A Natural Experiment on Discrimination in Elections

David E. Broockman, Stanford GSB
Evan J. Soltas, MIT

Motivation

Among elected officials worldwide, racial/ethnic minorities and women are underrepresented:

- Nonwhites in US:
    - 38% of 2015 US population (Census)
    - 17% of US Congress in 2015
    - 18% of US state legislatures in 2015

- Women in US:
    - 51% of 2015 US population (Census)
    - 19% of US Congress in 2015
    - 24% of US state legislatures in 2015

Motivation

Descriptive representation in politics matters for outcomes:

- **Political**: Politicians more closely reflect preferences of, obtain public goods for those who share their descriptive characteristics

  (Pande 2003, Chattopadhyay and Duflo 2014, Broockman 2013)

- **Economic**: Improvement in health, education, and criminal justice for descriptively represented groups

  (Beaman et al. 2012, Clots-Figueras 2012, Iyer et al. 2012, Fujiwara 2015, Ferreira and Gyourko 2014, Vogl 2014)

$\rightarrow$ What are the causes of underrepresentation in elected office?

Motivation

A potential explanation: voter discrimination, wherein voters do not vote for a nonwhite or female candidate but would have voted for an otherwise-identical white or male candidate

Motivation

A potential explanation: voter discrimination, wherein voters do not vote for a nonwhite or female candidate but would have voted for an otherwise-identical white or male candidate

- Two broad theoretical mechanisms could generate discrimination:

    1. **Statistical (Akerlof 1976, Phelps 1972):** Voters use politicians' characteristics to accurately infer other attributes, leading them to prefer whites or men

Motivation

A potential explanation: voter discrimination, wherein voters do not vote for a nonwhite or female candidate but would have voted for an otherwise-identical white or male candidate

- Two broad theoretical mechanisms could generate discrimination:

  1. **Statistical (Akerlof 1976, Phelps 1972):** Voters use politicians' characteristics to accurately infer other attributes, leading them to prefer whites or men

  2. **Taste-based (Becker 1957):** Voters have preferences over candidate race or gender, pay costs to discriminate
     - Voters vote against minority candidates they favor on non-demographic dimensions:
     - ↓ minority representation, ↓ appeal of elected officials on non-demographic dimensions

- → Policy relevance: What institutions and interventions can increase minority and female officeholding?

Overview of Paper

We use a natural experiment to study voter discrimination in elections:

- 4 Illinois Republican presidential primaries from 2000 to 2016
- Unique features of election design support credible inference and support tastes as likely mechanism

Overview of Paper

We use a natural experiment to study voter discrimination in elections:

- 4 Illinois Republican presidential primaries from 2000 to 2016
- Unique features of election design support credible inference and support tastes as likely mechanism
- We find:
  - Nonwhites receive **9%** fewer votes than otherwise identical whites
  - No gender discrimination: Small bias *for* women in some estimates
  - Heterogeneity further supports tastes as likely mechanism
  - All results survive a battery of robustness checks
- Conclusion: Evaluate substantive importance and external validity

Previous Literature

- Credible empirical evidence on voter discrimination is limited

  Washington (2006), Mas and Moretti (2009), Stephens-Davidowitz (2014), Kuziemko and Washington (2018)

- Many lab/field/quasi experiments on discrim. beyond elections

  Goldin and Rouse (2000), Bertrand and Mullainathan (2004), List (2004), Pager (2009), Price and Wolfers (2010), Doleac and Stein (2013), Rubinstein and Brenner (2014), Glover et al. (2017), Hedegaard and Tyran (2018)

- Many challenges prevent straightforward application of their research designs to real elections

Roadmap

Roadmap

Taste-Based Discrimination: A Definition

"If an individual has a '**taste for discrimination**,' he must act as if he were willing to pay something...to be associated with some persons instead of others. When actual discrimination occurs, he must, in fact, either pay or forfeit income for this privilege."

– Gary S. Becker,
The Economics of Discrimination (1957)

- Agent engaged in taste-based discrimination trades-off favored **demographic** for **non-demographic traits** to avoid "psychic cost"
- In politics instead of markets: individually inconsequential, collectively decisive

Identifying Voter Discrimination

- **What is the ideal experiment?** 3 criteria:

  1. Choice between otherwise-identical candidates but for race/gender
  2. Voter information sets restrict motives for statistical discrimination
  3. Costly for voters to discriminate

- **Real world:** Challenges in accomplishing all (any?) of 3

  1. Nonwhite/female candidates differ from white/male candidates
  2. Candidate multidimensionality and limited voter information often make statistical discrimination plausible, even likely
  3. Discrimination in lab/survey experiments is often costless

