



Late-Life Changes in Ethnoracial Self-identification: Evidence from Social Security Administrative Data

Casey F. Breen¹

Received: 11 May 2022 / Accepted: 27 November 2022
© The Author(s) 2023

Abstract

Researchers generally recognize that ethnoracial identification may shift over the life course. However, the prevalence of these shifts across cohorts and among older adults remains open questions. Using administrative data from Social Security applications from 1984 to 2007, we quantify the magnitude and direction of later-life shifts in ethnoracial self-identification between Black, White, Asian, American Indian, and Hispanic categories for the “Greatest Generation,” those born between 1901 and 1927. Overall, 2.3% of persons in these data changed their ethnoracial identification after the age of 57, with distinct patterns of change for ethnoracial subgroups. By linking to the 1940 Census, we find a positive and significant association between socioeconomic status in early life and a shift from non-White to non-Hispanic White identification in later life. We conclude that ethnoracial self-identification fluidity continues even among older adults, varying in response to social position, ethnoracial climate, and events in greater society.

Keywords Race · Racial fluidity · Identification · Full-count census · Record linkage

The author would like to thank participants at the 2020 UC Berkeley Record linkage conference, participants at the 2020 ASA Population Section “Immigration and Racial/Ethnic Inequality” Roundtable, Mara Loveman, Dennis M. Feehan, Joshua R. Goldstein, Leora Lawton, Leslie Root, Nathan Seltzer, and Robert Pickett for their helpful feedback on earlier drafts of this manuscript. C.F.B is supported by the National Institute of Aging R01AG058940 and National Institute of Aging T32-AG000246. Replication code is available on the Open Science Framework: <https://osf.io/>.

✉ Casey F. Breen
caseybreen@berkeley.edu

¹ Department of Demography, University of California, Berkeley, 2232 Piedmont Ave, Berkeley, CA 94720, USA

Introduction

Social scientists acknowledge that race and ethnicity are social constructions that have meaning in a specific time and place, and most contemporary scholars of race and ethnicity consider ethnoracial identification to be fluid and changeable over the life course. Several recent large-scale studies have found that ethnoracial classification (how an individual's ethnoracial identification is classified by others) and ethnoracial identification (how an individual self-identifies their race) may shift over time (Saperstein and Penner, 2012; Saperstein and Gullickson, 2013; Liebler et al., 2017; Dahis et al., 2019; Doyle and Kao, 2007; Saperstein et al., 2013; Davenport, 2020). These studies have documented distinct patterns of reclassification and response change across ethnoracial¹ groups using different methods and data collected from several different historical periods. Ethnoracial self-identification can shift over the life course, and patterns of ethnoracial identification and classification can shift at the societal level with changing cultural norms and beliefs.

Despite the importance of past research, several key questions remain unanswered. Relatively little is known about ethnoracial fluidity in older ages: is ethnoracial identification fluid for older ages, or does identification solidify at some point in the life course and remain stable at older ages? And if identification remains fluid in later life, how does social position in early adulthood correlate with these late-life shifts in identification? A more complete understanding of patterns of change in ethnoracial identification by age and cohort provides crucial insights into the strength of ethnoracial boundaries at a given time and the impact of social change on reconfigurations in the ethnoracial status hierarchy. However, data limitations have often precluded the analysis of ethnoracial fluidity in older ages, when cohorts have shrunk to a fraction of their original size. Further, late-life shifts in identification point away from theories suggesting that these shifts in identification are for instrumental or strategic reasons (e.g., qualifying for specific programs or avoiding discrimination based on their later-life identification); the incentives for shifting identification are plausibly lower after individuals have departed from the labor and marriage market. Life-course theories of racial identity formation emphasize childhood, adolescence, and young adulthood (Rivas-Drake et al., 2014; Umaña-Taylor et al., 2014). By studying older adulthood, we are able to see whether ethnoracial identification remains plastic and changeable at older ages.

This is the first work to study ethnoracial shifts among the “Greatest Generation,” who experienced the Great Depression, the economic impact of the New Deal, and the Second World War in early adulthood, and who are next observed as older Americans in the aftermath of the Civil Rights era. This is an especially interesting cohort to consider, as this group experienced Jim Crow laws, the Civil Rights era, and the racial and ethnic pride social movements of the last decades of the 20th century. Their cohort-specific experiences across the life course may have further

¹ We follow Alba (2020) in using the term ethnoracial to capture the major racial and ethnic groups.

entrenched and reified their views on race and ethnicity, making them less likely to shift their ethnoracial identification in older ages than other cohorts.

We use linked administrative Social Security SS-5 applications to study shifts in ethnoracial identification for Americans born between 1901 and 1927, who predominantly died between 1988 and 2007 and filled out two or more Social Security SS-5 applications after 1983 ($N = 448,827$). Individuals submitted Social Security applications for an original Social Security card, a replacement Social Security card, or to change their information on record (e.g., name change, corrected birth dates). We find that the highest rates of ethnoracial shifting for people who initially selected American Indian (14.1%) or Asian (8.3%) and lower rates of fluidity for those who initially identified as Hispanic (2.7%), White (1.8%), or Black (1.3%). We also link these individuals to the 1940 Census records to obtain socioeconomic status indicators in early adulthood, finding a statistically significant relationship between more advantaged social position and shifts toward a White identification.

The use of Social Security applications has several unique advantages when it comes to looking at changes in ethnoracial identification. First, other research that uses nominal record linkage can generate “false positives” for shifts in ethnoracial identification, because the incorrectly linked records can actually come from two different people who may report different ethnoracial identities. Even a very high match-accuracy rate (e.g., 97%) will mean that a large share of ethnoracial changes will be an artifact of false matches. Using Social Security records matched on Social Security numbers assures nearly perfect matches. Second, when individuals fill out official government forms, there should be a conservative tendency, so that, all else equal, the people trying to fill out the forms are trying to avoid changes so as to preserve the full record of their earnings history. Third, Social Security form applicants have no incentives in terms of desirability bias or access to benefits by marking a particular race or ethnicity. Fourth, the forms were generally filled out by the respondents themselves, who swore with their signature about the accuracy of the form. In this context, we argue, changes in race are likely to correspond to “real” changes in individuals’ perception and identification of themselves.

