Routledge
Taylor & Francis Group

Journal of Elections, Public Opinion and Parties

ISSN: 1745-7289 (Print) 1745-7297 (Online) Journal homepage: http://www.tandfonline.com/loi/fbep20

Does microtargeting matter? Campaign contact strategies and young voters

Kyle Endres & Kristin J. Kelly

To cite this article: Kyle Endres & Kristin J. Kelly (2017): Does microtargeting matter? Campaign contact strategies and young voters, Journal of Elections, Public Opinion and Parties

To link to this article: http://dx.doi.org/10.1080/17457289.2017.1378222

JOURNAL OF ELECTIONS, PUBLIC OPINION AND PARTIES, 2017
https://doi.org/10.1080/17457289.2017.1378222

Routledge
Taylor & Francis Group

Does microtargeting matter? Campaign contact strategies and young voters

Kyle Endres[a] and Kristin J. Kelly[b]

[a]Duke University, Durham, NC, USA; [b]Austin Community College, Austin, TX, USA

ABSTRACT

In recent elections, campaigns have based their contact decisions on individual microtargeted propensity scores generated using "Big Data" rather than the more traditional geographic-based contacting. Shifts in campaign strategy have implications for who is contacted and ultimately who participates in elections. As campaigns focus more of their outreach towards individuals who the data indicate are more likely to turn out and more likely to vote for their candidate, some groups may be systematically excluded from contact. We investigate this using voter files and survey data from the 2012 US elections to compare who the Republicans identified for campaign contact using microtargeted propensity scores and who would have been identified for contact if they used a strictly geographic-based approach. Our findings suggest that young people are much less likely to be designated for contact when campaigns rely on microtargeted data than older individuals, the latter of whom are more likely to be contacted under both geographic and microtargeting strategies.

The Help America Vote Act of 2002 (HAVA) mandates that states compile and maintain a comprehensive list of registered voters. These lists accelerated a shift from geographic-based targeting strategies to individual-level campaigns reliant on Big Data. Modern campaigns use state voter files supplemented with data from public and private sources to identify voters for persuasion and mobilization by running predictive models to estimate an individual's propensity to vote and propensity to support each party. The use of data-generated propensity scores to identify likely voters and potential supporters is commonly known as microtargeting.[1] As the use of microtargeted propensity scores have become more widespread, and data have become increasingly available, scholars have warned that segmentation of

CONTACT Kyle Endres kyle.endres@gmail.com Duke University, Durham, NC 27708, USA
Supplemental data for this article can be accessed at https://doi.org/10.1080/17457289.2017.1378222.
[1]The term "microtargeting" is sometimes used to refer any individual-level contact performed by campaigns and sometimes refers to the data-generated propensity scores used to guide individual-level contact decisions. In this study, we are referring to the latter.

this nature may have negative implications (Hillygus and Shields 2009, chap.7; Kreiss and Howard 2010; Barocas 2012; Serazio 2014).

Among these concerns is the idea that candidates will only communicate with their supporters and undecided individuals who are already predisposed to participate in the electoral process. Granted, campaigns have always been strategic about their contact decisions given their limited resources. Their overall strategy and who they contact matters. For example, in the 1980s and 1990s, campaigns began limiting contact to registered voters which raised concerns about participatory inequality and underrepresentation of unregistered segments of the population (Wielhouwer 2003). Subsequent studies found that there are, in fact, substantial differences between the registered and unregistered populations both in terms of demographic characteristics and policy attitudes (Jackman and Spahn 2015).

Advances in Big Data ushered in a similar shift in campaign strategy at the turn of the century (Panagopoulos 2016) as microtargeted propensity scores predicting turnout and vote choice became a key tool for campaigns to use when deciding which individuals to contact. This new reliance on microtargeted data may further exacerbate the gap in representation since registered citizens from groups with a history of low turnout may be disproportionately excluded from campaign mobilization and persuasion efforts based primarily on whether the data suggest they are less likely to vote. In this manuscript, we examine whether the shift in strategy affects the demographic composition of the electorate. To do so, we evaluate voter contact decisions based on microtargeted propensity estimates of voting against contact decisions based on a geographic strategy.

While most extant research on microtargeted data focuses on Democratic campaigns, this paper leverages rarely studied data from the Republican National Committee (RNC). The RNC provided their files for a glimpse into how registered voters are perceived based on their microtargeted propensity scores. Access to these scores allows us to determine whether some groups are less likely to be contacted when voter contact decisions are made using microtargeted estimates. We also presumed nonvoters when creating their microtargeted scores and discuss the practical impact that contacting voters based on their microtargeted scores may have had on their 2012 campaigns.

We argue that the transition from geographic-based targeting to strategies incorporating microtargeted propensity scores has exacerbated the inequality in contact for certain segments of the population. In contrast to a microtargeted strategy, when contact decisions are geographically based on the underlying partisanship of the precinct in which a registered voter lives, individuals ignored by one party might be mobilized by the other party. But one limitation of contacting individuals based on their microtargeted propensity

estimates is both parties may choose to routinely avoid the same registered voters based on their predicted low propensity to vote.