$\rightarrow$ **Our paper:** Try to hit 3 birds with one quasi-experimental stone

Portion of Ballot from McLean County, IL-18

**FOR DELEGATE TO THE
NATIONAL NOMINATING CONVENTION
EIGHTEENTH CONGRESSIONAL DISTRICT**
(PLEASE NOTE: Next to the name of each candidate for
delegate appears in parentheses the candidate's preference for
President of the United States or the word "uncommitted".)
(Vote for not more than three)

- ⬭ **ROBERT BROWNING** (CHRISTIE)
- ⬭ **MARY K. BROOKHART** (CHRISTIE)
- ⬭ **DONNA K. THOMPSON** (CHRISTIE)
- ⬭ **JIM EDGAR** (BUSH)
- ⬭ **BILL BRADY** (BUSH)
- ⬭ **RAYMOND POE** (BUSH)
- ⬭ **KENT GRAY** (TRUMP)
- ⬭ **SANDRA YEH** (TRUMP)
- ⬭ **WILLIAM GRAFF** (TRUMP)

Properties of Election Setting

- Presidential candidates nominate delegate candidates:

  - K (usually $K = 3$) delegate candidates per congressional district
  - Bound to support presidential candidate at convention
  - Usually recruited from campaign email lists (not a plum job)

Properties of Election Setting

- Presidential candidates nominate delegate candidates:
  - K (usually $K = 3$) delegate candidates per congressional district
  - Bound to support presidential candidate at convention
  - Usually recruited from campaign email lists (not a plum job)
- Voters can:
  - Vote for up to K delegates
  - Split votes among different presidential candidates' delegates
- Top K delegates win, even if split among presidential candidates
- No other information on ballot but delegate name and candidate

Properties of Election Setting

- Presidential candidates nominate delegate candidates:

  - K (usually $K = 3$) delegate candidates per congressional district
  - Bound to support presidential candidate at convention
  - Usually recruited from campaign email lists (not a plum job)

- Voters can:

  - Vote for up to K delegates
  - Split votes among different presidential candidates' delegates

- Top K delegates win, even if split among presidential candidates

- No other information on ballot but delegate name and candidate

  - But name reveals race and gender $\rightarrow$ opportunity for discrimination
  - Simple formalization in paper: voters trade-off "psychic cost" (Becker 1957) of voting for nonwhites and instrumental and expressive value of voting for preferred Presidential candidate

Identification Strategy

- Intuition:

  - Suppose Tom, Dick, and Harry are candidates for Mitt Romney delegates in IL-1. To maximize the value of his or her ballot, a Romney supporter should vote for Tom, Dick, and Harry.

  - Now imagine that Romney candidates are instead Tom, Dick, and José. Romney supporter should vote for Tom, Dick, and José.

  - But if the Romney supporter has a sufficiently strong racial taste, he or she may vote for Tom and Dick but not for José $\rightarrow$ José wins fewer votes than Tom and Dick.

Identification Strategy

- Intuition:

  - Suppose Tom, Dick, and Harry are candidates for Mitt Romney delegates in IL-1. To maximize the value of his or her ballot, a Romney supporter should vote for Tom, Dick, and Harry.

  - Now imagine that Romney candidates are instead Tom, Dick, and José. Romney supporter should vote for Tom, Dick, and José.

  - But if the Romney supporter has a sufficiently strong racial taste, he or she may vote for Tom and Dick but not for José $\rightarrow$ José wins fewer votes than Tom and Dick.

- Within candidate $\times$ county-district $\times$ year cells, do nonwhite delegates receive on average fewer votes than white delegates?

- 816 natural experiments of the form above

Identifying Voter Discrimination

Problem 1: Nonwhite/female politicians differ from whites/males

- Run in different areas, at different times, on different platforms, with different experience, with different resources, . . .

Our setting: Candidates within "cell" are plausibly comparable

- Same platform (tied to same presidential candidate)
- Face same voters on same ballot (all of whom should support)
- Whites/males are naturally-occurring control groups

Problem 2: Many plausible voter inferences in most real elections

- Hard to hold constant voter perceptions of all relevant non-race or non-gender dimensions, even in artificial survey experiments

  (Heckman 1998, Dafoe et al. 2016)

Our setting: Little room for problematic voter inferences

- Voters (and economists) get perfect signal of sole dimension
- No other relevant dimensions for voters to make inferences about

Identifying Voter *Taste-Based* Discrimination

Problem 3: What is the cost of discrimination?

