SOCIAL SCIENCES

Childhood cross-ethnic exposure predicts political behavior seven decades later: Evidence from linked administrative data

Jacob R. Brown[1], Ryan D. Enos[1]*, James Feigenbaum[2], Soumyajit Mazumder[1]

Does contact across social groups influence sociopolitical behavior? This question is among the most studied in the social sciences with deep implications for the harmony of diverse societies. Yet, despite a voluminous body of scholarship, evidence around this question is limited to cross-sectional surveys that only measure short-term consequences of contact or to panel surveys with small samples covering short time periods. Using advances in machine learning that enable large-scale linkages across datasets, we examine the long-term determinants of sociopolitical behavior through an unprecedented individual-level analysis linking contemporary political records to the 1940 U.S. Census. These linked data allow us to measure the exact residential context of nearly every person in the United States in 1940 and, for men, connect this with the political behavior of those still alive over 70 years later. We find that, among white Americans, early-life exposure to black neighbors predicts Democratic partisanship over 70 years later.

INTRODUCTION

The social and behavioral consequences of ethnic diversity are implicated in the long-term success of diverse societies and, consequently, are among the most important and long-standing topics across the social sciences. Scholars have argued that ethnic diversity leads to social inefficiencies, including discriminatory behavior (*1*–*3*) and, in the aggregate, social and political instability (*4*–*6*). Intergroup conflict may have been a crucial selective pressure in human evolution and is a nearly universal feature of human societies (*7*, *8*), leading to a "liberal dilemma" (*9*) of an association between diversity and social inefficiency in an increasingly diversifying world.

As an antidote to this dilemma, scholars have long argued that interpersonal relationships across social groups can reduce prejudice and lead to more aggregate harmony (*10*–*12*), especially when this contact occurs during adolescence. These claims from psychology and other fields have been implicated in consequential jurisprudence [e.g., (*13*)] and public policies (*14*) and are among the most publicly influential social scientific theories (*15*).

Yet, despite the important implications and voluminous research, there are severe limitations to the evidence on the effect of inter-ethnic contact on long-term outcomes (*16*). To date, studies have relied almost entirely on experiments where the outcome is measured nearly immediately or a matter of days after the treatment, longitudinal surveys over short time periods, or cross-sectional surveys where the persistence of the association cannot be adequately measured (*15*, *17*, *18*), thus limiting the scientific and policy relevance of the findings, and the long-term association between inter-ethnic contact and social harmony remains obscured.

To understand the long-term effects of intergroup contact, we undertake the first large-scale linking of full population, individual-level administrative data on adolescent inter-ethnic contact with records of later sociopolitical behavior. In doing so, we are able to construct a dataset that captures a number of early-life experiences, including exact residential context, of nearly every child living in the United States in 1940 and, for those still living and registered to vote, to match this context to political behavior over six and, in many cases, seven decades later. This linking yields a dataset of over 650,000 individuals who were living in nearly every U.S. county in 1940 and for whom we can observe political behavior much later in life. With these linkages, we are able to observe substantial individual variation in racial diversity, socioeconomic status, and other social indicators during childhood and link this variation to later individual political behavior, allowing us to test for the long-term relationship between political behavior and having a different race neighbor in 1940.

Using an empirical strategy of increasingly fine-grained geographic comparisons to account for sorting at small levels of geography, we show that white men who had a black neighbor in 1940, compared to white men who did not, are more likely to be associated with racially liberal politics, as indicated by their registration with the Democratic Party even as late as 2017. This relationship persists even when comparing whites living in the same neighborhood but with different levels of cross-racial exposure. We are further able to stratify our sample by individuals with like-age neighbors, by residential history, and by other criteria indicating that the likely mechanism for this association is the influence of exposure to a neighbor of a different race rather than transmission of attitudes through parents' political attitudes or other unobserved variables.

In the U.S. context, psychologists have argued that partisanship and racial attitudes are conditioned early in life and tend to be stable throughout a person's life span (*19*, *20*), with attitudes on inter-ethnic tolerance more strongly predicted by early, rather than later, life environments (*21*). Cross-ethnic exposure, including early in life, has been shown to liberalize short-term sociopolitical attitudes (*6*, *22*–*24*) perhaps because contact reduces the salience of group-based categories (*25*) or because positive experiences with individuals are generalized to the out-group (*26*). The long-term effect of these early-life experiences on racial liberalism and spillover into other political attitudes could explain why racial attitudes and partisanship are so highly correlated in the United States (*27*) and other

[1]Institute for Quantitative Social Science and Department of Government, Harvard University, 1737 Cambridge Street, Cambridge, MA 02138, USA. [2]Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, USA.
*Corresponding author. Email: renos@gov.harvard.edu

countries (*28*), with members of left-of-center parties consistently displaying less ethnic and racial prejudice than members of right-of-center parties (*29*). In the United States, anti-black prejudice was a major driver of the sorting of voters into parties in the mid-20th century (*30*, *31*) and remains among the strongest predictors of vote choice in recent elections (*32*–*34*).

The long-term endurance of these effects has important implications: If racial exposure is associated with behavior over seven decades later, especially on a characteristic with as much overtime stability as partisanship, it may be that the influence of these early-life experiences is also present in intervening decades, when people are active in politics and the workforce, and that the influence may be present in a range of sociopolitical behaviors and attitudes (*35*). Partisanship can be characterized as a social identity (*36*) that is closely tied to a person's self-image and, hence, is a powerful predictor of behavior (*37*). Party membership has been shown to induce a range of behaviors, including the type of group-based bias that characterizes race-based social identities (*38*). Partisanship in the United States is among the strongest predictors of nearly every political attitude (*39*) and many nonpolitical lifestyle choices (*40*, *41*), and thus, long-term influences on partisanship have sweeping implications.

Linking 1940 Census data to contemporary voter files

Census data

We draw data on early-life experiences from the 1940 U.S. Decennial Census, which attempted to record every person living in the United States in that year and likely reached around 99% of the population (*42*), including noncitizens. Enumerators recorded information on name, age, gender, race, place of birth, years of education, labor earnings, employment status, and many other characteristics. Individual census records are not available when initially recorded but are made available for public use in accordance with a statutory 72-year restriction to protect the privacy of respondents. The census data we draw on have been transcribed and organized into structured datasets through collaborative efforts of the Integrated Public Use Microdata Series (IPUMS) at the Minnesota Population Center (MPC), Ancestry.com, and FamilySearch.org. We access de-anonymized data through files deposited by MPC at the National Bureau for Economic Research. For 131,903,910 of the 132,164,569 individuals enumerated in the 1940 Census, we observe first and last names, as well as ages and places of birth, which we use to match to contemporary voter data.

Voter file data

In most U.S. states, citizens must register to be able to vote. When doing so, they usually provide basic information, including their full name, address, and date of birth. In most, but not all, states, voters also declare a party affiliation when registering. We use party registration in our analysis to measure psychological association with a party. Party registration has been shown to accurately measure self-reported association with a party, as well as policy and other beliefs (*43*). Because we can directly observe party registration using these administrative data, we largely avoid problems of measurement error associated with the surveys common in this field of study (*44*).

In three states, California, North Carolina, and Nebraska, citizens provide their place of birth when registering. This information facilitates very accurate matches, and so we focus on people currently living in those states. This population, however, was living in nearly every county in the United States in 1940 (3088 of 3108 counties, 99.4%), providing almost blanket coverage of U.S. locations during the time period. We measure contemporary partisan registration at two different points in time by pooling data from California and North Carolina in 2005 and 2009, respectively, and again to these states and Nebraska in 2017. The 2005/2009 data come directly from the state governments, while the 2017 data are aggregated by the commercial vendor L2. The inclusion of samples at two different points in time allows us to test the robustness of the results to different datasets and states and across different points in time. Because attrition via death and other causes will change the composition of the sample between 2005 and 2017, measurement at different time points also allows us to understand the influence of attrition on our inferences.

Record linkage

Many of the methods for record linkage rely on methods based on the matching of available variables, most commonly name, gender, race, and age. These methods are nonstatistical in the sense that they do not rely on a probabilistic model. Recent advances in the record linkage methodology argue for the use of probabilistic algorithms such as support vector machines (*45*), regression (*46*), and Bayesian inference (*47*). The core advantage of probabilistic models is that they allow the researcher to directly control the false-positive rate by tuning the acceptance threshold for match probabilities.