The remainder of the paper proceeds as follows. In Section 2, we give a brief overview of past research of the fluidity of ethnoracial identification and classification. In Section 3, we describe the data and methods used in our analysis. We then present our estimated rates of late-life shifts in ethnoracial identification and the socioeconomic factors predicting these shifts in Section 4. We conclude in Section 5 with a discussion of our findings and their broader implications for the ethnoracial fluidity literature.

Background

Ethnoracial fluidity can be understood as a boundary process. At any point in time, boundaries are either “bright” (distinct and unambiguous) or “blurred” (ambiguous) depending on how they have been institutionalized (Alba, 2005; Alba and Nee, 2009; Lichter and Qian, 2018). These boundaries can also move and become increasingly bright or blurred over time. For example, the Puerto Rican population

became significantly Whiter between the 1910 and 1920, likely because of changes in the sociocultural definition of Whiteness (Loveman and Muniz, 2007).

Ethnoracial self-identification reflects some combination of personal history, socialization, and ancestry that contributes to individual identity processes. Ethnoracial classification involves the evaluation and categorization of others based on one's perceived phenotype or social qualities. Both are fluid over the life course for the individual (Saperstein and Penner, 2012). Past work finds ethnoracial identification among those who originally self-identified as Black, White, or Asian remains consistent for at least 90% over a 10-year period (Liebler et al., 2017; Doyle and Kao, 2007). Those who selected other ethnoracial identities, such as Pacific Islander, American Indian, or Hispanics, have higher rates of fluidity in and out of these categories. The relatively low overall level of fluidity reflects the brightness of boundaries surrounding ethnoracial membership in the U.S.

Data limitations make the study of ethnoracial fluidity challenging, particularly for smaller population subgroups. Repeated cross-sectional data show net changes between categories but hide between-category flows. To overcome this, several large-scale studies have linked decennial censuses at the individual level to observe between-category population churning, large countervailing flows into and out of race and ethnic response categories. One previous study relying on linked full-count U.S. census data (Liebler et al., 2017) used probabilistic record linkage techniques to assign each person in the 2000 and 2010 decennial Censuses an anonymized Protected Identification Keys (PIKS) using name, sex, date of birth, and address. The PIKS were then used to link each individual's 2000 and 2010 Census records to identify patterns of race and ethnicity-response change. The study found large countervailing flows between race and ethnic response categories that are otherwise obscured in the cross-sectional data. Approximately 6.1% of people in the data reported a race and/or Hispanic response change, with significant variation by ethnoracial group. Notably, American Indian was a very fluid identification during this period; only one-third of individuals who identified as American Indian in either the 2000 Census or 2010 Census identified as American Indian in both censuses. Dahis et al. (2019) measured ethnoracial fluidity for men during a much earlier historical period by linking full-count records for the 1880–1940 decennial Censuses.² In their sample, the implied rates of intercensal Black Americans “passing” as White were 6.8% to 9.9%.

Other studies center on identification of the factors that predict a change in ethnoracial self-identification. Due in part to the persistent socioeconomic disadvantage of Hispanic Americans in the United States, selective ethnic attrition—the process by which

² Dahis et al. (2019) use a very conservative two-sided perfect link algorithm to minimize mismatch. The algorithm proceeds by identifying those records in census decade t that have one (and only one) exact match to Census decade $t + 10$. Then it takes the forward-matched record and works backward to check that it has one (and only one) exact match to Census decade t . This conservative strategy minimizes false positive matches, which is crucial for measuring ethnoracial fluidity because estimates of rare events, such as shifts in ethnoracial identification, are particularly sensitive to false matches. The analysis is limited to men to avoid the difficulties associated with linking women, who often change their name after marriage.

ethnic identification fades selective on higher socioeconomic status—has received much attention in the literature. Golash-Boza (2006) finds that Hispanic individuals who have gained more traditional markers of assimilation are less likely to experience ethnic discrimination and more likely to identify as non-Hispanic. Duncan and Trejo (2011) find evidence of sizable ethnic attrition among Hispanics in the Current Population Survey (CPS) that is correlated with socioeconomic attainment. They also find that Hispanics who marry White spouses are more likely to identify as White, and Hispanics with more education and higher income are more likely to marry a White spouse (Duncan and Trejo, 2011). Selective attrition is problematic because it downwardly biases cross-sectional standard measures of socioeconomic status and intergenerational attainment for Hispanic immigrants.

Macro-level social change in the ethnoracial and political landscape can lead to individuals changing or reclaiming their ethnoracial identification. The seminal work of Nagel (1995) shows that the number of Americans who reported an American Indian race more than tripled between the 1960 Census and 1990 Census, vastly exceeding what is feasible by population growth or changing enumeration instructions or definitions alone. During this period, individuals reclaimed their American Indian identification as part of the “Red Power” Indian political movement, which advocated for a reclamation of the American Indian identity. Similarly, the Hispanic category was first introduced in the 1970s by activist groups and media, with its eventual inclusion in the 1980 Census (Mora, 2014). Inclusion in the 1980 Census allowed for the “Hispanic” identity to rapidly expand to driver’s licenses, birth certificates, and Social Security Application forms.

Saperstein and Penner (2012) used panel data from the National Longitudinal Survey of Youth (NLSY) to quantify the relationship between racial fluidity and inequality for both racial classification and identification. The NLSY surveyed a nationally representative cohort of 12,686 U.S. men and women aged 14–21 in 1979, and re-interviewed annually until 1994, when they re-interviewed biennially. The interviewer classified the respondents’ race at the end of all interviews conducted between 1979 and 1998, except for the 1987 interview. Additionally, respondents self-identified their race or ethnic identification in the 1979 and 2002 interviews. The study finds evidence that racial classification does change over time, and that social indicators such as unemployment, incarceration, and marriage influence the way an interviewer classifies a person’s race. Their analysis of racial identification response change compares a respondent’s identified race in adolescence in 1979 to identified race in middle adulthood nearly a quarter-century later in 2002, and again finds higher social status to be a predictor of a shift from non-White to White self-identification. The present study and Saperstein and Penner (2012) have an overlapping observation period, but while Saperstein and Penner (2012) centered on changes in racial classification and identification between adolescence and middle adulthood, the present study centers on changes in older adulthood.