- Hard to say that voters for a non-minority candidate prefer the minority on non-demographic dimensions

- Hard to say how voters who discriminate in hypotheticals would behave in a real election

Our setting: Unambiguous and real-life trade-off for discriminators

- To discriminate, voters must abandon their presidential candidate (give up both expressive and instrumental gains)

- Clear meaning to concept that alternative must be inferior on attributes other than race or gender

Why Study Republican Primaries?

- **Ex ante:** reasons to suspect Republican primaries would be an important channel for nonwhite and female underrepresentation

  - Strong partisan preferences mean general election votes will be difficult to change; discrimination is "less costly" in primaries

  - Republican voters generally more racially conservative since the 1960s/70s

Why Study Republican Primaries?

Table: Racial and Gender Composition of Officeholders and Voters by Party

	% Nonwhite 2006-14		% Female 2016	
	Republicans	Democrats	Republicans	Democrats
U.S. House Members	4%	31%	9%	31%
Primary Voters	11%	35%	51%	60%
Party Identifiers	15%	43%	53%	59%

- Obviously, these descriptive patterns are merely suggestive.

Roadmap

1. Election Setting and Identification Strategy

2. **Data**

3. Results

4. Discussion

Data Sources

- **IL Board of Elections:** County-district vote totals by delegate
- 3 measures for delegate race/ethnicity:

  1. **Subjective perception, Amazon mTurk:** Paid 'master' Turk workers (Kuziemko et al. 2015) to guess the race of each delegate with a provided **full name**, yielding 30 guesses per delegate name: White, Black, Hispanic, East Asian, Indian, Middle Eastern

  2. **Objective, US Census:** Public-use file w/ racial composition of **surnames only** occurring $100+$ times ($N = 151,671$ names): White, Black, Hispanic, Asian

  3. **Objective, Onolytics:** Proprietary commercial database of 1.5 million **full names** matched to name origins: White, Black, Hispanic, East Asian, Indian, Middle Eastern

  $\rightarrow$ Combine these into PC1 index

- Gender: Use SSA first name file to predict gender (also: age)

Data Sources

For controls, we collected information voters might plausibly know (and more):

- **Background checks:** RAs scoured internet for officeholders
  - Mostly minor/local office: judge, town council, board of ed., etc.

- **Delegate addresses** from candidacy filings:
  - Will drop home counties as robustness check
  - Measure income and schooling of home Census block group

Some Facts About Delegate Candidates

$N = 2{,}386$ delegates across four elections

- **Racial composition:**
  - 94% non-Hispanic white
  - 4% Hispanic, 1% black, 1% Asian

- **Officeholding:** 39% are public/party officials (mostly local)

- Nonwhite delegates nominated by all candidates in our sample; results not driven by particular candidates

Some Facts About IL Republican Voters

- Almost entirely white:
    - IL voter file (Catalist): 96% white
    - 2008 CCES: 97% white
- 51% male
- Older, higher-income, more educated than IL population
- IL generally near median US state in racial attitudes according to several measures
    - Racially-charged search rates, self-administered IAT scores, per-capita rate of hate crimes / number of hate groups
- Voters seem to care about presidential primaries:
    - Turnout is much higher in presidential primaries than non-presidential primaries
    - Little roll-off from "beauty contest" to delegate selection

Estimation

$$\mathbb{E}[\text{Votes}_{ipct}] = \exp\left(\beta \cdot \text{Nonwhite}_i + \gamma \cdot \text{Female}_i + \mathbf{X}'_{ipct}\delta + \alpha_{pct}\right)$$

- Poisson regression: We estimate *share of potential votes* a delegate loses by being nonwhite or female by comparison to otherwise-identical white or male

- Unit of observation: Delegate i, county-district c, election t

- Votes$_{ipct}$: Vote count delegate i receives in c at t

- Nonwhite$_i$, Female$_i$: Delegate's perceived race and gender

- α_{pct}: Fixed effect for each cell of comparable delegates tied to p

- Cluster standard errors at delegate level

Roadmap

	MTurk		Census		Onolytics		Rescaled PC1	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	0.006 (0.006)	0.006 (0.006)	0.004 (0.006)	0.001 (0.006)	0.005 (0.006)	0.004 (0.006)	0.005 (0.006)	0.001 (0.006)
Nonwhite	-0.087*** (0.016)		-0.045*** (0.014)		-0.035*** (0.007)		-0.092*** (0.016)	
Black		-0.033 (0.050)		0.018 (0.025)		-0.040*** (0.015)		-0.094*** (0.032)
Hispanic/Latino		-0.057*** (0.018)		-0.061*** (0.018)		-0.045*** (0.012)		-0.079*** (0.018)
Any Asian				-0.069*** (0.014)				-0.110*** (0.022)
East Asian		-0.093*** (0.025)				-0.055*** (0.008)		
Indian		-0.174*** (0.040)				-0.076*** (0.013)		
Middle Eastern		-0.160*** (0.031)				-0.088*** (0.033)		
White Ethnic						-0.016* (0.009)		
N	18,958	18,958	16,945	11,166	18,639	18,639	16,668	11,049
Pseudo-R^2	0.991	0.991	0.991	0.992	0.991	0.991	0.991	0.992