Our broader goal is to demonstrate that while microtargeted data have allowed campaigns to more efficiently utilize their resources by targeting people with higher probabilities of voting, it may systematically fail to bring potential voters into the fold who might be contacted using a more traditional geographic targeting strategy. Past changes in contact strategies have had consequences for turnout (Rosenstone and Hansen 1993), and the use of microtargeted propensity scores may likewise have long-term effects on vote choice and participation. We proceed by first providing an overview of contact strategies used by campaigns past and present before discussing our data and results. As a preview, we find that young people are significantly less likely to be identified for contact when campaigns decide whom to contact using microtargeted propensity scores compared to more traditional contact strategies that rely primarily on precinct-level information. We also find that registered voters who were predicted to have a low propensity to vote reported being contacted at significantly lower rates than individuals estimated as higher propensity voters.

Microtargeting versus geographic targeting

The goal of political campaigns is to win on Election Day. The path to victory often means prioritizing some voters at the expense of others since campaigns have limited resources and consequently must focus more of their attention on a subset of states, precincts, or individuals. Even within battleground states, which tend to receive the lion's share of money and attention in presidential elections, campaigns strategically allocate resources (see Shaw 2006; Panagopoulos 2009). They routinely divide the electorate into broad segments – their base supporters, swing voters, and the opposition's base – and approach each group differently. They look to mobilize their base, persuade swing voters to vote for their candidate, and avoid wasting resources on the opposition's base (Burton and Shea 2010; Sides et al. 2015; Panagopoulos 2017).

When developing contact strategies, office seekers must rely on their perceptions of the electorate (Hersh 2015). While it is relatively easy to identify loyal supporters of each party, it is a more difficult task to identify individuals who are open to persuasion. For many decades, campaigns' perceptions about voters and their decision to contact certain individuals over others were based on the geographic unit (usually the precinct) in which individuals are registered to vote. Geographic targeting strategies were first utilized in the 1970s when Democrats began mapping political geography for their candidates to target every precinct for persuasion and turnout (Issenberg 2013, 47–48). Compiling precinct statistics helped strategists forecast how precincts

would perform in an election, allowing campaigns to use prior results to predict possible swing precincts (Blaemire 2013, 220). This enabled campaigns to predict aggregate levels of turnout and party loyalty which helped them properly allocate resources.

Individual-based data-driven contact strategies expanded further in the 1980s when firms began clustering people according to shared demographic data and behavioral patterns, with a focus on reaching individuals in ZIP codes that fell into a particular cluster, in a method called geodemographic targeting (Blaemire 2013, 221). This combined demographic data from the Census with geographic marketing data, leveraging the assumption that people choose their neighborhoods according to their lifestyles, and these characteristics could be used to draw inferences about the people living there (Burton and Shea 2010, 125). But campaigns and parties today utilize precise, individual-level data that are unmoored from geography, which allows them to more selectively target certain individuals.

The 2000 presidential election represented a turning point for political campaigns where they became more reliant upon individual-level data and subsequently became more strategic about whom they contacted and whom they ignored (Panagopoulos and Wielhouwer 2008; Panagopoulos 2016). Campaigns use data about the individual to deliver messages to the segments of the electorate who are thought to be the most receptive to their outreach based on their perceived partisanship and perceived issue positions (Hillygus and Shields 2009; Endres 2016). Selective targeting of individuals is particularly useful in precincts where the other party has the most support (Kramer 1966). When properly executed, microtargeted propensity scores allow campaigns to identify supporters and swing voters who might otherwise not be contacted under a precinct-based geographic strategy because they live in areas largely populated by the opposition. This gives campaigns the ability to mobilize likely supporters and swing voters while minimizing the risk of mobilizing the opposition's base.

Conventional wisdom suggests that microtargeted propensity scores[2] have greatly improved campaigns' perceptions about voters. The accuracy of these perceptions is important since a campaign strategy based on individual-level contact decisions is only superior to a geographic targeting strategy when their perceptions of voters are accurate (Hersh 2015). If perceptions are accurate, campaigns should make their contact decisions at the individual level instead of the precinct level. However, one potential limitation of

[2]Campaigns use political data to construct predictive models to make targeting communications more efficient, usually creating three categories of predictive scores (Nickerson and Rogers 2014). Behavior scores use past participatory behavior and demographic information to calculate the probability that a citizen will turn out to vote. Support scores predict aggregate candidate and issue preferences. Responsiveness scores estimate individual persuadability in response to particular targeted communications.

microtargeted propensity scores is campaigns struggle to develop a complete picture for some groups of registered voters and may be better off with a precinct strategy for difficult-to-estimate subgroups such as young people.

The process used to generate microtargeted propensity scores sheds some light as to why certain groups may fare worse than others under a microtargeted strategy. Publicly available voter files are combined with supplemental data, including other publicly available information and private consumer data, to sort potential voters (Sides et al. 2015; Panagopoulos 2017). The content of voter files varies by state, but they typically include the date of birth, gender, and turnout history (Hersh 2015). The data allow campaigns to model the characteristics of individuals who are registered to vote, and these models are then used to determine whom to target. Some voters are easier to categorize than others due to their lengthier voting histories and previous interactions with politics.

The targeting estimates allow campaigns to ignore predicted nonvoters and those believed to support the other party (Nickerson and Rogers 2014), but which voters a campaign chooses to ignore may depend on the election. Some voters are more responsive to outreach than others, which may justify a campaign's decision to ignore individuals who the data suggest are the least likely to vote. Enos, Fowler, and Vavreck (2014) find that high-propensity voters are more likely to favorably respond to campaign contact than low-propensity individuals – especially in low-salience, off-year elections – exacerbating existing participation gaps in the electorate. However, the disproportionate response to campaign contact is less evident in high-salience elections where the differences in turnout between low- and high-propensity voters are minor (Arceneaux and Nickerson 2009). This suggests that campaigns should cast a wider net in presidential election years and the exclusion of lower-propensity voters from contact may be more problematic in off-year elections.