Data and Methods

We use a novel data resource for our analysis of ethnoracial fluidity: U.S. Social Security Administration application (SS-5) record entries from the Social Security Administration's Numerical Identification File (Numident) (Finlay and Genadek, 2021). The publicly available subset of Numident application entries used in this analysis was released by the National Archives and Records Administration and contains 72.2 million Social Security application record entries.³ The Numident contains one record for each Social Security number ever assigned; each record can contain multiple application entries, which are added when a Social Security cardholder submits a new application.

Figure 1 shows the Social Security application form. Each application entry contains individual-level information from the form: full first name and last name, middle initial, date of birth, birthplace, parents' names, and basic demographics, such as race and sex (Breen and Goldstein, 2022). The "race/ethnic identification" question contains five categories: (1) Asian, Asian American, or Pacific Islander, (2) Hispanic, (3) Black (not Hispanic), (4) North American Indian or Alaskan Native, or (5) White (not Hispanic). For brevity, we will henceforth abbreviate these five ethnoracial categories as Asian, Hispanic, Black, American Indian, and White, respectively. The Social Security application form and its ethnoracial categories were updated in 1980. Prior to 1980, the Social Security application form included (1) White, (2) Black, and (3) Other (Scott, 1999).

For our analysis, we focus on the birth cohorts of 1901–1927. Because the ethnoracial coding schema on Social Security application forms changed in 1980 and the forms weren't fully phased out until 1983, we limit our analysis to individuals with two or more Social Security applications submitted after 1983. The most common reason for submitting an application form is a lost Social Security card or a change of name, sex, or date of birth information (Puckett, 2009). We link Social Security applications using Social Security number and define a shift in ethnoracial identification as a change in a response for the "race/ethnic identification" question between the first and last Social Security applications. The average window between the first and last application is 7 years, with a standard deviation of 6.1 years. The mean start date of an observation window is 1985, with a standard deviation of 3.8 years.

The Social Security data gives us overall estimates of rates of shifts in ethnoracial self-identification among members of the "Greatest Generation" in later life. We also link men and women in our Numident sample to the full-count IPUMS 1940 Census records (Ruggles et al., 2020) at the individual level using the CenSoc-Numident to obtain measures of social status in early adulthood (Goldstein et al., 2021). As there is not a shared common identifier available in both datasets, the CenSoc-Numident dataset links on key identifiers unlikely to change over the life course:

³ The publicly available Numident records used for this study are available here: <https://aad.archives.gov/aad/series-description.jsp?s=5057>.

SOCIAL SECURITY ADMINISTRATION **Application for a Social Security Card**

 Form Approved
 OMB No. 0960-0066

1	NAME → TO BE SHOWN ON CARD	First	Full Middle Name	Last
	FULL NAME AT BIRTH IF OTHER THAN ABOVE	First	Full Middle Name	Last
	OTHER NAMES USED			
2	MAILING ADDRESS → Do Not Abbreviate	Street Address, Apt. No., PO Box, Rural Route No.		
		City	State	Zip Code
3	CITIZENSHIP → (Check One)	<input type="checkbox"/> U.S. Citizen <input type="checkbox"/> Legal Alien Allowed To Work <input type="checkbox"/> Legal Alien Not Allowed To Work (See Instructions On Page 1) <input type="checkbox"/> Other (See Instructions On Page 1)		
4	SEX →	<input type="checkbox"/> Male <input type="checkbox"/> Female		
5	RACE/ETHNIC DESCRIPTION → (Check One Only - Voluntary)	<input type="checkbox"/> Asian, Asian-American or Pacific Islander <input type="checkbox"/> Hispanic <input type="checkbox"/> Black (Not Hispanic) <input type="checkbox"/> North American Indian or Alaskan Native <input type="checkbox"/> White (Not Hispanic)		
6	DATE OF BIRTH → Month, Day, Year	7 PLACE OF BIRTH → (Do Not Abbreviate) City State or Foreign Country FCI		
8	A. MOTHER'S MAIDEN NAME →	First	Full Middle Name	Last Name At Her Birth
	B. MOTHER'S SOCIAL SECURITY NUMBER →	<input type="text"/> <input type="text"/> <input type="text"/> - <input type="text"/> <input type="text"/> <input type="text"/> - <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>		
9	A. FATHER'S NAME →	First	Full Middle Name	Last
	B. FATHER'S SOCIAL SECURITY NUMBER →	<input type="text"/> <input type="text"/> <input type="text"/> - <input type="text"/> <input type="text"/> <input type="text"/> - <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>		
10	Has the applicant or anyone acting on his/her behalf ever filed for or received a Social Security number card before? <input type="checkbox"/> Yes (If "yes", answer questions 11-13.) <input type="checkbox"/> No (If "no", go on to question 14.) <input type="checkbox"/> Don't Know (If "don't know", go on to question 14.)			
11	Enter the Social Security number previously assigned to the person listed in item 1. →	<input type="text"/> <input type="text"/> <input type="text"/> - <input type="text"/> <input type="text"/> <input type="text"/> - <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>		
12	Enter the name shown on the most recent Social Security card issued for the person listed in item 1. →	First	Middle Name	Last
13	Enter any different date of birth if used on an earlier application for a card. →	Month, Day, Year		
14	TODAY'S DATE → Month, Day, Year	15 DAYTIME PHONE NUMBER → () Area Code Number		
16	I declare under penalty of perjury that I have examined all the information on this form, and on any accompanying statements or forms, and it is true and correct to the best of my knowledge.			
16	YOUR SIGNATURE →	17 YOUR RELATIONSHIP TO THE PERSON IN ITEM 1 IS: <input type="checkbox"/> Self <input type="checkbox"/> Natural Or Adoptive Parent <input type="checkbox"/> Legal Guardian <input type="checkbox"/> Other (Specify)		

Fig. 1 Social Security application form (post-1980)

first name, last name, birthplace, and birth year.⁴ To account for surname changes during marriage for women, we first identify marital status in the 1940 Census. For ever-married women, links are established using the reported last name in both the

⁴ The CenSoc dataset are available at <https://censoc.berkeley.edu/>. The CenSoc-Numident establishes individual-level matches between the 1940 Census and Numident records using the ABE exact record linkage algorithm, which requires an exact match on standardized first name, last name, and place of birth, but allows for flexibility ± 2 years on birth year (Abramitzky et al., 2019, 2012, 2014). The resulting dataset contains over 7.9 million records and allows researchers to take advantage of the rich measures available in the public 1940 Census: education, salary and wage income, geography, home ownership, income, occupation, family structure, parental birthplace, and more.