To construct our linked sample, we use the supervised machine learning procedure developed by (*46*). The key strength of this algorithm is that it allows the researcher to take advantage of the value of human coders by generating a hand-linked training set. Census data from 1940 were recorded in cursive by enumerators and transcribed 72 years later, introducing many layers of measurement error. Humans are quite good at comparing lists of names and making judgments across imperfect links, implicitly weighing differences in first and last names, as well as differences in year of birth or implied age. The machine learning approach makes those implicit weights on various record link features explicit and then replicably applies those weights to link very large datasets at scale.

We restrict our searches to men since the common practice of surname changes at marriage during the 20th century makes it quite difficult to link women accurately. We then further subset the data to examine possible matches from among those who were similar on first name, last name, and age, blocking on state of birth by looking at all men with matching states of birth, born within 5 years of the record in the voter file, and with a first and last name within 0.3 in Jaro-Winkler string distance. A research assistant then attempts to link a random sample of the data by hand. We estimated a statistical model to predict which records the human linked and which records the human did not link, using a number of constructed features or variables based on the first and last name strings and the years of birth. Last, we tuned a pair of hyperparameters to convert probit scores into matching rule decisions; the hyperparameters govern how strong a given match is both absolutely and relative to the next best alternative match. Trained on a small subset of the data, we then applied the linking algorithm to the rest of the data, creating matches from the voter files to the census (see Materials and Methods).

Overall, our linkage method performed quite well. The match rates—the shares of voter records of adequate age that we are confident in linking to a record in 1940—for California, Nebraska, and North Carolina were 52%, 46%, and 65%, respectively, for a sample of 672,318 individuals in 2005/2009 and 259,762 in 2017 (760,337 unique individuals) who were living across the United States in 1940 (Fig. 1). Almost half (46.6%) were living in California,

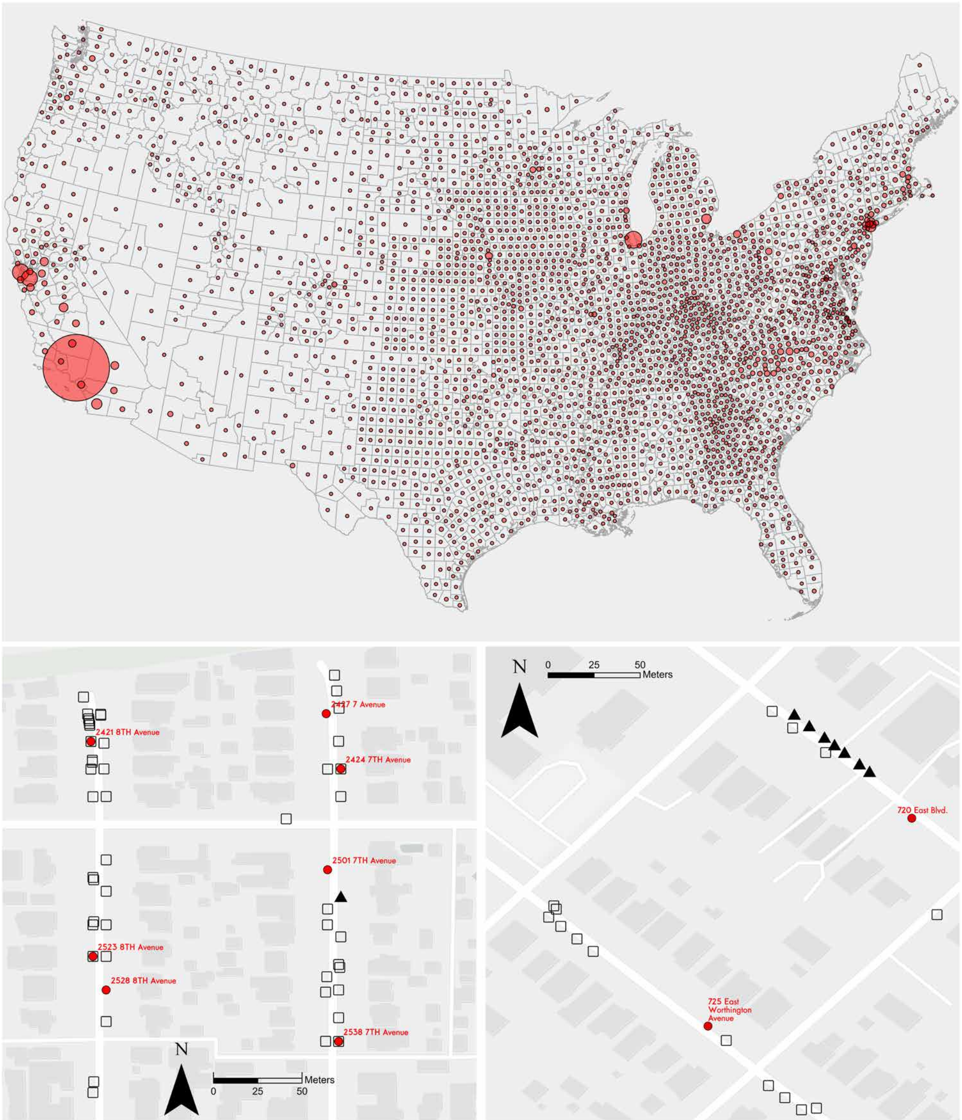

Fig. 1. Maps of linked sample and neighbors in 1940. Top panel is the counties represented in 1940 in the linked sample, with dots scaled by population represented in the contemporary sample. Bottom left panel is a neighborhood in Central Los Angeles in 1940. Bottom right panel is a neighborhood in Charlotte, NC in 1940. In these figures, red squares represent the white households found in the linked sample, with addresses labeled. Black triangles represent black households, and squares represent all other households. On the left, the subject at 2501 7th Avenue has a black neighbor, while subjects on 8th Avenue do not. On the right, the subject on East Boulevard has several black neighbors, while the subject on Worthington Avenue has none.

Nebraska, or North Carolina, the states in which we examine contemporary behavior, but the mid-century population flows in the United States are also evident, with the 1940 population also living in major metropolitan areas in the industrial Midwest and Northeast, such as Chicago, Detroit, and New York, from which they moved to California or North Carolina at a later time. We restrict the study to whites, yielding 618,712 subjects in 2005/2009 and 238,353 in 2017 (699,554 unique individuals).

In 1940, 3.75% of the sample had an immediate black neighbor. Because we can observe the exact location of individuals in 1940, we can distinguish between fine-grained differences in cross-ethnic exposure, even between people with different cross-ethnic exposure living in the same neighborhood (Fig. 1). The proportion of registered Democrats in the linked sample, around 42%, is quite close to the proportion of individuals today who identify with Democrats from these two states and were born before 1940 (based on authors' calculations from the 2016 Cooperative Congressional Election Study; see the Supplementary Materials). A large majority of the sample were younger than 10 years of age in 1940, and nearly the entire sample was younger than 20 years old (fig. S1). In the Supplementary Materials, we provide further descriptive statistics about our sample, documenting that our linked sample appears to be quite representative of the population of white children in 1940.

Measuring cross-ethnic exposure

To measure cross-ethnic exposure, we build on methods for estimating segregation developed by (*48*), exploiting the ordering of households on the 1940 U.S. Census enumeration sheets. When compiling the census, enumerators in 1940 went door to door down the street, recording information for each house. As a result, adjacent households are very likely to appear immediately next to each other on census pages, and we are able to identify the characteristics of an individual's immediate next-door neighbor, as well as the characteristics of their neighbors two doors down, three doors down, and so on. With this information, we construct measures of racial exposure for each individual in our dataset, specifically whether each individual had a neighbor of another race and their relative proximity to said neighbor.

The key assumptions for our measures of racial exposure are that proximity in census enumeration is a proxy for residential proximity in actual geography, and that geographic proximity serves as a reasonable proxy for racial exposure. This assumption is supported by evidence that census enumerators were exhaustive, in that they visited every recorded household, and that they visited and recorded households per their geographic ordering (*49*, *50*). Households geographically next door to each other are very likely to be recorded next to each other on census manuscript pages. Thus, as we move up or down the census reel away from an individual in our dataset, we locate increasingly more distant neighbors. We also assume that geographic distance moving up and down the census manuscript page is the same so that, say, three spaces up the page is, on average, the same relative position as moving three spaces down the page.

