109-020-09648-9>. <https://link.springer.com/10.1007/s11109-020-09648-9>.
- Yang, Alan, and Rodolfo O. de la Garza. “Americanizing Latinos, Latinoizing America: The Political Consequences of Latino Incorporation*: Americanizing Latinos, Latinoizing America”. *Social Science Quarterly* 98, no. 2 (June 2017): 690–727. ISSN: 00384941, visited on 04/02/2023. <https://doi.org/10.1111/ssqu.12345>.

[//doi.org/10.1111/ssqu.12325](https://doi.org/10.1111/ssqu.12325). <https://onlinelibrary.wiley.com/doi/10.1111/ssqu.12325>.

Zepeda-Millán, Chris. *Latino Mass Mobilization: Immigration, Racialization, and Activism*. 1st ed. Cambridge University Press, Sept. 2017. ISBN: 978-1-139-92471-9 978-1-107-07694-5 978-1-107-43412-7, visited on 08/19/2022. <https://doi.org/10.1017/9781139924719>. <https://www.cambridge.org/core/product/identifier/9781139924719/type/book>.

Appendix

Appendix A: Additional Variable Importance Plot

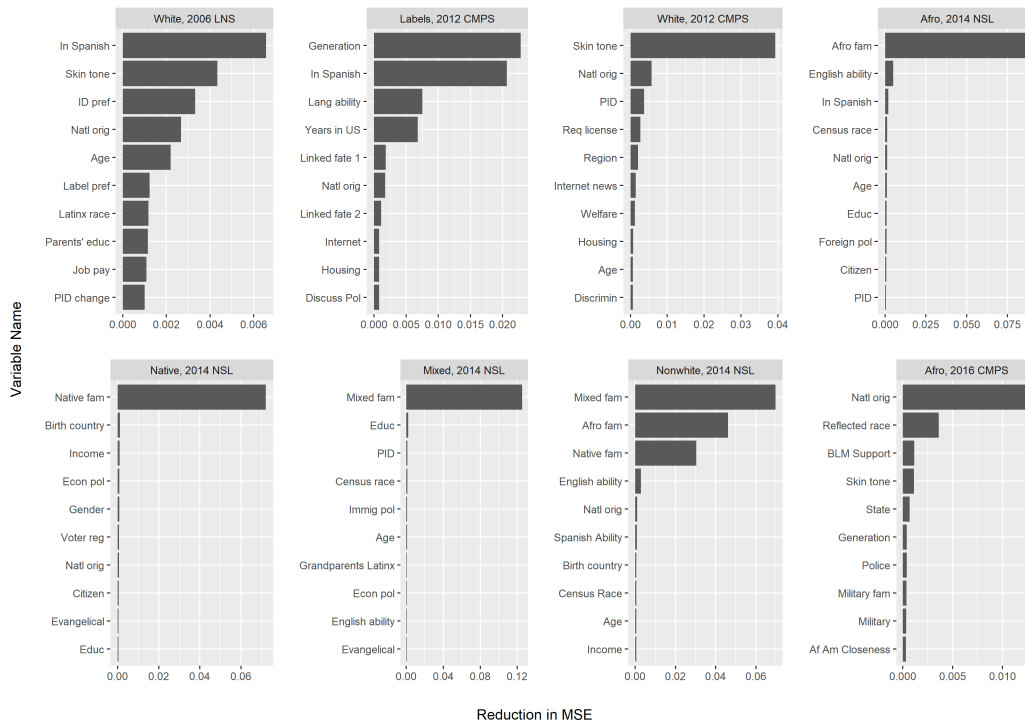


Figure A1: Top 10 predictors by variable importance for each DV. This figure displays the top 10 predictors of each dependent variable using permutation variable importance. The y-axis displays variable names, and the x-axis displays the reduction in mean-squared error (MSE) when the variable is randomly shuffled versus when it is held at its true values. The scale of MSE runs from 0 to 1, so a reduction of 0.01 represents a reduction in MSE of 1 percent of the full scale. More concretely, if the classification error of the model is 15 percent, then a variable with a value of 0.01 reduces the error of the model from 16 to 15 percent. Ancestry variables are included for the 2014 NSL.

