# Hacking the Electorate

How Campaigns Perceive Voters

How politicians, in the context of their *campaigns*, *perceive* voters and how those perceptions translate into the relationship politicians build with their electorates, specifically regarding direct contact.

# Motivation: 2 Common Misconceptions

1.  **The Information fallacy** - propagated by media and academia - that campaigns have encyclopedic knowledge of every voter
    - Not true!
        i.   Voters are constantly in transition
        ii.  Candidate-centered nature of American politics
        iii. Even the most sophisticated campaigns use primarily public records
            1. Cheap, readily available, and relatively predictive
    - Not necessarily useful
2.  Studied in academia - Inconsistency between actual voter and **perceived voter**

# Voter

## Nathan Fielder

Host of Nathan For You

Founder of an outdoor apparel company

5' 9''

Studied business at University of Victoria

Key policy concern is protecting small businesses

Enjoys thinly cut watermelon

# Perceived Voter

## Fielder, Nathan

Male

White

35

Vancouver, Canada

Democrat

# Policy behind Public Data

- Voter registration, census, and licensing data
- Help America Vote Act of 2002 (HAVA)
  - Requires every state to develop a "single, uniform, official, centralized, interactive computerized statewide voter registration list defined, maintained, and administered at the State level"
- Conflict of interest and potential abuse of power

The World

Data Collection & Formatting
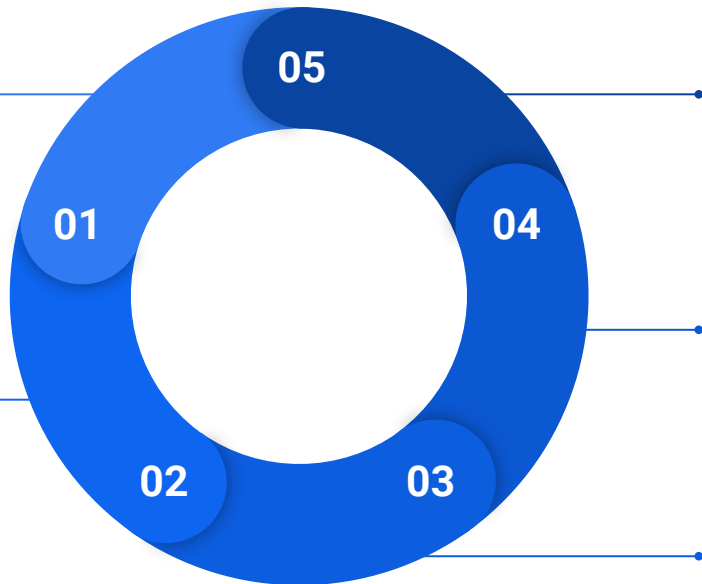
Decisions, Predictions, Actions

Model

ML Algorithm

01

02

03

04

05

**The Perceived Voter Model** draws attention to how and why the particular set of data available to campaigns affects their assessments of voters, which in turn guides strategic decisions

# Perceived Voter Model

- **Campaign =** the formal campaign organization of a political candidate seeking office
- Elite-Level Study - not focused on behavior of the public
- **Perceived voter** = the attributes of voters that campaigns consider when making strategic decisions
- **Direct Voter Contact** = door-to-door, telephone, and mail appeals that are largely volunteer based
  - Accommodates the finest-grained strategic choices
  - Particularly affected by variations in public data policy
  - Relies on databases also used in constituent services and in other governmental function
- Use of shortcuts or heuristics
  - Public data

# Hypotheses

Strategic decisions for mobilization and persuasion can be explained by variation in public data laws. Broadly, that the availability of public records that are predictive of partisan support will affect whether a campaign uses a geographic-level contacting strategy or an individual-level contact strategy.

# Hypotheses: Mobilization

1. If party affiliation is available, campaigns will focus more on…
   a. Mobilizing individual voters and less on geographic areas that are highly concentrated with partisans
   b. Mobilizing supporters and less on trying to persuade undecided voters
      i. Voting among partisans will be higher in these areas as compared to places where party affiliation is not available
2. If voting records contain a voter's race, campaigns will focus more on…
   a. Mobilizing voters based on their racial identity, and less on mobilizing voters because of the racial composition of a neighborhood
      i. Black voters in particular will be mobilized to vote more in states with public race data
         1. When racial data is not available, expect lower turnout of black voters, but relatively higher turnout in areas with a high concentration of black residents, since geographic rather than individual strategy was used

# Hypotheses: Persuasion

- As there is no data that successfully predicts voter persuadability, campaigns will not direct their contracting efforts to persuadable voters

# Data Accumulation

- Individual voter characteristics are stored in databases and every major campaign can engage with voters based on those characteristics.
- Catalist
  - Makes its prediction about voter partisanship based on over 150 variables
- NGP VAN
  - A company that provides a user interface that allows campaign workers to interact with databases

# Catalist

- Specializes in microtargeting for Democratic political campaigns.
- Claims to have data on 240 million unique individuals in the US.
- Role in the 2016 election:
  - Analyzed records from 10 battleground states and found a major influx of new voters who were responsible for the record-breaking turnout in the Republican primaries.
  - Found that Sanders supporters voted less frequently and were less reliably Democratic than Clinton supporters.