By measuring cross-ethnic exposure on the individual level using administrative records, we avoid having to use survey or aggregate data that can be subject to measurement error, aggregation error, and ecological fallacies (*9*). We construct an indicator variable representing the 10 most proximate pairs of neighbors for each individual in our dataset. Each indicator incorporates two variables because for any given position (closest neighbor, second closest neighbor, …) there are two households that are equally proximate to the individual, depending on whether you look up or down the census page. Hence, for $k \in [1,10]$, each indicator equals 1 if at least one of the two households located k doors down from the individual is inhabited by a head of the household who is black. In the Supplementary Materials, we also present analysis with k equal to 0, 1, or 2 to record the exact number of households with a black head of household out of the two possible households.

We measure the relationship for each $k \in [1,10]$ neighbors in 1940 and 2005/2009 or 2017 party registration by regressing party registration on indicators for a black neighbor as well as variables for the age and education [to measure social status not captured by race (*51*) and to control for this possible confounding with race] of both the subject and each neighbor. We estimate the following model

$$Y_{\mathrm{i,g}} = \alpha + \sum_{k=1}^{10} \beta_k D_{k,i} + \gamma X_i + \lambda_{\mathrm{g}} + \epsilon_{i,g}$$

where $Y_{i,g}$ measures Democratic Party registration in 2005/2009 or 2017 for individual i in geography g; α is the intercept; $D_{k,i}$ is a matrix of neighbor characteristics for neighbor k positions away on the census page—for each of these k neighbors, the matrix includes whether the neighbor is black and whether the head of the household for the neighbor has at least a high school education; and X_{i} is a matrix of individual-level covariates that may affect both residential location and eventual partisan registration: individual age, family size, whether their family had resided in the same residence 5 years previous, and the high school education, income, age, employment status, and hours worked on average per week and per week for the head of the household. λ_g is the geographic fixed effect, and $\epsilon_{\mathrm{i},g}$ is the error term. In the Supplementary Materials, we compare the balance of these covariates across our sample with and without black next-door neighbors. At smaller geographic fixed effects, most covariates are balanced across groups, but we include these covariates as controls to account for potential confounding in all specifications. We also show in the Supplementary Materials that our results are similar in models not including these covariates.

Using fixed effects, we examine this relationship at increasingly small geographic levels, beginning with state, county, and enumeration district fixed effects. Thus, all comparisons are between individuals living in the same geography, allowing us to account for residential sorting at increasingly fine-grained levels. In the Supplementary Materials, we present power analyses across fixed-effect specifications. Enumeration districts are subcounty geographic units defined as the area for which a census enumerator could complete a count of the area's population for a given census year. We then proceed to areas defined by groups of census pages, which capture geographic proximity because adjacent houses were recorded on consecutive page lines by enumerators (*49*, *50*). We define geographies by 10- and 5-page groups, with the smallest group representing a very small geography, approximately several city blocks. Thus, when comparing individuals at these smallest levels of geography, we have increased confidence that the white individuals with and without black neighbors are equal on observable characteristics and the difference in contemporary behavior is a result of exposure to these black neighbors.

RESULTS

Whites with a black next-door neighbor in 1940 are 1.5 percentage points (with district fixed effects) to 4.2 percentage points (with no geographic fixed effects) more likely to be Democrats in 2005/2009, and 2.8 percentage points (with 10-page fixed effects) to 5.3 percentage points (with no geographic fixed effects) more likely in 2017 than whites without a black next-door neighbor (Fig. 2) (results for neighbors $k > 1$ are in Materials and Methods). The relationship is stable within increasingly small geographies and when accounting for other characteristics of the individuals in 1940, and in the Supplementary Materials, we use sensitivity analysis (*52*) to show that the results are very unlikely to be explained by an unobserved confounding variable, suggesting that the relationship is the result of exposure to a black neighbor. In the Supplementary Materials, we also analyze Republican Party registration as the outcome. A black next-door neighbor predicts decreased likelihood of being a Republican in the 2005/2009 and 2017 samples and with the results generally mirroring the Democratic results in size.

The size of this association is similar to that reported by (*53*) for the effect of white children's exposure to school diversity on downstream partisanship and is consistent with recent research showing that partisanship can change in response to local context even for adults (*54*). A growing body of evidence shows that early-life context can have large effects on economic and other outcomes (*55*), and longstanding evidence in political science demonstrates that childhood is when sociopolitical attitudes are most malleable (*20*). Given these baselines, and that previous scholarship has demonstrated that even fleeting cross-ethnic exposure can alter sociopolitical attitudes (*2*, *56*), the size of these effects may be within an expected range for

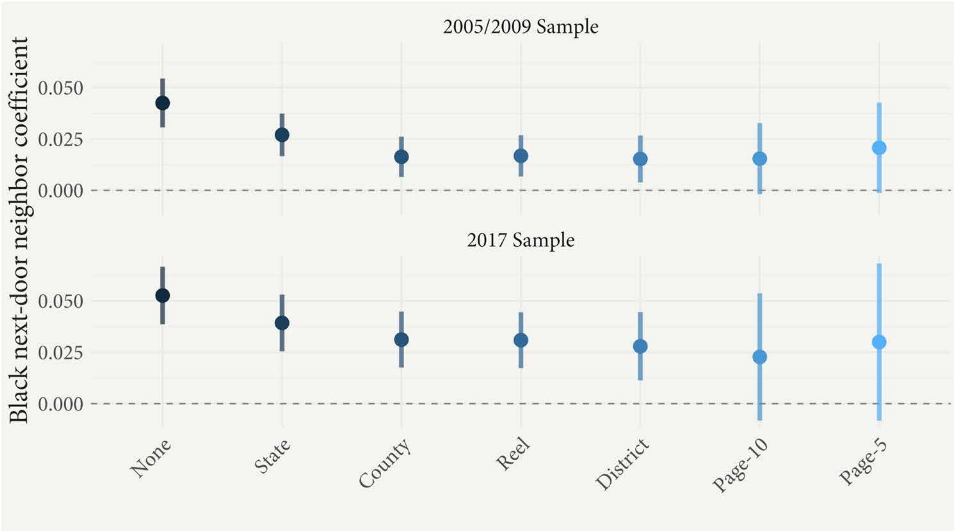

Fig. 2. Black next-door neighbor and Democratic partisanship association. Points represent the coefficient of a black next-door neighbor on Democratic partisanship. Coefficients are from separate specifications with different geographic fixed effects and are displayed in order of largest to smallest geographic comparison. Standard errors are clustered at the county level. Controls include individual age, family size, whether their family had resided in the same residence 5 years previous, as well as the high school education, income, age, employment status, and hours worked on average per week and per week for the head of the household, and whether or not each neighbor has a head of the household with at least a high school education. Differences between sample size in covariate models and full sample are due to missingness across covariates.

the treatment on children, given the intense interpersonal contact possible when interacting with neighbors.

The relationship among nonmovers, non-Southerners, and urban dwellers

We examine several alternative explanations, other than cross-ethnic contact, for the association between childhood exposure and long-term party registration. First, the relationship could be driven by white people with racially progressive attitudes who selected to live next to black people. Because we can compare individuals who lived within the same neighborhood, this geographic sorting is unlikely to drive the results, but to further investigate this possibility, we estimate specifications fitted to the subset of our sample that has lived in their same residence for at least 5 years as recorded in the 1940 Census.

Second, in the first half of the 20th century in the southern United States, racially conservative whites tended to be Democrats and were clustered around areas with large black populations (*57*), and thus, the results may reflect the persistence of this old-fashioned Democratic Party identification rather than racial liberalism. Hence, we subset our data by residents of the South and non-South in the 1940s, where residents of the former may reflect racial conservatism, while the latter reflects racial liberalism. The expectations for the direction of the effect among people living in the South in the 1940s are ambiguous because racially conservative whites may have switched party registration to Republican later in life (*30*), while others may have simply not changed their party affiliation to reflect the realignment of the parties on racial issues. Thus, we examine results for non-Southerners in 1940, for whom we should expect the association to more clearly reflect racial liberalism.

Last, because distances between neighbors will vary based on population density, and this likely influences frequency of contact between neighbors, we compare individuals in our samples living in urban areas in 1940 to those living in rural areas. The 1940 Census classifies a household as being in an urban area if it is located in a city or incorporated place of at least 2500 inhabitants.