Appendix B: Additional Partial Dependence Plots

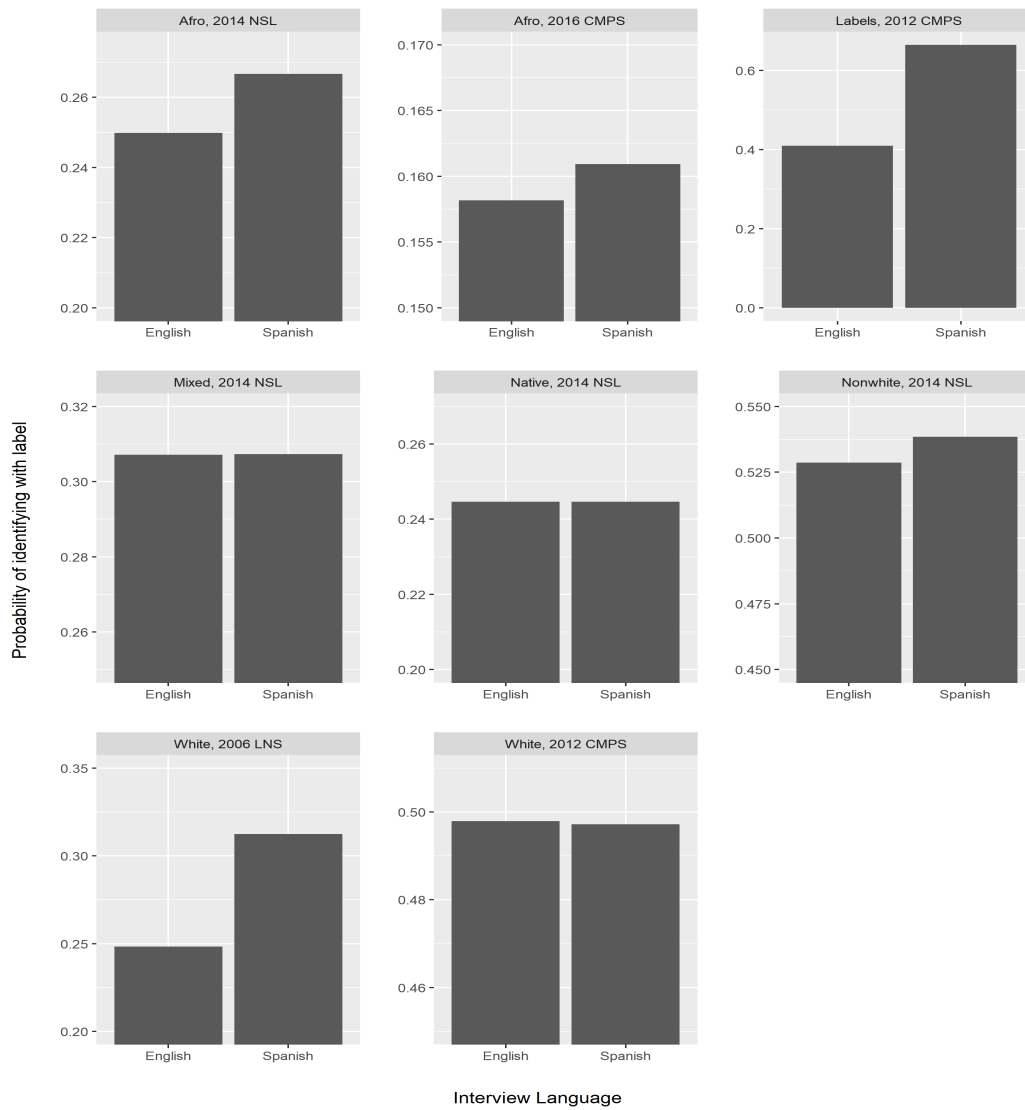


Figure B1: Partial dependence plot of interview language. This figure plots interview language on the x-axis and corresponding predicted probabilities of identifying with each DV – derived from random forest regressions – on the y-axis. For Latinos, the surveys could be completed either in English or Spanish.