Heterogeneity in Voter Discrimination

We present several cuts of the data:

- By race/gender of presidential candidate
- By competitiveness
- By candidate left–right ideology
- Across geography
- Over time

Main take-away: Results further support tastes as likely mechanism

- If racial tastes have stakes, should see voters with strong racial tastes sort out of voting for nonwhite presidential candidates

	(1)	(2)	(3)
Nonwhite Delegate			
$\times$ White Pres. Cand.	-0.101***		-0.104***
	(0.017)		(0.017)
$\times$ Nonwhite Pres. Cand.	0.052		0.053
	(0.064)		(0.069)
Female Delegate			
$\times$ Female Pres. Cand.		0.335***	0.335***
		(0.096)	(0.109)
$\times$ Male Pres. Cand.		0.003	0.003
		(0.006)	(0.007)
χ^2 test of equality	7.07***	12.49***	15.73***
N	17,126	18,958	16,668
Pseudo-R^2	0.990	0.991	0.991

Heterogeneity: By the "Cost of Discrimination"

- Voter demand for discrimination slopes down
- Non-zero equilibrium discrimination even in close elections

	Statewide Votes
Nonwhite $\times$ Below-Median Votes	-0.572***
	(0.145)
Nonwhite $\times$ Above-Median Votes	-0.093***
	(0.013)
N	16,668
Pseudo-R^2	0.992

Heterogeneity: By Presidential Candidate Ideology

- No heterogeneity by presidential candidate ideology (consistent with no statistical discrimination)

	Ideology
Nonwhite $\times$ Left of Median CFScore	-0.108***
	(0.019)
Nonwhite $\times$ Right of Median CFScore	-0.105***
	(0.019)
N	16,668
Pseudo-R^2	0.992

Other Heterogeneity Results

- **Time:** Flat time trends for discrimination by race and gender

- **Geography:**

  - Nonwhites face more discrimination in counties with:

    - Low share of college graduates

    - Low median household income

    - High Republican vote share in presidential elections

  - Same qualitative patterns for women, but substantially weaker

  - Conclusions robust in within-delegate, across-county specification

Robustness

- Already seen: Robust to multiple methods of coding the racial backgrounds implied by delegate names

- Results nearly identical with $ln(1 + \text{Votes})$ and vote share DVs

- Robust to controls for ballot order (not presented for time)
  - We find ballot order is uncorrelated with race or gender

- Robust to controls for signals of age in name and income/education of delegate's home Census block group

Robustness: Voter Prior Information

- Concern:

    - Voters may have information other than name they feel is relevant on some delegate candidates (e.g., current officeholders)

    - May be more such prior info for white delegates than nonwhites

    - Intuition: I see my mayor on the ballot, may vote for her even if she is not my preferred candidate's delegate

- We offer 3 robustness checks:

    1. Lightweight control for indicator of prior info: prior candidacy $\rightarrow$ results unchanged

    2. "Background check" control for delegate officeholding: $Corr(Office_i, Nonwhite_i) \approx 0 \rightarrow$ nonwhite estimate unchanged by control, female estimate becomes slightly *positive*

    3. Drop home-county observations $\rightarrow$ results unchanged

Other Alternative Explanations: Voter Discrimination

- Inferences about presidential candidates:
  - Absorbed into cell fixed effects
  - Would bias away from our finding if prejudiced voters select out

- Unobservable confounders in general:
  - Perform Oster (2019) sensitivity analysis
  - Confounder would need to be 25x more correlated with ethnicity than existing controls

Other Alternative Explanations: Taste

- Residual incentives for statistical discrimination

    - Example of implausibility: at least 9% of Romney voters would need to believe that the nonwhite Romney delegates were less likely to vote for Romney at the convention than the white Santorum delegates were to vote for Romney

    - In convenience sample, Facebook-recruited survey of self-reported IL primary voters, little difference in expected loyalty of white and nonwhite delegates