For each of these three moderating variables, we estimate specifications on subsets defined by each variable (Fig. 3). Across both samples, the predictive effect of a black next-door neighbor remains positive and significant in each of these subsamples.

In addition, in the Supplementary Materials, we also show that the result is stable if we limit the sample to individuals who live next door to neighbors with children similar in age to the individual, consistent with the mechanism of cross-ethnic interaction producing durable effects. We find that black neighbors in 1940 are not predictive of living in neighborhoods with larger black populations later in life (fig. S17). This indicates that the connection between early-life context and future party registration is more likely a product of the lasting impression left by cross-ethnic contact at a young age rather than a product of greater long-term exposure to diversity.

In the Supplementary Materials, we also show analysis aimed to further test the potential for parents' partisanship to confound the results, wherein more racially liberal white parents are more likely to move next to black neighbors and more likely to produce children who will become Democrats. While we cannot measure parents' partisanship directly with our data, we obtained summary data used in (*58*) detailing the proportion of Democrats in 105 different occupations in California in 1940. We merge these data with the Census-recorded occupations of the head of the household for each individual in our dataset and use this within-occupation proportion Democrat as a control. The results from this alternative specification are shown in comparison to the main effects in the Supplementary Materials, demonstrating that the inclusion of this proxy for parents' partisanship does not substantively change the results.

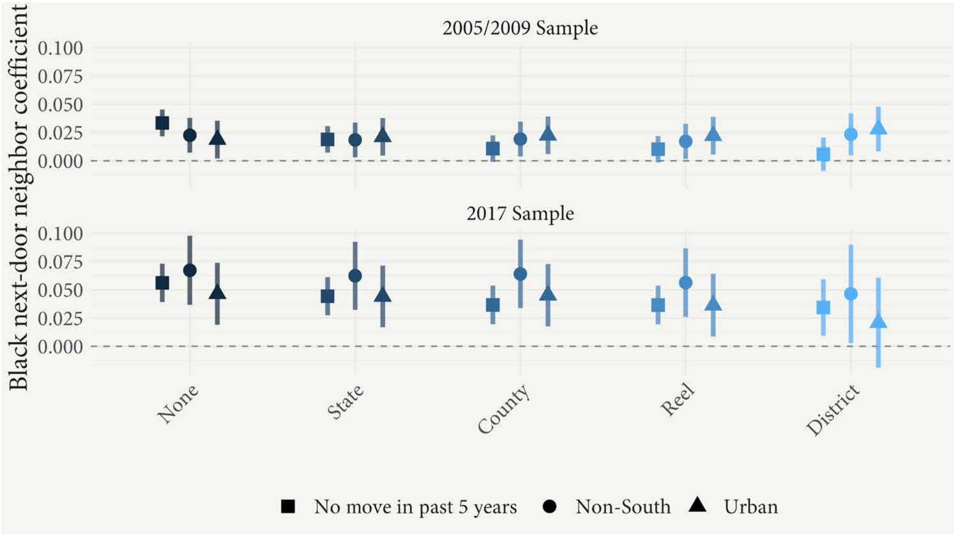

Fig. 3. Black next-door neighbor association on subsamples. Points represent the coefficient of the black next-door neighbor on future Democratic partisanship. Coefficients are from separate specifications with different geographic fixed effects and are displayed in order of largest to smallest geographic comparison. Coefficient estimates are from models fitted to subset of whites in the data who have either not moved in the past 5 years, lived outside the South in 1940, or lived in an urban area in 1940, respectively. Standard errors are clustered at the county level. Controls include individual age, family size, whether their family had resided in the same residence 5 years previous, as well as the high school education, income, age, employment status, and hours worked on average per week and per week for the head of the household, and whether or not each neighbor has a head of the household with at least a high school education.

DISCUSSION

We have demonstrated that cross-ethnic contact early in life is associated with sociopolitical behavior over 60 and 70 years later and that this relationship can be observed even when comparing subjects who lived in the same neighborhood but who had different levels of exposure. Furthermore, the association is consistent across different regions of the country and different levels of density and for people who had and had not recently moved during childhood, and when controlling for a range of characteristics of households, suggesting that socialization early in life has a long-term effect on sociopolitical attitudes.

Most research on contemporary diversity and political behavior uses aggregate data, such as a measure of county-level demographics, to estimate ecological correlations with the assumption that local diversity increases the probability of cross-ethnic contact. There are reasons to believe that a direct measure of exposure, such as the one we use here, is superior to or, at least, distinct from measures of contact taken from aggregate data. Further analysis in the Supplementary Materials supports our finding that one's closest neighbors, rather than aggregate racial context, has the strongest relationship with future party registration. However, the relationship between direct cross-ethnic exposure, early-life aggregate diversity, and future political behavior deserves future investigation, especially since recent research has demonstrated a relationship between racial diversity in elementary (*53*) or high school (*59*) and downstream sociopolitical attitudes for white Americans. Of course, future research examining the nature and valence of these cross-ethnic contacts will help with our understanding.

The relationship between early-life experiences and party registration persists over this long period despite the many intervening life experiences during this period—including the population of study being one of the more economically and geographically mobile populations in American history and with a large portion of the men experiencing the social disruption of military service. This persistence suggests a dominating socializing force of early-life experiences. The long-term persistence of this relationship suggests that, despite the short-term social inefficiencies associated with diversity, there may be long-term positive effects for social harmony.

MATERIALS AND METHODS

The study was reviewed by the Harvard and Boston University Committees on the Use of Human Subjects and was declared not human subject research.

Linking 1940 Census and contemporary voter file data

To link records from the voter file to the census, we train a linking algorithm on hand-coded examples of links. The algorithm uses the following features to build a linking score for each potential match from the voter file to the 1940 Census:

1) Jaro-Winkler string distance in first and last names. The Jaro-Winkler distance metric captures, on a scale from 0 to 1, how many edits have to be made to the characters in one string to convert it to another string. Edits include substitutions, deletions, and additions. Differences in strings earlier in the string are more heavily penalized, leading to larger string distances. Distance in first name and distance in last name are two key features in our linking procedure.

2) Absolute value difference in age in 1940. The 1940 Census asked for respondents' age as of April 1940. We use the dates of birth in the voter file to estimate how old people in our sample should have been in 1940. We calculate the number of years different between a record in the voter file and a record in the 1940 Census. Age or year of birth can be off for many reasons, including simple data entry error, misreporting, and age heaping.

3) Soundex agreement in first and last names. Soundex and other phonex coding schemes attempt to convert strings or words into

Table 1. Democratic partisanship by a black next-door neighbor, 2005/2009 sample. Table shows the coefficients from the main specifications estimating the effect of a black next-door neighbor on future Democratic partisanship for the 2005/2009 sample. Models in this table range from no fixed effects (FE) to district-level fixed effects. Cluster robust standard errors are shown in parentheses, with standard errors clustered at the county level. Models with and without covariates are shown. Both models include variables for black neighbors at each $k \in [1,10]$ position up and down the Census page. Specifications with covariates control for individual age, family size, whether their family had resided in the same residence 5 years previous, as well as the high school education, income, age, employment status, and hours worked on average per week and per week for the head of the household (HoH), and whether or not each neighbor has a head of the household with at least a high school education. Differences between sample size in covariate models and full sample are due to missingness across covariates. *$P < 0.1$; **$P < 0.05$; ***$P < 0.01$.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age		0.001***		0.001***		0.001***		0.001***
		(0.0002)		(0.0002)		(0.0002)		(0.0003)
High school		−0.004*		−0.003		−0.003		−0.005*
		(0.002)		(0.002)		(0.002)		(0.003)
Income		−0.00001***		−0.00000**		−0.00000***		−0.00000**
		(0.00000)		(0.00000)		(0.00000)		(0.00000)
Mover		0.002		0.001		−0.00005		0.001
		(0.002)		(0.002)		(0.002)		(0.002)
HoH age		−0.0001		0.0002**		0.0001		0.0001
		(0.0001)		(0.0001)		(0.0001)		(0.0001)
Homeowner		−0.0003		−0.007***		−0.004**		−0.001
		(0.002)		(0.002)		(0.002)		(0.002)
Family size		0.002***		−0.0003		0.0003		−0.0001
		(0.001)		(0.0005)		(0.0004)		(0.0005)
Employed		0.011***		−0.001		−0.003		0.001
		(0.004)		(0.004)		(0.004)		(0.005)
Hours per week		−0.0001***		−0.0001		−0.0001		−0.00005
		(0.00005)		(0.00004)		(0.00004)		(0.0001)
Weeks per year		0.0002***		0.00001		−0.00002		0.00004
		(0.0001)		(0.0001)		(0.0001)		(0.0001)
Black neighbor ($k = 1$)	0.046***	0.042***	0.028***	0.027***	0.018***	0.016***	0.015**	0.015**
	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.007)
Black neighbor ($k = 2$)	0.018***	0.015***	0.006	0.005	−0.002	−0.002	−0.006	−0.008
	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
Black neighbor ($k = 3$)	0.018***	0.019***	0.008*	0.011**	0.002	0.005	0.004	0.009
	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)	(0.006)
Black neighbor ($k = 4$)	0.014***	0.012**	0.006	0.007	0.003	0.005	−0.001	0.001
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Black neighbor ($k = 5$)	0.018***	0.013**	0.008*	0.005	0.004	0.0002	0.001	−0.004
	(0.005)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)	(0.006)
Black neighbor ($k = 6$)	0.017***	0.015***	0.007	0.006	0.003	0.002	−0.002	−0.003
	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)	(0.006)
Black neighbor ($k = 7$)	0.011***	0.011**	0.003	0.004	0.001	0.002	0.002	0.004
	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)	(0.006)
Black neighbor ($k = 8$)	0.020***	0.020***	0.008**	0.010*	0.001	0.002	−0.001	0.001
	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.006)
Black neighbor ($k = 9$)	0.023***	0.021***	0.011***	0.010**	0.007*	0.007	0.004	0.004
	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)	(0.006)