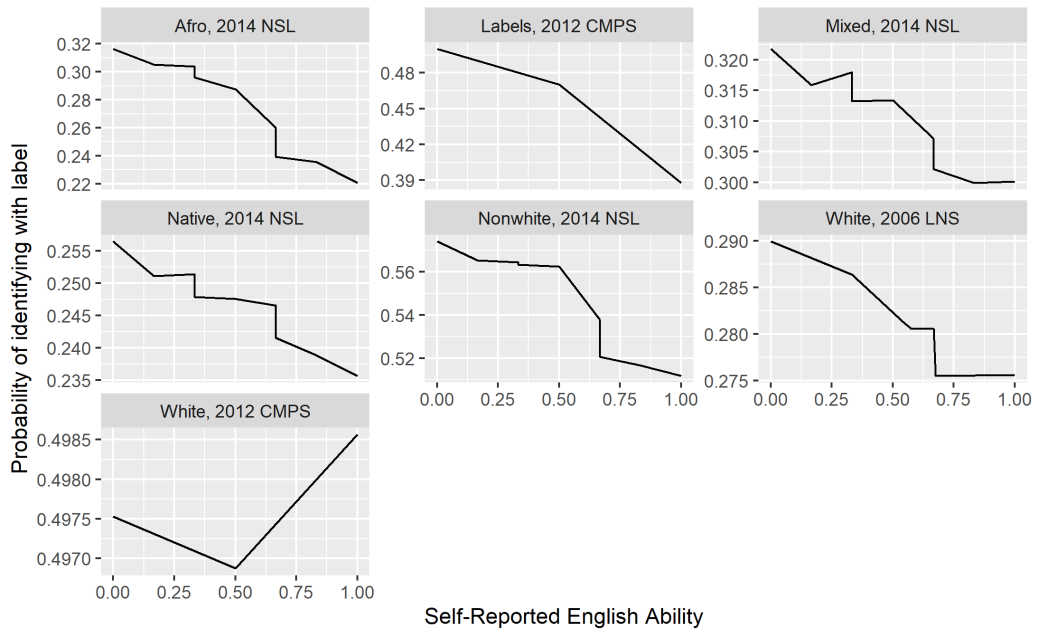


Figure B2: Partial dependence plot of English ability. This figure plots English ability on the x-axis and corresponding predicted probabilities of identifying with each DV – derived from random forest regressions – on the y-axis. Higher values indicate better self-reported English ability. The 2016 CMPS is missing because they did not ask self-reported English ability.