# NGP VAN

- Privately owned voter database and web hosting service provider.
- Partisan provider of campaign compliance software.
- Used by most Democratic members of Congress.
- Utilized by the Obama, Hillary Clinton, and Sanders campaigns as well as the British Liberal Democrats and Liberal Party of Canada.

# Databases

- Why did campaigns obtain data from intermediary vendors or political parties rather than retrieving a list directly from the election authority?
  - The reason campaigns use intermediary data suppliers is that registration lists can be augmented substantially to increase their usefulness.
- Current Democratic database is called VoteBuilder.
- Current Republican database is called Voter Vault.

# Implication

- "With Catalist I have a comprehensive database including hundreds of characteristics about every American voter.
- With the GCP I have a survey of campaign workers engaged in direct contact.
- With the NGP VAN query project I have thousands of list queries generated in one state.

# Importance

- These data resources together provide new insights into the strategic capabilities and perceptions of political campaigns.
- Through these resources I can measure how voters appear from the campaign's-eye view.
- I can examine the American public, not as voters, but as perceived voters - the avatars that exist in campaign databases."

# Results of Tested hypothesis

Perceptions and Availability of public data

1) Perceiving partisanship with public records
   a) Two forms of party information available to campaigns
      i) Party registration
      ii) Party primary data
   b) The public party information is highly predictive, but not they are not available in many states
2) Perceiving partisanship without public Identifiers
   a) Geographic level strategy - past election result by precinct
      i) Problem: most voters live in mixed-partisan precincts.
   b) Commercial microtargeting models (Catalist's model)
   c) Even with a model like Catalist's, the perceptions of voters look very different than in states with public records of partisanship → consequence of the information availability

# Results of Tested hypothesis

How do perceptions of partisanship affect strategies

1) In party registration states
   a) Target independent voters who have a high chance of voting → use direct contact strategies.
   b) Direct contact strategy was more focused on mobilizing voters (GOTV) based on partisanship
   c) Less focus on mobilizing voters in partisan neighborhoods
   d) Less accidental contact with out-partisans
2) In non-party registration states
   a) Focus more on persuading undecided voters because campaigns have less reliable data
      i) More people appears as persuadable from Catalist
   b) Focus more on persuading than mobilizing

# Results of Tested hypothesis

Downstream Effects of the strategies on Voters

Because campaigns stick to certain patterns of strategies, there are possible downstream effects

a) In non-party registration swing states, campaigns focus less on mobilizing → partisanship (according to self-reported data) was not correlated with turnout
b) While in party registration swing states, the partisanship was correlated with turnout → due to campaigns focusing on mobilization
c) Even though the overall turnout among partisans was lower in these states, turnout in overwhelmingly partisan precincts was higher → campaigns relied on geographic strategies

# Results of Tested hypothesis

Key Takeaways

1) Public data policies (deciding availability of public partisan data) affect how campaigns perceive their supporters and how they decide their strategies.
2) Campaigns' perceptions vary drastically across the country because of the difference in data availability.
3) Perceived Voter Model shows how campaigns interact with voters based on their perceptions, but the perceptions can be varied based on the dataset that campaigns have.
4) Therefore, the source data which dictate the campaign's ability to perceive voters are crucial

# Chapters 6-7
# Context Behind "Racialized Engineering"

- **Book was written in 2015**
  - Before Trump bid
  - Before Cambridge Analytica scandal
  - Before mainstream coverage of gerrymandering
  - Before DNC email hack (although the author does reference information published on WikiLeaks)
- **Analysis based on 2008 election**
  - Most of the analysis was conducted around 2010
  - There was no incumbent running in 2008

# Use of Public Record

**TABLE 6.1** *Percentage of Registrants Identifying Their Race on Voter Registration Forms*

| State | Race Listed on Voter File | Total Obs. in Sample |
|---|---|---|
| Alabama | 99.0% | 56,176 |
| Florida | 96.7 | 192,856 |
| Georgia | 96.1 | 100,936 |
| Louisiana | 98.1 | 51,789 |
| Mississippi | 38.1 | 37,148 |
| North Carolina | 98.0 | 113,497 |
| South Carolina | 99.7 | 51,033 |
| Tennessee | 67.5 | 71,330 |

*Source:* Two independently drawn Catalist 1% Analytics Samples, combined.