continued to next page

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Black neighbor ($k = 10$)	0.026***	0.020***	0.008**	0.004	0.001	−0.003	−0.001	−0.005
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
High school– educated neighbor ($k = 1$)		0.0003		0.001		0.001		0.001
		(0.002)		(0.002)		(0.002)		(0.002)
High school– educated neighbor ($k = 2$)		−0.001		−0.001		−0.001		−0.002
		(0.002)		(0.002)		(0.002)		(0.002)
High school– educated neighbor ($k = 3$)		0.0003		0.0004		0.0005		0.00005
		(0.002)		(0.002)		(0.002)		(0.002)
High school– educated neighbor ($k = 4$)		−0.002		−0.002		−0.001		−0.002
		(0.002)		(0.002)		(0.002)		(0.002)
High school–educated neighbor ($k = 5$)		−0.002		−0.002		−0.002		−0.001
		(0.002)		(0.002)		(0.002)		(0.002)
High school– educated neighbor ($k = 6$)		0.001		0.001		0.001		0.00003
		(0.002)		(0.002)		(0.002)		(0.002)
High school– educated neighbor ($k = 7$)		0.001		0.001		0.001		0.001
		(0.002)		(0.002)		(0.002)		(0.002)
High school– educated neighbor ($k = 8$)		−0.00001		−0.0002		−0.00001		−0.001
		(0.002)		(0.002)		(0.002)		(0.002)
High school– educated neighbor ($k = 9$)		−0.002		−0.002		−0.002		−0.002
		(0.002)		(0.002)		(0.002)		(0.002)
High school– educated neighbor ($k = 10$)		−0.001		−0.001		−0.001		0.001
		(0.002)		(0.002)		(0.002)		(0.002)
FE	None	None	State	State	County	County	District	District
Controls	No	Yes	No	Yes	No	Yes	No	Yes
N	609,878	512,558	609,878	512,558	609,878	512,558	609,800	512,493
R^2	0.005	0.005	0.015	0.014	0.025	0.025	0.198	0.215
Adjusted R^2	0.005	0.005	0.015	0.014	0.020	0.019	0.021	0.019

Table 2. Democratic partisanship by a black next-door neighbor, 2005/2009 sample. Table shows the coefficients from the main specifications estimating the effect of a black next-door neighbor on future Democratic partisanship for the 2005/2009 sample. Models in this table range from Census reel fixed effects to 5 Census page fixed effects. Cluster robust standard errors are shown in parentheses, with standard errors clustered at the county level. Models with and without covariates are shown. Both models include variables for black neighbors at each $k \in [1,10]$ position up and down the Census page. Specifications with covariates control for individual age, family size, whether their family had resided in the same residence 5 years previous, as well as the high school education, income, age, employment status, and hours worked on average per week and per week for the head of the household, and whether or not each neighbor has a head of the household with at least a high school education. Differences between sample size in covariate models and full sample are due to missingness across covariates. *$P < 0.1$; **$P < 0.05$; ***$P < 0.01$.

	(1)	(2)	(3)	(4)	(5)	(6)
Age		0.001***		0.001***		0.001***
		(0.0002)		(0.0004)		(0.001)
High school		−0.003		−0.003		−0.006
		(0.002)		(0.004)		(0.005)
Income		−0.00000**		−0.00000		−0.00000
		(0.00000)		(0.00000)		(0.00000)
Mover		−0.0001		0.001		0.001
		(0.002)		(0.003)		(0.003)

continued to next page

	(1)	(2)	(3)	(4)	(5)	(6)
HoH age		0.0001		0.0001		0.00005
		(0.0001)		(0.0002)		(0.0002)
Homeowner		−0.003**		0.001		−0.003
		(0.002)		(0.003)		(0.004)
Family size		0.0002		−0.00004		−0.00004
		(0.0004)		(0.001)		(0.001)
Employed		−0.001		0.001		0.0003
		(0.003)		(0.006)		(0.008)
Hours per week		−0.0001		−0.0001		−0.00004
		(0.00004)		(0.0001)		(0.0001)
Weeks per year		−0.00002		0.0001		0.0001
		(0.0001)		(0.0001)		(0.0001)
Black neighbor ($k = 1$)	0.018***	0.017***	0.016**	0.015*	0.017*	0.021*
	(0.005)	(0.005)	(0.008)	(0.009)	(0.009)	(0.011)
Black neighbor ($k = 2$)	−0.001	−0.001	−0.009	−0.011	−0.006	−0.012
	(0.004)	(0.004)	(0.006)	(0.007)	(0.008)	(0.010)
Black neighbor ($k = 3$)	0.003	0.005	0.006	0.014**	0.006	0.015
	(0.004)	(0.005)	(0.006)	(0.007)	(0.008)	(0.010)
Black neighbor ($k = 4$)	0.003	0.005	0.003	0.005	0.005	0.007
	(0.004)	(0.005)	(0.007)	(0.009)	(0.009)	(0.011)
Black neighbor ($k = 5$)	0.004	0.001	0.0003	−0.003	−0.001	−0.003
	(0.004)	(0.005)	(0.007)	(0.009)	(0.009)	(0.011)
Black neighbor ($k = 6$)	0.003	0.002	−0.002	−0.005	0.001	0.002
	(0.004)	(0.005)	(0.007)	(0.008)	(0.008)	(0.010)
Black neighbor ($k = 7$)	0.001	0.003	0.002	0.002	0.003	−0.0001
	(0.004)	(0.005)	(0.007)	(0.008)	(0.009)	(0.011)
Black neighbor ($k = 8$)	0.001	0.002	−0.001	0.004	−0.002	0.005
	(0.004)	(0.005)	(0.006)	(0.007)	(0.007)	(0.009)
Black neighbor ($k = 9$)	0.007*	0.007	0.008	0.007	0.011	0.009
	(0.004)	(0.005)	(0.007)	(0.008)	(0.009)	(0.010)
Black neighbor ($k = 10$)	0.002	−0.002	−0.003	−0.006	0.002	−0.003
	(0.004)	(0.004)	(0.006)	(0.007)	(0.007)	(0.009)
High school–educated neighbor ($k = 1$)		0.001		0.001		0.001
		(0.002)		(0.004)		(0.004)
High school–educated neighbor ($k = 2$)		−0.001		−0.002		−0.004
		(0.002)		(0.003)		(0.003)
High school–educated neighbor ($k = 3$)		0.0004		0.001		−0.001
		(0.002)		(0.003)		(0.004)
High school–educated neighbor ($k = 4$)		−0.002		0.0003		0.004
		(0.002)		(0.003)		(0.003)
High school– educated neighbor ($k = 5$)		−0.002		−0.001		−0.001
		(0.002)		(0.003)		(0.004)
High school– educated neighbor ($k = 6$)		0.001		−0.0001		−0.003
		(0.002)		(0.003)		(0.004)

continued to next page

	(1)	(2)	(3)	(4)	(5)	(6)
High school– educated neighbor ($k = 7$)		0.001		0.001		−0.0005
		(0.002)		(0.003)		(0.004)
High school– educated neighbor ($k = 8$)		0.0001		−0.0001		−0.003
		(0.002)		(0.003)		(0.004)
High school–educated neighbor ($k = 9$)		−0.002		−0.001		−0.002
		(0.002)		(0.003)		(0.004)
High school– educated neighbor ($k = 10$)		−0.001		0.001		0.003
		(0.002)		(0.003)		(0.004)
FE	Reel	Reel	Page 10	Page 10	Page 5	Page 5
Controls	No	Yes	No	Yes	No	Yes
N	609,878	512,558	609,878	512,558	609,878	512,558
R^2	0.028	0.028	0.372	0.398	0.495	0.520
Adjusted R^2	0.020	0.019	0.023	0.021	0.022	0.021