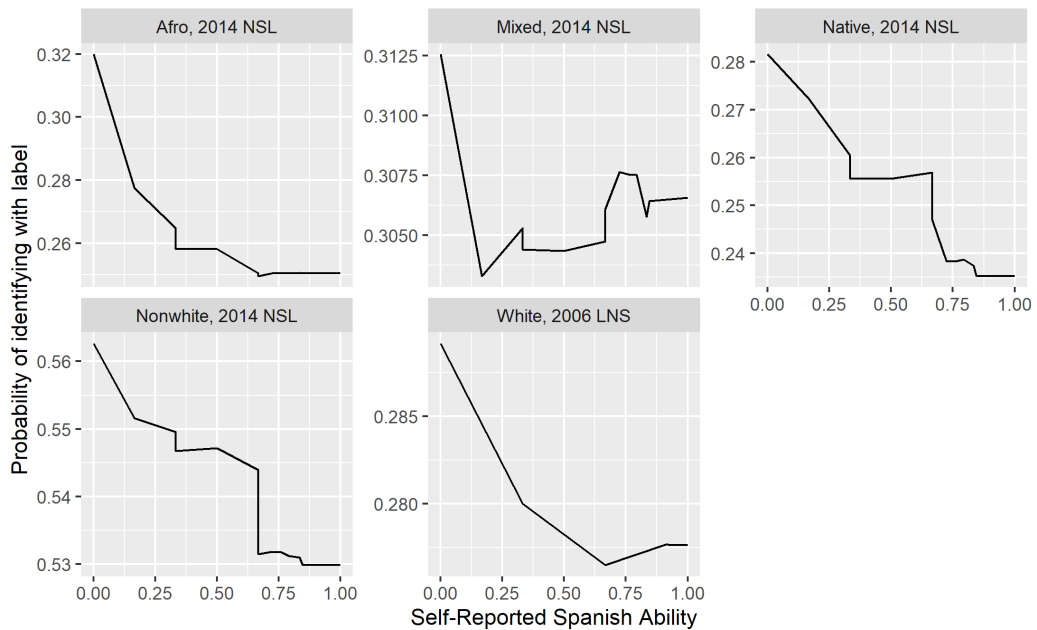


Figure B3: Partial dependence plot of Spanish ability. This figure plots Spanish ability on the x-axis and corresponding predicted probabilities of identifying with each DV – derived from random forest regressions – on the y-axis. Higher values indicate better self-reported Spanish ability. The 2012 and 2016 CMPS are missing because they did not ask self-reported Spanish ability.

