NBER WORKING PAPER SERIES

THE EFFECT OF CHILDHOOD ENVIRONMENT ON POLITICAL BEHAVIOR: EVIDENCE FROM YOUNG U.S. MOVERS, 1992–2021

Jacob R. Brown
Enrico Cantoni
Sahil Chinoy
Martin Koenen
Vincent Pons

Working Paper 31759
http://www.nber.org/papers/w31759

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2023

We thank Ryan Enos for sharing the TargetSmart voter file data and Opportunity Insights for sharing the Infutor data. For suggestions that have improved this article, we are grateful to Raj Chetty, Ryan DeTamble, Rafael DiTella, Mitch Downey, Nathaniel Hendren, Matthew Weinzierl, Jennifer Wolak, as well as seminar and conference participants at the Toronto Political Behaviour Workshop, HEC Montréal, Universite de Montréal, Goethe University Frankfurt, Universitat Autónoma de Barcelona, the 2023 Midwest Political Science Association Annual Meeting, the International Conference for Computational Social Science, the European Meeting of the Urban Economics Association in Milan, the Princeton Center for the Study of Democratic Politics Colloquium, the 2023 American Political Science Association Annual Meeting, and the Marco Fanno Workshop at Collegio Carlo Alberto. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w31759

The Effect of Childhood Environment on Political Behavior: Evidence from Young U.S. Movers, 1992–2021
Jacob R. Brown, Enrico Cantoni, Sahil Chinoy, Martin Koenen, and Vincent Pons
NBER Working Paper No. 31759
October 2023
JEL No. D72,P0

ABSTRACT

We ask how childhood environment shapes political behavior. We measure young voters' participation and party affiliation in nationally comprehensive voter files and reconstruct their childhood location histories based on their parents' addresses. We compare outcomes of individuals who moved between the same origin and destination counties but at different ages. Those who spend more time in the destination are more influenced by it: Growing up in a county where their peers are 10 percentage points more likely to become Republicans makes them 4.7 percentage points more likely to become Republican themselves upon entering the electorate. The effects are of similar magnitude for Democratic partisanship and turnout. These exposure effects are primarily driven by teenage years, and they persist but decay after the first election. They reflect both state-level factors and factors varying at a smaller scale such as peer effects.

Jacob R. Brown
Boston University
jbrown13@bu.edu

Enrico Cantoni
University of Bologna
Department of Economics
Piazza Scaravilli 2, Bologna
Italy
enrico.cantoni@unibo.it

Sahil Chinoy
Harvard University
schinoy@g.harvard.edu

Martin Koenen
Harvard University
martin_koenen@fas.harvard.edu

Vincent Pons
Harvard Business School
Morgan Hall 289
Soldiers Field
Boston, MA 02163
and CEPR
and also NBER
vpons@hbs.edu

1 Introduction

Our behavior is shaped by innate characteristics, contextual factors, as well as the long-lasting influence of our family upbringing and the environment in which we grew up. Much of the work of social scientists has been dedicated to assessing the contributions of these four sets of factors. In democracies, where the designation of political leaders is delegated to millions of citizens and ensuring strong electoral participation is a constant struggle, special attention has been given to the forces that make us favor one party over another and that bring us to vote instead of abstaining. In influential work, Fowler et al. (2008) establish the importance of genetic factors by comparing the correlation in political behavior between monozygotic vs. dizygotic twins. Beyond genes, studies such as Jennings et al. (2009) and Healy and Malhotra (2013) point to family transmission of political attitudes. Cantoni and Pons (2022) assess the share of cross-state variation in turnout and party affiliation due to people's *current environment* by tracking changes in the behavior of voters who move across states. Existing evidence assessing the influence of *childhood environment* on later political outcomes is less definitive. Recent studies have focused on specific components of childhood environment, including racial context (Billings et al., 2021; Brown et al., 2021; Kaplan et al., 2021), economic adversity (Chyn and Haggag, 2019; Elder et al., 2023; McNeil et al., 2023), school quality (Cohodes and Feigenbaum, 2021), and immigration (Bolotnyy et al., 2022), but we lack evidence on the combined effect of all such factors.

In this paper, we measure the *overall* impact of where voters grow up during childhood and adolescence on their partisanship and political behavior as adults. Our analysis first exploits administrative voter file data provided by TargetSmart, which include the exact residential address, individual turnout, and, in states recording it, party affiliation of all registered voters in the United States from 2012 through 2021. These data do not tell us where people grew up. To obtain that information, we identify the parents of young voters as individuals who share the same last name and address and are more than 15 years older, we retrieve parents' address histories going back to 1992 from a second dataset provided by Infutor, and we assume that children live and move with their parents before turning 18.

The strategy we use to estimate the causal effect of childhood environment closely follows Chetty and Hendren (2018a), who measure effects on economic outcomes such as adult earnings. We focus on people who relocated once during their childhood and compare individuals who moved between the same origin and destination counties but at different ages. Then, we ask whether the political behavior of those who moved earlier and thus spent more time in the destination county is more similar to the behavior of permanent residents in that place. For instance, differences in outcomes between children who moved at age six versus age seven enable us to measure the effect of spending one additional year (the seventh year) in the destination. Importantly, we do not need to assume that children's potential outcomes are independent from parents' choice of where to live, only that they are uncorrelated with the timing of moves.

Our main estimates measure effects in the first election in which people are eligible to vote after coming of age. We find that children who move to the destination county at a younger age are more influenced by that environment. Overall, our results imply that spending one's entire childhood (ages 0 to 19) in a county where permanent residents end up 10 percentage points (p.p.) more Republican in their first election increases one's own likelihood of registering as a Republican by 4.72 p.p. This effect is 4.05 p.p. for Democratic partisan registration. Turnout effects are similar, with a 10 p.p. increase in permanent resident outcomes corresponding to a 4.75 p.p. individual increase. In other words, the environment in which people grow up makes their future partisanship and political participation become 40 to 50% more similar to the people they grow up around. The influence of place persists beyond the first election but decays over time, with a stronger decay for turnout than for party affiliation.

We address concerns of endogeneity in the timing of moves by verifying that our estimates are robust to controlling for observable parent characteristics or including family fixed effects, thereby only using variation across siblings who moved at the same time but spent different portions of their childhood in the same origin and destination. Furthermore, we show that movers' outcomes are better predicted by the outcomes of permanent residents of the exact same age, gender, or race and infer that our results are unlikely to be driven by shocks affecting both parents' decision to relocate and their children's future outcome. A separate concern is that sample selection bias could arise from focusing on individuals that can be linked to parents and, for the most part, had to be registered on the voter rolls at least once to enter the data. Reassuringly, the magnitude of exposure effects does not vary much with the registration rate in the county. It also remains very similar when focusing on voters who turn 18 shortly before Election Day and are thus more likely to live with their parents and be linked to them around the time they register for their first election.

To understand how exactly the childhood environment shapes future outcomes, we first compare exposure effects across different ages. We observe a steeper gradient in the correspondence between the outcomes of one-time movers and permanent residents across teenage years compared to pre-teen years, indicating that much of the exposure effects occur during adolescence. In a separate analysis considering individuals who move multiple times, we compare individuals who spend the same number of childhood years in a county, but at different ages, and find larger exposure effects during adolescence again. On average, the influence of the environment in which one lives between ages 13 and 19 is four times as large as between ages 0 and 12.

Finally, we explore the mechanisms driving place effects. To check whether the effects on young adults' political behavior are mediated by effects on socioeconomic outcomes, we use sample restrictions that hold the effects of neighborhoods on future earnings and educational attainment constant across origin and destination locations. The magnitude of the effects remains almost identical, suggesting that childhood environments impact future political behavior directly. We further compare effects on people who move across states, across counties within the same state, and across ZIP codes within the same county, to disentangle the importance of factors varying at each of these levels. We find that state-level factors such as electoral procedures and the party exerting power account for much of the exposure effects on voter turnout, and for a smaller portion of the effects on party affiliation. Rather, young voters' partisanship is influenced by factors varying at the sub-state level, such as schools or peer effects.

Overall, we make two core contributions to the literature. First, we present the most comprehensive evidence to date on the causal effect of childhood location on adult political behavior. We build on recent work that documents large and long-lasting effects of childhood neighborhoods on other adult outcomes such as earnings, educational attainment, health, marriage, and criminal activity (e.g., Kling et al., 2007; Damm and Dustmann, 2014; Bell et al., 2018; Chetty and Hendren, 2018b; Chyn, 2018; Deutscher, 2020; Laliberté, 2021; Nakamura et al., 2021; Derenoncourt, 2022).[1] Our study is also closely related to Cantoni and Pons (2022), who measure the immediate influence of the context in which a voter currently lives (e.g., state voting rules and the political and economic environment) on their electoral participation and party affiliation, and find particularly strong effects for young adults.[2] We study similar outcomes but focus instead on the long-term influence of the context in which voters grew up on their behavior today. Most of the preexisting research on the impact of childhood context on political behavior focuses on specific components (e.g., Billings et al., 2021; Kaplan et al., 2021; Cohodes and Feigenbaum, 2021). In recent work, Chyn and Haggag (2019) and Elder et al. (2023) study individuals displaced from disadvantaged neighborhoods during childhood due respectively to the demolition of Chicago housing projects and the Moving to Opportunity experiment, and they find contrasting effects of these relocations on adult participation. Our quasi-experimental strategy pools information from all registered citizens in the U.S., hence speaking to a much broader population. In addition, we study neighborhood exposure effects not just on turnout but also on partisan affiliation.

Second, we build on previous work in economics and political science that has documented the specific importance of adolescent years (e.g., Sears and Valentino, 1997; Pacheco, 2008; Neundorf and Smets, 2017; Ghitza et al., 2023) and of the context of one's first election (e.g., Mullainathan and Washington, 2009; Dinas, 2014) for the formation of long-lasting partisan attachments and political behavior. Our empirical design enables us to compare the influence of each and every age and reveals that early childhood environment does influence future political behavior but that the impact of the environment in which one lives during adolescence years is much larger. This result constitutes comprehensive evidence on the importance of formative experiences during teenage years.

The rest of the paper proceeds as follows. Section 2 describes the data and the construction of our analysis sample and Section 3 presents our empirical strategy. In Section 4, we show the main results as well as robustness checks addressing concerns regarding the identification assumption and the risk of sample selection bias. Section 5 compares exposure effects at different ages using multiple-time movers, and Section 6 investigates the mechanisms responsible for childhood neighborhood effects. In Section 7, we conclude by discussing the broader implications of our results.

[1] See Chyn and Katz (2021) for a systematic review of the recent literature on neighborhood effects.

[2] Other work on the influence of contemporary context on adults' partisanship, political behavior, and a range of social and political attitudes includes Huckfeldt and Sprague (1987), Campbell (2006), Perez-Truglia (2018), Brown (2023), and Berry and Trounstine (2023).

2 Data and sample

Our analysis requires combining young voters' political behavior and their address history.[3] We measure their political behavior using data from TargetSmart, described in Section 2.1. We do not observe their address history before 18 directly, but we assume that it is shared with their parents. We measure the address history of their parents using Infutor data, described in Section 2.2. Section 2.3 explains how we match young voters with their parents, Section 2.4 describes our main outcomes and provides summary statistics, and Section 2.5 discusses possible sources of sample selection.

2.1 TargetSmart data

We obtained our voter data from TargetSmart, a vendor that collects public voter lists, combines them with commercial data, and sells these data to campaigns, non-profits, and researchers. The data consist of yearly nationwide snapshots covering all registered voters in the country as well as some unregistered people from 2012 to 2021. These snapshots contain each voter's name, residential address, gender, race, date of birth, whether or not they were registered and voted in prior elections, and their party affiliation. Gender is recorded in many states and imputed or sourced from commercial data by TargetSmart otherwise. Race is directly recorded for some or all years in eight states (Alabama, Florida, Georgia, Louisiana, North Carolina, South Carolina, Tennessee, and Texas), and sourced from commercial data or imputed by TargetSmart based on surname and local census demographics otherwise.[4] We observe the exact date of birth of the majority of individuals.[5] Thirty states and the District of Columbia record party affiliation in their voter data, while the other twenty do not offer voters the option of affiliating with a party when they register.[6] Party affiliation has been found repeatedly to be one of the strongest predictors of an individual's likelihood to vote for Republican or Democratic candidates (e.g., Bartels, 2000; Green et al., 2008; Guo, 2023).

We construct a panel of all individuals appearing in the TargetSmart data by matching records that correspond to the same person across states and over time. Our cleaning and de-duplicating procedure uses state-level administrative identifiers, personal information including name and date of birth, turnout history, as well as TargetSmart's own linkage model, which incorporates external information such as the USPS National Change of Address Database. The full procedure is explained in detail in Appendix C. There, we also present validation checks showing that the de-duplicated TargetSmart data mirror inde-

[3]Throughout, we will use "voters" to designate all individuals with the right to vote, including people who may actually abstain.

[4]Overall, 53.8% of individual-year observations include gender information sourced directly from state records, and 14.9% include race information sourced directly.

[5]The recorded birth date for 23% of voters in the sample is January 1. It is likely that only the year of birth, rather than the exact date of birth, was recorded for these voters. In Appendix Table A1, we show that estimates excluding these individuals are similar to the main results in Table 2.

[6]The states recording party affiliation are Alaska, Arizona, California, Colorado, Connecticut, Delaware, the District of Columbia, Florida, Iowa, Idaho, Kansas, Kentucky, Louisiana, Massachusetts, Maryland, Maine, North Carolina, Nebraska, New Hampshire, New Jersey, New Mexico, Nevada, New York, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Dakota, Utah, West Virginia, and Wyoming.

pendent sources of information about the number of registered voters, voter turnout, and Republican and Democratic vote shares in each state.[7]

2.2 Infutor data

We build the address histories of young voters' parents using data from the vendor Infutor, which aggregates public and proprietary data such as phone books, property deeds, voter files, magazine subscriptions, and credit header files.[8] These data include 168 million individuals born between 1947 and 1977, corresponding to the cohorts of likely parents in our data. Each record indicates the person's full name, current residential address, date of birth,[9] and gender. Crucially, these data also include longitudinal address information covering more than 1.3 billion address observations from 1992 to 2021.

Infutor's coverage of the adult population is relatively comprehensive, particularly so since 2000 (Diamond et al., 2019). In Appendix B, we show that the number of Infutor records lines up with counts of the adult population from the decennial census and that the data are representative across U.S. counties.[10] Coverage is particularly good for individuals in the likely age range of parents during young voters' childhood and adolescence (i.e., 30 to 49 years old), who are also the likely parents in our sample. Coverage is worse for younger individuals, since they are less likely to appear in some of the sources that Infutor relies on, such as credit header files and property deeds.

2.3 Merge between TargetSmart and Infutor data

To build our sample, we focus on young voters: individuals who appear in the TargetSmart data and are between 18 and 24 years old at any point between 2012 and 2021. We link

[7]Appendix Figure B1 plots the number of TargetSmart registered voters in each state in 2021 against the 2020 Current Population Survey (CPS) estimates of registered voters. State-level counts are plotted separately for four age groups: voters 18 to 24 years old, 25 to 34, 35 to 44, and 45 to 64. For each group, the counts in TargetSmart closely match those in the CPS, but they are generally larger, with slopes between 1.066 and 1.148. The fact that TargetSmart includes more registered voters is likely due to deadwood and imperfect de-duplication. Appendix Figure B2 plots the total number of votes in 2020 from the TargetSmart data against turnout from state-level administrative reports collected by the United States Election Project (McDonald, 2020). State-level correlations are again close to 1. Finally, Appendix Figure B3 plots the proportion of registered Democrats and Republicans in each state against the 2020 presidential vote share and shows a similarly high correspondence.

[8]For further information on the Infutor data and previous studies using them, see Diamond et al. (2019), Pennington (2021), and `https://infutor.com/on-premise-identities/total-research-education` (accessed September 19, 2023).

[9]Since the Infutor data do not include the day of birth, by date of birth we henceforth mean the month and year of birth when we refer to these data.

[10]Appendix Figure B4 shows counts of active Infutor records – i.e., individuals appearing in the dataset by a given date and not flagged as deceased – divided by the total adult population in each decennial census from 1990 to 2020. Data coverage in 2000 and later is close to or exceeds 100%. Further corroborating the good coverage of the Infutor data, Appendix Figure B5 reports estimated slopes and R^2 statistics from linear fits of county-level headcounts in the Infutor data on county-level total adult population in each decennial census. Both the slope and the R^2 approach one in 2000 and later years. The estimated slope for 1990 is lower (0.504), but the R^2 remains high, suggesting that, even in earlier years, the Infutor data are quite representative.

them to their parents: individuals with the same last name, living at the same address, more than 15 years older, and present either in the TargetSmart or Infutor data.[11] We provide more details about this procedure in Appendix C. In total, we link 62.4% of young voters to a parent. Young voters whom we fail to link to a parent include those who never appear at an address where we also observe their parents and those who live at the same address but have a different last name.

We then construct Infutor address histories for parents. The address history is immediately available for parents identified in Infutor, but parents identified in TargetSmart first need to be matched to Infutor records based on name, date of birth, and address. We successfully match 87.3% of TargetSmart parents to Infutor records.

We build parents' address histories using only Infutor data. Our dataset associates each parent and year to at most one ZIP code. We drop return moves that take place suspiciously close to an earlier move: if an individual moves from ZIP code A to ZIP code B and, within one year, relocates back to the exact same address in ZIP code A, we drop this return move.

We assume that children move with their parents until we observe them in the voter file data. We pick exactly one parent from whom to infer the address history of each young voter, using the following lexicographic order: the parent with the longest address history, the parent with the most recent address observation, the female parent, and the youngest parent. We define children who share the same primary parent as siblings and assign them to the same family unit.

Appendix Figure B6 plots the probability of moving across any two states any year for parents in the linked TargetSmart/Infutor data against the probability of a cross-state move from 2000 to 2019 in the American Community Survey (ACS), measured for a similar sample of adults with children. The slope is 0.710 and the R^2 is 0.444, indicating high correspondence between interstate mobility measured in our linked sample and in the ACS.

2.4 Outcomes and summary statistics

Our analysis focuses on 18 to 24 year old voters in TargetSmart whose parents' address history could be retrieved from Infutor and who were born between November 1992 and October 2002. This time frame ensures that we observe these voters in the first general election for which they were eligible to vote, since the TargetSmart data cover the 2012 to 2020 election cycles.

Our main sample includes a total of 15.5 million young voters: 11.0 million permanent residents, defined as voters whose parents never changed county between the year these voters were born and the year they reached age 24, and 4.5 million one-time movers, whose parents moved across counties exactly once before they reached 24.[12] Note that permanent

[11] For other papers using last name, address, and date of birth to build family links, see e.g., Hersh and Ghitza (2018) and Pons and Liegey (2018).

[12] In analyses shown in Section 5, we also include voters whose parents moved across counties twice or thrice. Furthermore, in Section 6.1, we study moves across ZIP codes, including some taking place within

residents and one-time movers are defined based only on the address history of their parents, irrespective of the residential moves that they may make themselves after age 18.

Table 1 presents summary statistics for both types of voters. The mean year of birth in our sample is 1997 for both permanent residents (Panel a) and one-time movers (Panel b). In each group, about 51% of individuals are male. Movers are somewhat more likely to be white and less likely to be Black than permanent residents. Permanent residents and movers are linked to parents associated with 1.6 and 1.5 children on average, respectively.

Our first set of outcomes is based on turnout as recorded in the voter file data.[13] Our main participation outcome indicates whether individuals voted in the first general election for which they were eligible. We find comparable values for permanent residents and one-time movers: 42.7% and 44.9%, respectively.

Next, we construct outcomes regarding party affiliation for voters in the states that record this information.[14] Our main outcomes indicate whether an individual was registered with the Democratic or the Republican party in their first general election. These outcomes are equal to zero if the individual is not registered on the voter rolls or if they are registered but not affiliated with the party. They are missing for individuals living in states that do not record party affiliation. Once again, we find comparable values for permanent residents and movers: 19.2% of permanent residents and 21.0% of movers register with the Republican Party in their first election, and the corresponding values for the Democratic Party are 26.3% and 24.1%. A small fraction of permanent residents and movers (2.1% and 2.0%) is affiliated with one major party in some elections and with the other major party in other elections.

Panel b of Table 1 presents additional statistics for one-time movers. On average, these individuals move at age 11, to a county located 377 miles away. (The *median* move distance is 61.7 miles.) We further characterize their counties of origin and destination by computing outcome means among permanent residents of the same birth cohort. Appendix Figure A1 plots the distributions of permanent resident outcomes at the county-by-cohort level and Appendix Figures A2 and A3 show maps of the variation in these outcomes across counties. The maps for party affiliation reveal familiar patterns of geographic segregation along party lines, with higher shares of Democrats in urban coastal counties and of Republicans in interior and rural counties. The turnout map, instead, reveals a North-versus-South gradient, with higher rates of voter participation in Northern states than in their Southern counterparts. On average, movers relocate to counties that differ meaningfully with respect to turnout

the same county.

[13]Our voter files report an individual's entire voter turnout history in that state up to the date the file was drawn. For example, the 2021 voter file includes turnout in 2018, 2016, 2014, etc. Unlike the outcomes related to party affiliation, we define turnout outcomes by checking whether *any* voter file snapshot linked to an individual indicates that they voted in a particular election. Indeed, states can take multiple years to fully update their turnout records. Furthermore, using all snapshots enables us to account for movers whose earlier turnout history may not be reflected in the most recent voter file record linked to them.

[14]In election years, voter files may have been drawn or updated months before Election Day. Thus, we assess individuals' party affiliation based on voter files from the year *after* that election. In addition, to assess whether an individual lived in a state that records party affiliation, we use the parents' address history until the individual is first observed in the voter file data, and the individual's address from the voter file thereafter.

Table 1: Summary Statistics for Permanent Residents and One-Time Movers

(a) Permanent residents

	Mean	Std. dev.	Num. of obs.
Birth year	1997.4	2.87	10,986,694
Male	0.505	0.500	10,319,042
White	0.707	0.455	10,634,844
Black	0.105	0.307	10,634,844
Hispanic	0.148	0.356	10,634,844
Asian	0.033	0.178	10,634,844
Other race	0.007	0.084	10,634,844
Number of children in family	1.55	0.74	10,986,694
Voted in first election	0.427	0.495	10,986,694
Republican in first election	0.192	0.394	6,250,638
Democrat in first election	0.263	0.440	6,250,638
Ever Republican and Democrat	0.021	0.144	6,403,265
State records party	0.575	0.494	10,985,015

(b) One-time movers

	Mean	Std. dev.	Num. of obs.
Birth year	1997.4	2.84	4,489,370
Male	0.509	0.500	4,261,865
White	0.813	0.390	4,349,884
Black	0.073	0.260	4,349,884
Hispanic	0.077	0.267	4,349,884
Asian	0.030	0.169	4,349,884
Other race	0.007	0.083	4,349,884
Number of children in family	1.49	0.70	4,489,370
Voted in first election	0.449	0.497	4,489,370
Republican in first election	0.210	0.408	2,501,946
Democrat in first election	0.241	0.428	2,501,946
Ever Republican and Democrat	0.020	0.142	2,610,052
Age at move	11.2	7.53	4,489,370
Move distance (mi.)	377.3	599.9	4,489,370
Origin and destination states record party	0.466	0.499	4,489,370
Abs. exposure difference in voted	0.065	0.062	4,488,514
Abs. exposure difference in Republican	0.088	0.077	2,089,713
Abs. exposure difference in Democrat	0.099	0.084	2,089,713

Notes: We show summary statistics for permanent residents (Panel a) and one-time movers (Panel b). The samples and variables are defined in the text.

and party affiliation among permanent residents of the same birth cohort: For turnout, the average absolute difference between origin and destination is 6.5 p.p., and for affiliation with the Republican (resp. Democratic) Party it is 8.8 p.p. (resp. 9.9 p.p.). Note that the latter statistics are computed on the subsample of 46.6% of one-time movers moving between states that both record party affiliation. We will make the same sample restriction when estimating childhood neighborhood effects on party affiliation, since our empirical strategy exploits outcome differences between the origin and destination counties.