Table 1 Sample sizes

Sample	Size
2+ SS Applications (1984–2007)	1,719,106
& Born 1901–1927	448,827
& Linked to 1940 Census	140,710

1940 Census and Social Security application. For never-married women, matches are established using last name in the 1940 Census and father's last name in the Numident sample, on the assumption that a woman's father's last name will be the last name they received at birth (and will match the last name recorded in their 1940 census record). Additionally, the ABE algorithm standardizes names to account for common misspellings or nicknames. Our linked sample contains 140,710 records, corresponding to a raw match rate of approximately 31.5% (see Table 1).

For individuals in our sample successfully linked to the 1940 Census, Table 2 compares the distribution of education, census race, marital status, a socioeconomic composite indicator, and gender. Individuals who submitted two or more Social

Table 2 Comparison of socioeconomic characteristics by number of Social Security applications submitted after 1983

	0 apps		1 app		2+ apps	
	No.	%	No.	%	No.	%
Gender						
Female	3635432	51.7	555642	52.9	75452	53.6
Male	3393408	48.3	494438	47.1	65274	46.4
Birth year						
1901–1909	533795	7.6	150499	14.3	13790	9.8
1910–1919	2964356	42.2	399456	38.0	49552	35.2
1920–1927	3510992	50.0	495849	47.2	76490	54.4
Education Level						
Less than high school	4707262	67.0	742768	70.7	105550	75.0
High school or some college	2011719	28.6	256925	24.5	29030	20.6
Bachelors degree or higher	208235	3.0	34895	3.3	3931	2.8
NA	101624	1.4	15492	1.5	2215	1.6
Marital status						
Married	2249441	32.0	386187	36.8	44201	31.4
Not married	4779399	68.0	663893	63.2	96525	68.6
SEI						
1–9	588459	8.4	96769	9.2	14610	10.4
10–14	220661	3.1	40435	3.9	5197	3.7
15–25	926582	13.2	134235	12.8	16512	11.7
26+	1142438	16.3	165164	15.7	18100	12.9
NA	4150700	59.1	613477	58.4	86307	61.3

Security applications after 1983 – the individuals used in our analysis – have slightly lower socioeconomic status and are slightly more likely to be women than those who submitted fewer than two Social Security applications.

To summarize, we use two samples for our analyses. For our main analysis of shifts in ethnoracial self-identification in later life, we use our restricted Numident sample including birth cohorts of 1901–1927 ($N = 448,827$). These individuals are all 57 or older when they are first observed in our data. As we are using Social Security applications linked on Social Security number, we have nearly perfect matches, and do not risk false matches upwardly biasing our estimates of ethnoracial fluidity. For our analysis of the earlier-life social status correlates of shifts in ethnoracial identification in later life, we use the subset of our restricted Numident sample successfully linked to the 1940 Census ($N = 140,710$).

Results

We observe relatively low overall rates of fluidity, with only 2.3% of individuals shifting their ethnoracial identification between Social Security applications. However, we find distinct patterns of response change across ethnoracial groups. Table 3 shows a cross-tabulation of the ethnoracial fluidity observed in the Social Security applications. Several substantive insights emerge from this table. First, the main diagonal of the matrix shows high response stability (consistent ethnoracial identification on both first and last Social Security applications) for those who initially identified as White (98.2%), Black (98.7%), or Hispanic (97.3%), and lower response consistency for those who initially identified as Asian (91.7%) or American Indian (85.9%). Those who initially identified as Black or White had the lowest probability of shifting their ethnoracial identification. Second, for those who initially identified as Hispanic, Black, Asian, or American Indian, the most common shift was toward a White identification on the last application. The most common shift for those who initially identified as White was toward a Hispanic identification (0.7%).

These trends are mostly constant across age and birth cohort. Figure 2 shows age-specific (panel a) and birth-cohort-specific (panel b) rates of ethnoracial response consistency. The age-specific analysis demonstrates that rates of self-identification are constant throughout the latter part of the life course. We see similar results for birth-cohort-specific analysis, with the exception of people who initially selected Asian being less likely to switch their identification if they were born later.

Finally, we investigate the relationship between early-life socioeconomic status and later-life shifts toward a White identification. Figure 3 shows the bivariate correlation between two measures of social status in 1940—educational attainment in years and wage and salary income—and a shift from a non-White to a White identification.⁵ For those who initially identified as Black, Hispanic, or American Indian,

⁵ Salary and wage income in the 1940 Census is an imperfect measure of income, because it was not collected for business owners or farmers, who composed 32% of the sample.