Table 3. Democratic partisanship by black next-door neighbor, 2017 sample. Table shows the coefficients from the main specifications estimating the effect of a black next-door neighbor on future Democratic partisanship for the 2017 sample. Models in this table range from no fixed effects to district-level fixed effects. Cluster robust standard errors are shown in parentheses, with standard errors clustered at the county level. Models with and without covariates are shown. Both models include variables for black neighbors at each $k \in [1,10]$ position up and down the Census page. Specifications with covariates control for individual age, family size, whether their family had resided in the same residence 5 years previous, as well as the high school education, income, age, employment status, and hours worked on average per week and per week for the head of the household, and whether or not each neighbor has a head of the household with at least a high school education. Differences between sample size in covariate models and full sample are due to missingness across covariates. *$P < 0.1$; **$P < 0.05$; ***$P < 0.01$.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age		0.004***		0.004***		0.004***		0.005***
		(0.0003)		(0.0003)		(0.0003)		(0.001)
High school		−0.021**		−0.019**		−0.018**		−0.012*
		(0.008)		(0.009)		(0.009)		(0.007)
Income		−0.00001***		−0.00000**		−0.00001***		−0.00000**
		(0.00000)		(0.00000)		(0.00000)		(0.00000)
Mover		−0.011***		−0.007**		−0.007**		−0.010**
		(0.003)		(0.003)		(0.003)		(0.004)
HoH age		−0.0001		0.0003*		0.0001		0.0001
		(0.0002)		(0.0002)		(0.0002)		(0.0002)
Homeowner		−0.022***		−0.032***		−0.028***		−0.018***
		(0.006)		(0.005)		(0.005)		(0.005)
Family size		0.011***		0.009***		0.009***		0.007***
		(0.002)		(0.002)		(0.002)		(0.002)
Employed		−0.020**		−0.027***		−0.029***		−0.026***
		(0.008)		(0.007)		(0.007)		(0.009)
Hours per week		−0.0005***		−0.0004***		−0.0003***		−0.0003***
		(0.0001)		(0.0001)		(0.0001)		(0.0001)
Weeks per year		−0.0004***		−0.001***		−0.001***		−0.0004**
		(0.0001)		(0.0001)		(0.0001)		(0.0002)
Black neighbor ($k = 1$)	0.060***	0.053***	0.043***	0.039***	0.034***	0.031***	0.027***	0.028***
	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)	(0.008)	(0.010)

continued to next page

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Black neighbor ($k = 2$)	0.021***	0.013*	0.010	0.004	0.003	−0.002	−0.004	−0.003
	(0.006)	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)	(0.008)	(0.010)
Black neighbor ($k = 3$)	0.020***	0.018**	0.011	0.010	0.006	0.005	0.008	0.009
	(0.007)	(0.008)	(0.007)	(0.008)	(0.007)	(0.008)	(0.010)	(0.010)
Black neighbor ($k = 4$)	0.016**	0.009	0.007	0.003	0.005	0.002	−0.006	−0.009
	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)	(0.008)	(0.010)
Black neighbor ($k = 5$)	0.022***	0.023***	0.012**	0.015**	0.009	0.011*	0.003	0.002
	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)	(0.007)	(0.009)
Black neighbor ($k = 6$)	0.023***	0.013*	0.014**	0.006	0.009	0.0004	0.007	−0.001
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.009)	(0.010)
Black neighbor ($k = 7$)	0.015**	0.018***	0.008	0.012*	0.008	0.012*	−0.003	0.002
	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)	(0.009)	(0.010)
Black neighbor ($k = 8$)	0.018***	0.016**	0.007	0.007	0.003	0.002	0.002	0.005
	(0.007)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)	(0.008)	(0.009)
Black neighbor ($k = 9$)	0.020***	0.017**	0.009	0.008	0.004	0.003	0.003	0.005
	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)	(0.009)	(0.011)
Black neighbor ($k = 10$)	0.024***	0.021***	0.006	0.006	0.002	0.001	−0.003	0.0001
	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.008)
High school– educated neighbor ($k = 1$)		−0.006		−0.006		−0.005		−0.004
		(0.004)		(0.004)		(0.004)		(0.004)
High school– educated neighbor ($k = 2$)		−0.002		−0.002		−0.002		0.002
		(0.004)		(0.004)		(0.004)		(0.005)
High school– educated neighbor ($k = 3$)		−0.005		−0.006*		−0.006*		−0.005
		(0.003)		(0.003)		(0.003)		(0.004)
High school– educated neighbor ($k = 4$)		−0.009***		−0.011***		−0.010***		−0.005
		(0.003)		(0.003)		(0.003)		(0.005)
High school– educated neighbor ($k = 5$)		−0.006*		−0.007**		−0.007**		−0.005
		(0.004)		(0.003)		(0.004)		(0.004)
High school– educated neighbor ($k = 6$)		0.0004		−0.001		−0.001		0.004
		(0.003)		(0.003)		(0.003)		(0.004)
High school– educated neighbor ($k = 7$)		−0.007**		−0.008***		−0.009***		−0.007
		(0.003)		(0.003)		(0.003)		(0.005)
High school– educated neighbor ($k = 8$)		0.002		0.001		0.001		0.001
		(0.005)		(0.005)		(0.005)		(0.006)
High school– educated neighbor ($k = 9$)		−0.006**		−0.006**		−0.007**		−0.004
		(0.003)		(0.003)		(0.003)		(0.004)
High school– educated neighbor ($k = 10$)		−0.005*		−0.006**		−0.006**		−0.003
		(0.003)		(0.003)		(0.003)		(0.004)
FE	None	None	State	State	County	County	District	District
Controls	No	Yes	No	Yes	No	Yes	No	Yes
N	238,344	203,915	238,344	203,915	238,344	203,915	238,318	203,892
R^2	0.008	0.015	0.022	0.027	0.046	0.052	0.341	0.361
Adjusted R^2	0.007	0.014	0.022	0.027	0.034	0.038	0.043	0.044

Table 4. Democratic partisanship by black next-door neighbor, 2017 sample. Table shows the coefficients from the main specifications estimating the effect of a next-door black neighbor on future Democratic partisanship for the 2005/2009 sample. Models in this table range from Census reel fixed effects to 5 Census page fixed effects. Cluster robust standard errors are shown in parentheses, with standard errors clustered at the county level. Models with and without covariates are shown. Both models include variables for black neighbors at each $k \in [1,10]$ position up and down the Census page. Specifications with covariates control for individual age, family size, whether their family had resided in the same residence 5 years previous, as well as the high school education, income, age, employment status, and hours worked on average per week and per week for the head of the household, and whether or not each neighbor has a head of the household with at least a high school education. Differences between sample size in covariate models and full sample are due to missingness across covariates. *$P < 0.1$; **$P < 0.05$; ***$P < 0.01$.