2.5 Sample selection

Sample selection may affect our analysis data for two reasons. First, the TargetSmart data mostly cover individuals who are registered on the voter rolls at least once during our sample period. The data do include some individuals who never register, identified through commercial records provided by data aggregation firms. Furthermore, we include observations for individuals who register at some point also in years in which they were not registered but eligible to vote. In the average election, 23.8% of 18 to 24 year olds in the sample are not registered. It remains that our coverage of unregistered citizens is incomplete. Second, we lose young voters that we fail to link to their parents or whose parents we fail to link to the Infutor address data.

To gauge the extent of sample selection in our data, Appendix Table A2 shows median county (Panel a) and census block group (Panel b) ACS characteristics using different weights: ACS estimates of counts of 18 to 29 year old citizens (column 1), counts of young voters in the TargetSmart data (column 2), counts of young voters in the TargetSmart data linked to parents (column 3), and counts of young voters in the TargetSmart data linked both to parents and to Infutor (column 4). The weighted ACS characteristics using counts of young voters in the ACS data and in the TargetSmart data are remarkably similar, especially at the county level (columns 1 and 2). By contrast, relative to the ACS, individuals in the linked samples (columns 3 and 4) live in places with slightly more white, high-income, and college-educated residents, and lower population density. This fact is consistent with white people and richer individuals having relatively higher registration rates (U.S. Census Bureau, 2021), which makes it easier to link young voters and parents registered at the same location in the TargetSmart data.[15] We address concerns related to sample selection in Section 4.3.

3 Empirical strategy

To estimate the causal effect of childhood neighborhoods, our strategy is to measure the extent to which a voter whose family moves to a new neighborhood during their childhood adopts a political behavior similar to their permanent-resident peers in that neighborhood. The stronger the effect of neighborhoods, the more similar mover and permanent-resident political behavior should become. To identify these effects, we leverage variation in the

[15] Appendix Figure A4 shows how our sample compares to the broader population, reporting coefficients from univariate regressions predicting the ratio between block group-level counts of individuals in our sample and ACS population counts using block group characteristics such as racial composition and household income.

timing of moves: we compare voters who move to similar neighborhoods but spend different proportions of their childhood in those neighborhoods. This empirical strategy largely follows Chetty and Hendren (2018a).

To fix ideas, focus on one outcome: whether an individual registers as a Republican in the first general election for which they are eligible to vote. We define the exposure effect at age m as the impact of spending year m of childhood or adolescence in an area where the proportion of permanent residents who register as Republicans in their first election is 1 p.p. higher.

The ideal experiment to identify this effect would be to randomly assign children to new neighborhoods starting at different ages for the rest of their childhood. For each assignment age m, we would then regress y_i, an indicator for whether mover i registers as a Republican in their first eligible election, on $\overline{y}_{ds}$, the average outcome among permanent residents of the same birth cohort s in destination county d:

$$y_i = \alpha + \beta_m \overline{y}_{ds} + \epsilon_i. \tag{1}$$

The exposure effect at age m would then be obtained by computing $\beta_m - \beta_{m+1}$. Of course, this experiment is not feasible, and parents' choice of neighborhood is likely correlated with children's potential outcomes: parents who move to a more Republican place may have a latent ideology θ_i that affects their children's likelihood of registering as Republicans. Thus, estimating equation (1) in observational data would yield

$$b_m = \beta_m + \delta_m,$$

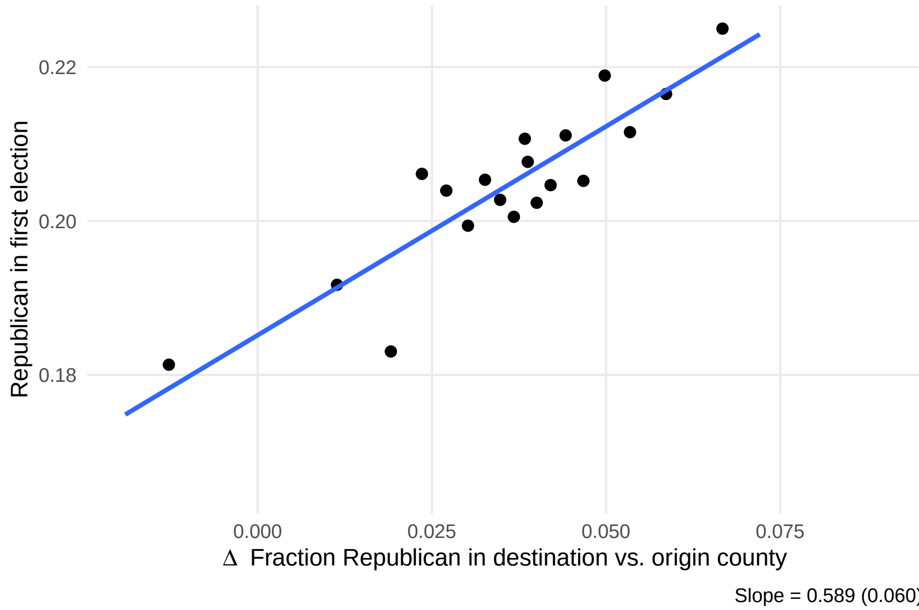
where $\delta_m = \frac{Cov(\theta_i, \overline{y}_{ds})}{Var(\overline{y}_{ds})}$ is a selection term. The magnitude of δ_m is proportional to the covariance between parents' Republican ideology and the proportion of Republicans among permanent residents in the destination neighborhood. The more likely Republican-leaning families are to move to Republican places, the greater the selection.

To account for selection, we leverage variation in the timing of moves. We compare voters who spend different amounts of time growing up in similar neighborhoods because they moved there at different points in their childhood. In this setup, identification of exposure effects does not require that *where* families choose to move is unrelated to their children's potential outcomes. It only requires that the *timing* of move is orthogonal to potential outcomes. Assuming that selection effects do not vary with child's age at move, i.e., that

$$\delta_m = \delta \ \forall m,$$

$\gamma_m = b_m - b_{m+1}$ is an unbiased estimate of the exposure effect at age m. Although this assumption is substantially weaker than assuming location choice is unrelated to political ideology, it could still be violated. For example, parents with a stronger Republican ideology may be more likely to move to Republican places when their children are young or entering

Figure 1: Movers At Age Six



Notes: On the vertical axis, we show the likelihood that an individual who moved to a different county at age six registers with the Republican Party in their first election. On the horizontal axis, we show the difference in that outcome among permanent residents in the destination versus origin county. We control for origin-by-cohort fixed effects and origin-by-destination fixed effects. The slope corresponds to b_6 in equation (2). $N = 73{,}599$.

.

primary school. We return to validating the identification assumption in Section 4.2, after presenting our main results.

We now exploit this identification strategy to derive our baseline regression specification. First consider families who move with a child who is six years old. We can regress y_i, the indicator for whether the child registers as a Republican in their first election, on $\Delta\overline{y}_{dos}$, the difference between the destination (d) and origin (o) counties in the proportion of permanent residents of the same birth cohort s who register as Republicans in their first election. The regressors also include origin-by-cohort fixed effects α_{os} to control for common characteristics of voters born in the same year and coming from the same county, and origin-by-destination fixed effects μ_{od}, to control for common characteristics of voters coming from the same county who move to the same destination county:

$$y_i = \alpha_{os} + \mu_{od} + b_6 \Delta\overline{y}_{dos} + \epsilon_{1i}. \tag{2}$$

Figure 1 is a binned scatter plot of this relationship. Controlling for the proportion of Republicans among permanent residents in the origin, children who move at age six to a county where 1 p.p. more permanent residents of the same birth cohort register as Republicans are 0.589 p.p. more likely to register as Republicans themselves.

As discussed above, b_6 does not have a causal interpretation as it combines exposure effects with selection, but the differences with the coefficients for other ages do. Therefore, we repeat this exercise for children who moved at each age from 0 to 24.

We estimate these parameters simultaneously using a single regression as follows:

$$y_i = \alpha_{osm} + \mu_{od} + \sum_{m=0}^{24} b_m (I_m = m)\Delta\overline{y}_{dos} + \sum_{s=1992}^{2002} \lambda_s (I_s = s)\Delta\overline{y}_{dos} + \epsilon_{2i}. \quad (3)$$

The coefficients of interest b_m are estimated in the first sum, which interacts indicators for age at move I_m with the destination-minus-origin difference in permanent residents' average outcomes $\Delta\overline{y}_{dos}$. Because we include origin-by-destination fixed effects (μ_{od}), exposure effects are identified by comparing individuals who move between the same origin and destination county, just at different ages. We note two differences from equation (2). First, we include the interaction of cohort indicators I_s and the exposure term $\Delta\overline{y}_{dos}$ to identify b_m purely from within-cohort variation. We could not include these terms in equation (2), where they would be colinear with $\Delta\overline{y}_{dos}$.[16] Second, we replace the origin-by-cohort fixed effects α_{os} with fixed effects specific to each age at move, α_{osm}. By controlling flexibly for age at move, we account for the possibility that moving at different ages generates different disruptions that are independent from the characteristics of the origin or destination county.

Our main specifications use unclustered heteroskedasticity-robust standard errors to simplify computation. As shown in Appendix Table A3, the standard errors are virtually unchanged when we cluster them at the destination-origin county pair level (to correspond to the level of the treatment), at the origin-by-age-at-move level (to account for the assignment mechanism of the treatment), and at the family level (to account for potential correlation between siblings).

Under our identifying assumption that the unobserved determinants of children's probability of registering Republican are uncorrelated with the age that they move, $\gamma_m = b_m - b_{m+1}$ captures the causal effect on a voter's likelihood of registering as a Republican of one extra year of exposure at age m to a destination neighborhood where permanent residents of the same cohort are 1 p.p. more likely to register as Republicans themselves.

4 Results

4.1 Main results

4.1.1 Graphical analysis

We now report our main results on the effects of childhood neighborhood exposure on young voters' political behavior in the first general election in which they are eligible to vote,

[16] The inclusion of these terms controls for differential measurement error across cohorts. For instance, our ability to observe moves at early ages varies across cohorts due to differences in the quality of our address history over time. Furthermore, for the youngest cohorts in the sample, we cannot observe moves at older ages, since our data stop in 2021.

beginning with a graphical analysis. We plot the b_m coefficients from estimating equation (3) on the sample of one-time cross-county movers for ages at move 0 to 24 years old. Differences between adjacent coefficients from 0 to 19 years old indicate how one additional year spent in the destination increases the correlation between movers' behavior and permanent resident outcomes in the destination county. Since elections take place every other year, a person's first eligible election occurs at age 18 or 19, and individuals whose parents move after they are age 20 had their behavior in this election measured before the move. The time spent in the destination cannot, by construction, affect the outcome for these voters, so we should not expect much difference across coefficients post-20. However, the *level* of these coefficients remains informative: given the absence of any destination effect, it can be entirely attributed to the selection term δ discussed in Section 3.

The results are shown in Figure 2. Each panel corresponds to a different outcome: Republican Party affiliation (Panel a), Democratic Party affiliation (Panel b), and voter turnout (Panel c). All series exhibit three key features.

First, all coefficients are strictly positive and statistically significant, including for moving ages beyond 20. For instance, in Panel a, $b_{20} = 0.150$, which indicates that voters moving at age 20 to a place where permanent residents are 1 p.p. more likely to be affiliated with the Republican Party are themselves 0.150 p.p. more likely to be affiliated with the Republican Party in their first election. The positive coefficients above 20 confirm that parents moving to more Republican areas have some latent, unobserved characteristics that make their children more likely to be Republicans in adulthood. The same goes for parents moving to more Democratic areas or areas with higher participation. Our empirical strategy controls for this selection effect by comparing the magnitude of coefficients for different moving ages.

Second, in all three panels, b_m decreases as m increases from 0 to 19. For the Republican Party (Panel a), we find that $b_0 = 0.622$ and $b_{19} = 0.152$, with relatively similar values for the Democratic Party (Panel b). Intermediate b_m values for individuals moving at ages between 1 and 18 largely lie between these extremes and exhibit a nearly monotonic decline as age at move increases. In other words, children who move at later ages – and thus spend less time in the destination county – are less likely to mirror the party affiliation of permanent residents of the same birth cohort in the destination than children who move at younger ages. This pattern is consistent with childhood neighborhoods affecting an individual's partisan affiliation in adulthood, beyond the influence of their family and innate characteristics.

While a comparison of any two adjacent coefficients gives the effect of spending one extra year in a given place, subtracting the b_{20} estimate from the very first estimate (i.e., b_0) gives the total effect of childhood place on political behavior. For affiliation with the Republican (resp. Democratic) Party, we find that this total effect is 0.472 (resp. 0.405) implying that spending one's childhood in a place that is 10 p.p. more Republican (resp. Democratic) makes people 4.72 (resp. 4.05) p.p. more likely to register as a Republican (resp. Democrat). Assuming that these impacts also hold for non-movers, a claim we return to in Section 4.4, we infer that 40 to 50% of the cross-county variation in these outcomes is explained by the causal effect of where one grows up.

The influence of childhood neighborhoods on participation is equally strong. Like for party affiliation, the coefficients shown in Panel c are strictly positive and their magnitude declines monotonically with age at move. Comparing the b_{20} and b_0 estimates indicates that spending one's childhood in a county where permanent residents are 10 p.p. more likely to participate in their first election increases one's own likelihood to vote by 4.75 p.p. Thus, 48% of the cross-county variation in turnout among young registered adults is driven by the causal effect of childhood location.

Third, while the b_m parameters decline with age at move across childhood, the decrease is more pronounced during teenage years. This implies that the effect of an additional year of exposure is larger in these years compared to earlier in childhood. We now test formally whether the slopes differ before and after age 13.

4.1.2 Parametric results

We turn to a specification that fits three separate lines to plots of the form in Figure 2, one for ages 12 and below, one for ages 13 to 19, and one for ages 20 and above. We estimate separate linear splines for these three age ranges because of the patterns observed in Figure 2 and discussed above. Formally, we run the following regression, which includes the same variables as equation (3) but restricts the relationship between age at move m_i and the exposure difference in the origin and destination neighborhoods to be piecewise linear instead of using non-parametric interactions with a full set of age at move fixed effects:[17]

$$\begin{aligned} y_i =& \alpha_{osm} + \mu_{od} + \sum_{s=1992}^{2002} \lambda_s (I_s = s)\Delta\overline{y}_{dos} + I(m_i \leq 12)(\zeta' - \zeta m_i)\Delta\overline{y}_{dos} \\ &+ I(12 < m_i \leq 19)(\gamma' - \gamma m_i)\Delta\overline{y}_{dos} + I(m_i > 19)(\delta' - \delta m_i)\Delta\overline{y}_{dos} + \epsilon_{3i}. \end{aligned} \tag{5}$$

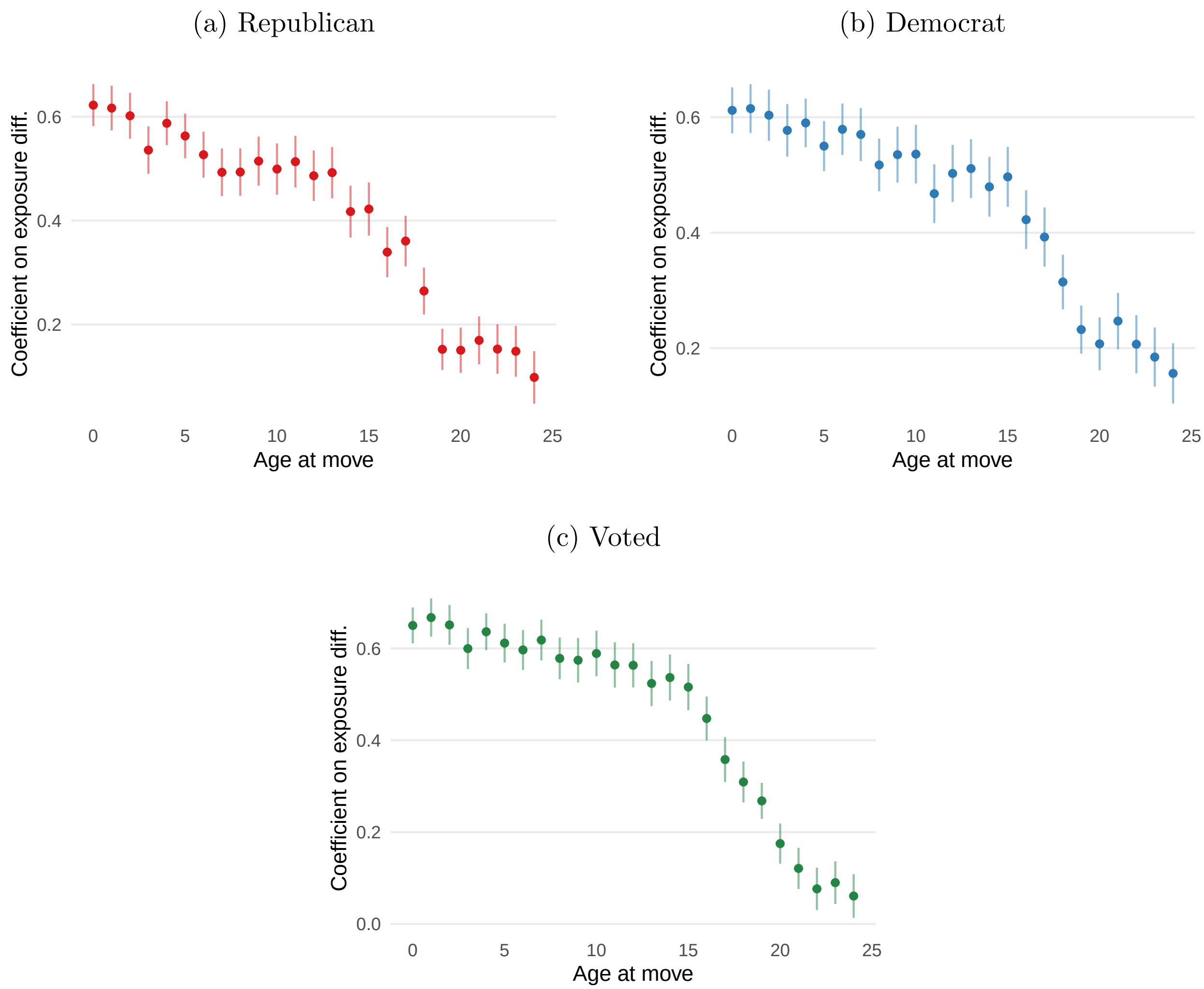
Table 2 reports estimates of the average yearly exposure effects ζ, γ, and δ. One additional year of exposure between ages 0 and 12 to a county where the proportion of Republicans (resp. Democrats) is 1 p.p. higher makes a person 0.012 (0.010) p.p. more likely to register as Republican (resp. Democrat) in their first election (columns 1 and 4). Consistent with the patterns in Figure 2, the yearly exposure effects between ages 13 and 19 for Republican and Democratic party affiliation are much larger: 0.052 and 0.047 p.p. Similarly, exposure effects on turnout are larger after 13 than before: 0.051 versus 0.008 p.p. (column 7). Differences between exposure effects in childhood and in adolescence are statistically significant at the 1% level for all three outcomes. These results suggest that where people live during their adolescence affects their political behavior in adulthood more than where they live during childhood. We return to a more formal assessment of this claim in Section 5.

[17]In some analyses, we use the following specification to more parsimoniously compare average yearly exposure effects over the full period from 0 to 19 years old across subsamples:

$$\begin{aligned} y_i =& \alpha_{osm} + \mu_{od} + \sum_{s=1992}^{2002} \lambda_s (I_s = s)\Delta\overline{y}_{dos} + I(m_i \leq 19)(\zeta' - \zeta m_i)\Delta\overline{y}_{dos} \\ &+ I(m_i > 19)(\delta' - \delta m_i)\Delta\overline{y}_{dos} + \epsilon_{4i}. \end{aligned} \tag{4}$$

The results from estimating this equation on the full sample are shown in Appendix Table A4.

Figure 2: Nonparametric Estimates of Childhood Exposure Effects



Notes: We show age-specific exposure effects corresponding to the estimates of b_m in equation (3). All outcomes are measured for the first election in which an individual is eligible to vote. Vertical lines show 95% confidence intervals. N = 1,927,376 for party affiliation and N = 4,073,607 for turnout.

Table 2: Parametric Estimates of Exposure Effects

	Republican			Democrat			Voted		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Exposure effect, 0–12	0.0120***	0.0117***	0.0083	0.0103***	0.0096***	0.0011	0.0082***	0.0067***	0.0146***
	(0.0014)	(0.0019)	(0.0057)	(0.0014)	(0.0019)	(0.0057)	(0.0015)	(0.0022)	(0.0053)
Exposure effect, 13–19	0.0520***	0.0430***	0.0487***	0.0471***	0.0424***	0.0588***	0.0493***	0.0458***	0.0569***
	(0.0037)	(0.0051)	(0.0099)	(0.0039)	(0.0054)	(0.0106)	(0.0039)	(0.0058)	(0.0107)
Exposure effect, 20–24	0.0114*	0.0033	0.0030	0.0147**	0.0206**	0.0073	0.0265***	0.0160*	0.0078
	(0.0060)	(0.0081)	(0.0138)	(0.0064)	(0.0088)	(0.0148)	(0.0061)	(0.0090)	(0.0158)
Difference $[13, 19]$ vs. $[0, 12]$	0.040	0.031	0.040	0.037	0.033	0.058	0.041	0.039	0.042
t-statistic $[13, 19]$ vs. $[0, 12]$	10.0922	5.6501	3.3889	8.8223	5.7358	4.5552	9.8025	6.2669	3.3503
p-value $[13, 19] = [0, 12]$	0.0000	0.0000	0.0007	0.0000	0.0000	0.0000	0.0000	0.0000	0.0008
Control for income in origin		✓			✓			✓	
Family fixed effects			✓			✓			✓
Number of observations	1,927,376	1,569,149	750,764	1,927,376	1,569,149	750,764	4,073,607	3,203,012	1,613,101
Dependent variable mean	0.202	0.193	0.219	0.245	0.250	0.235	0.450	0.452	0.460
R^2	0.215	0.385	0.738	0.177	0.340	0.704	0.282	0.441	0.689

Notes: We report annual childhood exposure effects, measuring the effect of spending an additional year of childhood in a place where the permanent residents are 1 p.p. more likely to affiliate with the Republican Party (columns 1–3), Democratic Party (columns 4–6), or vote (columns 7–9). Estimates are based on equation (5). All outcomes are measured for the first election for which an individual is eligible to vote. For each outcome, the second column adds origin by destination by block group-level median income decile fixed effects as well as origin by birth cohort by age at move by block group-level median income decile fixed effects. Block group-level median income deciles are constructed from 2016–2021 ACS data for the origin block group. The third column adds family fixed effects and thus compares siblings making the same move with different amounts of exposure to the destination county. We also show the difference between the coefficient estimates for ages 0 to 12 and 13 to 19, as well as t-statistics and p-values corresponding to testing the null of equality of those two estimates. Heteroskedasticity-robust standard errors are in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively.