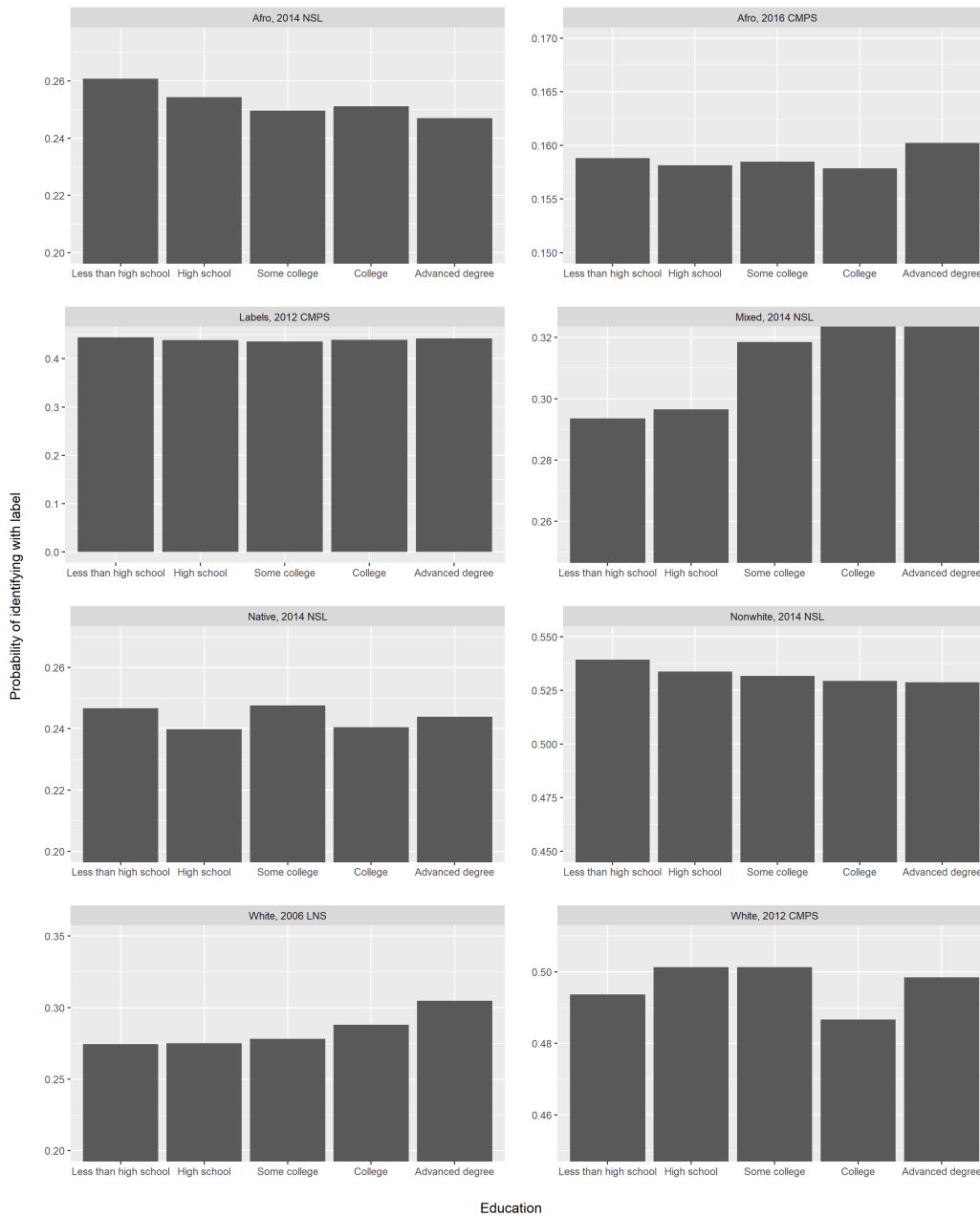


Figure B4: Partial dependence plot of education. This figure plots education on the x-axis and corresponding predicted probabilities of identifying with each DV – derived from random forest regressions – on the y-axis.

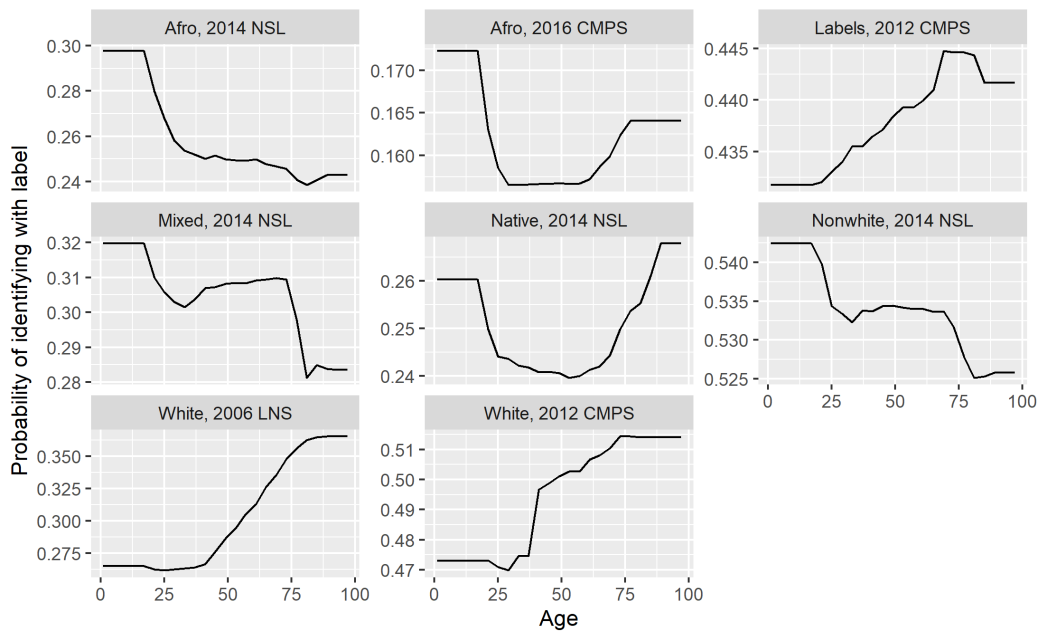


Figure B5: Partial dependence plot of age. This figure plots age on the x-axis and corresponding predicted probabilities of identifying with each DV – derived from random forest regressions – on the y-axis. Age is measured in years.