Table 3 Rates of response change in ethnoracial identification between first and last Social Security application. Numbers in parentheses show raw counts. The main diagonal shows response stability for each ethnoracial group

Identification first SS App	Identification Last SS App				
	American Indian	Asian	Black	Hispanic	White
American Indian	85.9% (3439)	0.9% (35)	2.9% (118)	1.0% (39)	9.3% (371)
Asian	0.2% (56)	91.7% (23469)	0.5% (123)	0.8% (197)	6.8% (1743)
Black	0.1% (118)	0.1% (113)	98.7% (108466)	0.3% (374)	0.7% (784)
Hispanic	0.1% (51)	0.4% (208)	0.8% (410)	97.3% (50530)	1.4% (737)
White	0.2% (464)	0.6% (1592)	0.3% (763)	0.7% (1915)	98.2% (252712)

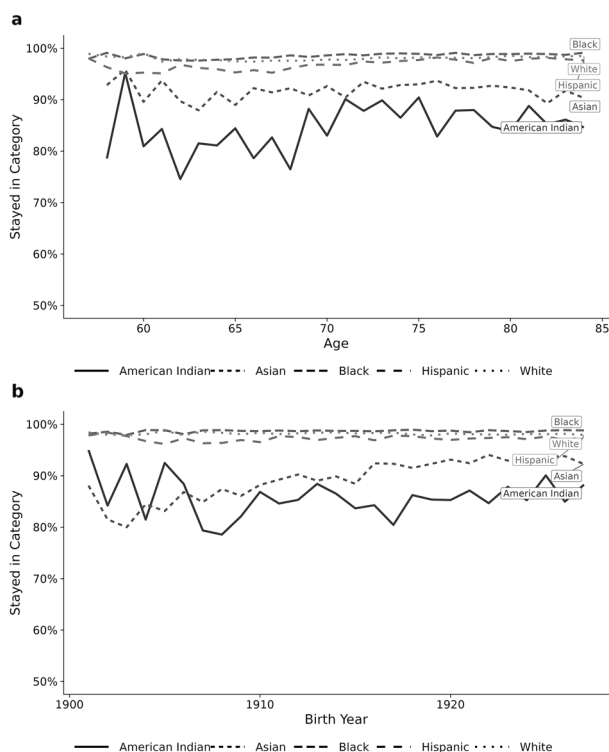


Fig. 2 **a** Shows age-specific trends in ethnoracial identification shifts over the observed window, where age is defined as a person's age when they submitted their last Social Security application. **b** Shows birth-cohort-specific trends in identification shifts for all birth cohorts included in this analysis. The higher variation in response change for the Asian and American Indian groups is due to smaller sample size

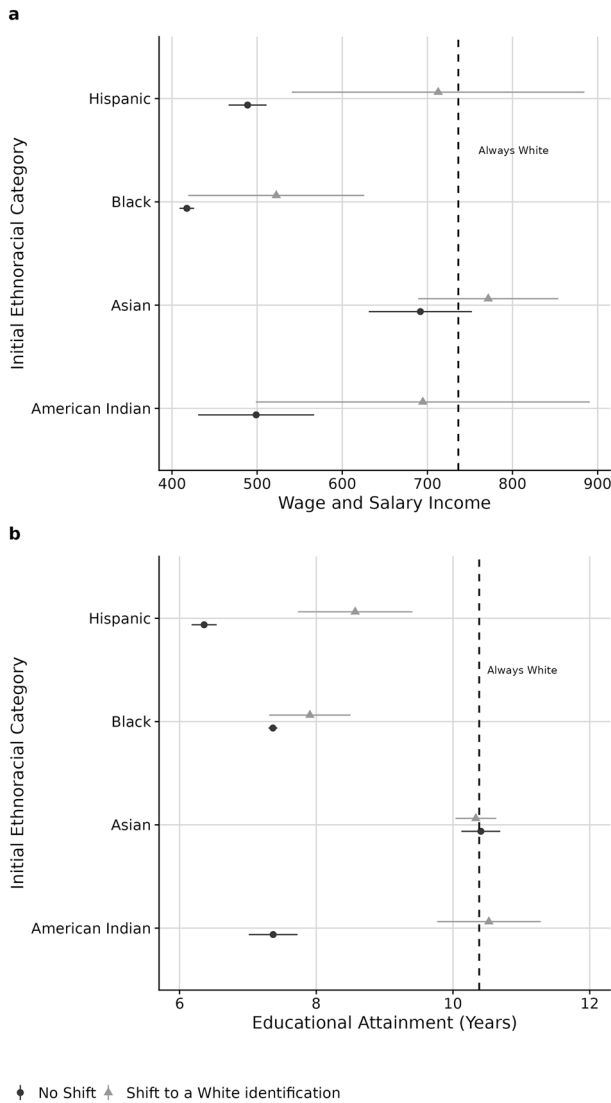


Fig. 3 Bivariate plot showing the association between early socioeconomic status and a shift from a non-White to a White identification in our linked sample. **a** shows mean pre-tax salary and wage income in 1940 ± 1.96 SE for persons 18+. **b** Shows mean years of education ± 1.96 SE in 1940 for persons 18+. Vertical dashed line denotes mean wage income (**a**) or educational attainment (**b**) for individuals identifying as White in both first and last application

higher educational attainment and wage and salary income was associated with a shift toward a White identification. For those who initially identified as Asian, neither measure of social status is strongly associated with shifts toward a White identification.

Discussion and Conclusion

We use novel administrative data from the Social Security administration to document shifts in ethnoracial self-identification for older Americans.⁶ Our analysis focuses on the previously unexamined birth cohorts of 1901–1927, children of the Great Depression who experienced the economic consequences of the New Deal and the aftermath of World War II. Our findings reveal several new empirical insights into rates of late-life shifts in ethnoracial self-identification and the factors predicting these shifts.

We observe relatively low overall rates of fluidity (2.3%), with significant variation in the direction and magnitude of these changes across ethnoracial groups. Late-life shifts in ethnoracial identification for this cohort are predominantly from non-White to White, upward in the ethnoracial hierarchy. Those who initially identified as the most historically advantaged (White) or disadvantaged (Black) groups have the lowest probability of shifting their identification. For White Americans, the most privileged in the U.S. racial hierarchy, there are few incentives to shift; for Black Americans, lived experience within the racial order defined by a rigid Black-White boundary might strengthen ethnoracial identity. Those who initially identified as American Indian or Asian are the most likely to shift their ethnoracial self-identification. This is consistent with previous evidence that the U.S. ethnoracial hierarchy provides different incentives and opportunities for individuals to cross ethnoracial boundaries (e.g., Nagel, 1995; Saperstein et al., 2013; Loveman and Muniz, 2007). The most common shift for those who originally identified as White was toward a Hispanic identification, likely reflecting the increasing demographic importance of Hispanics as a distinct and recognized ethnic identity, and the concurrent rise in ethnic pride movements of this time (Mora, 2014).