4.2 Robustness checks: differential age selection effects

4.2.1 Time-invariant family characteristics

The effects shown above can be interpreted causally under the assumption that the timing of moves during childhood is uncorrelated with any factor affecting young adults' political behavior other than the neighborhood. This assumption would be violated if, e.g., families moving to a more Republican destination earlier than other families had some latent political ideology making their children more likely to register as Republicans in adulthood.

Controlling for observable parent characteristics One way to address this concern would be to directly control for parents' party affiliation and level of participation. However, recall that the voter file data begin in 2012, while the address history data start in 1992. For most moves, we thus only observe and could only control for parents' political behavior several years after the move. But parents' behavior is likely to be affected by their neighborhood of residence, making this behavior a "bad" (post-treatment) control. This concern is especially relevant in our context given evidence in Cantoni and Pons (2022) showing that adults' turnout and partisanship is affected by their place of residence.

An alternative way to account for parent-specific factors that can be observed before the family moves is to use parental income, inferred from where parents live before the move. To do so, we map street addresses to census block groups and assign parents to national income deciles based on the median household income in their block group. We then augment equation (5) and further interact the origin-by-cohort-by-age-at-move fixed effects – as well as the origin-by-destination fixed effects – with parental income rank r, i.e., we replace α_{osm} and μ_{od} with α_{osmr} and μ_{odr}, respectively.

Intuitively, we explore how the political behavior of young people moving at different ages between similar origin-destination pairs *and with parents who have similar incomes* relates to the destination-minus-origin difference in behavior among permanent residents of the same age. Given the relationship between income and partisan affiliation documented in prior research (e.g., Verba et al., 1995) and shown in Appendix Figure A5, these narrower comparisons help address the concern that parents who move when their children are younger may be different – along economic and political dimensions – from parents who move later.

Results from this exercise are shown in Table 2, columns 2, 5, and 8. Reassuringly, we obtain results that are very similar to the baseline estimates for all three outcomes.

Comparing siblings Parents who move to a certain destination when their children are of different ages may still be *un*-observably different. To address this concern, we restrict the sample to young voters with at least one sibling and we add family fixed effects to equation (5), thus comparing siblings with the same parents. Consider, for instance, a family with two children that moves across counties when one child is 12 years old and the other is 15, so that the younger child spends three more years in the destination before becoming eligible to vote. The family fixed effects specification identifies exposure effects from this difference alone, while maintaining all time-invariant family characteristics constant.

We show the results in Table 2, columns 3, 6, and 9. While we lose more than half the sample, the point estimates for exposure effects from ages 13 to 19 across outcomes are very similar to the main specification. Effects above 20 years old are small and non-significant, and the 0 to 12 estimates for Democratic and Republican party affiliation are smaller than in the main specification and no longer statistically distinguishable from zero. For turnout, the effect increases in size but it remains much smaller than the 13 to 19 effect. As shown in Appendix Figure A6, non-parametric estimates controlling for family fixed effects also feature declining patterns consistent with those in Figure 2. Overall, even these much more narrow comparisons reveal substantial childhood exposure effects, with much of the effects occurring during teenage years. Furthermore, we provide evidence that differences between the magnitude of the effects in the main specification and in the family fixed effects specification are partly driven by sample differences. In Appendix Table A5, we restrict the sample to children with at least one sibling in regressions both with and without family fixed effects and obtain estimates that are remarkably close to each other. This suggests that time-invariant family characteristics, which are only controlled for when family fixed effects are included, do not create important concerns for the causal interpretation of our estimates in the full sample.

4.2.2 Time-varying shocks

The preceding exercises show that our results are robust to more demanding specifications that explicitly account for time-invariant family characteristics correlated both with children's age at move and with their political behavior as adults. However, one could still be concerned about time-*varying* shocks; that is, events that are correlated both with parents' decision to relocate and with their children's future political behavior. Consider, for instance, a family with two children that moves to a more Republican area when the parents become richer. Compared to their older sibling, the younger child will not only spend more time in the destination neighborhood, they will also spend more time growing up in a richer familial environment, making it difficult to isolate the neighborhood exposure effect even when family fixed effects are included.

To address this concern, we follow Chetty and Hendren (2018a) and present a series of overidentification tests. First, we test whether movers' outcomes are better predicted by the outcomes of permanent residents of similar cohorts than by those of older or younger cohorts. Second, we exploit variation in permanent residents' outcomes between individuals of different genders and different races within cohort-by-county cells. The logic underpinning these tests is that neighborhoods' impact on political behavior may lead movers' outcomes to converge to the outcomes of permanent residents of the exact same age, gender, and race, i.e., the people in their destination neighborhood who were most likely to experience the local context in the same way and to interact with them and influence them. By contrast, unobserved time-varying shocks would not generate this specific pattern of convergence. For example, family income shocks correlated with the timing of the move should not lead male movers' outcomes to converge to the outcomes of male permanent residents to a greater degree than to those of female permanent residents.

Cohort For each mover, we first average the exposure differences based on the outcomes of permanent residents in three groups: those in birth cohorts two to four years older, one year older and one year younger (including the exact same cohort as the mover), and two to four years younger. We replace the exposure measures $\Delta\overline{y}_{dos}$ with these three measures $\Delta\overline{y}_{dos}^{-4 \text{ to } -2}, \Delta\overline{y}_{dos}^{-1 \text{ to } +1}$, and $\Delta\overline{y}_{dos}^{+2 \text{ to } +4}$ in equation (4). As shown in Figure 3, Panel a, the exposure effect estimates largely load onto the outcomes of the permanent residents of similar cohorts, and the coefficient estimates on adjacent cohorts are substantially smaller and often non-significant. This result is all the more striking as permanent resident outcomes are highly correlated across consecutive cohorts.[18] It implies that, to bias our baseline estimates, unobserved time-varying confounders would have to cause parents to move to a neighborhood not just where children in general grow up to be Democrats, but where children *born around the same time as their own child* grow up to be Democrats. As party affiliation and participation are only realized later in life, once children become eligible to vote, it is unlikely that families can accurately predict the evolution of political geography and make relocation decisions based on this information.

Gender We then compute average outcomes for permanent residents of the mover's gender and of the opposite gender within county-by-cohort cells and replace the exposure measures $\Delta\overline{y}_{dos}$ with $\Delta\overline{y}_{dos}^{g}$ and $\Delta\overline{y}_{dos}^{-g}$. As expected, the outcomes of permanent residents of the same gender are much better predictors of the mover's behavior than the outcomes of permanent residents of the opposite gender (Figure 3, Panel b).

Race Finally, we compute average outcomes for permanent residents of the mover's race and of the other races within county-by-cohort cells and replace $\Delta\overline{y}_{dos}$ with $\Delta\overline{y}_{dos}^{r}$ and $\Delta\overline{y}_{dos}^{-r}$. Once again, the outcomes of permanent residents of the same race are much better predictors of the mover's future party affiliation and participation than the outcomes of permanent residents of other races (Figure 3, Panel c). Overall, the consistent results of these demanding overidentification tests alleviate the concern that time-varying shocks could bias our estimates.

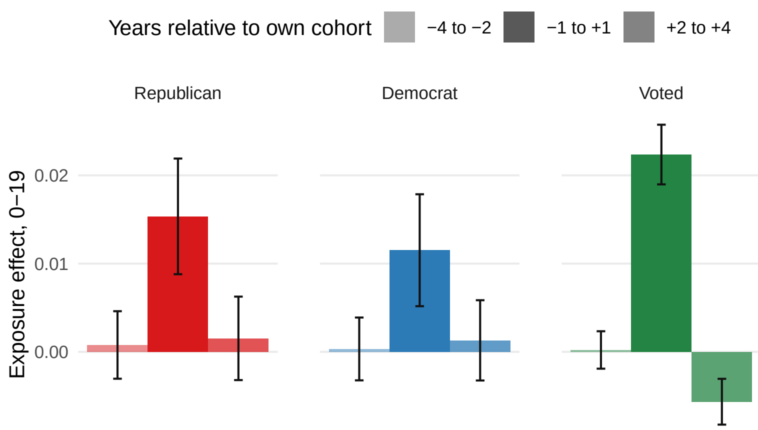
4.3 Robustness checks: sample selection bias

A second threat to the validity of our results is sample selection bias, which could arise from mostly focusing on individuals who are registered at least once during our sample period, or from restricting our sample to individuals who can be linked to parents and whose parents can in turn be linked to the Infutor data. In what follows, we discuss both concerns and present evidence to assuage them.

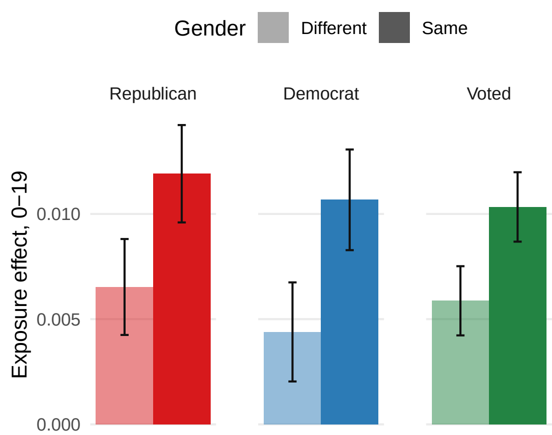
[18] The population-weighted autocorrelation between adjacent cohorts within counties is 0.880 for Republican Party affiliation, 0.846 for Democratic Party affiliation, and 0.472 for turnout in the first eligible election. Similarly, permanent residents' outcomes within county-by-cohort cell are strongly correlated across genders. The two-way correlation across genders is 0.839 for Republican Party affiliation, 0.863 for Democratic Party affiliation, and 0.917 for turnout. Within county-by-cohort cells, permanent residents' outcomes are also correlated across races, albeit to a lesser extent. The two-way correlation (comparing individuals of one race to individuals of all other races) is -0.096 and 0.194 for Republican and Democratic Party affiliation, and 0.727 for turnout.

Figure 3: Overidentification Tests

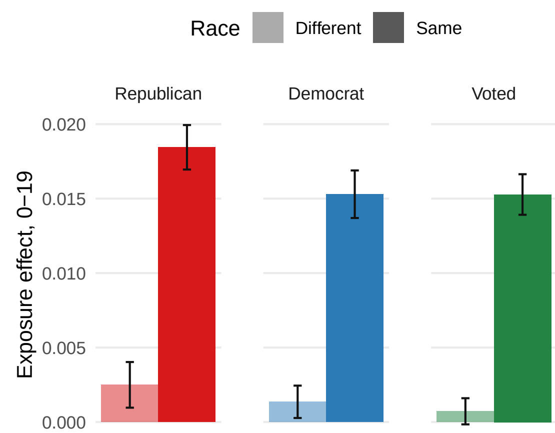
(a) Cohort



(b) Gender



(c) Race



Notes: We show overidentification tests that change the exposure measures from the baseline specification. In Panel a, we replace the cohort-specific exposure measures in the baseline specification (equation (4)) with exposure measures for three cohort groups: permanent residents between two and four years older than the mover, permanent residents between one year older and one year younger, and permanent residents between two and four years younger. In Panel b, we use exposure measures specific to permanent residents of the same gender as the mover and of the opposite gender. In Panel c, we use exposure measures specific to permanent residents of the same race as the mover and of another race, where race is collapsed into four categories: white, Black, Hispanic, and other. We exclude voters missing race information. Vertical lines show 95% confidence intervals. Appendix Table A6 reports these coefficient estimates along with standard errors. In Panel a, N = 1,357,158 for party affiliation and N = 2,838,093 for turnout. In Panel b, N = 1,924,178 for party affiliation and N = 4,064,053 for turnout. In Panel c, N = 1,766,839 for party affiliation and N = 3,712,474 for turnout.

Focus on registered individuals Our analysis sample comprises predominantly individuals who register on the voter rolls at least once between 2012 and 2021.[19] If neighborhood exposure influences the likelihood of becoming registered and, thus, of appearing in the sample, our results may be biased.

Previous research demonstrates the sensitivity of voter file analyses to differential post-treatment registration (Nyhan et al., 2017). In our case, low-propensity voters who would otherwise remain unregistered may be pushed into the electorate – and into the sample – if they spend time in a neighborhood that induces electoral participation. Our estimates of effects on voter turnout would then be biased downwards, since they would sum the exposure effect of interest (the effect of a destination county on the participation of individuals who were going to register regardless), and a counteracting selection effect (the inclusion of marginal registrants with low propensity to vote in counties inducing high participation).

The same mechanism may bias effects on party affiliation downwards, in that a county conducive to strong political engagement may increase inframarginal registrants' likelihood to affiliate with a party while also bringing into the electorate marginal voters who are less interested in politics and thus less likely to affiliate with a major party. On the other hand, a neighborhood that influences residents to become Republicans may dissuade voters that lean towards the Democratic Party from registering on the voter rolls. Because these voters would not be included in the sample, our estimates would overstate the true size of neighborhood exposure effects on party affiliation. Thus, conditioning on registration could bias estimates for these outcomes in either direction.

We address these concerns by testing whether the magnitude of exposure effects varies depending on the overall voter registration rate in the county. We group counties into quartiles based on registration rates among individuals aged 18 to 29,[20] and estimate exposure effects separately for movers relocating to counties in different quartiles of voter registration. If conditioning our sample on those who register to vote at least once biased our main estimates, we would observe systematically different magnitudes across the four quartiles. For instance, in the above example for turnout, we would expect to measure larger exposure effects for the highest quartile. Instead, Figure 4 shows that the magnitude of the effects does not vary in any systematic or statistically significant way for any of the outcomes. We interpret this as evidence that restricting the sample to individuals who register at least once is not a major source of bias.

Linked sample The fact that all our analyses are specific to young people whom we can link to parents and to the Infutor data creates an additional source of selection and

[19] In Appendix Table A7, we show that excluding the individuals who never appear as registered voters on voter rolls (and were thus likely included by TargetSmart based on complementary datasets) yields virtually identical results to those in Table 2.

[20] We construct this registration rate by combining county-level ACS data for the years 2015–2019 with the TargetSmart voter file data. We count the number of registered voters aged 18 to 29 in the TargetSmart data in each county and divide it by the number of citizens of the same age range and in the same county in the ACS data. We choose the age range 18 to 29 rather than 18 to 24 because the ACS only includes these statistics for the former.

potential bias. In particular, the likelihood of attending college may be affected by the number of years spent in the destination county. But young voters who move out to attend college may be less likely to be observed at the same address as their parents and to enter our linked sample.

To address this concern, we focus on voters who turn 18 shortly before Election Day. Indeed, many of these voters will be induced to register before the election, at a younger age than voters who turn 18 shortly after the election and thus have nearly two additional years to register before the next one. But young citizens are more likely to live with their parents at 18 years old than at 19 or 20. Therefore, the registration address of movers who turn 18 before an election should be more likely to match that of their parents, increasing the likelihood that they appear in our linked sample regardless of their destination county, and reducing the risk of endogenous sample selection.

Panel a of Figure 5 presents evidence for this idea. It shows the number of voters in our linked analysis sample by the day on which they turn 18 relative to Election Day. That is, we show the number of people in our sample who turned 18 shortly before or after the 2012 election. We then repeat the same exercise for all elections until 2018 and stack these counts of voters in our sample. The figure shows that our linked analysis sample contains substantially more people turning 18 right before a general election than right after: As expected, turning 18 right before an election greatly increases the probability of appearing in our sample.

Building on this evidence, Panel b of Figure 5 and Appendix Table A8 report exposure effects estimated on the sample of voters who turned 18 in the 100 days before a general election between 2012 and 2020, and compare them to estimates for the full sample.[21,22] If sample selection driven by the imperfect linkage of voters to parents biases our results, we would expect effects on the full sample to differ substantially from those estimated using voters who turned 18 right before a general election. Reassuringly, point estimates for all outcomes remain positive, quantitatively similar, and statistically significant in that subsample. For Republican Party affiliation and Democratic Party affiliation, point estimates in the subsample and the full sample are statistically indistinguishable. The effect for turnout is larger in the subsample, indicating that if we are endogenously losing people who register to vote after starting college and moving away from home, the resulting bias may cause us to underestimate the true turnout effect.

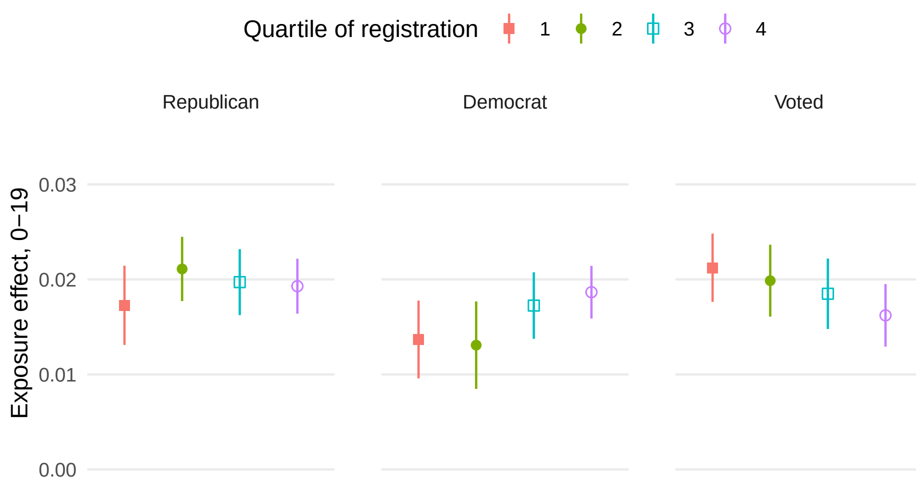
4.4 External validity

Setting aside internal validity concerns, our estimates may still be unrepresentative of the broader U.S. population. As discussed in Section 2.4 and shown in Appendix Table A2, there

[21] While exact dates of birth are reported for most of our sample, for a subset of voters we only observe month and year (or just year) of birth, as discussed in Section 2.1. For this exercise, we exclude voters with reported dates of birth ending in 01, as those could be accurate only to the year.

[22] In Appendix Figure A7, we further restrict the sample to individuals born after the state-specific school entry cutoff (where available) to focus on individuals who are likely still in high school at the time of their first election. We expect these individuals to be more likely to still live with their parents. We obtain estimates similar to the ones reported here but much less precise, due to the smaller sample size.

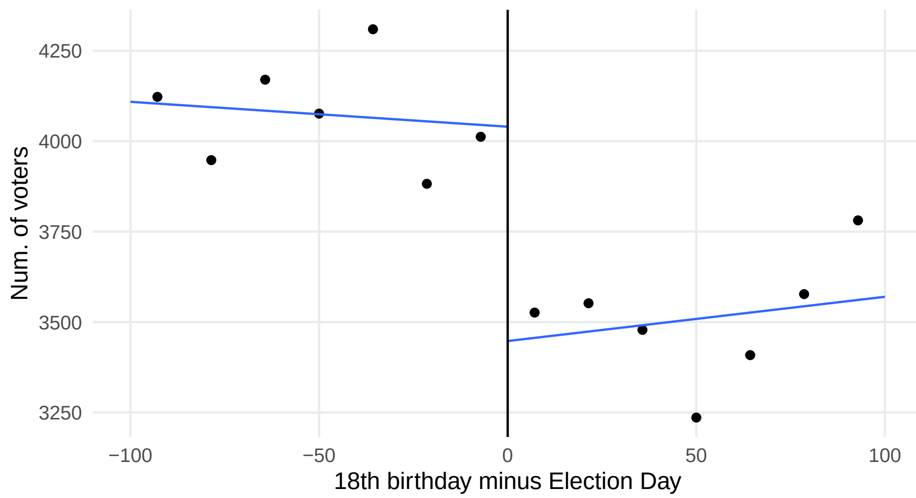
Figure 4: Splitting Sample by Quartile of Registration



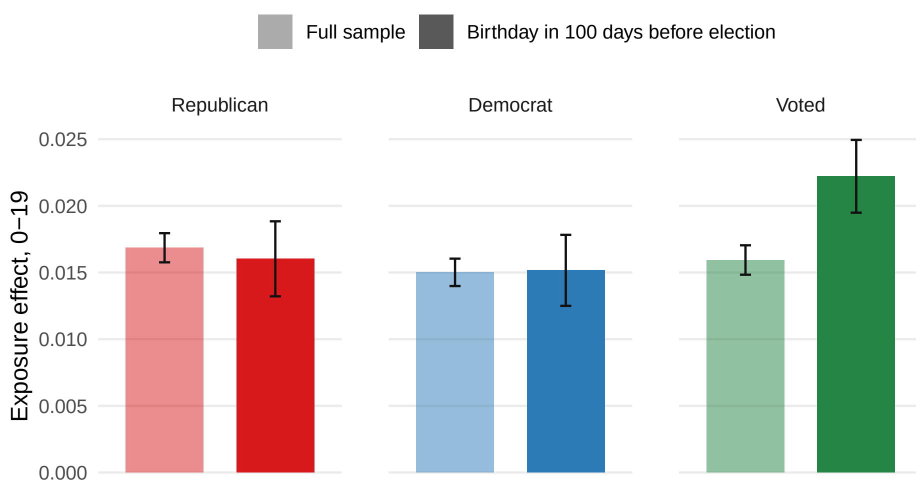
Notes: We split our sample into four quartiles based on the voter registration rate among individuals aged 18 to 29 in the mover's destination county. We construct this registration rate by combining county-level ACS data for the years 2015 to 2019 (source table: Citizen, Voting-Age Population by Age) with the TargetSmart voter file data. We count the number of registered voters aged 18 to 29 in each county in the TargetSmart data and divide it by the number of citizens of the same age range and in the same county in the ACS data. We choose the age range 18 to 29 rather than 18 to 24 because the ACS only includes these statistics for the former. We then assign counties to quartiles of the resulting registration rate and estimate equation (4) separately for each quartile. Q1 and Q4 correspond to counties with the lowest and highest registration rates, respectively. Vertical lines show 95% confidence intervals. N = 1,823,422 for party affiliation and N = 3,815,284 for turnout.