Expectations

Some demographic groups such as highly educated, older, high-income, and white voters have historically participated in politics at higher rates than less-educated, younger, lower-income, and non-white individuals (Rosenstone and Hansen 1993). We expect that microtargeted propensity scores disproportionately categorize individuals with these latter characteristics as unlikely to vote which increases the prospects of their exclusion from campaign contact. Aside from age (and race in a handful of states), none of these variables are included in state voter files and cannot be directly incorporated into a contact strategy by campaigns (Hersh 2015). Past participation, however, is directly integrated into the models and tends to lead to high predictive accuracy. Younger voters (18–29 year olds) in particular should be negatively

impacted by the microtargeted models caused by their lack of a (lengthy) voting history and gaps in their voting history due to high rates of mobility.

If campaigns pursued a geographic-based contact strategy, individuals who are registered to vote in a precinct would have a higher likelihood of being contacted (by both parties) if their voting precinct is designated as a competitive or swing precinct. Similarly, if their home precinct routinely votes in one party's favor, registered individuals should be contacted for turnout by that party as they work to mobilize their base. Fewer voters in these precincts should be contacted when microtargeted propensity scores are used because the campaigns can selectively contact (or ignore) some individuals based on their scores. We expect that many individuals who are ignored under a microtargeted strategy would have been designated for contact under the precinct-targeting strategies from earlier campaigns.

Data and methods

We use an original data-set created by merging the RNC's voter files with survey data from the 2012 Cooperative Congressional Election Study (CCES) for two states, Florida and Virginia. Both states were considered presidential battlegrounds in 2012 and had competitive down-ballot races as well, including some congressional seats. The voter files include microtargeted propensity scores for partisanship and turnout. The estimates and the exact models used are considered proprietary information by the data firms, and thus are rarely evaluated by anyone external to the campaign.

The CCES is a national survey, with large state-by-state sample sizes, administered by YouGov (Ansolabehere 2012). We provided YouGov with the RNC voter files so that they could be merged with the CCES[3] using the personal identifying information that it collects from survey participants. The process used to match the voter files and survey data is similar to the matching process performed to supplement the CCES with validated turnout for participants (see Ansolabehere and Hersh 2012). The major difference is that the Republican Party provided the voter files for this study, while a commercial vendor that works closely with the Democratic Party matches and validates turnout for CCES participants.

There are some differences between the RNC's microtargeted propensity estimates for individuals in Florida and their propensity estimates for Virginia. In Virginia, the format of the microtargeted propensity scores divided registered voters into the three broad categories based on their estimated

[3]For a detailed explanation of the design and methodology of the 2012 CCES, see Ansolabehere 2013. For a discussion of the combined data-set and merging process, see Endres 2016.

likelihood to vote – high, mid, and low propensity.[4] The microtargeted propensity scores for partisanship took the form of a five-point ordinal variable ranging from strong Democrat to strong Republican. The microtargeted propensity scores for Florida were provided in a different format. The estimated turnout variable was delivered as a scale ranging from 1 (least likely to turnout) to 10 (most likely), and the partisanship variable ranged from 1 (most Democratic) to 60 (most Republican). We collapsed the Florida variables into the same categories as Virginia to analyze the two states collectively. To ensure our results are not a consequence of measurement decisions, we also analyze Florida separately and include the results in the supplemental appendix when applicable.

To determine whether some demographic groups were more likely to be categorized as low propensity relative to others, we run a multinomial logistic regression using data for both states under investigation. The dependent variable for this portion of our study is the likelihood of voting based on the propensity estimates used by the RNC. Individuals estimated to be high- or mid-propensity voters were collapsed into a single "likely" voter category, resulting in a dependent variable with three categories – likely voters, low-propensity voters, and individuals without a microtargeted turnout estimate. Explanatory variables in our model include gender (female = 1), race/ethnicity (Black, Latino, and Asian/multi-racial/other with White as the base category), age, education, and income. We also include an indicator variable for Florida to account for possible differences between the states.

After demonstrating that some demographic groups were significantly more likely to have been categorized as low-propensity voters, we turn our attention to self-reported contact. During the post-election wave, CCES respondents were asked "Did a candidate or political campaign organization contact you during the 2012 election?" Responses were coded (1) if individuals reported campaign contact and (0) if they reported they were not contacted. We regress self-reported contact on the same set of demographic explanatory variables and indicators for classification as low-propensity voters, mid-propensity voters, or not having an estimated propensity to turnout in the Republican voter files.

We follow this analysis with a side-by-side comparison of the demographic characteristics of individuals who were identified for contact based on their microtargeted propensity scores and those who would have been contacted under a precinct-based geographic strategy. Precinct-level data (Ansolabehere and Rodden 2011a, 2011b; Virginia State Board of Elections 2012) from the Office of the Secretary of State in Florida and the Department of Elections

[4]Microtargeting models produce predicted probabilities of turnout and partisanship for each individual. These predicted probabilities are then simplified into categorical or ordinal variables so the data are easier for campaigns and activists to use when making contact decisions.

in Virginia were used in combination with the CCES data to simulate whether low-propensity individuals would have been contacted through a geographic targeting strategy. Precincts within these battleground states were deemed "competitive" (at two different levels) or "not competitive" based on their vote share in the previous presidential election – the 2008 race between Republican John McCain and Democrat Barack Obama.[5]

We examine two levels of competitiveness because exact geographic targeting strategies can vary based on a campaign's resources and overall strategy. Precincts were designated as competitive if the 2008 vote difference between Obama and McCain was within 10 points (McCain's share between 45% and 55%) and designated as "highly" competitive if the vote difference between Obama and McCain was within six points in 2008 (McCain's share between 47% and 53%).