We find that higher social status in early adulthood is associated with higher rates of change to a White identification in late-life for those who originally identified as Black, Hispanic, or American Indian. The association is particularly pronounced for those who initially identified as Hispanic or American Indian, for whom identification is overall more fluid in our data, and less pronounced for those who initially identified as Black, for whom identification is more rigid. It is noteworthy that despite our findings for other non-White groups, we find no statistically significant association between social position and a shift toward a White identification for those who originally identified as Asian. While the average 1940 socioeconomic status of those who initially identified as Asian is quite similar to those who initially identified as White, this alone cannot account for the lack of association between early adulthood social position and shifts toward a White identification in later life.

As expected, our overall estimates of rates of ethnoracial fluidity differ from previous studies considering earlier or later birth cohorts. Comparisons of our estimated rates of ethnoracial fluidity to those by Liebler et al. (2017) for the most

⁶ Additionally, we analyzed agreement between an individual's race in the 1940 Census, which was classified by a Census enumerator, and the self-identified race on the Social Security Applications. See Section A.1 for full analysis.

directly comparable age category (ages 45+) show slight disagreement: those who originally identified as White are less likely in our data to shift identities (0.8% vs. 1.8%), while those who originally identified as Black are more likely to shift identities (2.3% vs. 1.3%). Both estimates are substantially lower than the Dahis et al. (2019) estimates of 6.8% to 9.9% of Americans classified as Black being classified as White ten years later in the next decennial Census. This suggests that rates of fluidity depend on each cohort's unique experience, some of whom lived through times of heightened identity. For example, a shift to a White identification may have bestowed greater advantage to young adults during the Jim Crow era, which may explain the higher rates observed by Dahis et al. (2019). The birth cohorts studied in this paper also experienced Jim Crow as young adults, but we study them at older ages, after they have experienced the Civil Rights era and the racial and ethnic pride social movements of the last decades of the 20th century. Their cohort-specific experiences across the life course may have further entrenched and reified their views on race by the time they reached older ages. Despite these differences, the overall patterns of shifts echo those of previous work, supporting the reciprocal relationship between social position and ethnoracial identification in the U.S. (Saperstein and Penner, 2012).

Importantly, our results indicate that ethnoracial identification does not solidify completely at older ages. This is the strongest evidence to date that instrumental reasons, such as responses to affirmative-action policies or identifying with higher-status ethnoracial group to benefit from economic, political, or social privileges afforded to that group, are only one of several driving factors of shifts in ethnoracial identification (Francis and Tannuri-Pianto, 2012; Antman and Duncan, 2015). The incentives for shifting identification are likely lower after individuals have departed from the labor and marriage market. Rather, our research suggests it is that the confluence of an individual's socioeconomic status and perceived ethnoracial identity continue to shape the likelihood of shifting ethnoracial identification in later life. Further, while much of ethnoracial identity formation emphasizes the early-life course (childhood, adolescence, and young adulthood) (Rivas-Drake et al., 2014; Umaña-Taylor et al., 2014), our results show that for some individuals, ethnoracial identity formation can be a lifelong process that continues throughout the most advanced ages. This may also reflect shifts in the boundaries separating ethnoracial groups over time (Loveman and Muniz, 2007; Alba, 2005). These findings also have implications for researchers studying these cohorts. While ethnoracial identification for Blacks and Whites is highly stable, movement across ethnoracial categories due to selective attrition for other ethnoracial groups will systematically downwardly bias estimates of socioeconomic status for non-Whites observed for these cohorts. In sum, we argue that late-life ethnoracial fluidity is both real and contextual, as the potential for re-imagining one's ethnoracial identification and the incentives, opportunities, and constraints for doing so depend on the specific historical context.

There are several limitations of this study that warrant discussion. First, our analysis is restricted to individuals in the Numident file who submitted two or more Social Security applications after 1983. The most common reason for submitting an application form is a lost Social Security card (Puckett, 2009), which is likely not related to a shift in identification. Other common reasons for filling out a second

application form include corrections or changes to name, sex, or date of birth. If shifts in race and ethnicity coincide with other shifts in identification, it is possible that our sample may slightly overrepresent individuals who shift identification. Second, the ethnoracial category that is checked on a Social Security application form reflects the public presentation of one's ethnoracial identity, which may differ from their internal self-identity (Davenport, 2020). In the context of Social Security applications, people are trying to work within a bureaucratic system to get something they need. With this goal in mind, people may try to put down the same race that they did for their last application in order to maximize their success of receiving a new Social Security card. If this were the case, our estimates would be lower than rates of ethnoracial fluidity in other settings.

Third, any estimate of the rate of ethnoracial fluidity from censuses or surveys inherently depends on the ethnic and racial categories available on the form or survey. For instance, the Social Security application form didn't differentiate between race and Hispanic ethnicity. Moreover, there are no multiracial identifiers in the data, a limitation shared by much of the previous work in this area.⁷ Fourth, there may be slight selection by race and socioeconomic status by who lives long enough to submit a Social Security application. Finally, it is possible that some of the Social Security application forms were filled out by a proxy, such as a family member or home health aide.⁸ However, the age-specific trends within the focal birth cohorts show that fluidity rates are mostly constant, and do not indicate any systematic difference between the youngest old and oldest old, suggesting this has little impact on the observed rates of ethnoracial fluidity.

We see several avenues for future work. We note that a direct comparison between our study and others is challenging because of differential lengths of observation window, different ethnoracial coding schema, and data limitations. An important avenue for future research is teasing apart these potential explanations for the high amount of relative variation in ethnoracial fluidity found across studies. Additionally, estimating the effect of selective attrition on measures of attainment could shed light on the extent of the bias it introduces. The geographic variation in ethnoracial shifts could be quantified using the rich geographic information in the Numident records. Future work could also explore the effect of social origins, such as social status of parents, on ethnoracial shifting. The puzzling lack of association between socioeconomic status and shifts toward a White identification for those who initially identified as Asian could be further investigated. Most importantly, a more complete examination of ethnoracial fluidity over the life course and across birth cohorts will shed light on historical change in opportunity structures and racialized institutions in the U.S.