Figure 5: Estimates for Voters With a Birthday Within 100 Days of Election Day

(a) Number of Voters in the Sample By Birth Date

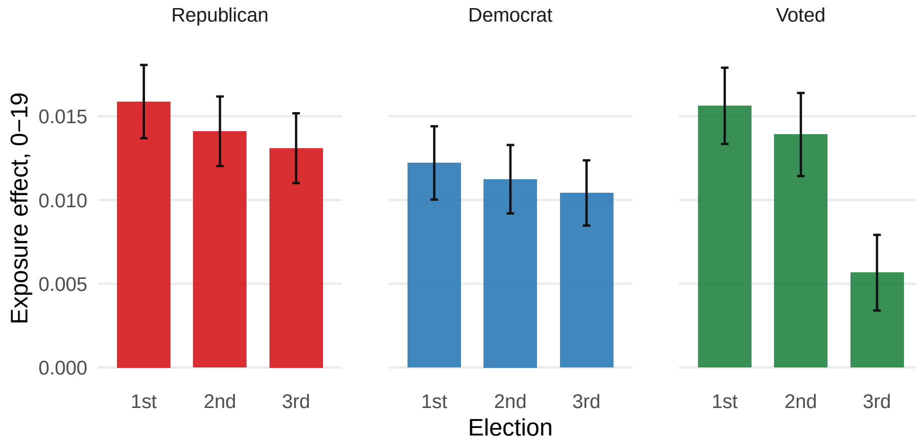


(b) Exposure Effects: Full Sample vs. Specific Birthday Sample



Notes: Panel a shows the number of voters in our sample by the day on which they turned 18 relative to Election Day. For instance, we show the number of people in our sample who turned 18 shortly before or after the 2012 election. Those born to the left of the cutoff – the barely 18-year olds – were eligible to vote in the 2012 election, while those to the right of the cutoff turned 18 only after Election Day, making 2014 the first election in which they were eligible to vote. We repeat this for all elections until 2018 and stack the counts of voters in our sample. We exclude the 2020 election from the exercise in this panel since our voter file data end in 2021 and thus we do not observe individuals who turned 18 shortly after the 2020 election in any subsequent election. $N = 844{,}524$. Panel b shows exposure estimates separately for the full sample (in lighter colors) and for the sample of barely 18-year-olds, i.e., those born at most 100 days before Election Day (in darker colors). Vertical lines show 95% confidence intervals. For the full sample, $N = 1{,}635{,}379$ for party affiliation and $N = 3{,}415{,}738$ for turnout. For the restricted sample, $N = 233{,}445$ for party affiliation and $N = 493{,}251$ for turnout.

Figure 6: Persistence of Childhood Exposure Effects



Notes: We show estimates of exposure effects for our three primary outcomes in each of the first three elections for which an individual is eligible to vote. The sample is restricted to individuals for whom we can observe all three elections, i.e., birth cohorts 1992 to 1998. Vertical lines show 95% confidence intervals. Appendix Table A9 reports these coefficient estimates along with standard errors. N = 1,122,485 for party affiliation and N = 2,361,712 for turnout.

are a few observable differences between our sample and the U.S. voting-eligible population. For instance, relative to ACS data, our sample slightly over-represents residents of census block groups with high median incomes and large proportions of white residents.

To address external validity concerns, we weight movers by the ratio between counts of voting-eligible individuals aged 18 through 29 in the county of origin or (separately) in the county of destination, according to the ACS, and the number of individuals in the same county in our linked sample of movers and permanent residents. In Appendix Figure A8, we compare point estimates from the weighted samples to our baseline estimates from the unweighted sample and find that they are remarkably similar and statistically indistinguishable from each other for all outcomes.

4.5 Effect persistence

Our main effects demonstrate the influence of childhood environment on political behavior when voters first become eligible to vote. We now ask whether this influence persists through subsequent elections. When measuring effects in the first eligible election, the influence of family and genes can mitigate the influence of the area in which people grew up. When looking at effects on later outcomes, a third type of factor can attenuate the influence of the childhood context: all factors taking place after one comes of age, such as where one lives and the friends one makes as an adult. Therefore, we should expect smaller effects of childhood location down the line.

The persistence results are presented in Figure 6 and in Appendix Table A9. To facilitate comparisons, all estimates are based on the subsample of people whose outcomes we observe in each of the first three elections after they turn 18. While exposure effects do shrink over time, they remain significant at the 1% level in the second and third eligible elections for all outcomes. Exposure effects on an individual's Republican or Democratic party affiliation in their second eligible election are more than 90% the size of the effects for the first observable election. By the third eligible election, exposure effects on party affiliation are still around 85% those for the first election. For turnout, effects for the second election are still about 90% the size of those for the first, but the decline is faster: the effect size for the third election is just under 40% the size of the first. Thus, where one grows up influences political behavior even several years into adulthood, after people have lived through formative experiences such as attending college and starting their first job. Still, we observe a decline in the influence of childhood environment across elections, particularly for voter turnout.

5 Formative years

5.1 Empirical strategy

In this section, we ask whether some ages are more formative than others, in the sense that where people live at these ages exerts a stronger influence on their future political behavior. In particular, one may expect adolescence years to be more impactful. Adolescents are generally more interested in politics and pay more attention to it than younger children, which may make them more malleable (Janmaat and Hoskins, 2021). In addition, the proximity of their first election means that the influence of political events and experiences is less likely to dissipate by the time they are eligible to register and vote. The fact that yearly exposure effects are larger between ages 13 and 19 than before, as shown in Figure 2 and Table 2, is in line with that hypothesis. However, that difference could also be due to diminishing marginal exposure effects. Indeed, exposure effects after 13 are estimated based on individuals who moved to the destination county after that age and spent no more than six years before their first eligible election in that county. By contrast, exposure effects in early childhood are estimated based on individuals who moved to the destination county at a younger age and spent much more time there. For instance, the effect at age four is estimated by comparing the behavior of individuals who moved to the destination county at age four rather than five and who thus spent 16 rather than 15 years in that county before reaching age 20. But spending one additional year in a location may have less impact when the total time one will spend there is 16 years instead of seven years or less.

The number of years that one-time movers live in an area is colinear with the ages that they live there, making it impossible to distinguish between the two explanations. Thus, we now shift attention to people who move not just once but two or three times across counties between ages 0 and 24, and we assess whether exposure spells of a given length are more impactful during teenage years than at earlier ages.[23] For instance, we consider two individuals who spend four years in the same county during childhood and adolescence, respectively, and we ask whether the latter behaves more similarly to the county's permanent residents.

[23] In Appendix Table A10, we report summary statistics for the sample of multiple-time movers.

Formally, we estimate the following specification:

$$
\begin{aligned}
y_i = \alpha_{os} + \sum_{a \in [0-12, 13-19]} I(a_{i1} = a)\beta_a \Delta\overline{y}_{dos_1} + \sum_{e_1=2}^{6} I(e_{i1} = e_1)\mu_{e_1} \\
+ \sum_{j=2}^{3} \sum_{e_j=0}^{17} I(e_{ij} = e_j)(\mu_{e_j} + \nu_{e_j} \Delta\overline{y}_{dos_j}) + \epsilon_{5i}.
\end{aligned}
\tag{6}
$$

Like in equation (3), y_i designates the outcome of interest. The main regressor is the difference in permanent residents' average outcomes between the origin and the *first* destination, $\Delta\overline{y}_{dos_1}$, which we interact with indicators for whether i lived in the first destination during their childhood (0–12) or teenage (13–19) years, $I(a_{i1} = a)$.

The coefficients of interest, β_a, indicate the extent to which the behavior of movers living in the first destination at ages a resembles the permanent resident outcomes in that destination, controlling for permanent resident outcomes in the origin. Like in the one-time movers design, the individual β_a coefficients combine exposure effects with selection. However, assuming that the unobserved determinants of political behavior are uncorrelated with age at move, as we did for one-time movers, the difference between β_{13-19} and β_{0-12} captures the differential impact of adolescence years relative to earlier childhood years.

In the above regression, we further control for μ_{e_1}, μ_{e_2}, and μ_{e_3}, fixed effects for the number of years of exposure (e_1, e_2, and e_3) that individual i had in the first, second, and third destinations before age 20. We also control for interactions between the fixed effects for the second and third destinations and the difference in permanent resident outcomes between the origin and these destinations ($\Delta\overline{y}_{dos_2}$ and $\Delta\overline{y}_{dos_3}$), to account for exposure effects in these places.

Beyond these regressors, the set of fixed effects we include is more parsimonious than in equation (3), due to the smaller number of observations.[24]

Finally, in our preferred specification, we only include individuals who spent two to six years in the first destination. Indeed, exposure spells shorter than two years may reflect measurement error in the address history, and six years corresponds to the total number of years between 13 and 19. We also exclude individuals who lived in the first destination during both age ranges (0–12 and 13–19) and those who appear to have returned to a previous location, which may also reflect measurement error.

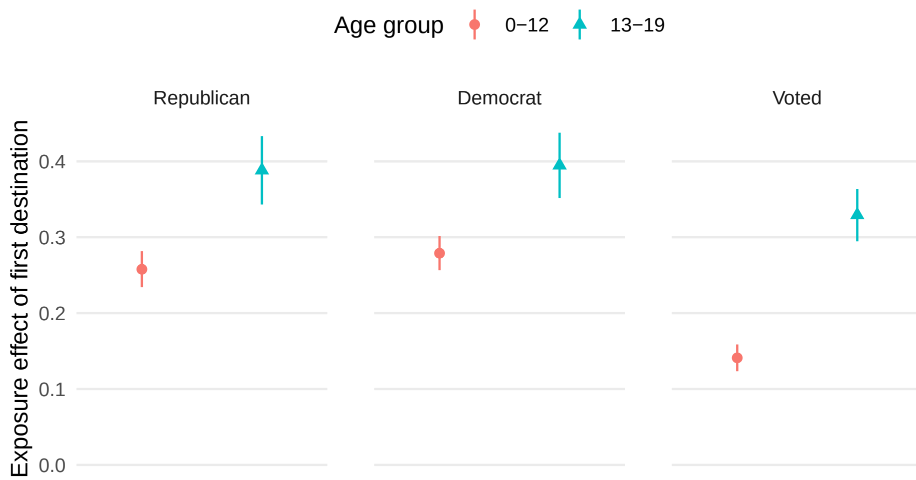
5.2 Results using multiple movers

Panel a of Figure 7 shows the estimates β_{13-19} and β_{0-12} for affiliation with the Democratic and Republican parties and turnout in one's first election.

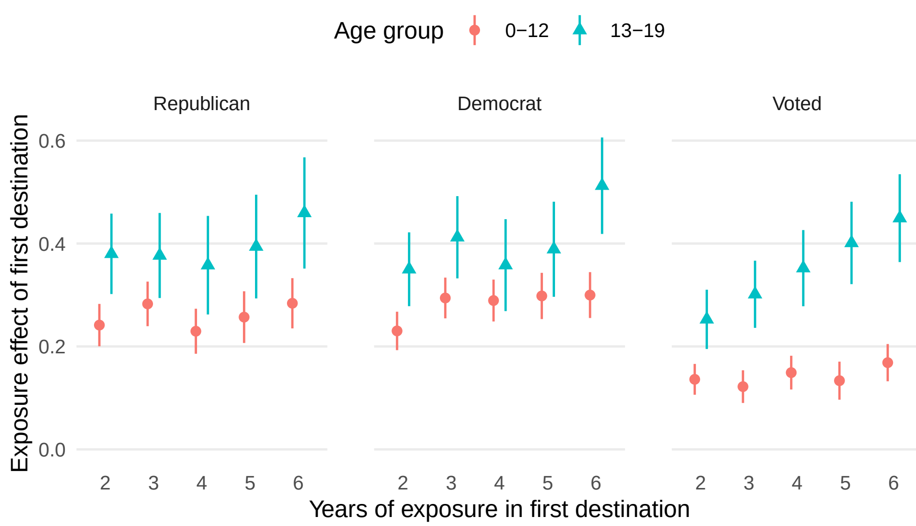
[24]Specifically, we include origin-by-cohort fixed effects α_{os} instead of fixed effects specific for each age at move, we omit origin-by-destination fixed effects, and we omit interactions between cohort fixed effects and the difference in permanent resident outcomes.

Figure 7: Effect of First Destination by Age Group for Multiple-Time Movers

(a) 2 to 6 Years of Exposure, Pooled



(b) 2 to 6 Years of Exposure, Separately



Notes: We estimate equation (6) on the sample of two or three-time movers who spent two to six years in the first destination, either during ages 0 to 12 or 13 to 19. Individuals living in the destination during both age ranges are excluded. The estimates measure the extent to which the political behavior of young movers aligns with the difference in the permanent resident outcomes between the origin and the first destination. Panel a pools individuals who spent two to six years in the destination, while Panel b shows estimates separately for each number of years of exposure. Vertical lines show 95% confidence intervals. N = 176,144 for party affiliation and N = 577,966 for turnout.

For all three outcomes, the estimates for those living two to six years in the first destination between ages 13 and 19 are substantially larger than those for individuals living there between ages 0 and 12, indicating that their behavior resembles the behavior of permanent residents in the destination more closely. For affiliation with the Republican Party, β_{0-12} and β_{13-19} are around 0.258 and 0.388. The estimates are similar for affiliation with the Democratic Party. For turnout, the levels are somewhat smaller (0.141 and 0.329), but the difference between the two estimates is larger.

While we restrict our analysis to younger and older movers spending two to six years in the first destination, the average number of years could differ slightly between the two subsamples. To address this concern, Panel b of Figure 7 shows separate estimates for exposure spells of 2, 3, 4, 5, and 6 years. We observe that later movers are more similar to permanent residents than earlier movers for every exposure spell length.

In Table 3, we present the point estimates corresponding to Panel a of Figure 7 as well as the difference between β_{13-19} and β_{0-12}. Under the assumption of constant selection, these differences recover the differential causal effect of living in the destination during adolescence rather than during childhood. For the Republican Party, that differential impact is 0.130, meaning that spending two to six years between ages 13 and 19 in a destination with a 10 p.p. higher proportion of Republicans rather than between 0 and 12 makes one 1.30 p.p. more likely to register as Republican in adulthood. The differential impact of time spent in the first destination is similar for affiliation with the Democratic Party (0.118), and slightly larger for turnout (0.188). All of these differential effects are significant at the 1% level. They indicate that neighborhood exposure during teenage years is substantially more formative of political behavior.

Table 3 also contains a robustness check: In columns 2, 4, and 6, we add fixed effects for the second and third destinations interacted with the mover's birth cohort. All estimates remain nearly unchanged and significant at the 1% level.

These estimates based on multiple-time movers can help us interpret the non-linearity that we observe for one-time movers in Figure 2 and the corresponding parametric results in Table 2. For comparison, we construct the differential impact of neighborhood exposure during adolescence years based on Table 2: We take the difference between the exposure effects during childhood and adolescence years (0 to 12 and 13 to 19) and multiply it by four, which is the average number of years of exposure among multiple-time movers. We obtain magnitudes of 0.160 ($= 4 \times (0.0520 - 0.0120)$) and 0.147 ($= 4 \times (0.0471 - 0.0103)$) for affiliation with the Republican and Democratic parties, respectively, and 0.164 ($= 4 \times (0.0493 - 0.0082)$) for turnout. These values are quite similar to and within the confidence interval of the differential impact of adolescence years in Table 3, suggesting that the non-linearity in Figure 2 is driven by the important role of formative years effects during adolescence rather than diminishing marginal exposure effects. Thus, we interpret the fact that the estimates in Table 2 are about four times as large for ages 13 to 19 as for ages 0 to 12 as evidence that exposure effects during teenage years are about four times the size of exposure effects during childhood years.

Lastly, to allow for more granular age patterns, we split our sample of multiple-time movers

Table 3: Multiple Movers, Effect of First Destination by Age Group

	Republican		Democrat		Voted	
	(1)	(2)	(3)	(4)	(5)	(6)
Exposure effect, 0-12	0.258***	0.266***	0.279***	0.280***	0.141***	0.125***
	(0.012)	(0.013)	(0.011)	(0.013)	(0.009)	(0.010)
Exposure effect, 13-19	0.388***	0.384***	0.395***	0.382***	0.329***	0.293***
	(0.023)	(0.025)	(0.022)	(0.024)	(0.018)	(0.019)
Difference $[13, 19]$ vs. $[0, 12]$	0.130	0.118	0.116	0.102	0.188	0.169
t-statistic $[13, 19]$ vs. $[0, 12]$	5.229	4.409	4.926	4.019	9.901	8.308
p-value $[13, 19] = [0, 12]$	0.000	0.000	0.000	0.000	0.000	0.000
Destination 2 & 3 FE		✓		✓		✓
Number of observations	176,144	171,707	176,144	171,707	577,966	567,711
Dependent variable mean	0.210	0.207	0.233	0.235	0.461	0.462
R^2	0.118	0.204	0.105	0.172	0.169	0.230

Notes: We estimate equation (6) on the sample of two or three-time movers who spent two to six years in the first destination, either during ages 0 to 12 or 13 to 19. Individuals living in the destination during both age ranges are excluded. The estimates measure the extent to which the political behavior of young movers aligns with the difference in the permanent resident outcomes between the origin and the first destination. Columns 1–2 (resp. 3–4) show estimates for affiliation with the Republican (resp. Democratic) Party, and columns 5–6 present estimates for turnout. Columns 1, 3, and 5 show baseline estimates corresponding to equation (6). Columns 2, 4 and 6 add destination fixed effects for the second and third destination interacted with birth cohort. We also report the difference between the estimates for ages 13 to 19 and 0 to 12, as well as the corresponding t-test statistics and p-values. This difference represents the effect of spending between two to six years (on average four years) in the first destination during ages 13 to 19 compared to spending the same amount of time in the first destination between ages 0 to 12. Heteroskedasticity-robust standard errors are in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively.

into four parts instead of two: those who lived in the first destination at some point between ages 0 to 4, 5 to 9, 10 to 14, and 15 to 19.[25] These results are shown in Appendix Figure A9. For all three outcomes, the early childhood years (0 to 4) are the least influential, and the age ranges 10 to 14 and 15 to 19 the most. The similarity between the estimates for the two older ranges enables us to rule out a possible interpretation of the patterns in Figure 7: that young individuals' behavior in their first election is mostly determined by where they live exactly when they become eligible to register and vote. This interpretation would be consistent with observing larger exposure effects for ages 13 to 19, since many of the individuals living in a given place during these ages will be there for their first election. However, it is inconsistent with the fact that the estimates for ages 10 to 14 are roughly as large as the ones for 15 to 19 in Appendix Figure A9. Overall, our evidence points to the importance of formative experiences throughout adolescence, not just contextual effects at the time of a voter's first election.

6 Mechanisms

Finally, we investigate the mechanisms responsible for the effects of childhood neighborhoods on adult partisanship and political participation shown in the previous sections. We first compare the magnitudes of exposure effects for cross-state, cross-county, and cross-ZIP code moves to assess the role of explanatory factors varying at each of these levels. We then ask whether the influence of childhood neighborhoods on future political behavior is direct or mediated by effects on socioeconomic outcomes such as income and education.

6.1 State, county, and ZIP code factors

We first consider the possibility that our effects reflect the influence of the electoral context: rules that govern how people register and vote, the competitiveness of elections, the intensity of political campaigns, and which political parties hold power. For example, in battleground states, frequent competitive presidential elections and intense campaign activity may make future voters more likely to be involved in politics (Hersh, 2015). Likewise, the political party that governors and representatives belong to and the policies they implement may shape voters' assessment of parties and their partisan affiliations (Fiorina, 1981).

These factors primarily vary at the state level. Thus, to assess the importance of the electoral context, we compare effects for people who move across states versus people who move across counties within the same state. The second group is always exposed to the same state-level factors, no matter how many years they spend in the origin versus destination county, so exposure effects on their behavior can only reflect county-specific factors such as neighborhood demographics, the type and quality of schools, the local labor market, and the local political context. Naturally, differences between the two groups of movers may also reflect differences between the types of families that move across versus within states and should thus be interpreted with caution.

[25] We continue to require that individuals are observed in the first destination in only one of these age ranges. Therefore, we estimate effects for individuals who spent between two and four (rather than six) years in the destination.

Table 4: Parametric Estimates of Exposure Effects for Moves Within and Across States

	Republican		Democrat		Voted	
	(1)	(2)	(3)	(4)	(5)	(6)
Exposure effect, 0–19	0.0215***	0.0158***	0.0199***	0.0128***	0.0217***	0.0064***
	(0.0012)	(0.0011)	(0.0013)	(0.0011)	(0.0012)	(0.0016)
Across states	✓		✓		✓	
Within state		✓		✓		✓
Mean move distance (mi.)	816.0	48.3	816.0	48.3	808.2	46.6
Number of observations	625,673	1,265,755	625,673	1,265,755	1,614,937	2,362,653
Dependent variable mean	0.189	0.206	0.255	0.241	0.459	0.444
R^2	0.303	0.203	0.266	0.160	0.364	0.259

Notes: We report annual childhood exposure effects based on equation (4), or the effects of spending an additional year of childhood in a place where the permanent residents are 1 p.p. more likely to affiliate with the Republican Party (columns 1–2), Democratic Party (columns 3–4), or vote (columns 5–6). For each outcome, the first column reports estimates restricting the sample to individuals who move across states, while the second column restricts the sample to individuals who move within the same state. All outcomes are measured for the first election for which an individual is eligible to vote. Heteroskedasticity-robust standard errors are in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively.

As shown in Table 4, exposure effects on all outcomes are larger in the cross-state sample than in the within-state sample. The within-state effects on Republican and Democratic party affiliation are about 73% and 65% the size of the cross-state effects, respectively. The difference is more pronounced for voting, where the within-state effect is just 29% the size of the cross-state effect. These results suggest that state-level factors including the electoral context and state-wide economic conditions account for some of the exposure effects on party affiliation and for much of the effect on voter turnout. However, even when these factors are held constant, the county in which people grow up affects their future political behavior.

To further identify the types of factors responsible for exposure effects, we zoom in again and compare the magnitudes of the effects for cross-county moves and within-county moves across ZIP codes.[26] Within-county moves hold all county-level economic and political factors constant, including local labor market conditions and political culture. Yet, as shown in Appendix Table A11, exposure effects remain substantial in this subsample, indicating that environmental factors varying at a small geographical scale such as the towns that children and adolescents live in, the schools they attend, and their daily interactions with peers and adults from their neighborhood greatly influence their future political behavior.

[26] For these analyses, moves are defined at the ZIP code level, rather than the county level. Therefore, one-time movers are now defined as individuals who moved exactly once across ZIP codes, and the sample includes individuals who were considered permanent residents in the county-level analysis.

6.2 Income and education

Next, we examine whether childhood environment effects on political behavior are mediated by effects on socioeconomic indicators. What makes this mechanism plausible is prior evidence that turnout and partisanship are strongly correlated with income and education (e.g., Wolfinger and Rosenstone, 1980; Verba et al., 1995; Milligan et al., 2004; Gelman et al., 2008) and that childhood environments have strong effects on the latter outcomes (Chetty and Hendren, 2018a; Chetty et al., 2018).

To test for this mediating relationship, we use data from Chetty et al. (2018) and restrict the sample to voters who moved between counties belonging to the same nationwide quartiles of expected college attendance rate or expected income, both defined for individuals with parents at the median of the income distribution. These sample restrictions should thus reduce the effects of childhood environment on political behavior if education and income are important mediating factors. In a third restriction, we focus on moves between counties in the same quartile of both the expected college attendance rate and expected income. Figure 8 shows the results. The estimates in the subsamples are generally very similar to those in the full sample. We only find one statistically significant difference: the effect on Republican Party affiliation is slightly smaller for the sample of voters moving between counties with similar effects on college mobility. We infer that place effects on future partisanship and political participation are direct rather than mediated by effects on socioeconomic outcomes.