Furthermore, we also estimate a hybrid strategy that incorporates the "known" partisanship of individuals in the precincts identified for contact under our simulated geographic targeting strategies since the post-HAVA lists of registered voters supplied by the states provide some insight into the partisanship of voters who are registered with a party (in states like Florida) or who routinely vote in the same party's primaries (in states like Virginia). In Florida, where voters can declare their party affiliation when registering to vote, individuals who are registered with the Democratic Party may be excluded from contact by the Republican Party and its candidates who do not want to risk mobilizing Democratic voters. In Virginia, a state without party registration, individuals who have voted exclusively in recent Democratic primaries would similarly be excluded. We compare both levels of competitiveness and the hybrid strategy to see what proportion of "low-propensity" voters would have been contacted in 2012 had Republican campaigns made their targeting decisions without the assistance of microtargeted propensity scores.

Results

Young, less-educated, and lower-income registered voters were all classified as unlikely to turnout at higher rates than older, higher socioeconomic status individuals based on the microtargeted propensity estimates generated for the RNC, as expected (column 1 in Table 1). Each of these demographic variables – age, education, and income – has a significant effect on the classification of registered individuals as low-propensity voters relative to the base category of high-propensity voters. The negative coefficients for these variables suggest that as individuals get older and levels of education and

[5]Because redistricting occurred between the 2008 and 2012 elections, and there were some incomplete data within the state voter files, individuals in precincts where the lines changed quite drastically between the two elections were excluded from the analysis. The majority of these precincts were in Miami-Dade, Osceola, and Broward Counties in Florida.

Table 1. Demographic differences in estimated propensity to turnout.

	Low propensity	No microtargeting
Women	−0.01	−0.21
	(0.11)	(0.13)
Age (18+)	−0.07*	−0.06*
	(0.00)	(0.00)
Education (1 to 5)	−0.31*	−0.20*
	(0.06)	(0.07)
Race: Black	0.37*	−0.44*
	(0.15)	(0.19)
Race: Latino	0.19	0.08
	(0.16)	(0.22)
Race: Other	0.07	−0.21
	(0.26)	(0.30)
Income (1 to 16)	−0.11*	−0.17*
	(0.02)	(0.02)
Florida	−0.27*	−2.50*
	(0.12)	(0.14)
Constant	3.80*	4.76*
	(0.32)	(0.38)
N	3354	
Pseudo R^2	0.22	

Notes: Entries are multinomial logistic model regression coefficients and standard errors. The base group is "likely voters" combining estimated high- and mid-propensity voters.
*$p < .05$, two-tailed.

income rise, the probability of being classified by the RNC as a low-propensity voter declines. The statistically significant coefficient for the Black indicator variable signals that Blacks were more likely to be designated as low propensity relative to Whites.

Looking next at the "No Microtargeting" column in Table 1, we see that even having a microtargeted propensity estimate of turnout created by the party also increased as age, education, and income rose. The negative coefficients for these variables indicate that older, more educated, and higher income registrants were all more likely to have microtargeted propensity estimates created for them and to be identified as likely to vote relative to younger individuals and those with lower levels of education and income. Of note is the fact that the indicator variable for African-Americans in this model is statistically significant and negative, suggesting that Black registered voters were more likely than White registered voters to have microtargeted propensity estimates created for them, holding all other variables constant.

Next, we turn our attention to self-reported contact rates to explore if 2012 campaigns utilized these propensity estimates. Our results (Table 2) show individuals who were estimated as low-propensity voters, or who did not have estimated microtargeted propensity scores, were less likely to recall contact[6] than those estimated as high propensity. Individuals who were

[6]However, we must exercise some caution with our interpretation as measures of self-reported contact tend to be unreliable because recall abilities are fallible (Ansolabehere, Iyengar, and Simon 1999;

Table 2. Self-reported contact in the 2012 election.

	Contacted
No turnout estimate	−0.35*
	(0.15)
Estimated low propensity	−0.48*
	(0.12)
Estimated mid propensity	−0.11
	(0.12)
Women	0.16
	(0.08)
Age (18+)	0.03*
	(0.00)
Education (1 to 5)	0.22*
	(0.05)
Race: Black	−1.38*
	(0.13)
Race: Latino	−0.94
	(0.13)
Race: Other	−0.12
	(0.21)
Income (1 to 16)	0.06
	(0.02)
Florida	−0.11
	(0.11)
Constant	−1.27*
	(0.29)
N	3354
Pseudo R^2	.13

Notes: Entries are coefficients and standard errors from a logistic regression model. The outcome variable is self-reported contact: 1 = contacted, 0 = not contacted.
*$p < .05$, two-tailed.

scored as mid-propensity were less likely to report contact than high-propensity voters, but the difference was not statistically significant. Based on self-reports, both parties appear to have devoted fewer resources to mobilizing and persuading low-propensity voters. This is consistent with an examination of Democratic contact records that found Democratic campaigns to be less likely to have contacted low-propensity voters in 2012 (Nickerson and Rogers 2014).