⁷ An estimated 3.1–6.6% of the population would mark multiple races in the context of a survey or census form in 2000 (Goldstein and Morning, 2000).

⁸ The Social Security SS-5 form (Figure 1) includes a question about the relationship between the applicant and the person who filled out the SS-5 Social Security form (e.g., self, parent or guardian, etc.). This variable was not available in the digitized and publicly released data for our observation window.

Appendices

A: Additional Analyses

In our main analysis of rates of ethnoracial fluidity, we include foreign-born individuals in our sample. In Table 4, we restrict to those born in the U.S. Our estimates of rates of ethnoracial fluidity are highly comparable to our estimates including the foreign born (reported in Table 3) with one exception: rates of ethnoracial fluidity are significantly higher for those who originally identified as Asian.

A.1: Ethnoracial shifts between 1940 Census and Social Security applications

Table 5 shows the rates of shifts between the enumerator-classified race in the 1940 Census and the self-identified ethnoracial identification in the Social Security applications. We limit this analysis to the Black and White categories to ensure comparability between the Census and Social Security application ethnoracial coding schema. While ethnoracial identification and classification are highly stable for those who initially identified as White (98%), there is less stability for those who initially identified as Black (84.1%). The lower rates of agreement for those who originally identified as Black likely reflects a methodological limitation of this analysis: estimates of rare events, such as shifts in racial identification, are particularly sensitive to false matches. In similar settings, the false match rate was 10% or higher (Abramitzky et al., 2019). We conclude the higher rates of fluidity for those who originally identified as Black is likely driven by false matches upwardly biasing rates of racial fluidity.

A.2: Logistic Regressions

We fit a series of logistic regression models to investigate the relationship between social position and a shift from a non-White to a White identification. The models adjust for gender and birth year to account for gender differences and the different years for which we observe each birth cohort. Figure 4 visualizes the odds ratios of these regression models.⁹ Across most ethnoracial groups, higher education and higher income are associated with an increased likelihood of shifting toward a White identification. The model-based results presented here closely mirror the bivariate associations shown in Figure 3 (see Table 6 for representativeness of the matched sample).

⁹ We note that there is likely a small proportion of false matches between the Numident records and the 1940 Census. False matches introduce measurement error that results in attenuation bias in regression models (Ruggles et al., 2018; Abramitzky et al., 2019). Therefore, our estimates of the association between social status and shifts in identification are likely underestimated. However, our estimates of racial fluidity draw on the full Numident sample and are not subject to bias due to the false matches.

Table 4 Rates of response change in ethnoracial identification between first and last Social Security application for individuals born in the U.S.

Identification first SS App	Identification last SS App				
	American Indian	Asian	Black	Hispanic	White
American Indian	87.0% (3373)	0.4% (14)	2.9% (114)	0.7% (29)	8.9% (346)
Asian	0.4% (16)	65.0% (2727)	2.1% (90)	1.6% (69)	30.8% (1291)
Black	0.1% (114)	0.1% (93)	98.8% (105111)	0.2% (256)	0.7% (767)
Hispanic	0.2% (40)	0.2% (53)	1.2% (304)	96.5% (25146)	2.0% (519)
White	0.2% (444)	0.5% (1246)	0.3% (723)	0.5% (1205)	98.5% (230987)

Numbers in parentheses show raw counts. The main diagonal shows response stability for each ethnoracial group

Table 5 Shifts between enumerator-classified race in the 1940 Census and the self-identified race on the most recent Social Security application

Racial classification: 1940 Census	Racial self-identification: most recent social security application	
	Black	White
Black	84.7% (16823)	15.3% (3044)
White	1.3% (1007)	98.7% (74604)

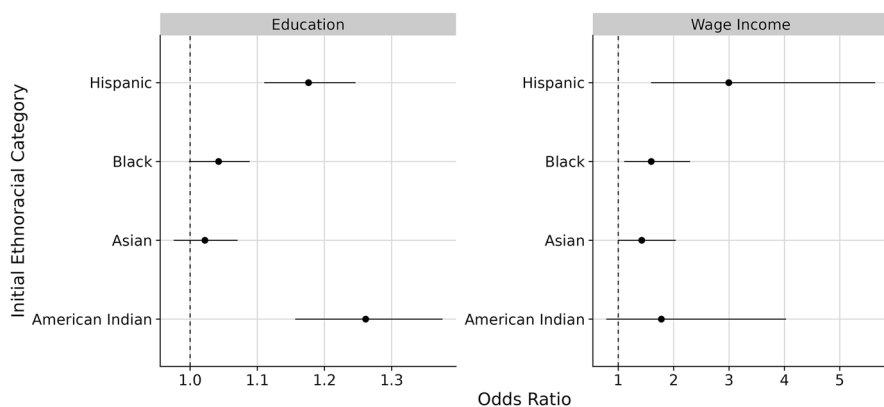
**Fig. 4** Odds ratios from eight separate logistic regression models predicting a shift from one of four non-White identifications on the first application toward a White identification on the last application. For each race, separate models were fit for education and wage and salary income. All models include gender and birth cohort fixed effects

Table 6 Demographic characteristics of individuals in the Numident sample (N = 448,827) that were successfully matched (N = 140,695) and unmatched (N = 308,132) to the 1940 Census

	Matched to census		Unmatched to census	
	No.	%	No.	%
Gender				
Female	75442	53.6	164743	53.5
Male	65253	46.4	143389	46.5
Birth year				
1901–1909	16299	11.6	31116	10.1
1910–1919	51965	36.9	116818	37.9
1920–1927	72431	51.5	160198	52.0
First ethnoracial self-identification				
American Indian	1148	0.8	2980	1.0
Asian	1290	0.9	24127	7.8
Black	33833	24.0	76047	24.7
Hispanic	4181	3.0	48874	15.9
White	100243	71.2	156104	50.7
Last ethnoracial self-identification				
American Indian	1108	0.8	2894	0.9
Asian	1294	0.9	24294	7.9
Black	33832	24.0	76023	24.7
Hispanic	4071	2.9	47865	15.5
White	100390	71.4	157056	51.0