7 Conclusion

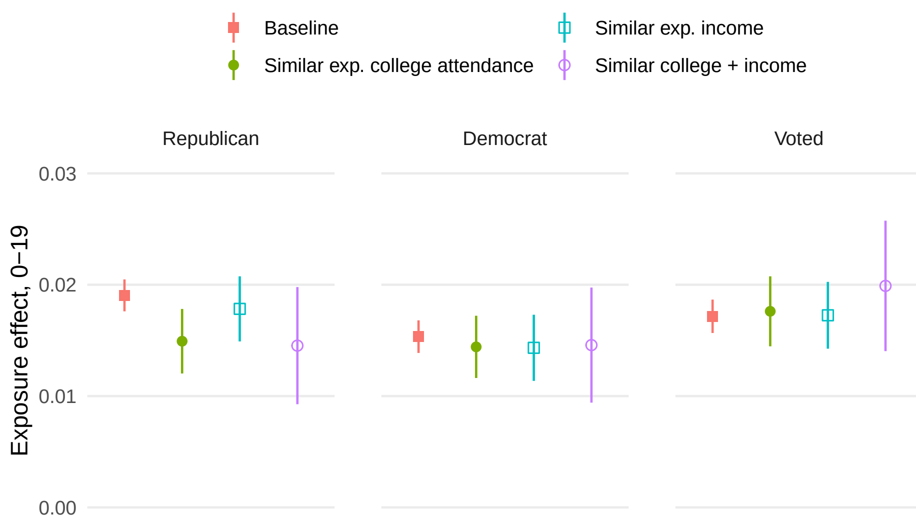
In this paper, we estimate the total impact of childhood environment on young adults' party affiliation and political participation. Our analysis combines data on all registered voters in the U.S. since 2012 with comprehensive address history data going back to 1992. We compare individuals who moved between identical counties before they turned 18 and find that those who moved earlier and spent more time in the destination behave more similarly to the permanent residents of the same cohort in that place.

Neighborhood exposure effects on future political behavior are strongest during adolescence and much more limited during early childhood. These patterns are not due to diminishing marginal returns of exposure, but to the existence of formative years in which individuals are most influenced by their environment.

Overall, growing up in a county where one's peers are 10 p.p. more likely to become Republicans makes someone 4.72 p.p. more likely to become a Republican themselves upon entering the electorate. The effects are of similar magnitude for Democratic partisanship (4.05 p.p.) and turnout in the first election (4.75 p.p.). The effects on partisanship decline slightly but persist over time, despite the competing influence of other factors. For turnout, we observe a far more substantial decline in effect size by the third election.

Our estimates imply that 40 to 50% of the cross-county variation in the partisanship and turnout of voters entering the electorate is due to the effects of childhood environment. In Chetty and Hendren (2018a), the share of cross-county variation in future earnings that

Figure 8: Exposure Effects for Moves Between Places with Similar Characteristics



Notes: We show estimates of equation (4) in samples restricted to individuals moving between counties belonging to the same nationwide quartile of expected college attendance rate, the same quartile of expected income, and the same quartiles of both expected college attendance rate and expected income. Expected college attendance rate and expected income are from Chetty et al. (2018) and are measured for individuals with parents at the median of the income distribution. Vertical lines show 95% confidence intervals. $N = 3{,}665{,}703$ for party affiliation and $N = 7{,}377{,}997$ for turnout.

is due to childhood place effects is somewhat larger: 70%. However, we do not find any evidence that the effects on political behavior are mediated by effects on income or education. Instead, our investigation of mechanisms suggests that these effects are direct. State-level factors such as electoral context account for much of the impact on participation. Small-scale environmental factors varying at the ZIP code level, such as peer effects, explain a larger share of the impact on their future party affiliation.

Overall, these results point to the importance of geographic context during childhood in shaping future political behavior. Despite strong influences of family and individual characteristics on future political outcomes, childhood location is a powerful determinant of who voters become politically – which parties they adopt as their own and how engaged they become with the democratic process. These findings also have broader implications for contemporary American politics, particularly in light of an increasingly segregated electorate (Brown and Enos, 2021; Kaplan et al., 2022). The more unevenly distributed supporters of different parties are across places, the more consequential residential choice becomes for shaping future political behavior. Today, the effects of childhood environment may be augmented by residential segregation and contribute to ongoing geographic polarization.

References

Bartels, Larry M., "Partisanship and Voting Behavior, 1952-1996," *American Journal of Political Science*, 2000, *44* (1), 35–50.

Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen, "Who Becomes an Inventor in America? The Importance of Exposure to Innovation," *The Quarterly Journal of Economics*, 2018, *134* (2), 647–713.

Berry, Christoper and Jessica Trounstine, "Does Where You Live Shape Your Politics? Measuring the Effect of Neighborhood on Political Identity and Behavior," 2023. Working Paper, University of Chicago.

Billings, Stephen B., Eric Chyn, and Kareem Haggag, "The Long-Run Effects of School Racial Diversity on Political Identity," *American Economic Review: Insights*, 2021, *3* (3), 267–84.

Bolotnyy, Valentin, Mayya Komisarchik, and Brian Libgober, "How Does Childhood Environment Shape Political Participation? Evidence from Refugees.," 2022. Working Paper.

Brown, Jacob R., "Partisan Conversion Through Neighborhood Influence: How Voters Adopt the Partisanship of their Neighbors," 2023. Working Paper.

_ and Ryan D. Enos, "The measurement of partisan sorting for 180 million voters," *Nature Human Behaviour*, 2021, *5* (8), 998–1008.

_ , _ , James Feigenbaum, and Soumyajit Mazumder, "Childhood cross-ethnic exposure predicts political behavior seven decades later: Evidence from linked administrative data," *Science Advances*, 2021, *7* (24).

Campbell, David E., *Why We Vote: How Schools and Communities Shape Our Civic Life* Princeton Studies in American Politics, Princeton University Press, 2006.

Cantoni, Enrico and Vincent Pons, "Does Context Outweigh Individual Characteristics in Driving Voting Behavior? Evidence from Relocations within the United States," *American Economic Review*, 2022, *112* (4), 1226–72.

Chetty, Raj and Nathaniel Hendren, "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects," *The Quarterly Journal of Economics*, 2018, *133* (3), 1107–1162.

_ and _ , "The impacts of neighborhoods on intergenerational mobility II: County-level estimates," *The Quarterly Journal of Economics*, 2018, *133* (3), 1163–1228.

_ , John N. Friedman, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter, "The Opportunity Atlas: Mapping the childhood roots of social mobility," Technical Report, National Bureau of Economic Research 2018.

Chyn, Eric, "Moved to opportunity: The long-run effects of public housing demolition on children," *American Economic Review*, 2018, *108* (10), 3028–56.

— **and Kareem Haggag**, "Moved to Vote: The Long-Run Effects of Neighborhoods on Political Participation," Technical Report, National Bureau of Economic Research 2019.

— **and Lawrence F. Katz**, "Neighborhoods Matter: Assessing the Evidence for Place Effects," *Journal of Economic Perspectives*, 2021, *35* (4), 197–222.

Cohodes, Sarah and James J Feigenbaum, "Why Does Education Increase Voting? Evidence from Boston's Charter Schools," Working Paper 29308, National Bureau of Economic Research September 2021.

Damm, Anna Piil and Christian Dustmann, "Does Growing Up in a High Crime Neighborhood Affect Youth Criminal Behavior?," *American Economic Review*, June 2014, *104* (6), 1806–32.

Derenoncourt, Ellora, "Can You Move to Opportunity? Evidence from the Great Migration," *American Economic Review*, 2022, *112* (2), 369–408.

Deutscher, Nathan, "Place, Peers, and the Teenage Years: Long-Run Neighborhood Effects in Australia," *American Economic Journal: Applied Economics*, April 2020, *12* (2), 220–49.

Diamond, Rebecca, Tim McQuade, and Franklin Qian, "The effects of rent control expansion on tenants, landlords, and inequality: Evidence from San Francisco," *American Economic Review*, 2019, *109* (9), 3365–94.

Dinas, Elias, "Does Choice Bring Loyalty? Electoral Participation and the Development of Party Identification," *American Journal of Political Science*, 2014, *58* (2), 449–465.

Elder, Elizabeth Mitchell, Ryan D. Enos, and Tali Mendelberg, "The Long-Term Effects of Neighborhood Disadvantage on Voting Behavior: The "Moving to Opportunity" Experiment," *American Political Science Review*, 2023, p. 1–17.

Fiorina, Morris P., *Retrospective Voting in American National Elections*, New Haven, CT: Yale University Press, 1981.

Fowler, James H., Laura A. Baker, and Christopher T. Dawes, "Genetic variation in political participation," *American Political Science Review*, 2008, *102* (2), 233–248.

Gelman, Andrew, Boris Shor, Joseph Bafumi, and David Park, "Rich State, Poor State, Red State, Blue State: What's the Matter with Connecticut?," *Quarterly Journal of Political Science*, 2008, *2* (4), 345–367.

Ghitza, Yair, Andrew Gelman, and Jonathan Auerbach, "The Great Society, Reagan's Revolution, and Generations of Presidential Voting," *American Journal of Political Science*, 2023, *67* (3), 520–537.

Goel, Sharad, Marc Meredith, Michael Morse, David Rothschild, and Houshmand Shirani-Mehr, "One Person, One Vote: Estimating the Prevalence of Double Voting in U.S. Presidential Elections," *American Political Science Review*, 2020, *114* (2), 456–469.

Green, Donald, Bradley Palmquist, and Eric Schickler, *Partisan hearts and minds*, Yale University Press, 2008.

Guo, Fang, "Partisan Registration: A Truthful Statement or A Strategic Choice?," Working Paper 2023.

Healy, Andrew and Neil Malhotra, "Childhood socialization and political attitudes: Evidence from a natural experiment," *The Journal of Politics*, 2013, *75* (4), 1023–1037.

Hersh, Eitan, *Hacking the Electorate: How Campaigns Perceive Voters*, Cambridge, UK: Cambridge University Press, 2015.

_ **and Yair Ghitza**, "Mixed partisan households and electoral participation in the United States," *PLOS ONE*, 10 2018, *13* (10), 1–18.

Huckfeldt, Robert and John Sprague, "Networks in Context: The Social Flow of Political Information," *American Political Science Review*, 1987, *81* (4), 1197–1216.

Janmaat, Jan Germen and Bryony Hoskins, "The Changing Impact of Family Background on Political Engagement During Adolescence and Early Adulthood," *Social Forces*, 09 2021, *101* (1), 227–251.

Jennings, M. Kent, Laura Stoker, and Jake Bowers, "Politics across Generations: Family Transmission Reexamined," *The Journal of Politics*, 2009, *71* (3), 782–799.

Kaplan, Ethan, Jörg L. Spenkuch, and Cody Tuttle, "School Desegregation and Political Preferences: Long-Run Evidence from Kentucky," 2021. Working Paper, University of Marylabnd.

_ **, Jörg L. Spenkuch, and Rebecca Sullivan**, "Partisan spatial sorting in the United States: A theoretical and empirical overview," *Journal of Public Economics*, 2022, *211*, 104668.

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz, "Experimental Analysis of Neighborhood Effects," *Econometrica*, 2007, *75* (1), 83–119.

Laliberté, Jean-William, "Long-Term Contextual Effects in Education: Schools and Neighborhoods," *American Economic Journal: Economic Policy*, 2021, *13* (2), 336–77.

McDonald, Michael P., "Presidential Voter Turnout Rates, 2020," Technical Report, United States Election Project 2020.

_ , "Voter Turnout Data, 2020," 2022.

McNeil, Andrew, Davide Luca, and Neil Lee, "The long shadow of local decline: Birthplace economic adversity and long-term individual outcomes in the UK," *Journal of Urban Economics*, 2023, *136*, 103571.

Milligan, Kevin, Enrico Moretti, and Philip Oreopoulos, "Does education improve citizenship? Evidence from the United States and the United Kingdom," *Journal of Public Economics*, 2004, *88* (9), 1667–1695.

Mullainathan, Sendhil and Ebonya Washington, "Sticking with Your Vote: Cognitive Dissonance and Political Attitudes," *American Economic Journal: Applied Economics*, 2009, *1* (1), 86–111.

Nakamura, Emi, Jósef Sigurdsson, and Jón Steinsson, "The Gift of Moving: Intergenerational Consequences of a Mobility Shock," *The Review of Economic Studies*, 09 2021, *89* (3), 1557–1592.

Neundorf, Anja and Kaat Smets, "Political Socialization and the Making of Citizens," in "Oxford Handbook Topics in Politics," Oxford University Press, 2017.

Nyhan, Brendan, Christopher Skovron, and Rocío Titiunik, "Differential registration bias in voter file data: A sensitivity analysis approach," *American Journal of Political Science*, 2017, *61* (3), 744–760.

Pacheco, Julianna Sandell, "Political Socialization in Context: The Effect of Political Competition on Youth Voter Turnout," *Political Behavior*, 2008, *30* (4), 415–436.

Pennington, Kate, "Does Building New Housing Cause Displacement?: The Supply and Demand Effects of Construction in San Francisco," Working Paper 2021.

Perez-Truglia, Ricardo, "Political conformity: Event-study evidence from the United States," *Review of Economics and Statistics*, 2018, *100* (1), 14–28.

Pons, Vincent and Guillaume Liegey, "Increasing the Electoral Participation of Immigrants: Experimental Evidence from France," *The Economic Journal*, 12 2018, *129* (617), 481–508.

Sears, David O. and Nicholas A. Valentino, "Politics Matters: Political Events as Catalysts for Preadult Socialization," *The American Political Science Review*, 1997, *91* (1), 45–65.

U.S. Census Bureau, "Voting and registration in the election of November 2020," October 2021.

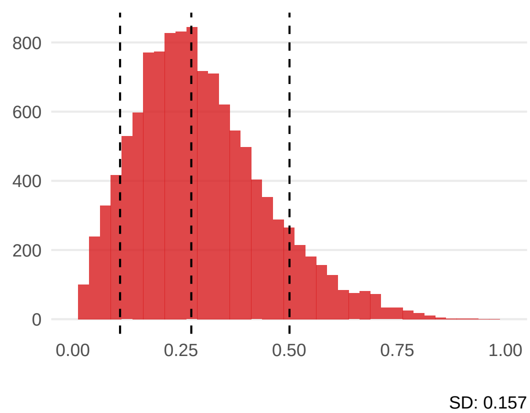
Verba, Sidney, Kay Lehman Schlozman, and Henry E. Brady, *Voice and Equality: Civic Voluntarism in American Politics*, Cambridge, MA: Harvard University Press, 1995.

Wolfinger, R.E. and S.J. Rosenstone, *Who Votes?* Yale Fastback Series, Yale University Press, 1980.

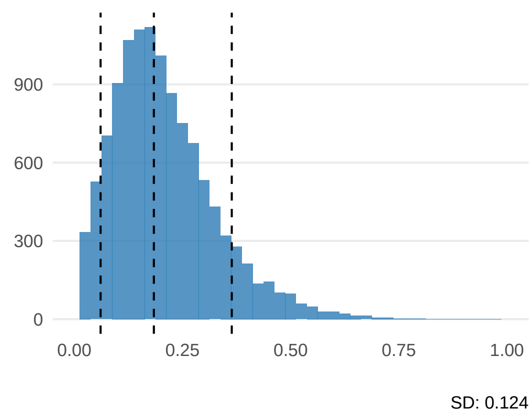
A Additional tables and figures

Figure A1: Distributions of Permanent Resident Outcomes

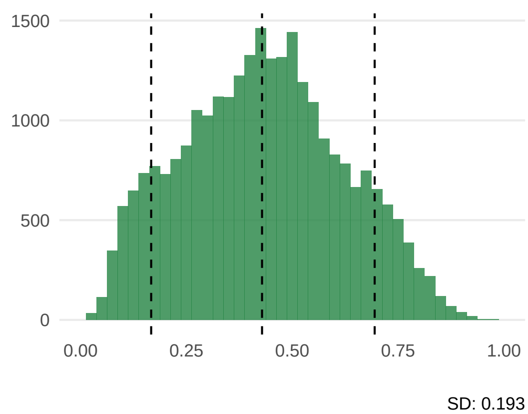
(a) Republican



(b) Democrat



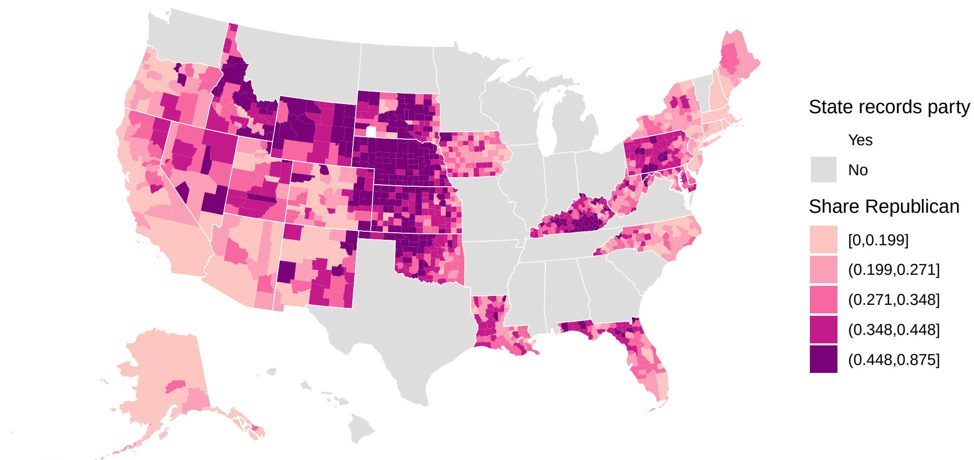
(c) Voted



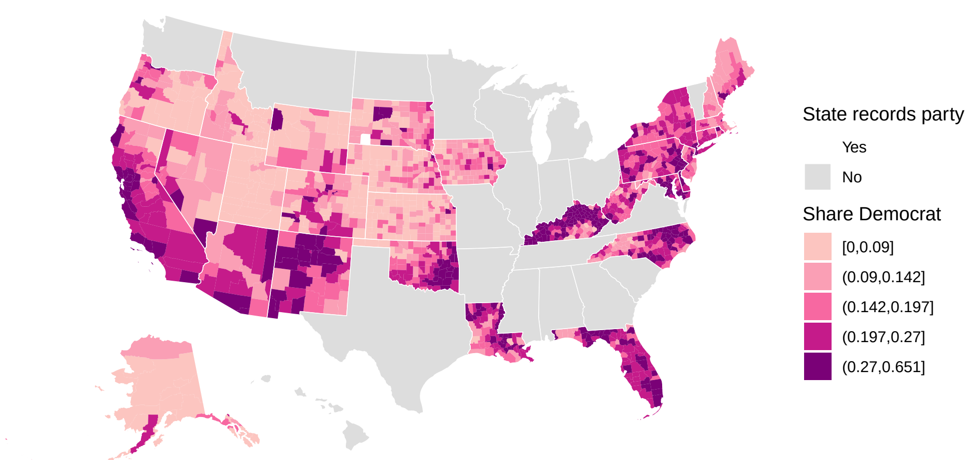
Notes: We show the distributions of mean outcomes for permanent residents in each county-by-birth cohort bin. All outcomes are measured for the first election for which an individual is eligible to vote. We restrict the sample to county-by-cohort bins with more than 25 people. Vertical lines show the (unweighted) 10th percentile, median, and 90th percentile of each distribution, and the (unweighted) standard deviation is reported at the bottom right of each panel. $N = 6{,}844{,}388$ for party affiliation and $N = 11{,}884{,}645$ for turnout.

Figure A2: Exposure Maps for Party Affiliation

(a) Republican

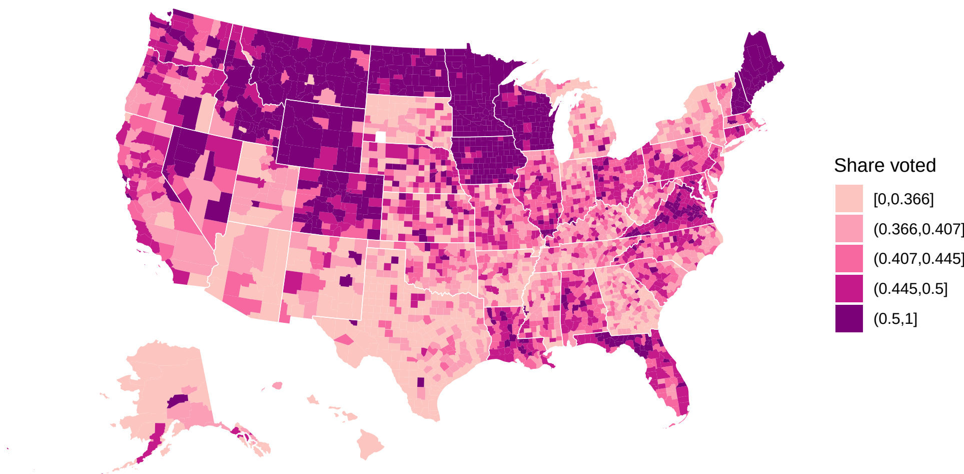


(b) Democrat



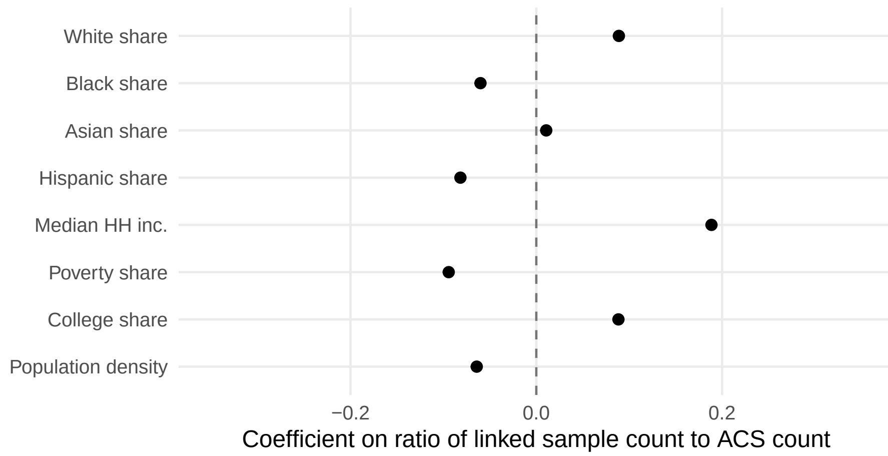
Notes: Panel a shows the proportion of permanent residents in each county who registered as Republican in their first eligible election. Panel b shows the analogous statistic for registration with the Democratic Party. These exposure measures are averaged over all birth cohorts in the sample of permanent residents. States in gray do not release party affiliation in the voter file. $N =$ 6,886,465.

Figure A3: Exposure Map for Turnout



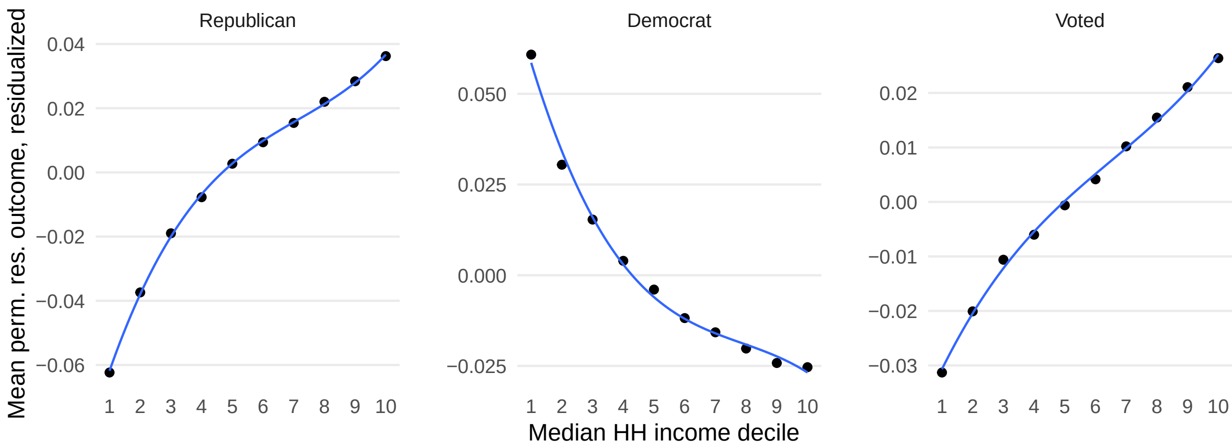
Notes: We show the proportion of permanent residents who voted in their first eligible election. These exposure measures are averaged over all birth cohorts in the sample of permanent residents. N = 11,970,722.