Afterward, we calculate the proportion of low-propensity voters who would have been contacted had Republican campaigns followed geographic contact approaches based on the competitiveness of the precincts in which citizens live compared to a contact strategy based on microtargeted estimates to better evaluate the implications of the evolution to campaign strategies guided by microtargeted propensity scores. To do so, we classified each

Bradburn, Rips, and Shevell 1987; Vavreck 2007), and individuals who are more involved in politics are the most likely to recognize and remember political messages (Bartels 1993; Higgins, Kuiper, and Olson 1981).

Table 3. Proportion of low-propensity voters contacted under geographic strategies.

Competitive (45–55%) and GOP precincts	61%
Highly competitive (47–53%) and GOP precincts	56%
Highly competitive (47–53%) and GOP precincts minus *known* Democrats	42%

Notes: Competitiveness levels based on 2008 presidential vote share in the precinct. Highly competitive precincts are those in which the 2008 vote difference between Obama and McCain was within 6 percentage points. Competitive precincts were those within 10 percentage points. *Known* Democrats are registered Democrats in Florida and individuals who have voted exclusively in recent Democratic primaries.

voting precinct in Florida and Virginia as solidly Democratic, solidly Republican, or competitive based on the two-party vote share from the 2008 presidential election.

Under a geographic strategy, Republican candidates would be expected to focus their attention and resources on solid-Republican and competitive voting districts. Table 3 presents the percentage of low-propensity voters who would have been designated for contact had Republican campaigns pursued a purely geographic-based strategy using the specified levels of precinct competitiveness to guide their contact decisions.[7] Our simulation suggests that roughly 61% of the voters designated as low propensity would have been contacted under the moderate conceptualization of what constitutes a competitive precinct (labeled Competitive) had a campaign attempted to contact every person registered to vote in either competitive (Grand Old Party (GOP) 2008 vote 45–55%) or Republican precincts (GOP 2008 vote >55%). As we tighten the restriction for what constitutes a competitive precinct to a six-point difference between the presidential candidates in 2008 (labeled Highly Competitive), the proportion of low-propensity respondents who would have been contacted under a geographic strategy decreases to 56%. Under the hybrid strategy that excludes "known" Democrats based on their official party registration or participation in past Democratic primaries in highly competitive or GOP precincts, many of the low-propensity voters (42%) still would have been designated for contact. If we were to assume that these low-propensity voters were not contacted using a strategy reliant upon microtargeted propensity estimates of voting, this simulation clearly demonstrates that the transition to an individualized contact strategy based on microtargeted propensity scores has implications for which voters are mobilized.

The question then becomes whether the move to microtargeted propensity scores affects some demographic groups more than others. We calculated the proportions of our sample that would have been contacted under both hypothetical geographical approaches and the hybrid approach, and

[7]It is important to note that all individual who are identified for contact will not necessarily be contacted. Some individuals are more likely to read/open their mail, have a phone number on file, answer an unknown number, and/or engage in a conversation with someone from the campaign either in person or on the phone than others.

Table 4. Contact rates by race/ethnicity, age, and strategy.

	Geographic targeting GOP + competitive precincts			Microtargeted propensity scores		
	Competitive	Highly competitive	Highly competitive, excluding Democrats	All likely voters	GOP, swing, and weak Democrats	GOP and swing
All Registered	67%	62%	46%	72%	57%	55%
White	71%	66%	51%	76%	63%	60%
Black	40%	36%	21%	55%	27%	22%
Latino	60%	53%	34%	63%	49%	48%
Young Registered	58%	53%	44%	22%	18%	17%
White	67%	62%	56%	25%	24%	23%
Black	31%	29%	19%	14%	3%	2%
Latino	59%	50%	30%	24%	15%	15%

Notes: Entries are the percent of respondents who would have been contacted if the campaign pursued the strategy indicated in the column heading. Competitive represents the percentages of registered voters residing in a precinct where the 2008 Republican presidential candidate received at least 45% of the two-party vote. Highly competitive narrows the range of competitiveness to precincts where the Republican presidential candidate received at least 47% of the vote. The highly competitive, excluding Democrats column uses the same 47%, but excludes individuals in Florida who are registered Democrats and individuals in Virginia who voted in Democratic primaries. The "All likely voters" column represents the percentage of registered voters who would have been contacted based on being estimated as either a high- or mid-propensity voter, regardless of partisanship. The "GOP, swing, and weak Democrats" column narrows this by excluding individuals who were estimated to be strong Democrats. The "GOP and swing" column restricts contact estimates to individuals estimated to be Republicans and swing voters.

compared these to the proportions that would have been contacted using microtargeted propensity estimates of turnout and partisanship for individuals of each racial/ethnic group, and by age. Regardless of how wide of a net that the Republicans may have cast when using the microtargeted propensity scores, Republican campaigns were less likely to identify young people (18–29 year olds) as targets for campaign outreach. Table 4 displays the proportion of White, African-American, and Latino registered voters who would have been contacted under each of the possible geographic strategies, as well as under three possible strategies based on the microtargeted propensity scores. Similar to the geographic strategies, we present multiple scenarios for microtargeted strategies as we do not know the criteria used by each of the Republican campaigns that had access to these propensity estimates.