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abramitzky, R., Boustan, L. P., & Eriksson, K. (2012). Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration. *American Economic Review*, 102(5), 1832–1856.
- Abramitzky, R., Boustan, L. P., & Eriksson, K. (2014). A nation of immigrants: Assimilation and economic outcomes in the age of mass migration. *Journal of Political Economy*, 122(3), 467–506.
- Abramitzky, R., Boustan, L. P., Eriksson, K., Feigenbaum, J., & Pérez, S. (2019). *Automated linking of historical data*. Technical Report w25825 National Bureau of Economic Research Cambridge, MA.
- Alba, R. (2005). Bright vs. blurred boundaries: Second-generation assimilation and exclusion in France, Germany, and the United States. *Ethnic and Racial Studies*, 28(1), 20–49.
- Alba, R. (2020). *The great demographic illusion: Majority, minority, and the expanding American mainstream*. Princeton University Press.

- Alba, R. D., & Nee, V. (2009). *Remaking the American mainstream: Assimilation and contemporary immigration*. Harvard University Press.
- Antman, F., & Duncan, B. (2015). Incentives to identify: Racial identity in the age of affirmative action. *Review of Economics and Statistics*, 97(3), 710–713.
- Breen, C. F., & Goldstein, J. R. (2022). Berkeley unified numident mortality database: Public administrative records for individual-level mortality research. *Demographic Research*, 47, 111–142.
- Dahis, R., Nix, E., & Qian, N. (2019). *Choosing racial identity in the United States, 1880-1940*. Technical Report w26465 National Bureau of Economic Research Cambridge, MA.
- Davenport, L. (2020). The fluidity of racial classifications. *Annual Review of Political Science*, 23(1), 221–240.
- Doyle, J. M., & Kao, G. (2007). Are racial identities of multiracials stable? Changing self-identification among single and multiple race individuals. *Social Psychology Quarterly*, 70(4), 405–423.
- Duncan, B., & Trejo, S. J. (2011). Tracking intergenerational progress for immigrant groups: The problem of ethnic attrition. *American Economic Review*, 101(3), 603–608.
- Finlay, K., & Genadek, K. R. (2021). Measuring all-cause mortality with the census numident file. *American Journal of Public Health*, 111(S2), S141–S148.
- Francis, A. M., & Tannuri-Pianto, M. (2012). Using Brazil's racial continuum to examine the short-term effects of affirmative action in higher education. *Journal of Human Resources*, 47(3), 754–784.
- Golash-Boza, T. (2006). Dropping the hyphen? Becoming Latino(a)-American through racialized assimilation. *Social Forces*, 85(1), 27–55.
- Goldstein, J. R., & Morning, A. J. (2000). The multiple-race population of the United States: Issues and estimates. *Proceedings of the National Academy of Sciences*, 97(11), 6230–6235.
- Goldstein, J. R., Alexander, M., Breen, C. F., Miranda-González, A., Menares, F., Osborne, M., & Yildirim, U. (2021). CenSoc Mortality File: Version 2.0.
- Lichter, D. T., & Qian, Z. (2018). Boundary blurring? Racial identification among the children of interracial couples. *The Annals of the American Academy of Political and Social Science*, 677(1), 81–94.
- Liebler, C. A., Porter, S. R., Fernandez, L. E., Noon, J. M., & Ennis, S. R. (2017). America's churning races: Race and ethnic response changes between census 2000 and the 2010 census. *Demography*, 54(1), 259–284.
- Loveman, M., & Muniz, J. O. (2007). How Puerto Rico became White: Boundary dynamics and intercensus racial reclassification. *American Sociological Review*, 72(6), 915–939.
- Mora, G. C. (2014). *Making Hispanics: How activists, bureaucrats, and media constructed a New American*. University of Chicago Press.
- Nagel, J. (1995). American Indian ethnic renewal: Politics and the resurgence of identity. *American Sociological Review*, 60(6), 947.
- Puckett, C. (2009). The story of the social security number. *Social Security Bulletin*, 69(2), 21.
- Rivas-Drake, D., Syed, M., Umaña-Taylor, A., Markstrom, C., French, S., Schwartz, S. J., Lee, R., & Ethnic and Racial Identity in the 21st Century Study Group. (2014). Feeling good, happy, and proud: A meta-analysis of positive ethnic-racial affect and adjustment. *Child Development*, 85(1), 77–102.
- Ruggles, S., Fitch, C. A., & Roberts, E. (2018). Historical census record linkage. *Annual Review of Sociology*, 44(1), 19–37.
- Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J., & Sobek, M. (2020). IPUMS USA: Version 10.0 [Dataset]. Minneapolis, MN: IPUMS. <https://doi.org/10.18128/D010.V10.0>.
- Saperstein, A., & Gullickson, A. (2013). A "Mulatto Escape Hatch" in the United States? Examining evidence of racial and social mobility during the Jim Crow Era. *Demography*, 50(5), 1921–1942.
- Saperstein, A., & Penner, A. M. (2012). Racial fluidity and inequality in the United States. *American Journal of Sociology*, 118(3), 676–727.
- Saperstein, A., Penner, A. M., & Light, R. (2013). Racial formation in perspective: Connecting individuals, institutions, and power relations. *Annual Review of Sociology*, 39(1), 359–378.
- Scott, C. G. (1999). *Identifying the Race or Ethnicity of SSI Recipients* (p. 12).
- Umaña-Taylor, Adriana J., Stephen M. Quintana, Richard M. Lee, William E. Cross, Deborah Rivas-Drake, Seth J. Schwartz, Mo'in Syed, Tiffany Yip, Eleanor Seaton and Ethnic and Racial Identity in the 21st Century Study Group. (2014). Ethnic and Racial Identity During Adolescence and Into Young Adulthood: An Integrated Conceptualization. *Child Development*, 85(1), 21–39.