Figure A4: Number of Individuals in Each Block Group, Linked Sample vs. American Community Survey



Notes: Each row shows the coefficient from a univariate regression at the block group level. The dependent variable is a ratio, where the numerator is the number of individuals in the linked sample of young people in a given block group between 2015 and 2019. Individuals who appear in more than one voter file snapshot are counted in the location where they register each year that they do so. The denominator is the 2015–2019 ACS 5-year estimate of the number of individuals aged 18 to 29 in that block group. The independent variables come from the 2015–2019 ACS. The dependent variable and all independent variables are standardized using population weights. A coefficient of zero means that the covariate does not predict the ratio between counts in our sample and the ACS counts, indicating high correspondence. A positive coefficient means that our sample over-counts people in block groups with higher values of the ACS variable, while a negative coefficient means the opposite. Horizontal lines show 95% confidence intervals. N = 22,587,485.

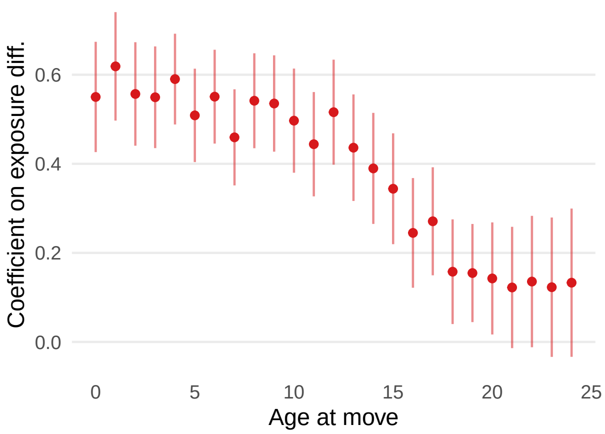
Figure A5: Relationship Between Permanent Resident Outcomes and Parent Income Decile



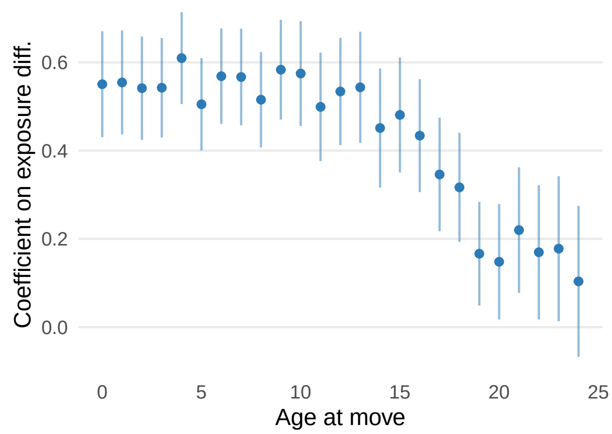
Notes: We show the relationships between the three primary outcomes and parent income decile, inferred from the median income of their block group, in the sample of permanent residents. All outcomes are measured at the first election for which an individual is eligible to vote, and are demeaned by county and birth cohort. Cubic polynomial fits are shown in blue. $N = 6{,}715{,}927$ for party affiliation and $N = 11{,}690{,}477$ for turnout.

Figure A6: Sibling Comparisons: Nonparametric Estimates

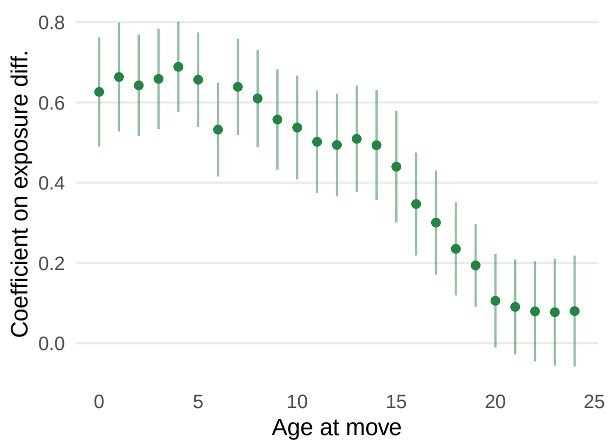
(a) Republican



(b) Democrat

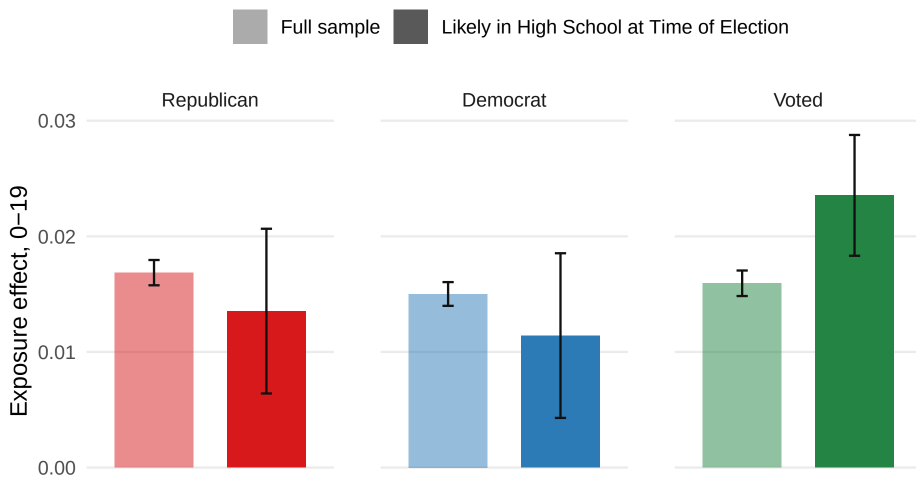


(c) Voted



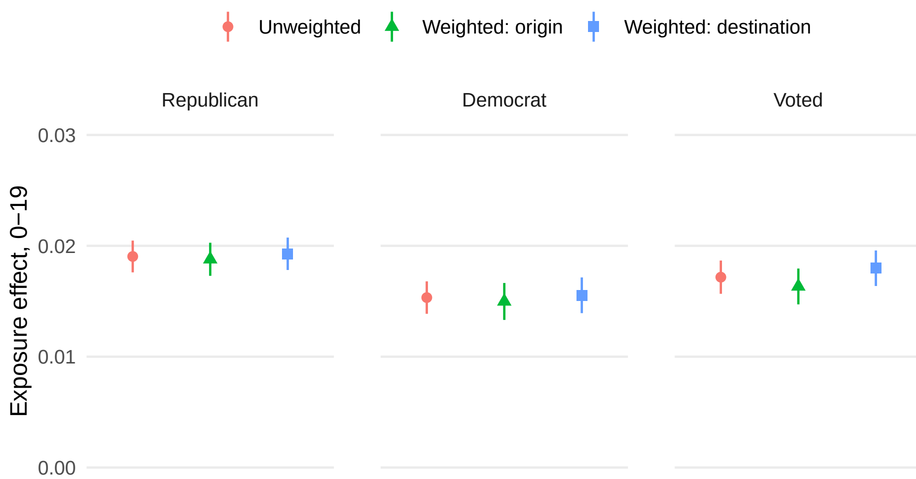
Notes: We show annual exposure effects identified using siblings of different ages with the same parent. See Section 4.2.1 for details. All outcomes are measured for the first election for which an individual is eligible to vote. Vertical lines show 95% confidence intervals. N = 750,764 for party affiliation and 1,613,101 for turnout.

Figure A7: Exposure Effects: Full Sample versus Likely High School Sample



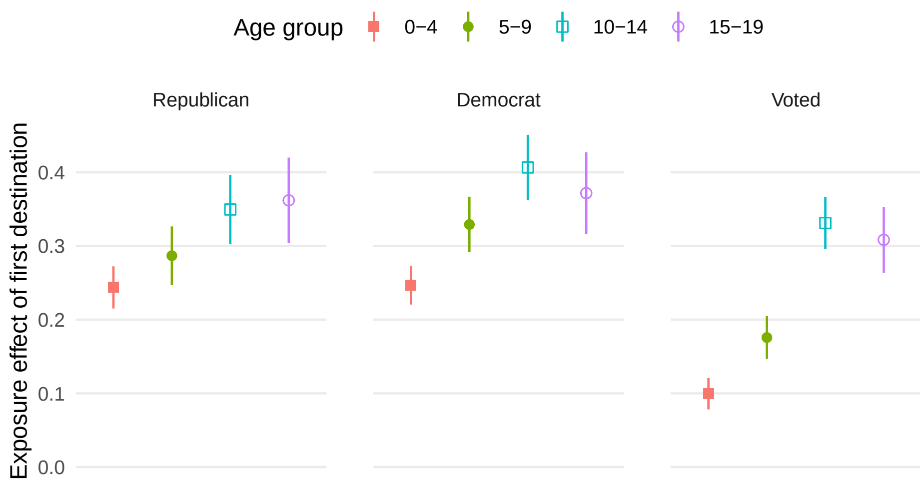
Notes: We show exposure estimates separately for the full sample (in lighter colors) and for the sample of barely 18-year olds who were born after the state-specific school entry age cutoff (in darker colors). If they follow the normal trajectory, these individuals were likely still in high school at the time of their first election. Vertical lines show 95% confidence intervals. For the full sample, N = 1,635,379 for party affiliation and N = 3,415,738 for turnout. For the restricted sample, N = 35,489 for party affiliation and N = 128,930 for turnout.

Figure A8: Weighting Observations to Match County Populations



Notes: We show exposure effects after weighting observations to match the number of voting-eligible individuals between 18 and 29 in each county according to the 2015–2019 5-year ACS, source table: Citizen, Voting-Age Population by Age. The orange series shows the baseline (unweighted) estimates, the green series shows estimates weighted by movers' counties of origin, and the blue series shows estimates weighted by movers' destination counties. Vertical lines show 95% confidence intervals. See Section 4.4 for details. N = 1,927,349 for party affiliation and 4,073,352 for turnout.

Figure A9: Effect of First Destination by Age Group: Four Age Groups



Notes: We show estimates corresponding to those in Figure 7, except that we divide the 0 to 19 range into four age groups instead of two. That is, we estimate equation (6) on the sample of two or three-time movers who spent two to four years in the first destination, either during ages 0 to 4, 5 to 9, 10 to 14, or 15 to 19. Individuals living in the destination during more than one of these age ranges are excluded from the analysis. Vertical lines show 95% confidence intervals. $N = 176{,}144$ for party affiliation and $N = 577{,}966$ for turnout.

Table A1: Parametric Estimates of Exposure Effects – Exact Birth Date Sample

	Republican			Democrat			Voted		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Exposure effect, 0–12	0.0118***	0.0125***	0.0124**	0.0114***	0.0095***	0.0079	0.0109***	0.0100***	0.0169***
	(0.0016)	(0.0021)	(0.0062)	(0.0016)	(0.0021)	(0.0064)	(0.0018)	(0.0027)	(0.0065)
Exposure effect, 13–19	0.0491***	0.0367***	0.0432***	0.0481***	0.0443***	0.0585***	0.0485***	0.0435***	0.0418***
	(0.0041)	(0.0056)	(0.0104)	(0.0045)	(0.0061)	(0.0117)	(0.0046)	(0.0070)	(0.0129)
Exposure effect, 20–24	0.0104	0.0068	-0.0003	0.0182***	0.0270***	0.0262	0.0250***	0.0207**	0.0146
	(0.0064)	(0.0087)	(0.0144)	(0.0070)	(0.0095)	(0.0160)	(0.0068)	(0.0101)	(0.0178)
Difference $[13, 19]$ vs. $[0, 12]$	0.037	0.024	0.031	0.037	0.035	0.051	0.038	0.034	0.025
t-statistic $[13, 19]$ vs. $[0, 12]$	8.4944	3.9935	2.3515	7.6883	5.3944	3.4850	7.6163	4.4431	1.6372
p-value $[13, 19] = [0, 12]$	0.0000	0.0001	0.0187	0.0000	0.0000	0.0005	0.0000	0.0000	0.1016
Control for income in origin		✓			✓			✓	
Family fixed effects			✓			✓			✓
Number of observations	1,532,038	1,262,502	586,633	1,532,038	1,262,502	586,633	3,062,534	2,381,290	1,154,296
Dependent variable mean	0.194	0.186	0.211	0.256	0.261	0.247	0.446	0.447	0.457
R^2	0.208	0.365	0.735	0.170	0.323	0.701	0.284	0.438	0.695

Notes: We report annual childhood exposure effects, excluding voters whose recorded birth dates indicate that they were born on January 1. It is likely that only the year of birth, rather than the exact date of birth, was recorded for these voters. This restriction applies to both movers and permanent residents. The regression specification is equation (5), and the estimates can be compared to those in Table 2. Heteroskedasticity-robust standard errors are in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively.

Table A2: Median ACS Characteristics for Various Samples

(a) County

	ACS	TargetSmart	Linked to parents	Linked to Infutor
White share	0.748	0.747	0.754	0.755
Black share	0.083	0.084	0.083	0.083
Hispanic share	0.111	0.118	0.108	0.107
Asian share	0.034	0.037	0.036	0.036
College share	0.605	0.609	0.612	0.612
Median HH income	61,705	63,299	64,335	64,304
Poverty rate	0.138	0.136	0.128	0.127
Pop. density (per sq. mi.)	555	614	614	610
Num. of obs.	49,855,197	16,006,592	10,910,603	10,596,560

(b) Block Group

	ACS	TargetSmart	Linked to parents	Linked to Infutor
White share	0.776	0.801	0.831	0.833
Black share	0.05	0.039	0.03	0.03
Hispanic share	0.089	0.081	0.069	0.068
Asian share	0.016	0.017	0.018	0.017
College share	0.602	0.619	0.646	0.648
Median HH income	59,773	65,750	72,708	72,986
Poverty rate	0.116	0.094	0.076	0.075
Pop. density (per sq. mi.)	3,246	2,681	2,304	2,281
Num. of obs.	49,855,197	16,002,639	10,907,687	10,593,710

Notes: We show weighted medians of 2015–2019 ACS 5-year estimates at the county and block group levels. In the first column, we weight each county or block group by the ACS count of 18 to 29 year old citizens. In the second column, we weight observations by the number of individuals in the 2015 through 2019 TargetSmart snapshots between ages 18 and 24. Individuals who appear in more than one voter file snapshot are counted in the location where they register each year that they do so. In the third column, we weight by the number of 18 to 24 year old individuals in the 2015 through 2019 TargetSmart snapshots whom we are able to link to a parent. In the last column, we weight by the number of 18 to 24 year old individuals in the 2015 through 2019 TargetSmart snapshots linked to a parent for whom we are able to retrieve address data from Infutor. This corresponds to a subset of our analysis sample of permanent residents and movers for 2015 through 2019.

Table A3: Parametric Estimates of Exposure Effects – Clustering of Standard Errors

	Republican			Democrat			Voted		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Exposure effect, 0–12	0.0120***	0.0120***	0.0120***	0.0103***	0.0103***	0.0103***	0.0082***	0.0082***	0.0082***
	(0.0015)	(0.0015)	(0.0014)	(0.0015)	(0.0014)	(0.0014)	(0.0015)	(0.0015)	(0.0016)
Exposure effect, 13–19	0.0520***	0.0520***	0.0520***	0.0471***	0.0471***	0.0471***	0.0493***	0.0493***	0.0493***
	(0.0037)	(0.0039)	(0.0038)	(0.0039)	(0.0040)	(0.0039)	(0.0040)	(0.0039)	(0.0039)
Exposure effect, 20–24	0.0114*	0.0114*	0.0114*	0.0147**	0.0147**	0.0147**	0.0265***	0.0265***	0.0265***
	(0.0060)	(0.0061)	(0.0060)	(0.0064)	(0.0062)	(0.0066)	(0.0063)	(0.0063)	(0.0060)
Difference $[13, 19]$ vs. $[0, 12]$	0.040	0.040	0.040	0.037	0.037	0.037	0.041	0.041	0.041
t-statistic $[13, 19]$ vs. $[0, 12]$	9.8499	9.5466	9.8286	8.6811	8.6823	8.8682	9.6209	9.7085	9.8190
p-value $[13, 19] = [0, 12]$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Cluster by family	✓			✓			✓		
Cluster by origin × destination		✓			✓			✓	
Cluster by origin × age at move			✓			✓			✓
Number of observations	1,927,376	1,927,376	1,927,376	1,927,376	1,927,376	1,927,376	4,073,607	4,073,607	4,073,607
Dependent variable mean	0.202	0.202	0.202	0.245	0.245	0.245	0.450	0.450	0.450
R^2	0.215	0.215	0.215	0.177	0.177	0.177	0.282	0.282	0.282

Notes: We report annual childhood exposure effects, clustering the standard errors at various levels. In columns 1, 4, and 7, standard errors are clustered by destination-origin county pair to correspond to the level of the treatment. In columns 2, 5, and 8, they are clustered by origin-by-age-at-move to account for the assignment mechanism of the treatment. In columns 3, 6, and 9, they are clustered by family to account for potential correlation between siblings. The regression specification is equation (5), and the estimates can be compared to those in Table 2. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively.

Table A4: Parametric Estimates of Exposure Effects – Two and Three Age Groups

	Republican		Democrat		Voted	
	(1)	(2)	(3)	(4)	(5)	(6)
Exposure effect, 0–19	0.0190***		0.0153***		0.0172***	
	(0.0007)		(0.0007)		(0.0008)	
Exposure effect, 0–12		0.0120***		0.0103***		0.0082***
		(0.0014)		(0.0014)		(0.0015)
Exposure effect, 13–19		0.0520***		0.0471***		0.0493***
		(0.0037)		(0.0039)		(0.0039)
Exposure effect, 20–24	0.0129**	0.0114*	0.0160**	0.0147**	0.0280***	0.0265***
	(0.0060)	(0.0060)	(0.0064)	(0.0064)	(0.0061)	(0.0061)
Number of observations	1,927,376	1,927,376	1,927,376	1,927,376	4,073,607	4,073,607
Dependent variable mean	0.202	0.202	0.245	0.245	0.450	0.450
R^2	0.215	0.215	0.177	0.177	0.282	0.282

Notes: We report annual childhood exposure effects, constraining the exposure effect to be the same for ages 0 to 19. The regression specification for columns 1, 3, and 5 is equation (4). For columns 2, 4, and 6, the regression specification is equation (5), and the estimates are identical to those in Table 2. Heteroskedasticity-robust standard errors are in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively.

Table A5: Parametric Estimates of Exposure Effects – Siblings Sample

	Republican		Democrat		Voted	
	(1)	(2)	(3)	(4)	(5)	(6)
Exposure effect, 0–12	0.0099***	0.0083	0.0077***	0.0011	0.0055*	0.0146***
	(0.0027)	(0.0057)	(0.0026)	(0.0057)	(0.0028)	(0.0053)
Exposure effect, 13–19	0.0474***	0.0487***	0.0514***	0.0588***	0.0581***	0.0569***
	(0.0063)	(0.0099)	(0.0067)	(0.0106)	(0.0067)	(0.0107)
Exposure effect, 20–24	0.0051	0.0030	0.0079	0.0073	0.0161	0.0078
	(0.0101)	(0.0138)	(0.0108)	(0.0148)	(0.0103)	(0.0158)
Difference $[13, 19]$ vs. $[0, 12]$	0.037	0.040	0.044	0.058	0.053	0.042
t-statistic $[13, 19]$ vs. $[0, 12]$	5.3690	3.3889	6.0407	4.5552	7.1581	3.3503
p-value $[13, 19] = [0, 12]$	0.0000	0.0007	0.0000	0.0000	0.0000	0.0008
Family fixed effects		✓		✓		✓
Number of observations	750,764	750,764	750,764	750,764	1,613,101	1,613,101
Dependent variable mean	0.219	0.219	0.235	0.235	0.460	0.460
R^2	0.300	0.738	0.248	0.704	0.358	0.689

Notes: We report annual childhood exposure effects, restricting the sample to individuals with siblings. The regression specification is equation (5), and the estimates can be compared to those in Table 2. Heteroskedasticity-robust standard errors are in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively.

Table A6: Overidentification Tests

(a) Cohort

	Republican		Democrat		Voted	
	(1)	(2)	(3)	(4)	(5)	(6)
-4 to -2		0.0008 (0.0020)		0.0003 (0.0018)		0.0002 (0.0011)
-1 to +1	0.0178*** (0.0010)	0.0154*** (0.0033)	0.0133*** (0.0010)	0.0115*** (0.0032)	0.0183*** (0.0012)	0.0224*** (0.0017)
+2 to +4		0.0015 (0.0024)		0.0013 (0.0023)		-0.0057*** (0.0013)
Num. of obs.	1,357,158	1,357,158	1,357,158	1,357,158	2,838,093	2,838,093
R^2	0.215	0.215	0.173	0.173	0.249	0.249

(b) Gender

	Republican		Democrat		Voted	
	(1)	(2)	(3)	(4)	(5)	(6)
Same gender	0.0171*** (0.0007)	0.0119*** (0.0012)	0.0146*** (0.0007)	0.0107*** (0.0012)	0.0142*** (0.0006)	0.0103*** (0.0008)
Other gender		0.0065*** (0.0012)		0.0044*** (0.0012)		0.0059*** (0.0008)
Num. of obs.	1,924,178	1,924,178	1,924,178	1,924,178	4,064,053	4,064,053
R^2	0.220	0.220	0.185	0.186	0.285	0.285

(c) Race

	Republican		Democrat		Voted	
	(1)	(2)	(3)	(4)	(5)	(6)
Same race	0.0189*** (0.0007)	0.0184*** (0.0008)	0.0160*** (0.0008)	0.0153*** (0.0008)	0.0158*** (0.0007)	0.0153*** (0.0007)
Other race		0.0025*** (0.0008)		0.0014** (0.0006)		0.0007 (0.0004)
Num. of obs.	1,766,839	1,766,839	1,766,839	1,766,839	3,712,474	3,712,474
R^2	0.217	0.217	0.189	0.190	0.275	0.275

Notes: We show the results of outcome-based overidentification tests described in Section 4.2.2 and summarized in Figure 3. There are three distinct tests exploiting heterogeneity in permanent resident outcomes by cohort in Panel a, gender in Panel b, and race in Panel c. For each outcome, the first column shows the exposure effect estimated using permanent resident outcomes specific to the mover's cohort group, gender, or race. The second column adds exposure measures for the other cohort groups, gender, or races. Race is collapsed into four categories: white, Black, Hispanic, and other. We exclude voters missing race information. Heteroskedasticity-robust standard errors are in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively.