    - Null by candidate ideology also inconsistent with statistical discrimination

Other Alternative Interpretations: Taste

- Do voters understand that delegate voting has stakes?
  - Voters who do not see incentive to vote for delegates should select out
  - Unclear why we would see the heterogeneity we did
  - In convenience sample, Facebook-recruited survey of self-reported IL primary voters, 72% understood (over a year after the election) and those who did still discriminated

Other Alternative Interpretations: Taste

- Do voters understand that delegate voting has stakes?
  - Voters who do not see incentive to vote for delegates should select out
  - Unclear why we would see the heterogeneity we did
  - In convenience sample, Facebook-recruited survey of self-reported IL primary voters, 72% understood (over a year after the election) and those who did still discriminated
- Are voters signaling to presidential candidates?
  - Cannot rule this out definitively, but seems extremely unlikely; Unaware of any other evidence voters do this
  - In survey, few say they would do this; $> 2x$ as many (9%) say voting for nonwhites makes them "uncomfortable"
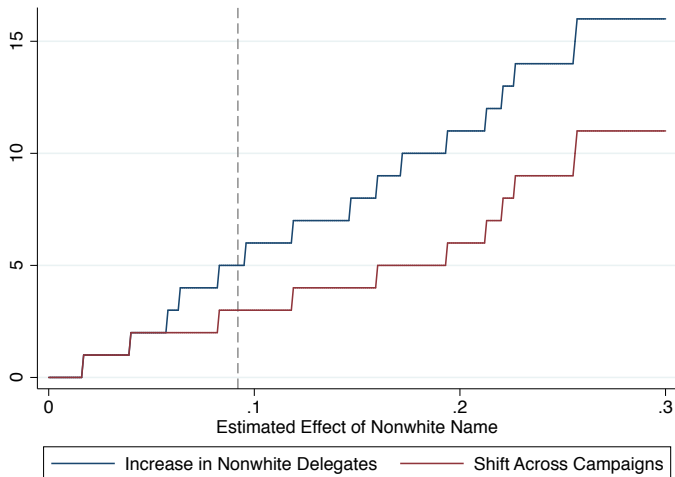 - Voters have other opportunities to signal if they wish

Roadmap

How relevant is discrimination to nonwhite underrepresentation?

Table: Example Delegate Selection Outcome: Top Six Delegates by Vote, Illinois 6th Congressional District 2016

Delegate Name	Presidential Candidate	# Votes	Won Delegate Election?
Paul Minch	Trump	37,150	Won
Barbara Kois	Trump	36,838	Won
Patrick Brady	Kasich	34,072	Won
Ronald Sandack	Kasich	33,538	Lost
Aaron Del Mar	Kasich	32,228	Lost
Nabi Fakroddin	Trump	32,136	Lost

Discussion: IL Counterfactual

- Discrimination reduces wins for nonwhite delegates by 20%
- Voters paid costs in equilibrium: Multiple favored delegates lose

Estimated Effect of Nonwhite Name

Increase in Nonwhite Delegates — Shift Across Campaigns

Q: How many House seats do nonwhites win if we "undo" 9% penalty?

Discussion: US House Counterfactual

Q: How many House seats do nonwhites win if we "undo" 9% penalty?

A: About 3 p.p. more, from 17% in 2015 (Many caveats to this, including potential additional psychic costs of voting for nonwhites)

	All Primaries		Favored Open Seats	
	Rep.	Dem.	Rep.	Dem.
# with Nonwhite <9 p.p. from White Winner	19	14	6	6
# Won by Nonwhites	218	372	9	21
# with Nonwhite Candidates	378	540	35	50
# of Close Primaries (<9 p.p. Margin)	160	130	32	24
# of Primaries (Total)	4,300	4,302	239	188
Increase in Nonwhite Winners (%)	8.7%	3.8%	66.7%	28.6%
Increase in Nonwhite Win Rate (p.p.)	5.6%	2.8%	20.0%	14.0%
% of Close Primaries Changed	11.9%	12.0%	18.75%	25.0%
% of All Primaries Changed	0.5%	0.3%	**2.9%**	**3.7%**

Discussion

- 9% of IL Republican primary voters avoid voting for their favored presidential candidate to avoid voting for a nonwhite delegate

- Unique evidence of discrimination against nonwhites

- Many results consistent with predictions of taste-based interpretation. Framework indicates two consequences:

  1. Reduction in non-white representation, hurting *nonwhites'* interests

  2. White candidates who win as a result also *represent discriminating voters* worse, potentially reducing quality of officeholders

- Implications for VRA, efforts to increase nonwhite representation

- Suggests party elites have an incentive to avoid nominating nonwhites, even if they could perfectly signal their attributes

- Consistent with other research, suggests *voter* tastes not main barrier to women's representation