	(1)	(2)	(3)	(4)	(5)	(6)
Age		0.004***		0.005***		0.006***
		(0.0003)		(0.001)		(0.001)
High school		−0.015**		−0.012		−0.005
		(0.007)		(0.010)		(0.013)
Income		−0.00000***		−0.00000		−0.00001
		(0.00000)		(0.00000)		(0.00000)
Mover		−0.007***		−0.009		−0.015*
		(0.003)		(0.006)		(0.008)
HoH age		0.0001		0.0001		0.0001
		(0.0002)		(0.0003)		(0.0004)
Homeowner		−0.023***		−0.023***		−0.021***
		(0.004)		(0.005)		(0.007)
Family size		0.009***		0.006***		0.005**
		(0.001)		(0.002)		(0.002)
Employed		−0.021***		−0.029**		−0.027
		(0.006)		(0.014)		(0.021)
Hours per week		−0.0003***		−0.0004***		−0.0003*
		(0.0001)		(0.0001)		(0.0002)
Weeks per year		−0.0005***		−0.0003		−0.0003
		(0.0001)		(0.0003)		(0.0003)
Black neighbor ($k = 1$)	0.033***	0.031***	0.025**	0.023	0.028*	0.030
	(0.006)	(0.007)	(0.013)	(0.016)	(0.017)	(0.020)
Black neighbor ($k = 2$)	0.002	−0.002	0.002	0.003	0.004	0.004
	(0.006)	(0.007)	(0.011)	(0.014)	(0.016)	(0.020)
Black neighbor ($k = 3$)	0.005	0.005	0.008	0.012	0.015	0.018
	(0.007)	(0.008)	(0.013)	(0.014)	(0.015)	(0.018)
Black neighbor ($k = 4$)	0.004	0.001	−0.009	−0.011	−0.016	−0.014
	(0.006)	(0.007)	(0.012)	(0.014)	(0.015)	(0.020)
Black neighbor ($k = 5$)	0.008	0.011*	0.001	0.004	0.001	0.001
	(0.006)	(0.007)	(0.010)	(0.013)	(0.014)	(0.018)
Black neighbor ($k = 6$)	0.007	−0.002	0.017	0.008	0.018	0.009
	(0.007)	(0.007)	(0.012)	(0.015)	(0.016)	(0.018)
Black neighbor ($k = 7$)	0.006	0.011	−0.001	0.002	−0.003	0.007
	(0.006)	(0.007)	(0.012)	(0.015)	(0.015)	(0.019)
Black neighbor ($k = 8$)	0.0001	0.0004	0.0003	0.006	−0.001	−0.001
	(0.006)	(0.007)	(0.010)	(0.011)	(0.015)	(0.017)
Black neighbor ($k = 9$)	0.003	0.003	0.004	0.011	0.015	0.023
	(0.006)	(0.007)	(0.011)	(0.014)	(0.016)	(0.018)
Black neighbor ($k = 10$)	0.001	0.001	−0.002	−0.001	−0.009	−0.007
	(0.006)	(0.006)	(0.010)	(0.012)	(0.015)	(0.018)

continued to next page

	(1)	(2)	(3)	(4)	(5)	(6)
High school– educated neighbor ($k = 1$)		−0.005		−0.003		0.001
		(0.003)		(0.006)		(0.012)
High school– educated neighbor ($k = 2$)		−0.0003		0.007		0.007
		(0.004)		(0.008)		(0.010)
High school– educated neighbor ($k = 3$)		−0.005		−0.0003		−0.002
		(0.003)		(0.006)		(0.010)
High school– educated neighbor ($k = 4$)		−0.009***		0.0002		0.002
		(0.003)		(0.007)		(0.009)
High school– educated neighbor ($k = 5$)		−0.006**		−0.007		−0.008
		(0.003)		(0.007)		(0.008)
High school– educated neighbor ($k = 6$)		0.001		0.003		0.003
		(0.003)		(0.006)		(0.009)
High school– educated neighbor ($k = 7$)		−0.008**		−0.004		−0.002
		(0.003)		(0.006)		(0.010)
High school– educated neighbor ($k = 8$)		0.001		−0.002		−0.003
		(0.004)		(0.009)		(0.011)
High school– educated neighbor ($k = 9$)		−0.006**		0.001		0.002
		(0.003)		(0.008)		(0.008)
High school– educated neighbor ($k = 10$)		−0.005*		−0.002		0.001
		(0.003)		(0.006)		(0.008)
FE	Reel	Reel	Page 10	Page 10	Page 5	Page 5
Controls	No	Yes	No	Yes	No	Yes
N	238,344	203,915	238,344	203,915	238,344	203,915
R^2	0.058	0.063	0.524	0.548	0.641	0.664
Adjusted R^2	0.039	0.042	0.044	0.045	0.048	0.052

codes such that names that sound the same (John and Jon, for example) get the same coding. Although Soundex is quite brittle—typos or transcription errors in strings will often "break" Soundex—it has some predictive power for which records should match, particularly because enumerators wrote down peoples' names as they heard them when recording the 1940 Census. We include two indicator variables for Soundex agreement, one for first names agreeing and one for last names agreeing.

4) Number of potential census matches (logged). In the first step of the linking procedure, we identify records that could possibly match a given person in the voter file, loosely restricting based on string distance in first and last names, age, sex, and state of birth. The number of possible hits for a given record is indicative of the commonness of his or her name, and the more common a name, the more likely we are to make a false-positive error in making a match. By including the number of potential or possible links, the algorithm can adjust matches accordingly.

5) Agreement on specific characters in first and last names can be a signal that two records are the same. We include four indicator variables for first letter of first name agreement, first letter of last name agreement, last letter of first name agreement, and last letter of last name agreement.

6) Middle initial agreement. Distinguishing between two people with the same or similar first and last names is often done with middle names or initials. In both the voter file and the 1940 Census, we see middle initials in some cases and use them when available.

7) Birthplace and sex. Our census links block on birthplace, requiring two records to have the same reported birthplace (state for the U.S.-born). We limit our sample to people in the voter file born in the United States and who are men to facilitate linking over so many decades.

Main result tables

Tables 1 to 4 are the full results for the model regressing Democratic partisanship in 2005/2009 and 2017 on characteristics of neighbors in position $k \in [1,10]$. Cells are ordinary least squares coefficients, with standard errors (clustered at the county level) in parentheses.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/7/24/eabe8432/DC1

REFERENCES AND NOTES

1. M. Alexander, F. Christia, Context modularity of human altruism. *Science* **334**, 1392–1394 (2011).
2. R. D. Enos, Causal effect of intergroup contact on exclusionary attitudes. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3699–3704 (2014).
3. M. Sherif, *Intergroup Conflict and Cooperation: The Robbers Cave Experiment* (University Book Exchange, 1961), vol. 10.