Table A7: Parametric Estimates of Exposure Effects – Registered Voters Sample

	Republican			Democrat			Voted		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Exposure effect, 0–12	0.0120***	0.0116***	0.0086	0.0100***	0.0094***	0.0007	0.0080***	0.0065***	0.0145***
	(0.0014)	(0.0019)	(0.0057)	(0.0014)	(0.0019)	(0.0057)	(0.0015)	(0.0022)	(0.0053)
Exposure effect, 13–19	0.0517***	0.0424***	0.0468***	0.0471***	0.0424***	0.0596***	0.0493***	0.0451***	0.0559***
	(0.0037)	(0.0051)	(0.0099)	(0.0039)	(0.0054)	(0.0106)	(0.0039)	(0.0058)	(0.0107)
Exposure effect, 20–24	0.0113*	0.0041	0.0010	0.0148**	0.0200**	0.0086	0.0255***	0.0165*	0.0070
	(0.0059)	(0.0081)	(0.0137)	(0.0064)	(0.0087)	(0.0147)	(0.0061)	(0.0090)	(0.0157)
Difference $[13,19]$ vs. $[0,12]$	0.040	0.031	0.038	0.037	0.033	0.059	0.041	0.039	0.041
t-statistic $[13,19]$ vs. $[0,12]$	10.0231	5.5538	3.2132	8.9092	5.7852	4.6619	9.8235	6.1886	3.2820
p-value $[13,19]=[0,12]$	0.0000	0.0000	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000	0.0010
Control for income in origin		✓			✓			✓	
Family fixed effects			✓			✓			✓
Number of observations	1,906,924	1,550,262	738,481	1,906,924	1,550,262	738,481	4,029,921	3,163,750	1,586,954
Dependent variable mean	0.204	0.195	0.221	0.247	0.253	0.237	0.454	0.456	0.465
R^2	0.216	0.386	0.740	0.178	0.341	0.707	0.283	0.441	0.691

Notes: We report annual childhood exposure effects, restricting the sample to individuals who register to vote at least once between 2012 and 2021. This restriction applies to both movers and permanent residents. The regression specification is equation (5), and the estimates can be compared to those in Table 2. Heteroskedasticity-robust standard errors are in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively.

Table A8: Exposure Effects Among Those Turning 18 Shortly Before Election Day

	Republican		Democrat		Voted	
	(1)	(2)	(3)	(4)	(5)	(6)
Exposure effect, 0–19	0.0169***	0.0160***	0.0150***	0.0152***	0.0159***	0.0222***
	(0.0006)	(0.0014)	(0.0005)	(0.0014)	(0.0006)	(0.0014)
Barely 18-year-olds		✓		✓		✓
Number of observations	1,635,379	233,445	1,635,379	233,445	3,415,738	493,251
Dependent variable mean	0.200	0.200	0.252	0.239	0.445	0.481
R^2	0.075	0.084	0.067	0.070	0.139	0.163

Notes: We report annual childhood exposure effects, restricting the sample in two ways. In all columns, the regression excludes individuals who were born on the first day of any given month as such dates of birth could be accurate only to the year, as discussed in Section 2.1. The second column for each outcome further subsets to *barely 18-year olds*, i.e., individuals who turned 18 within 100 days before Election Day and were hence just old enough to register to vote before the election. The regression specification is equation (4). Heteroskedasticity-robust standard errors are in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively.

Table A9: Persistence of Exposure Effects

	Republican			Democrat			Voted		
	1st (1)	2nd (2)	3rd (3)	1st (4)	2nd (5)	3rd (6)	1st (7)	2nd (8)	3rd (9)
Exposure effect, 0–19	0.0159***	0.0141***	0.0131***	0.0122***	0.0112***	0.0104***	0.0156***	0.0139***	0.0057***
	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0010)	(0.0010)	(0.0012)	(0.0013)	(0.0012)
Number of observations	1,122,485	1,122,485	1,122,485	1,122,485	1,122,485	1,122,485	2,361,712	2,361,712	2,361,712
Dependent variable mean	0.188	0.214	0.212	0.217	0.271	0.312	0.395	0.362	0.531
R^2	0.220	0.215	0.210	0.178	0.179	0.181	0.274	0.251	0.211

Notes: We report annual childhood exposure effects for outcomes measured at the first, second, and third election for which an individual is eligible to vote. The sample is restricted to individuals whom we can observe in all three elections, i.e., from birth cohorts 1992 to 1998. The regression specification is equation (4). Heteroskedasticity-robust standard errors are in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively.

Table A10: Summary Statistics for Multiple Movers

	Mean	Std. dev.	Num. of obs.
Birth year	1997.4	2.79	1,543,798
Male	0.510	0.500	1,476,462
White	0.860	0.347	1,498,080
Black	0.058	0.234	1,498,080
Hispanic	0.046	0.209	1,498,080
Asian	0.028	0.166	1,498,080
Other race	0.008	0.088	1,498,080
Number of children in family	1.50	0.71	1,543,798
Voted in first election	0.449	0.497	1,543,798
Republican in first election	0.221	0.415	822,423
Democrat in first election	0.221	0.415	822,423
Ever Republican and Democrat	0.020	0.140	880,670
Two-time mover	0.791	0.407	1,543,798
Age at move 1	5.8	5.66	1,543,798
Age at move 2	13.4	6.76	1,543,798
Age at move 3	16.2	5.76	323,165
Move 1 distance (mi.)	452.4	649.3	1,543,798
Move 2 distance (mi.)	511.6	655.0	1,543,798
Move 3 distance (mi.)	562.8	664.8	323,165
Origin and all destination states record party	0.325	0.469	1,543,798

Notes: We show summary statistics for the subsample of individuals who moved across counties two or three times during childhood.

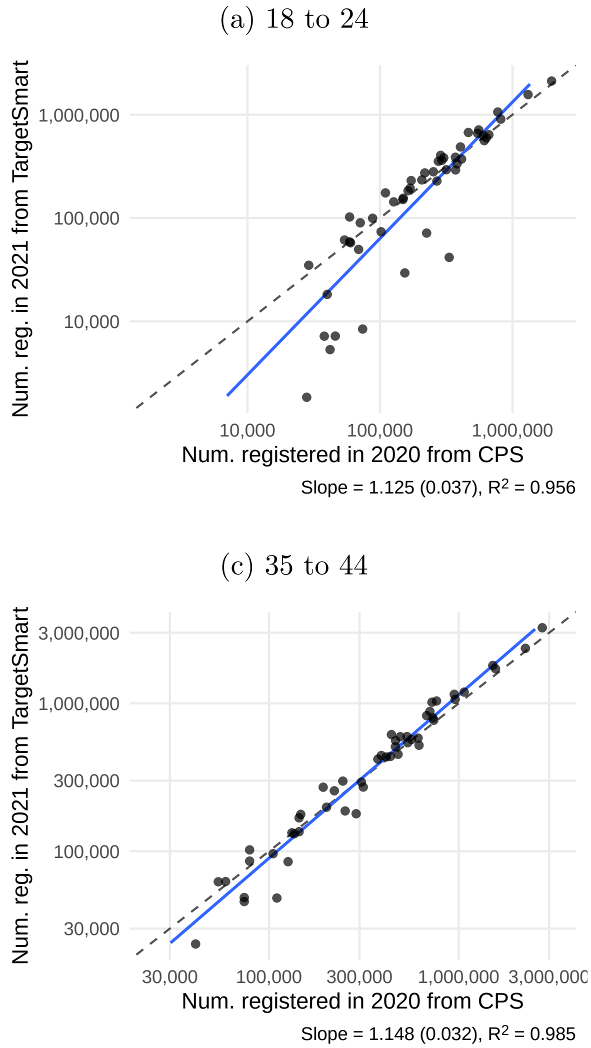
Table A11: Parametric Estimates for Moves Between ZIP Codes

	Republican			Democrat			Voted		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Exposure effect, 0–19	0.0067***	0.0075	0.0048***	0.0058***	0.0090*	0.0053***	0.0027***	0.0028	0.0024*
	(0.0014)	(0.0052)	(0.0018)	(0.0013)	(0.0050)	(0.0017)	(0.0010)	(0.0034)	(0.0013)
Across counties		✓			✓			✓	
Within county			✓			✓			✓
Mean move dist. (mi.)	131.0	297.0	7.2	131.0	297.0	7.2	181.1	361.9	7.3
Number of observations	1,912,529	522,683	1,118,529	1,912,529	522,683	1,118,529	3,589,231	1,153,941	1,912,445
Dependent variable mean	0.190	0.200	0.180	0.263	0.252	0.272	0.452	0.464	0.447
R^2	0.548	0.765	0.489	0.524	0.746	0.478	0.567	0.769	0.512

Notes: We report annual childhood exposure effects measuring exposure at the ZIP code level instead of the county level. For each outcome, the second column reports estimates restricting the sample to individuals who move across counties, while the third column restricts the sample to individuals who move within the same county. The regression specification is equation (4). Heteroskedasticity-robust standard errors are in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively.

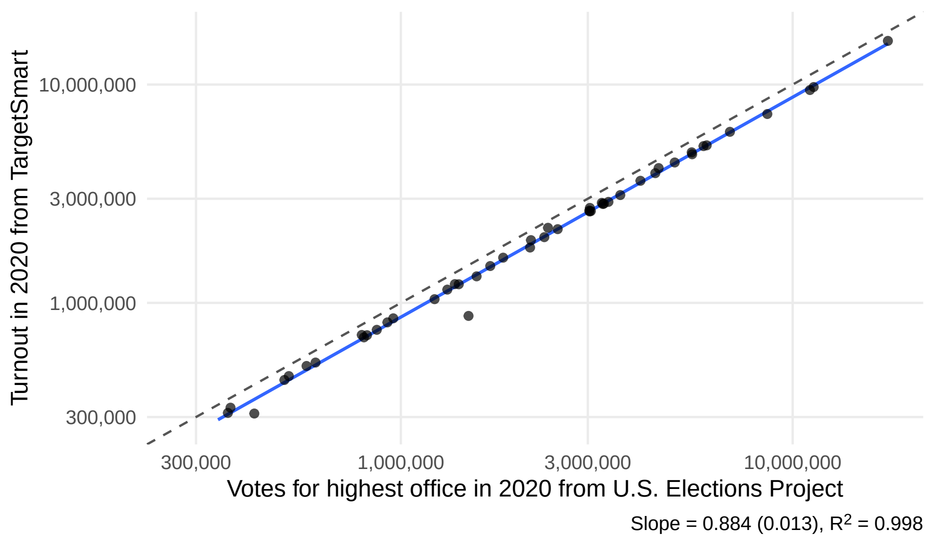
B Data quality checks

Figure B1: Registered Voters by State and Age, TargetSmart vs. Current Population Survey



Notes: On the vertical axis, we show the number of registered voters in the TargetSmart data in a particular state. On the horizontal axis, we show the estimated number of registered voters in that state according to the 2020 CPS Voting and Registration Supplement. We do this for each of four age groups: 18 to 24 in Panel a, 25 to 34 in Panel b, 35 to 44 in Panel c, and 45 to 64 in Panel d. The scales are logarithmic. In each panel, the blue line is a linear fit, and we present its slope and R^2 in the bottom right. The dashed reference lines have slopes of 1.

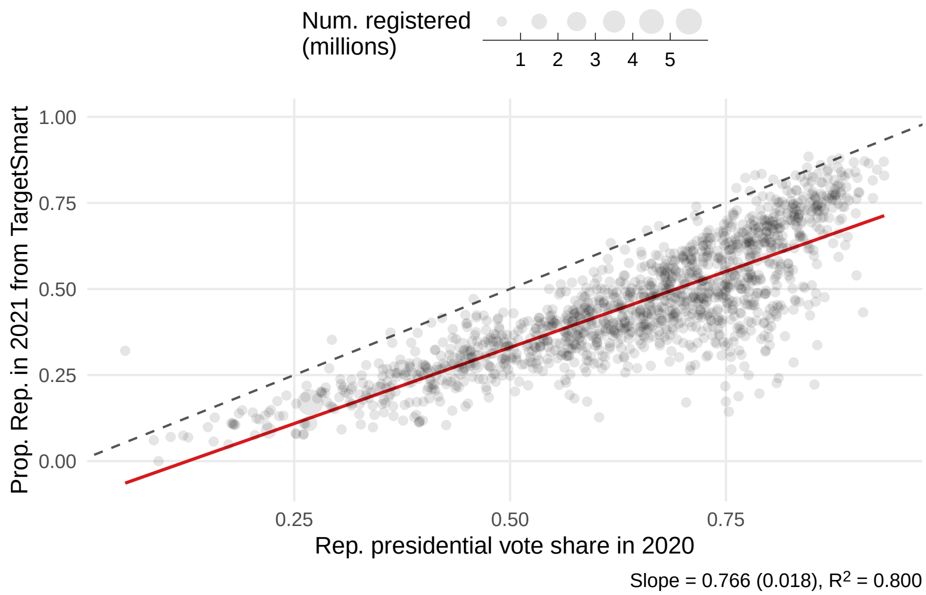
Figure B2: Turnout in 2020, TargetSmart vs. United States Elections Project



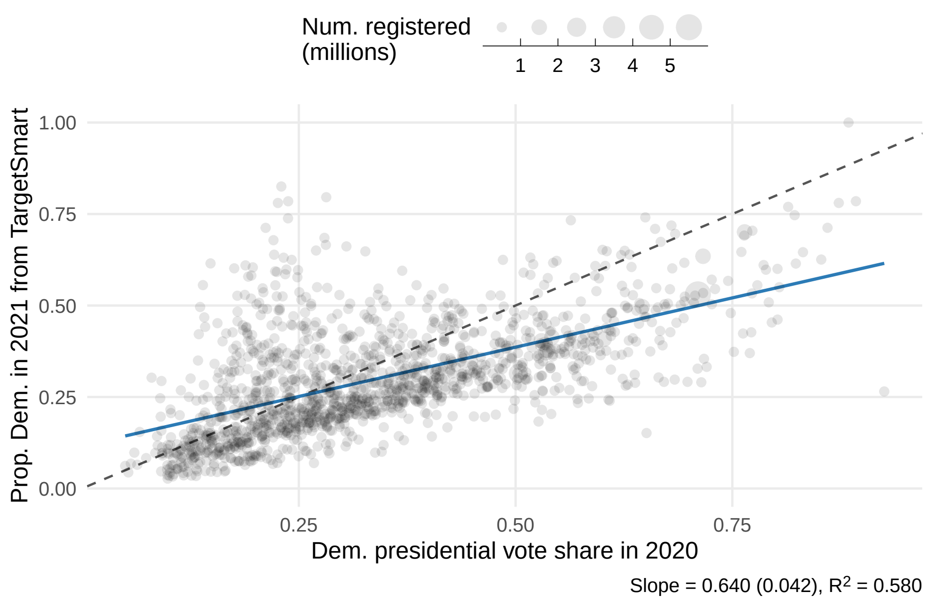
Notes: On the vertical axis, we show the number of people who voted in the 2020 general election in each state according to the 2021 TargetSmart voter file. On the horizontal axis, we show the number of votes cast for the highest office on the ballot in each state in the 2020 general election according to the United States Elections Project (McDonald, 2022). The scales are logarithmic. The blue line is a linear fit, and we present its slope and R^2 in the bottom right. The dashed reference line has a slope of 1.

Figure B3: Share of Partisans in TargetSmart vs. Presidential Vote Share in 2020

(a) Republican



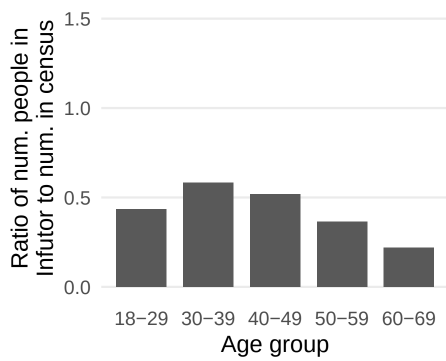
(b) Democrat



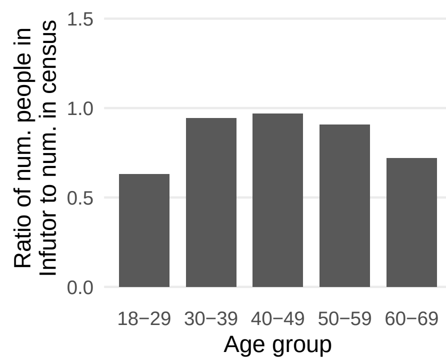
Notes: In Panel a, we show the proportion of people registered with the Republican Party in the 2021 TargetSmart voter file in each county on the vertical axis. On the horizontal axis, we show the proportion of the two-party vote for the Republican candidate (Donald Trump) in the 2020 general election in that county. In Panel b, we repeat this exercise for the Democratic candidate (Joe Biden). The red and blue lines are linear fits weighted by the number of registered voters in each county, and we present their slopes and R^2 in the bottom right of each panel. The dashed reference lines have slopes of 1.

Figure B4: Number of People in Infutor Relative to Census Population

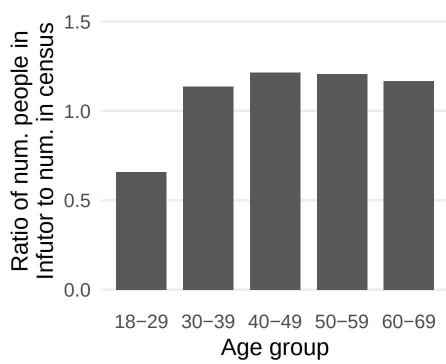
(a) 1990



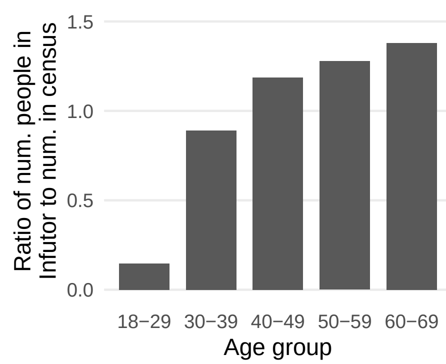
(b) 2000



(c) 2010



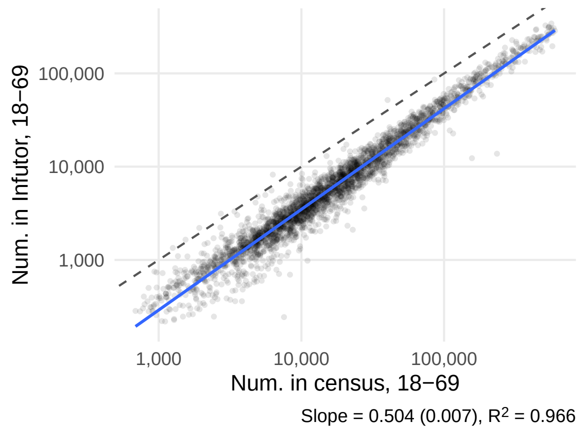
(d) 2020



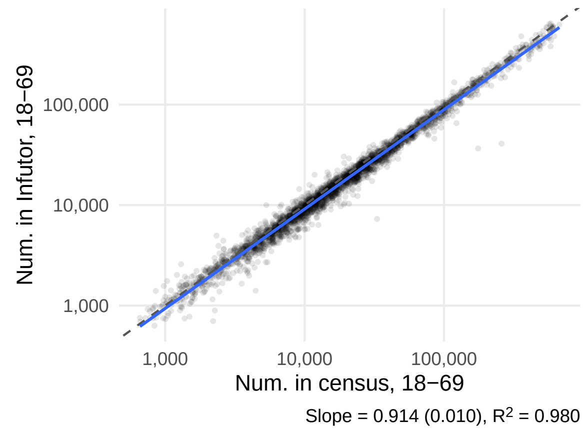
Notes: We show the ratio between the number of people in the Infutor data and the total number of people in the census by age group, for four decennial censuses: 1990 in Panel a, 2000 in Panel b, 2010 in Panel c, and 2020 in Panel d. We exclude individuals listed as “deceased” in Infutor.

Figure B5: County-Level Counts of Infutor vs. Census Population

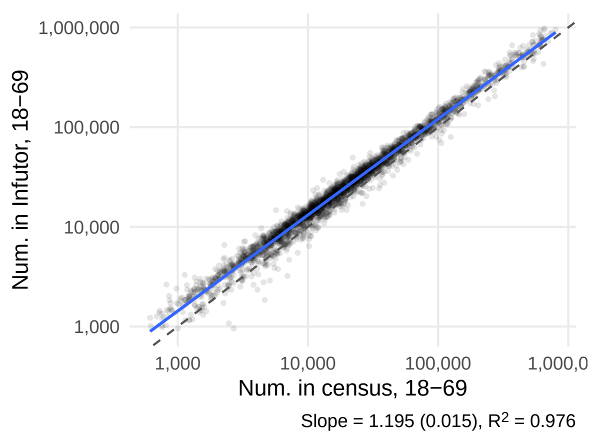
(a) 1990



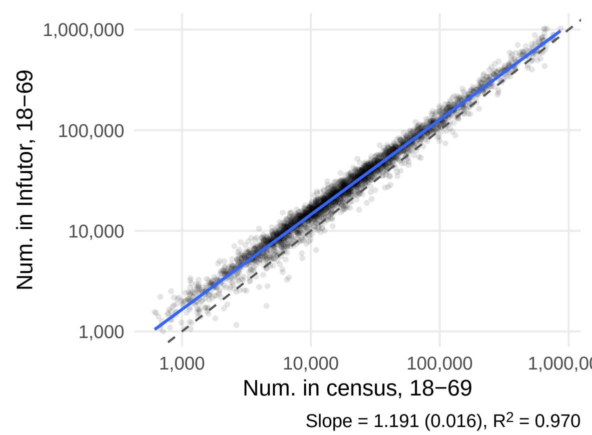
(b) 2000



(c) 2010

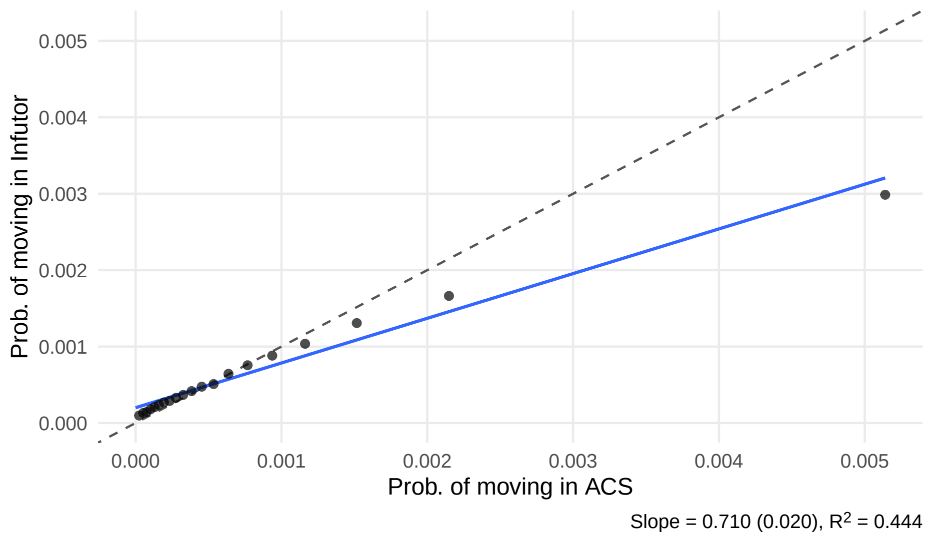


(d) 2020



Notes: On the vertical axis, we show the number of people between 18 and 69 in the Infutor data in a particular county, as of a particular decennial census year. On the horizontal axis, we show the number of people between 18 and 69 recorded in that year's census in the same county. We do this for four decennial censuses: 1990 in Panel a, 2000 in Panel b, 2010 in Panel c, and 2020 in Panel d. The scales are logarithmic. In each panel, the blue line is a linear fit, and we present its slope and R^2 in the bottom right. The dashed reference lines have slopes of 1. We exclude individuals listed as "deceased" in Infutor, and we exclude counties below the 1st percentile and above the 99th percentile of the Infutor or census population distributions.