The first two columns in Table 4 indicate the percentage of individuals who would have been contacted under the two geographic targeting strategies and column 3 displays the hybrid strategy. Columns 4–6 present the percentages for the microtargeted propensity estimates. The fourth column lists the proportion of registered voters who would have been contacted had the campaign reached out to everyone estimated to have a high or middle propensity to turnout, regardless of estimated partisanship. The fifth column is a microtargeted column that excludes predicted strong Democrats. The last column further excludes individuals predicted to be either weak or strong Democrats.

Among all voters in our sample, the differences in contact rates between the microtargeted scores and geographic targeting strategies appear relatively minor when looking only at estimates of turnout propensities (the "All likely voters" column against the two geographic targeting columns). But when we include estimates of partisanship in our models (columns 3, 5, and 6), contact rates drop among the full sample of registered voters. This is expected since Republicans would not want to risk mobilizing Democrats.

Breaking down respondents by race/ethnicity allows for a more nuanced look at who would have been designated for contact based on estimated turnout and estimated partisanship. Comparing the rates of contact between the more restrictive geographic targeting strategy (highly competitive) with the narrowest microtargeted column (likely GOP and swing voters), we see a 6% drop in contact among Whites, a five-point decline among Latinos, and a 14% drop among Blacks. The larger drop among Black registered voters is not surprising given that African-Americans have been a strong voting bloc in the Democratic Party's coalition for decades. When the precinct strategy is further narrowed to exclude "known" Democrats ("Highly Competitive, excluding Dems"), the Republican Party appears to reach out to a larger swath of registered voters when relying on the microtargeted propensity estimates. But when we drill down and simulate contact strategies among specific demographic groups, there are substantial differences in the percentage of young registered voters who are identified for contact based only on the microtargeted propensity estimates relative to the geographic and hybrid strategies. Comparing the first and fourth columns, we see a 36% decline in predicted contact rates between a campaign pursuing the competitive geographic strategy and a campaign that is contacting all registered young people estimated to be "likely voters." These proportions fall even further when microtargeted estimates of partisanship are considered. Only 18% of 18–29 year olds would have been contacted based on microtargeted propensity estimates of turnout and partisanship if candidates contacted Republicans, swing voters, and weak Democrats. This falls to 17% if campaigns were to exclude all estimated Democrats.

Young voters fared much better under a geographic-based strategy than a strategy based on microtargeted propensity estimates when further disaggregating by race. Among young White voters, 62% would have been designated for contact under a geographic strategy focusing on highly competitive precincts, and an even greater proportion would have been contacted under the more relaxed level of competitiveness. This is much higher than the 25% of young whites who would have been contacted if the party reached out to those designated as "likely voters" regardless of partisanship, and the 23% who would have been contacted had Republican candidates limited outreach to predicted Republican and swing voters. Young Blacks were the group least likely to be contacted under a highly competitive geographic contacting

strategy at 29%, and this contact rate drops further to 14% among all young Blacks designated as likely to turn out using microtargeted estimates. Projected contact plummets to 2% when only estimated Republicans and swing voters are contacted. Contact for young Hispanics is estimated at 50% under the highly competitive geographic approach examined and falls to 24% if everyone designated as likely to turnout were contacted under a microtargeted approach using the propensity scores. When looking at outreach for only those estimated to be Republican or swing voters based on the microtargeted propensity scores, this proportion declines to 15%. These microtargeted estimates of turnout and partisanship certainly narrow the playing field and greatly reduce the percentage of young voters identified for contact compared to the precinct-based strategies.

Implications

The goal of campaigns is to turn out enough voters to win the election. But this myopic focus by campaigns has both short- and long-term implications in terms of what the electorate looks like and how they vote. In the short run, Republican contact strategies may have hurt Romney's chances of winning the presidency in 2012. Nationally, 60% of young voters preferred Obama at the ballot box, while only 37% voted for Romney (CIRCLE 2013). Yet, vote margins were very close in some battleground states. If the RNC and the Romney campaign relied as heavily on these microtargeted propensity scores when strategically contacting voters as the self-reports of contact suggest they did, the microtargeted propensity scores for young people may have led to missed opportunities for persuasion and turnout among this critical voting bloc. Had the vote choice of young people more closely resembled older voters, enough battlegrounds states would have flipped to give Romney the Electoral College victory.

While Republican campaigns focus on the short-term imperative to win, they may be hurting their long-term prospects among this age group. Young voters become older voters, and young voters today – absent Republican outreach – have disproportionately supported Democratic candidates at the ballot box since 2004 (Kohut et al. 2011; Pew 2012). Democratic campaigns have dominated outreach among the youngest segment of the electorate, making concerted efforts to contact as many young voters as possible in 2008 and 2012 (Issenberg 2013; Hersh 2015). It is unsurprising, then, that they are winning both the mobilization and persuasion game among the young.

Republicans recognize that this Democratic preference in vote choice among the young is a problem (Barbour et al. 2013), and it may be exacerbated by their microtargeted propensity estimates disproportionately categorizing this group as unlikely to vote. As this youngest generation replaces

older generations of voters, this Democratic preference may be a bane to the electoral prospects of future Republican candidates.