4. A. Alesina, E. Spolaore, *The Size of Nations* (MIT Press, 2005).
5. J. Habyarimana, M. Humphreys, D. N. Posner, J. M. Weinstein, Why does ethnic diversity undermine public goods provision? *Am. Polit. Sci. Rev.* **101**, 709–725 (2007).
6. R. D. Putnam, *E Pluribus Unum*: Diversity and community in the twenty-first century the 2006 Johan Skytte Prize Lecture. *Scand. Polit. Stud.* **30**, 137–174 (2007).
7. I. D. Couzin, C. C. Ioannou, G. Demirel, T. Gross, C. J. Torney, A. Hartnett, L. Conradt, S. A. Levin, N. E. Leonard, Uninformed individuals promote democratic consensus in animal groups. *Science* **334**, 1578–1580 (2011).
8. M. Lim, R. Metzler, Y. Bar-Yam, Global pattern formation and ethnic/cultural violence. *Science* **317**, 1540–1544 (2007).
9. R. D. Enos, *The Space Between Us: Social Geography and Politics* (Cambridge Univ. Press, 2017).
10. G. W. Allport, *The Nature of Prejudice* (Addison-Wesley, 1954).
11. T. F. Pettigrew, L. R. Tropp, A meta-analytic test of intergroup contact theory. *J. Pers. Soc. Psychol.* **90**, 751–783 (2006).
12. M. R. Ramos, M. R. Bennett, D. S. Massey, M. Hewstone, Humans adapt to social diversity over time. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 12244–12249 (2019).
13. K. B. Clark, M. K. Clark, The development of consciousness of self and the emergence of racial identification in negro preschool children. *J. Soc. Psychol.* **10**, 591–599 (1939).
14. S. W. Cook, Experimenting on social issues: The case of school desegregation. *Am. Psychol.* **40**, 452–460 (1985).
15. E. L. Paluck, S. A. Green, D. P. Green, The contact hypothesis re-evaluated. *Behav. Public Policy* **3**, 129–158 (2018).
16. E. L. Paluck, D. P. Green, Prejudice reduction: What works? A review and assessment of research and practice. *Annu. Rev. Psychol.* **60**, 339–367 (2009).
17. M. Hewstone, S. Lolliot, H. Swart, E. Myers, A. Voci, A. al Ramiah, E. Cairns, Intergroup contact and intergroup conflict. *Peace Conflict* **20**, 39–53 (2014).
18. N. M. Wittlin, J. F. Dovidio, S. E. Burke, J. M. Przedworski, J. Herrin, L. Dyrbye, I. N. Onyeador, S. M. Phelan, M. van Ryn, Contact and role modeling predict bias against lesbian and gay individuals among early-career physicians: A longitudinal study. *Soc. Sci. Med.* **238**, 112422 (2019).
19. A. Campbell, P. E. Converse, W. E. Miller, D. E. Stokes, *The American Voter* (University of Chicago Press, 1960).
20. D. O. Sears, C. L. Funk, Evidence of the long-term persistence of adults' political predispositions. *J. Theor. Polit.* **61**, 1–28 (1999).
21. S. D. Miller, D. O. Sears, Stability and change in social tolerance: A test of the persistence hypothesis. *Am. J. Polit. Sci.* **30**, 214–236 (1986).
22. R. Brown, M. Hewstone, *Adv. Exp. Soc. Psychol.* **37**, 255 (2005).
23. R. D. Enos, What the demolition of public housing teaches us about the impact of racial threat on political behavior. *Am. J. Polit. Sci.* **60**, 123–142 (2016).
24. D. P. Green, J. S. Wong, Tolerance and the contact hypothesis: A field experiement, in *The Political Psychology of Democratic Citizenship* (Oxford Univ. Press, 2009), pp. 1–23.
25. N. Miller, M. B. Brewer, *Prejudice, Discrimination, and Racism*, J. F. Dovidio, S. L. Gaertner, Eds. (Academic Press, 1986).
26. M. Hewstone, R. Brown, *Contact and Conflict in Intergroup Encounters*, M. Hewstone, R. Brown, Eds. (Basil Blackwell, 1986).
27. R. S. Erikson, M. MacKuen, J. A. Stimson, Cambridge studies in political psychology and public opinion, in *The Macro Polity* (Cambridge Univ. Press, 2002).
28. J. Sidanius, F. Pratto, *Social Dominance: An Intergroup Theory of Social Hierarchy and Oppression* (Cambridge Univ. Press, 2001).
29. J. T. Jost, J. Glaser, A. W. Kruglanski, F. J. Sulloway, Political conservatism as motivated social cognition. *Psychol. Bull.* **129**, 339–375 (2003).
30. E. Carmines, J. Stimson, Political science, in *Issue Evolution: Race and the Transformation of American Politics* (Princeton Univ. Press, 1990).
31. E. Schickler, Princeton studies in American politics: Historical, international, and comparative perspectives, in *Racial Realignment: The Transformation of American Liberalism, 1932–1965* (Princeton Univ. Press, 2016).
32. M. Tesler, D. O. Sears, *Obama's Race* (University of Chicago Press, 2010).
33. M. Tesler, Chicago studies in american politics, in *Post-Racial or Most-Racial?: Race and Politics in the Obama Era* (University of Chicago Press, 2016).
34. J. Sides, M. Tesler, L. Vavreck, *Identity Crisis: The 2016 Presidential Campaign and the Battle for the Meaning of America* (Princeton Univ. Press, 2019).
35. R. S. Erikson, L. Stoker, Caught in the draft: The effects of vietnam draft lottery status on political attitudes. *Am. Polit. Sci. Rev.* **105**, 221–237 (2011).
36. D. P. Green, B. Palmquist, E. Schickler, *Partisan Hearts and Minds* (Yale Univ. Press, 2004).
37. S. T. Fiske, Stereotyping, prejudice, and discrimination at the seam between the centuries: Evolution, culture, mind, and brain. *Eur. J. Soc. Psychol.* **30**, 299–322 (2000).
38. S. Iyengar, S. J. Westwood, Fear and loathing across party lines: New evidence on group polarization. *Am. J. Polit. Sci.* **59**, 690–707 (2014).
39. C. H. Achen, L. M. Bartels, *Democracy for Realists: Why Elections Do Not Produce Responsive Government* (Princeton Univ. Press, 2017).
40. D. R. Carney, J. T. Jost, S. D. Gosling, J. Potter, The secret lives of liberals and conservatives: Personality profiles, interaction styles, and the things they leave behind. *Polit. Psychol.* **29**, 807–840 (2008).
41. M. K. Chen, R. Rohla, The effect of partisanship and political advertising on close family ties. *Science* **360**, 1020–1024 (2018).
42. R. H. Steckel, The quality of census data for historical inquiry: A research agenda. *Soc. Sci. Hist.* **15**, 579–599 (1991).
43. J. R. Brown, R. D. Enos, The measurement of partisan sorting for 180 million voters. *Nat. Hum. Behav.* , (2021).
44. J. R. Zaller, *et al.*, *The Nature and Origins of Mass Opinion* (Cambridge Univ. Press, 2012).
45. R. Goeken, L. Huynh, T. Lenius, R. Vick, New methods of census record linking. *Hist. Methods* **44**, 7–14 (2011).
46. J. J. Feigenbaum, A machine learning approach to census record linkage (Technical Report, National Bureau of Economic Research, 2016).
47. T. Enamorado, B. Fifield, K. Imai, Using a probabilistic model to assist merging of large-scale administrative records. *Am. Polit. Sci. Rev.* **113**, 353–371 (2019).
48. T. D. Logan, J. M. Parman, The national rise in residential segregation. *J. Econ. Hist.* **77**, 127–170 (2017).
49. D. L. Magnuson, M. L. King, Comparability of the public use microdata samples: Enumeration procedures. *Hist. Methods* **28**, 27–32 (1995).
50. A. Grigoryeva, M. Ruef, The historical demography of racial segregation. *Am. Sociol. Rev.* **80**, 814–842 (2015).
51. C. D. Goldin, L. F. Katz, *The Race Between Education and Technology* (Harvard Univ. Press, 2009).
52. C. Cinelli, C. Hazlett, Making sense of sensitivity: Extending omitted variable bias. *J. R. Stat. Soc. Ser. B* **82**, 39–67 (2020).
53. S. B. Billings, E. Chyn, K. Haggag, The long-run effects of school racial diversity on political identity. *Am. Econ. Rev. Insights* (2021).
54. G. J. Martin, S. W. Webster, Does residential sorting explain geographic polarization? *Polit. Sci. Res. Methods* **8**, 215–231 (2020).
55. R. Chetty, N. Hendren, L. Katz, The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *Am. Econ. Rev.* **106**, 855–902 (2016).
56. D. J. Hopkins, The upside of accents: Language, inter-group difference, and attitudes toward immigration. *Br. J. Polit. Sci.* **45**, 531–557 (2015).
57. V. O. Key, *Southern Politics in State and Nation* (University of Tennessee Press, 1949).
58. B. Spahn, Before the American voter (2019); https://dx.doi.org/10.2139/ssrn.3478473.
59. S. Goldman, D. J. Hopkins, Past place, present prejudice: The impact of adolescent racial context on white racial attitudes. *J. Theor. Polit.* **82**, 529–542 (2020).

Acknowledgments: We thank B. Spahn for providing data on California occupations in the 1940s and L. Stoker for comments on an earlier version of the paper. We also thank seminar participants at the London School of Economics, University College London, the Russel Sage Foundation, and the 2018 American Political Science Association conference. **Funding:** The project is generously supported by Harvard's Inequality in America Competitive Research Fund. **Author contributions:** All authors contributed equally to the design, implementation, and analysis of this project. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. All replication materials are available in the Harvard University Dataverse: https://dataverse.harvard.edu/.

Submitted 17 September 2020
Accepted 26 April 2021
Published 11 June 2021
10.1126/sciadv.abe8432

Citation: J. R. Brown, R. D. Enos, J. Feigenbaum, S. Mazumder, Childhood cross-ethnic exposure predicts political behavior seven decades later: Evidence from linked administrative data. *Sci. Adv.* **7**, eabe8432 (2021).

Science Advances

Childhood cross-ethnic exposure predicts political behavior seven decades later: Evidence from linked administrative data

Jacob R. Brown, Ryan D. Enos, James Feigenbaum and Soumyajit Mazumder

Sci Adv **7** (24), eabe8432.
DOI: 10.1126/sciadv.abe8432