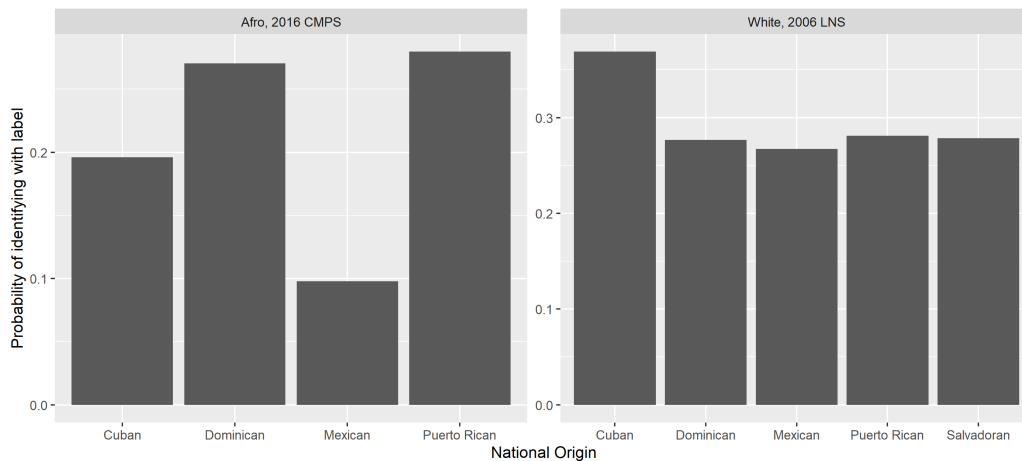


Figure B6: Partial dependence plot of national origin. This figure plots national origin on the x-axis and corresponding predicted probabilities of identifying with each DV – derived from random forest regressions – on the y-axis. Results are presented only for the 2016 CMPS and 2006 LNS because other surveys do not have sufficient sample sizes for national origin analyses.