In addition, young people today are not participating in the electoral process at their full potential, and the relatively low rates of turnout among young Americans may in part be explained by the lack of mobilizing communications directed their way due to current campaign strategies. In 2012, only 50% of 18–29 year olds voted in the presidential election (CIRCLE 2013). If microtargeted propensity scores continue to disproportionally designate young people as low propensity, and continue to be used with high frequency by campaigns, the youngest segment of the electorate will likely vote at relatively low rates in the future. Voting is a habit that takes a few election cycles to form (Plutzer 2002), and new voters who might have been brought into the fold of electoral participation through more traditional geographic targeting strategies may never be invited to participate, resulting in long-term effects such as increased cynicism towards government, lower levels of internal and external efficacy, and lower rates of participation as younger generations replace older generations who participate at higher levels.

Conclusion

The capabilities of microtargeted propensity scores and the potential repercussions of a data-driven approach to campaigning have been largely speculative up to this point given the proprietary nature of the industry. The preceding analyses are the first known examination of the effects of microtargeted propensity estimates on the demographic composition of the American electorate. Our results indicate that a strict reliance on microtargeted propensity estimates is particularly damaging if Republicans are trying to make inroads with young Americans. Not only are young registered voters less likely to have microtargeted propensity scores calculated for them than older registered voters due to their shorter vote histories, but those that do are more likely to be categorized as "unlikely" to turn out. Relying on microtargeted propensity scores is only superior to geographic targeting strategies when the propensity scores are accurate (Hersh 2015). This categorization of young people is perhaps one limitation of the shift to data-driven campaigns. Our findings suggest that political campaigns should bypass these scores in lieu of a more traditional contact strategy for young registered voters.

Campaigns exist to win elections, and as such it is sometimes rational for Republican campaigns not to target young people in certain electoral contests. But while it may be in the short-term interests of Republican campaigns not to do so, it may be in the party's long-term interests to contact young registrants and others predicted as less likely to turnout. Voting is a habit that an individual develops as he or she ages, and young people have yet to form this habit due to their newness to the voting process (Plutzer 2002).

Furthermore, while party identification is strongly influenced by early childhood socialization (Jennings and Niemi 1968; Stoker and Jennings 2008), it generally does not crystallize until the mid-twenties as young adults respond to their political environment (Franklin 1984). Mobilizing people to the polls in early adulthood is critical to ensure that they develop this habit for healthy democratic participation and representation, and contacting young people for persuasion may impact their party loyalties later in life. In fact, it may be imperative for Republicans if they wish to make inroads with the Millennial and post-Millennial cohorts, and if campaigns on both sides opt not to contact those deemed unlikely to vote, this decision could have important implications for democracy. From a policy perspective, it may lead to less representation of the views of these segments of the population, growing dissatisfaction towards our system of government, and growing cynicism in the populace.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Kyle Endres is a postdoctoral associate at the Duke Initiative on Survey Methodology (DISM) at Duke University.

Kristin J. Kelly is Associate Professor of Government at Austin Community College.

References

Ansolabehere, Stephen. 2013. "Guide to the 2012 Cooperative Congressional Election Survey." Harvard University.

Ansolabehere, Stephen. 2012. COOPERATIVE CONGRESSIONAL ELECTION STUDY, 2012: COMMON CONTENT. V7, April 15, 2013. Cambridge, MA: Harvard University [producer] http://cces.gov.harvard.edu

Ansolabehere, Stephen, and Eitan Hersh. 2012. "Validation: What Big Data Reveal About Survey Misreporting and the Real Electorate." *Political Analysis* 20 (4): 437–459.

Ansolabehere, Stephen D., Shanto Iyengar, and Adam Simon. 1999. "Replicating Experiments Using Aggregate and Survey Data: The Case of Negative Advertising and Turnout." *American Political Science Review* 93 (4): 901–909.

Ansolabehere, Stephen, and Jonathan Rodden. 2011a. "Virginia Data Files." http://hdl.handle.net/1902.1/15551 V3 [Version].

Ansolabehere, Stephen, and Jonathan Rodden. 2011b. "Florida Data Files." http://hdl.handle.net/1902.1/16797 UNF:5:4UlVSNNWWtboES03i623sA== V1 [Version].

Arceneaux, Kevin, and David W. Nickerson. 2009. "Who Is Mobilized to Vote? A Re-Analysis of 11 Field Experiments." *American Journal of Political Science* 53 (1): 1–16.

Barbour, Henry, Sally Bradshaw, Ari Fleischer, Zori Fonalledas, and Glenn McCall. 2013. "Growth and Opportunity Project." *Republican National Committee*. Accessed August 3, 2015. http://goproject.gop.com/rnc_growth_opportunity_book_2013.pdf.

Barocas, Solon. 2012. "The Price of Precision: Voter Microtargeting and its Potential Harms to the Democratic Process." In Proceedings of the First Edition Workshop on Politics, Elections, and Data. ACM, New York, 31–36.

Bartels, Larry M. 1993. "Messages Received: The Political Impact of Media Exposure." *American Political Science Review* 87 (2): 267–285.

Blaemire, Robert. 2013. "The Evolution of Microtargeting." In *Campaigns and Elections American Style*, edited by James A. Thurber and Candice J. Nelson, 217–236. Boulder: Westview Press.

Bradburn, Norman M., Lance J. Rips, and Steven K. Shevell. 1987. "Answering Autobiographical Questions: The Impact of Memory and Inference on Surveys." *Science* 236 (4798): 157–161.

Burton, Michael J., and Daniel M. Shea. 2010. *Campaign Craft: The Strategies, Tactics, and Art of Political Campaign Management*. 4th ed. Santa Barbara, CA: Praeger.