Claim:
**Campaigns focus more attention on voters' races when public race data are available.**

# Use of Public Record

Vote sample of North Carolina:

Voter information fee for Alabama:

# How Catalist Makes Predictions About Race

TABLE 6.2 *Commercial Predictions of Race are Less Accurate than Public Records of Race*

|  | Commercial Data | | Voter File Data | |
|---|---|---|---|---|
|  | Pct. Correctly Predicted | Obs. | Pct. Correctly Predicted | Obs. |
| Race = White | 90% | 18,528 | 94% | 3,296 |
| Race = Black | 68 | 2,403 | 95 | 841 |
| Race = Hispanic | 73 | 1,898 | 92 | 262 |

*Note:* Data are generated from the 2008 Cooperative Congressional Election Study matched to commercial data by Catalist. Catalist supplies a prediction of race for voters who are not listed in the public record by race. Because most voters do not live in one of the eight states that collect racial data in public records, the sample size for the commercial predictions is much larger than the sample size for the voter file data.

- Voters' names
- Racial composition of voters' neighborhoods
- …

# Why Is This Important: Downstream Effects

Data environment

⬇

Campaign's perceptions

⬇

Campaign strategies

⬇

Demographics of voters mobilized into the political process

**Republican strategy:**
Ignore voters listed as African-American

**Democratic strategy:**
Exclude outreach to voter with public records about gun ownership

# Claim: Race on Voter File Causes Higher Turnout

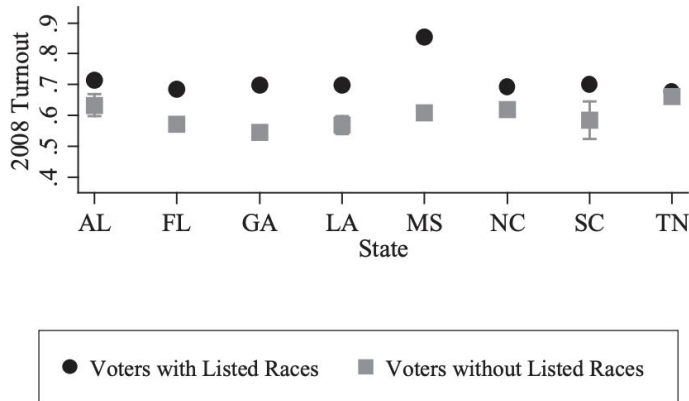Figures: voter turnout discrepancy between voter with and without listed races



TABLE 6.3 *Difference in Voting Rates between Treatment Group (Race Listed on Voter File) and Control Group (Race Not Listed on Voter File)*

| Voter Turnout | Race Listed on Voter File | Race Not Listed on Voter File | Dif. Of Means p-Value |
|---|---|---|---|
| **Blacks** | | | |
| 2008 General | 76% (8,887) | 72 (2,903) | 0.00 |
| 2006 General | 50 (7,124) | 44 (2,219) | 0.00 |
| 2004 General | 75 (6,411) | 72 (1,992) | 0.02 |
| **Whites** | | | |
| 2008 General | 81 (128,812) | 79 (34,745) | 0.00 |
| 2006 General | 62 (119,583) | 60 (21,502) | 0.00 |
| 2004 General | 84 (112,973) | 83 (20,562) | 0.00 |

*Note:* Observations are in parentheses. Comparison is between matched groups.

# Commercial Data

- Commercial Data is not very effective
- Gives slight (~2-4%) information about turnout
- Machine learning is only as effective as the correlations
- This is encouraging, it means that public data holds most of the power

# Social Networks

- Social networking is not very effective either
- This is because:
  - People don't want to campaign to their social circles
  - Most circles are homogenous
- Again, public data proves far more important

# Normative Questions

1. Is it good that campaigns collect microtargeting data from administrative databases?
2. Is microtargeting bad for democracy?
   a. Public policy acts as a lever for this

# Perverse Incentives

- Constituent databases are merged with campaign information
  - Congress has bias when they interact with their constituents
- Political policy can create databases useful for electioneering, giving inappropriate incentives
- Stricter policy is necessary

# Microtargeting: good or bad?

- Pros:
  - Campaigns can connect with the electorate
  - They can know what voters care about, and make changes based on that
- Cons:
  - Do not need to appeal to the whole electorate
  - Increases political echo chambers

# Solutions

- Campaigns releasing their databases semi-publically
- Privacy constraint that a voter can only look up their own information
- Helps keep transparency and voter interaction

# Thank you!