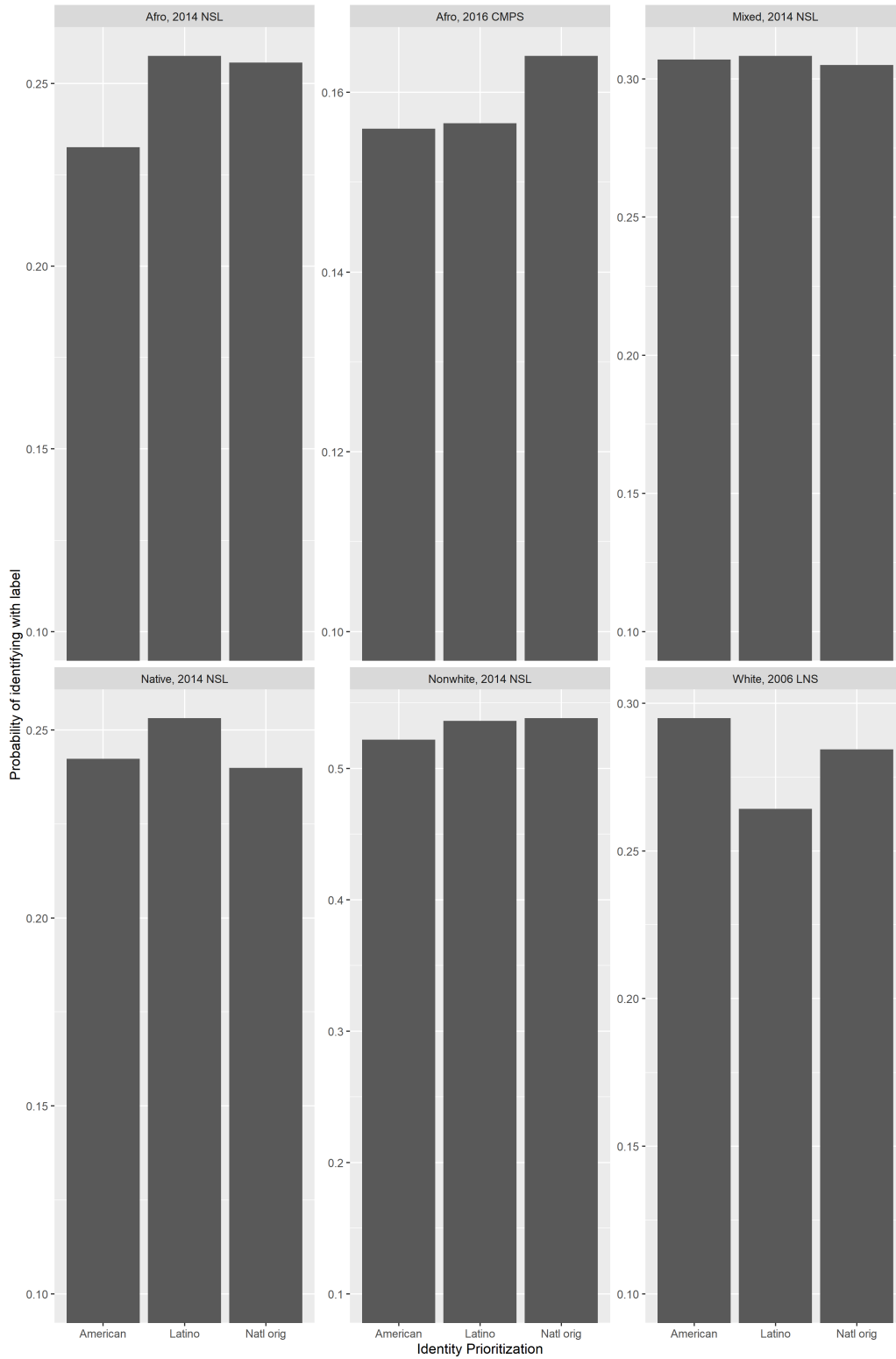


Figure B7: Partial dependence plot of identity prioritization. This figure plots identity prioritization on the x-axis and corresponding predicted probabilities of identifying with each DV – derived from random forest regressions – on the y-axis. The x-axis indicates which identity respondents prioritize most. DVs from the 2012 CMPS are missing because they did not ask this question.

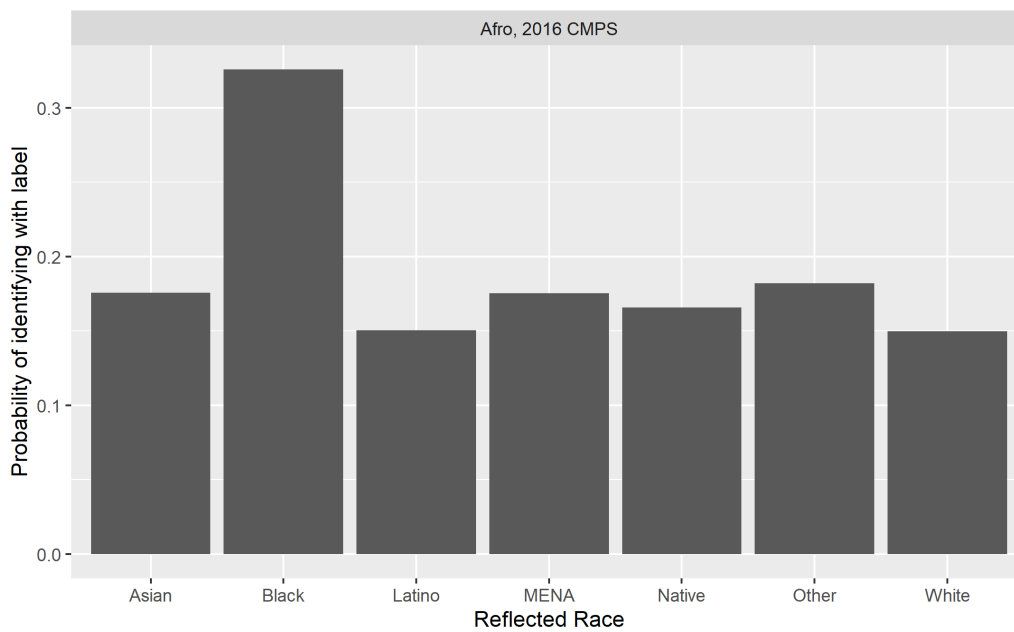


Figure B8: Partial dependence plot of reflected race. This figure plots reflected race on the x-axis and corresponding predicted probabilities of identifying with each DV – derived from random forest regressions – on the y-axis. The x-axis indicates which race respondents think others perceive them to be. Only the 2016 CMPS asked this question.