Center for Information & Research on Civil Learning & Engagement (CIRCLE). 2013. *The Youth Vote in 2012*. Medford, MA: CIRCLE. May 10. Accessed March 18, 2015. http://www.civicyouth.org/wpcontent/uploads/2013/05/CIRCLE_2013FS_youthVoting2012FINAL.pdf.

Endres, Kyle. 2016. "The Accuracy of Microtargeted Policy Positions." *PS: Political Science & Politics* 49 (4): 771–774.

Enos, Ryan D., Anthony Fowler, and Lynn Vavreck. 2014. "Increasing Inequality: The Effect of GOTV Mobilization on the Composition of the Electorate." *The Journal of Politics* 76 (1): 273–288.

Franklin, Charles. 1984. "Issue Preferences, Socialization, and the Evolution of Party Identification." *American Journal of Political Science* 28 (3): 459–478.

Hersh, Eitan D. 2015. *Hacking the Electorate: How Campaigns Perceive Voters*. New York, NY: Cambridge University Press.

Higgins, E. Tory, Nicholas A. Kuiper, and James M. Olson. 1981. "Social Cognition: A Need to Get Personal." In *Social Cognition*, edited by E. Tory Higgins, C. Peter Herman, and Mark P. Zanna, 395–420. Hillsdale, NJ: Lawrence Erlbaum.

Hillygus, D. Sunshine, and Todd G. Shields. 2009. *The Persuadable Voter: Wedge Issues in Presidential Campaigns*. Princeton, NJ: Princeton University Press.

Issenberg, Sasha. 2013. *The Victory Lab.* New York: Broadway Books.

Jackman, Simon, and Bradley Spahn. 2015. Unlisted in America. Working paper, Stanford University.

Jennings, M. Kent, and Richard G. Niemi. 1968. "The Transmission of Political Values from Parent to Child." *American Political Science Review* 62 (1): 169–184.

Kohut, Andrew, Paul Taylor, Scott Keeter, Carroll Doherty, Michael Dimock, and Kim Parker. 2011. "Angry Silents, Disengaged Millennials: The Generation Gap and the 2012 Election." *Pew Research Center for the People & the Press*, November 3. Accessed January 13, 2017. http://www.people-press.org/2011/11/03/the-generation-gap-and-the-2012-election-3/, November 3.

Kramer, Gerald H. 1966. "A Decision Theoretic Analysis of a Problem in Political Campaigning." In *Mathematical Implications in Political Science*. Vol. 2. edited by Joseph L. Bernd, 137–160. Dallas: Southern Methodist University Press.

Kreiss, Daniel, and Philip N. Howard. 2010. "New Challenges to Political Privacy: Lessons from the First US Presidential Race in the Web 2.0 Era." *International Journal of Communication* 4: 1032–1050.

Nickerson, David W., and Todd Rogers. 2014. "Political Campaigns and Big Data." *Journal of Economic Perspectives* 28 (2): 51–74.

Panagopoulos, Costas. 2009. "Campaign Dynamics in Battleground and Non-Battleground States." *Public Opinion Quarterly* 73 (10): 119–129.
Panagopoulos, Costas. 2016. "All About That Base." *Party Politics,* 22 (2): 179–190.
Panagopoulos, Costas. 2017. *Political Campaigns: Concepts, Context, and Consequences*. New York: Oxford University Press.
Panagopoulos, Costas, and Peter W. Wielhouwer. 2008. "Polls and Elections The Ground War 2000-2004: Strategic Targeting in Grassroots Campaigns." *Presidential Studies Quarterly* 38 (2): 347–362.
Pew Research Center. 2012. "Young Voters Supported Obama Less, But May Have Mattered More." *Pew Research Center for the People & the Press.* Accessed January 13, 2017. http://www.people-press.org/2012/11/26/young-voters-supported-obama-less-but-may-have- mattered-more/.
Plutzer, Eric. 2002. "Becoming a Habitual Voter: Inertia, Resources, and Growth in Young Adulthood." *American Political Science Review* 96 (1): 41–56.
Rosenstone, Steven J., and John Mark Hansen. 1993. *Mobilization, Participation, and Democracy in America*. New York: MacMillan.
Serazio, Michael. 2014. "The New Media Designs of Political Consultants: Campaign Production in a Fragmented Era." *Journal of Communication* 64 (4): 743–763.
Shaw, Daron R. 2006. *The Race to 270: The Electoral College and the Campaign Strategies of 2000 and 2004*. Chicago: University of Chicago Press.
Sides, John, Daron Shaw, Matt Grossman, and Keena Lipsitz. 2015. *Campaigns and Elections*. 2nd ed. New York: W.W. Norton and Company.
Stoker, Laura, and M. K. Jennings. 2008. "Of Time and the Development of Partisan Polarization." *American Journal of Political Science* 52 (3): 619–635.
Vavreck, Lynn. 2007. "The Exaggerated Effects of Advertising on Turnout: The Dangers of Self-Reports." *Quarterly Journal of Political Science* 2 (4): 325–343.
Virginia State Board of Elections. 2012. "November 6, 2012 General Election Official Results." Accessed February 26, 2014. https://www.voterinfo.sbe.virginia.gov/election/DATA/2012/68C30477-AAF2-46DD-994E-5D3BE8A89C9B/Official/1_s.shtml.
Wielhouwer, Peter W. 2003. "In Search of Lincoln's Perfect List." *American Politics Research* 31 (6): 632–669.