Figure B6: Probability of Moving in Infutor vs. American Community Survey



Notes: On the vertical axis, we show across-state move probabilities in the Infutor data for parents of young voters. On the horizontal axis, we show the corresponding probabilities from the ACS. To construct move probabilities from the Infutor data, we focus on parents (either identified directly in the Infutor data or identified in the TargetSmart data and subsequently linked to Infutor), count the number of people moving across states in each year, and divide it by the number of individuals in the origin state. To construct move probabilities from the ACS, we use annual individual-level data from the ACS from 2000 to 2019 and construct a sample of individuals that corresponds to the parents in the Infutor data. Specifically, we subset to U.S. citizens residing in the U.S. for at least one year, who have at least one own child living in the same household, who are at least 15 years older than their oldest child, and whose children's ages align with the children's ages in our sample, i.e., their oldest child was born in the youngest cohort observed in our sample or before. To obtain move probabilities, we then count the number of people moving across states within a given year and divide it by the number of people matching the above criteria in the origin state, where both counts are weighted by the appropriate sampling weights. We then match the move probabilities from the two sources by origin, destination, and year. To construct the figure, we bin the probabilities into ventiles, weighting by the total number of people living in the origin state. The blue line is a linear fit, and we present its slope and R^2 in the bottom right. The dashed reference line has a slope of 1.

C Data construction

This appendix describes how we de-duplicate and process the TargetSmart and Infutor data, how we match parents with their children, and how we build and clean parents' residential history, which we use to infer their children's residential history.

C.1 Processing TargetSmart data

The raw TargetSmart data consist of 510 state-by-year snapshots, one for each of the 50 states and the District of Columbia, for each year between 2012 and 2021. These snapshots contain each voter's name, residential address, gender, race, date of birth, whether or not they were registered and voted in prior elections, and their party affiliation.

TargetSmart also provides a "voterbase ID" (henceforth VBID), which uniquely identifies each row for a given state and year, and an "exact track ID" (henceforth ETID), which represents its efforts to link individuals *across* years and states. We use this information as well as voters' first name, middle name, last name, date of birth, and vote history to de-duplicate the TargetSmart data, so that for each state-by-year file, the observation used in the analysis is the most likely current record for each voter. We also use this information to expand on TargetSmart's linkage model to further link records that likely correspond to the same individual – first within states, then across states.

C.1.1 Cleaning the raw TargetSmart state-by-year files

To begin, we take the following steps to clean and de-duplicate the raw TargetSmart state-by-year files:

1. We remove hyphens and spaces from first and last names. We capitalize all letters of first and last names.

2. We recode invalid ZIP codes and census block group IDs as missing.

3. We use TargetSmart's information on whether a voter is found in Social Security death records to drop deceased voters.[C1]

4. We use TargetSmart's information from the United States Postal Service National Change of Address database to drop voters that no longer reside at their listed residence.

5. We de-duplicate records with the same ETID, first name, and last name, giving preference to the record whose registration status is "Registered" (vs. "Unregistered"), whose voter status is "Active" (vs. "Inactive," based on recent election participation), and with the most recent registration date. We do this because, for a given individual, outdated records may exist alongside the active, most recent record in the same state voter file snapshot. When de-duplicating, we harmonize the vote history so that if

[C1]Specifically, we drop voters indicated as "deceased" and who did not vote in the most recent general election before the year of the voter file.

any of the dropped records indicate that the person voted in a particular election, we maintain this turnout history in the record that we preserve for that individual.

6. We repeat the previous step, de-duplicating records with the same first name, middle initial, last name, date of birth, and ZIP code.[C2]

7. We drop any record where the voter's age is listed as under 18 and the individual is listed as "Registered."

C.1.2 Linking records within states

To link rows *within the same state* corresponding to the same individual but across multiple years – in other words, to assign a state unique identifier (henceforth "SUID") – we take the following steps:

1. We split the date of birth field into year, month, and day. The raw voter lists that TargetSmart receives from some states do not record the day of birth (only the month and year) and, in some cases, they only record year of birth. Even though TargetSmart uses commercial data to reconstruct the exact day of birth for voters from these states, the TargetSmart data feature high frequencies of dates of birth ending in `01` (i.e., indicating a voter born on the first day of a given month) and `0101` (i.e., indicating a voter born on January 1). When the date of birth ends in `01`, we set the *day* of birth to missing. When the date of birth ends in `0101`, we set the *month and day* of birth to missing.[C3]

2. The VBID uniquely identifies a row in each state-by-year snapshot, but when the same individual appears in multiple years in the same state, they typically appear with the same VBID. Therefore, we assume records that share a VBID are the same person, and assign them the same SUID. However, if TargetSmart assigned the same VBID to multiple rows where the first names, last names, and dates of birth are *all* different, or where the maximum difference in birth years is more than five years *and* the months and days of birth are different too, we break this link.

3. We group observations by ETID.

   - If *at least one* of the first names, last names, and dates of birth are the same among all observations sharing the same ETID, there is only one record per year, and the maximum age difference is less than or equal to five years, then we assign the same SUID to all these observations.

   - If not everyone in that set of observations shares *either* a first name, last name, or date of birth, we group them by name and date of birth and – as long as there is only one record per year – assign rows within each group the same SUID.

[C2]We consider groups of rows in which the middle initial is missing for one or more and the rest of these variables are identical to be duplicates.

[C3]We lose information by setting the month or day of birth to missing for some people who were actually born on the first of the month or on January 1. However, there is no reliable way to determine whether a date of birth ending in `0101` actually corresponds to a January 1 birthday, or whether it indicates that the month and day of birth are missing.

4. We group by first four letters of first name, last name, year of birth, and address. We require that:
   - Each record has non-missing information for all of the grouping variables.
   - Each record is unique *within a year* by these variables.

   If so, we assign these rows the same SUID.
5. We repeat the previous step using the following sets of grouping variables:[C4]
   - First four letters of first name, last name, year of birth, and address.[C5]
   - First name, last name, and address.
   - First name, last name, and date of birth.
   - First name, last name, and year and month of birth.
   - First name, last name, and year of birth.
   - First name, middle name, last name, and date of birth.
   - First name, middle name, last name, and year and month of birth.
   - First name, middle name, last name, and year of birth.

Appendix Table C1 summarizes the results of this procedure.

C.1.3 Linking records across states

To link rows corresponding to the same individual *across* states and assign them a nationally unique identifier (henceforth "UID"), we take the following steps.

1. We drop rows missing first name, last name, or date of birth.
2. As above, if the date of birth ends in `01`, we set the day of birth to missing. If the date of birth ends in `0101`, we set the month and day of birth to missing.
3. We group by ETID and check that the maximum vote count for any election is one.[C6] If so, we assign these rows the same UID.
4. We group by first name, last name, and date of birth. We require that:

[C4]If month and day of birth are not part of the grouping variables, we require that when non-missing, they do not conflict. We also require that middle initials, when non-missing, do not conflict, except if the set of grouping variables includes the address and year of birth, or if it includes the exact date of birth.

[C5]We only de-duplicate records if no more than 15 people ever lived at the same address, and if the difference between the maximum and minimum years of birth is no more than five years.

[C6]The "vote count" for a given election is the sum across all observations with the same ETID of an indicator equal to one if the person is indicated as having voted. We therefore require that, for each person we propose to de-duplicate, at most one record linked to that person is recorded as having voted in each election. In other words, we do not de-duplicate rows if this would imply that the same individual voted more than once in the same election in two different states, which would be illegal. We impose this restriction because double voting is extremely rare: Goel et al. (2020) estimate that about 0.02% or less of votes cast in 2012 were double votes.

Table C1: Assigning State UIDs

Step	Num. unique	Prop. of original	Complete
Start (unique VBIDs)	433,696,587	1.000	-
Break links with different name/DOB	433,782,739	1.000	-
Break links born > 5 years apart	433,865,721	1.000	-
Link by ETID	384,053,648	0.886	-
Link by f4, last, birth yr., exact address	376,154,793	0.867	0.865
Link by first, last, exact address	367,012,275	0.846	0.993
Link by first, last, DOB	355,670,978	0.820	0.583
Link by first, last, birth yr./mo.	350,356,912	0.808	0.733
Link by first, last, birth yr.	345,007,396	0.796	0.870
Link by first, middle, last, DOB	344,942,651	0.795	0.515
Link by first, middle, last, birth yr./mo	344,810,508	0.795	0.637
Link by first, middle, last, birth yr.	344,141,448	0.794	0.755

Notes: We describe the results of our procedure to link records that correspond to the same person within a state in the TargetSmart data. The pipeline proceeds sequentially through each row of the table. "Num. unique" refers to the number of unique individuals at each step, and "prop. of original" refers to the ratio between the number of unique individuals at each step and the original number of unique individuals. "Complete" is the proportion of rows with non-missing information for the relevant variables at each step. "f4" refers to the first four letters of the individual's first name. Although we perform this procedure separately for each state, the table presents numbers aggregated across states.

- Each record has non-missing information for all grouping variables.
- Each record is unique by these variables within state.
- The group has a record from at least two states.
- The maximum vote count among records in the group for any election is one.

If so, we assign these rows the same UID.

5. We repeat the previous step using the following sets of grouping variables:[C7]
    - First name, last name, and date of birth.
    - First name, last name, and year and month of birth.
    - First name, last name, and year of birth.
    - First name, middle name, last name, and date of birth.
    - First name, middle name, last name, and year and month of birth.
    - First name, middle name, last name, and year of birth.

[C7] As above, if month and day of birth are not part of the grouping variables, we require that when non-missing, they do not conflict. We also require that middle initials, when non-missing, do not conflict, *except* if the set of grouping variables includes exact date of birth.

Table C2: Assigning National UIDs

Step	Num. unique	Prop. of original	Complete	Unique
Start (unique SUIDs)	344,141,448	1.000	-	-
Link by ETID	330,242,382	0.960	-	-
Link by first, last, DOB	320,139,992	0.930	0.619	0.995
Link by first, last, birth yr./mo.	317,109,902	0.921	0.752	0.969
Link by first, last, birth yr.	312,909,936	0.909	0.877	0.864
Link by first, middle, last, DOB	312,839,558	0.909	0.539	0.998
Link by first, middle, last, birth yr./mo.	312,413,545	0.908	0.641	0.996
Link by first, middle, last, birth yr.	310,637,778	0.903	0.747	0.986

Notes: We describe the results of our procedure to link records that correspond to the same person across states in the TargetSmart data. The pipeline proceeds sequentially through each row of the table. "Num. unique" refers to the number of unique individuals at each step, and "prop. of original" refers to the ratio between the number of unique individuals at each step and the original number of unique individuals. "Complete" is the proportion of rows with non-missing information for the relevant variables. "Unique" is the proportion of rows with non-missing information for the relevant variables that are unique within state by those variables. At each step, in order to be linked to another observation, we require that an observation has complete information for the relevant variables and is unique within each state by those variables.

Appendix Table C2 summarizes the results of this procedure.

C.2 Processing Infutor data

The raw Infutor data consist of 51 files, one for each of the 50 states and the District of Columbia. The data contain the name, gender, year and month of birth, and up to 10 addresses associated with each individual/row, along with a date (year and month) at which Infutor considers that address effective.

We use two methods to de-duplicate the Infutor records, which are uniquely identified (across all address observations for a given individual) by "persistent ID" (henceforth PID). First, we group observations by first four letters of first name, last name, year of birth, and address. We require that:

- There are no more than 10 distinct PIDs in the group.
- Name suffixes, when non-missing, do not conflict.
- No more than 15 people ever live at the address.

If so, we assign these rows the same PID. Then, we repeat the same procedure, this time grouping by first name, last name, and address. In addition to the above restrictions, we require that:

- Middle initials, when non-missing, do not conflict.

Table C3: Infutor De-Duplication

Step	Num. unique	Prop. of original
Start (unique IDs)	644,033,829	1.000
De-duplicate by f4, last name, address, birth yr.	623,140,262	0.968
De-duplicate by first name, last name, address	571,289,316	0.887

Notes: We describe the results of our procedure to de-duplicate records that correspond to the same person within the Infutor data. The pipeline proceeds sequentially through each row of the table. "Num. unique" refers to the number of unique individuals at each step, and "prop. of original" refers to the ratio between the number of unique individuals at each step and the original number of unique individuals. "f4" refers to the first four letters of the individual's first name.

- The difference between the maximum and minimum years of birth is no more than five years.

Appendix Table C3 summarizes the results of this procedure.

C.3 Linking children to parents

To link children (individuals who are ever 18 to 24 years old in the TargetSmart data over our sample period, 2012 through 2021) to their parents, we employ two methods, using respectively (1) residential addresses from TargetSmart and Infutor, and (2) TargetSmart's household identifier field. We proceed as follows:

Method 1: Address

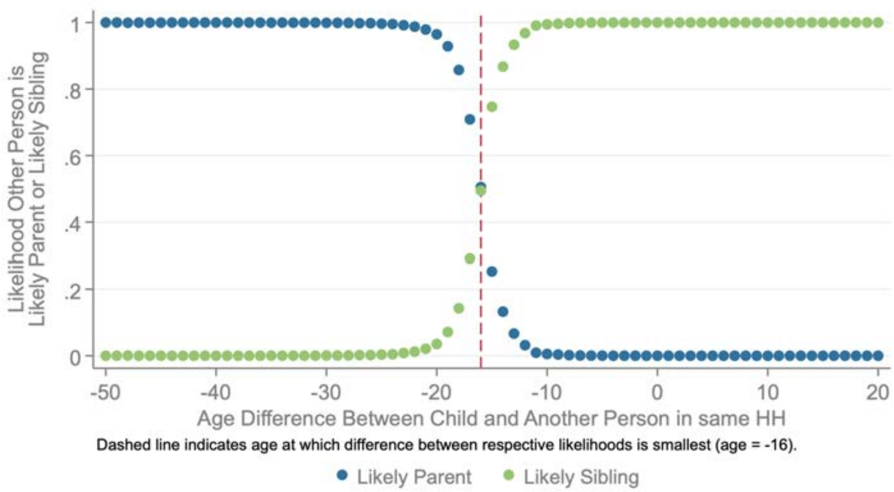
1. We stack the TargetSmart and Infutor addresses.
2. We merge children (18 to 24 years old in the TargetSmart data) with potential parents (25 years or older in either the TargetSmart or Infutor data) by address (but not including unit or apartment number) and first four letters of the last name. We keep these matches only if:
   - Parents are at least 16 years older than their children.[C8]
   - Either there are fewer than 15 people who ever live at that address, or (1) the child and parent's last names match exactly, (2) the child and parent's addresses have the same unit number, and (3) fewer than 15 people share that address, unit number, and last name combination.

Method 2: Household identifier

1. We merge children (18 to 24 years old in the TargetSmart data) with potential parents (25 years or older in the TargetSmart data) by TargetSmart's household ID. We keep

[C8] According to ACS data, this is the age difference at which it becomes more likely that two individuals in the same household have a parent-child relationship as opposed to a sibling relationship. See Appendix Figure C1.

Figure C1: Probability of Family Relationship vs. Age Difference



Notes: We use data from the 2021 ACS to show the probability that an individual living in the same household as a child is a "likely parent" or "likely sibling" by age difference between the child and the other person. Specifically, we subset the full ACS to individuals who are citizens, at most 21 years old, and who are classified as children. We then interact these individuals with all other people in the same household who are either classified as household head or spouse – i.e., individuals whom we refer to as "likely parents" – or children, whom we refer to as "likely siblings." For a given age difference between two individuals, we then calculate the probability that these two individuals have either a parent-child relationship or a child-child relationship.

these matches only if:

- Parents are at least 16 years older than their children.
- Fewer than 15 people share that household ID.

Appendix Table C4 summarizes the results of these linking processes. Overall, 62.4% of children are linked to at least one parent. We explain later how, in cases where a child is linked to multiple parents, we choose which parent from whom we infer the child's location history.

C.4 Creating the ZIP code panel

C.4.1 Merging Infutor and TargetSmart data

Next, we use the Infutor data to reconstruct young voters' residential histories by assembling residential histories for their parents. To do this, for children whom we link to a parent in TargetSmart (and not Infutor directly), we need to identify the parent's Infutor address history. We do this as follows:

1. As previously, in TargetSmart, if the date of birth ends in `0101`, we set the month of

Table C4: Linking Children to Parents

	Num.	Prop. of total
Total number of children	47,094,767	1.000
Linked using TargetSmart only	4,880,170	0.104
Linked using TargetSmart only, by address	4,224,870	0.090
Linked using TargetSmart only, by HH ID	4,594,408	0.098
Linked using Infutor only	1,826,565	0.039
Linked using TargetSmart and Infutor	22,699,996	0.482
Linked using TargetSmart or Infutor	29,406,731	0.624

Notes: We describe the results of our procedure to link children to their parents. "Num." refers to the number of unique individuals in each category, and "prop. of total" refers to the ratio between the number of individuals in each category and the total number of children, i.e., those who are ever 18 to 24 years old in the TargetSmart data over our sample period, 2012 through 2021. The groups are not mutually exclusive – that is, an individual can be linked using more than one method. "HH ID" refers to the household identifier assigned by TargetSmart.

birth to missing. In Infutor, if the date of birth ends in `01`, we set the month of birth to missing.[C9]

2. We merge the set of records identified as parents in the TargetSmart data with the full Infutor address history by first name, last name, address, and month and year of birth.
3. We repeat the previous step using the following sets of merging variables. In each round, all rows are candidates to be merged (not only previously un-merged rows). We keep all the unique Infutor PIDs matched to a given parent in the TargetSmart data in any of the merge rounds.[C10]
   - First name, last name, address, and month and year of birth.
   - First name, last name, address, and year of birth.
   - First four letters of first name, first four letters of last name, and address.
   - First four letters of first name, first four letters of last name, month and year of birth, and ZIP code.
   - First four letters of first name, first four letters of last name, year of birth, and ZIP code.
   - First name, last name, and ZIP code.
4. We remove TargetSmart-Infutor matches with conflicting dates of birth. Specifically, we drop matches if the Infutor date of birth falls outside a 12-month window around

[C9] Remember that the Infutor data only report month and year of birth.

[C10] For each merge round, the eligible TargetSmart rows must be uniquely identified by the merging variables. However, we allow one TargetSmart ID to be merged to multiple Infutor PIDs. We require that middle initials, when non-missing, do not conflict, except if the set of grouping variables includes exact address and year of birth.

Table C5: Merging TargetSmart and Infutor

	Num.	Prop. of total
Total number of TS parents	33,273,138	1.000
Merged using first, last, birth yr./mo., address	19,936,173	0.599
Merged using first, last, birth yr., address	23,979,063	0.721
Merged using f4, l4, address	26,892,125	0.808
Merged using f4, l4, birth yr./mo., ZIP	21,159,157	0.636
Merged using f4, l4, birth yr., ZIP	25,440,464	0.765
Merged using first, last, ZIP	27,275,598	0.820
Merged using any method	29,057,091	0.873

Notes: We describe the results of our procedure to merge parents identified in the TargetSmart data to Infutor. "Num." refers to the number of unique individuals merged using that set of variables, and "prop. of total" refers to the ratio between the number of individuals in that group and the total number of parents we identify in the TargetSmart data. The groups are not mutually exclusive – that is, an individual can be merged using more than one set of variables. "f4" refers to the first four letters of the individual's first name and "l4" refers to the first four letters of the last name.

the TargetSmart date of birth.[C11]

Appendix Table C5 summarizes the results of these merges. Overall, 87.3% of TargetSmart UIDs are successfully merged to an Infutor PID.

C.4.2 Creating ZIP code histories for parents

Our analysis uses location information from the Infutor data alone, not TargetSmart. Therefore, we drop the 1.6% of children whose parents were identified only in the TargetSmart data and not linked to the Infutor data. The remaining children – who constitute 60.7% of the total number of 18 to 24 year olds in the TargetSmart data – have at least one parent either identified directly in the Infutor data or identified in the TargetSmart data and successfully linked to the Infutor data.[C12]

We define moves at the ZIP code or county level: We do not analyze moves to different addresses in the same ZIP code. Thus, we build a panel of parents' residential histories at the ZIP code level by stacking all ZIP codes associated with all Infutor records linked to a

[C11]Like for TargetSmart, Infutor dates of birth are sometimes approximate (e.g., January of the birth year). Thus, using the exact month and year of birth to resolve conflicts may lead to incorrectly dropping too many matches. We treat such records differently depending on whether there are other matched Infutor records with a non-missing date of birth that conflicts with TargetSmart's. Specifically, if all non-missing Infutor dates of birth are within ±12 months of TargetSmart's, we also keep Infutor records with missing date of birth; otherwise, we drop them (along with Infutor records with non-missing dates of birth conflicting with the TargetSmart date of birth).

[C12]If a young voter is linked to the same parent in the Infutor data both directly and indirectly through the TargetSmart-Infutor merge, we prioritize the latter, so we preserve information from the voter file data about parent demographics and political behavior. We also require that each Infutor PID is matched to no more than one TargetSmart UID.

parent. We require that each parent is observed at only one ZIP code within a month. Since Infutor combines multiple sources of address information, the data may contain duplicate address observations with slightly different dates or variants of the way the address was recorded. Thus, for each parent, we de-duplicate the ZIP code history as follows:

1. If a person is observed first in ZIP code A and then in ZIP code B in the same month, and they are next observed in ZIP code A, we drop the observation in ZIP code B.
2. For remaining duplicates by month, we keep the ZIP code that appears more often in the future. For example, if a person is observed in ZIP codes A and B in the same month and is later observed three times in ZIP code A but only once in ZIP code B, we drop the first, conflicting observation in ZIP code B.
3. For remaining duplicates by month, we keep the ZIP code that is geographically closer to the future modal ZIP code, that is, the one that has the smaller centroid-to-centroid geodesic distance to the ZIP code that appears most often in the list of ZIP codes after that month.

Finally, we drop return moves that happen suspiciously close – within one year – to earlier observations. That is, if an individual moves from ZIP code A to ZIP code B and relocates back to the exact same address in ZIP code within one year, we drop this return move from B to A.

Ultimately, our analysis uses a location panel measured yearly, not monthly. Thus, if a person is associated with two or more ZIP codes in the same year, we keep the one closest to November (the month in which elections occur). If there are two ZIP codes in a given year, one in October or November and one in December, we keep the former ZIP code for the current year and the December ZIP code for following years, until the next move.

C.4.3 Creating ZIP code histories for children

Finally, we use the processed Infutor ZIP code histories for parents to infer children's residential histories. Since a child may have been matched to multiple parents, for each child, we choose one parent from whom to infer the address history using the following lexicographic rules: (1) the parent with the longest address history (time between the first and last Infutor address observations), (2) the parent who was most recently observed, (3) the female parent, and (4) the younger parent.[C13]

Here, we also create a family ID: we say that children are in the same family if they share the same "primary" parent (from whom we infer the address history), except if this would imply that more than 12 children are part of the same family.

[C13]We first standardize identifying information for each parent by choosing the modal value of name, date of birth, and gender. We prioritize information from TargetSmart over Infutor.