Ballot Design and Electoral Outcomes: The Role of Candidate Order and Party Affiliation

Alessandro Arlotto,[a,*] Alexandre Belloni,[a,c] Fei Fang[b] and Saša Pekeč[a]

[a]The Fuqua School of Business, Duke University, 100 Fuqua Drive, Durham, 27708, NC, USA, [b]Yale School of Public Health, Yale University, 60 College St, New Haven, 06510, CT, USA and [c]WW FBA Selling Partner Services, Amazon, 320 108th Ave NE, Bellevue, 98004, WA, USA

[*]To whom correspondence should be addressed: aa249@duke.edu

Abstract

We study how designing ballots with and without party designations impacts electoral outcomes when partisan voters rely on party-order cues to infer candidate affiliation in races without designations. If the party orders of candidates in races with and without party designations differ, these voters might cast their votes incorrectly. We identify a quasi-randomized natural experiment with contest-level treatment assignment pertaining to North Carolina judicial elections and use double machine learning to accurately capture the magnitude of such incorrectly cast votes. Using precinct-level election and demographic data, we estimate that 12.08% (95% confidence interval: 4.95%, 19.20%) of democratic partisan voters and 13.63% (95% confidence interval: 5.14%, 22.10%) of republican partisan voters cast their votes incorrectly due to the difference in party orders. Our results indicate that ballots mixing contests with and without party designations mislead many voters, leading to outcomes that do not reflect true voter preferences. To accurately capture voter intent, such ballot designs should be avoided.

Keywords. Partisan and nonpartisan elections, Ballot design, Causal inference, Conditional average treatment effect

History. This version: October 8, 2024

Introduction

Voting is fundamental to the proper functioning of a democratic society and its institutions. It provides elected officials with the legitimacy and authority to govern, shape policy, make laws, and administer them while allowing voters to hold these officials accountable. While many aspects—such as partisanship [see, e.g., 2, 10, 4, 11] and incumbency [see, e.g. 15, 7, 6]—can affect voters' preferences, a voting process that ensures the accurate communication of voters' intent is crucial to prevent misinterpretation and ensure correct electoral outcomes. Election ballots play a crucial role and, paraphrasing from [26], ballot design should not determine the outcome of an election under any reasonable standard of fairness. One infamous example is the butterfly-ballot design used in Palm Beach County, FL, which misinterpreted the intent of thousands of voters, potentially altering the outcome of the 2000 U.S. Presidential Election in which over a hundred million votes were cast [see, e.g., 22, 27, 23]. Other examples involve ballot effects pertaining the length and complexity of the ballot, the order of candidates within a race, and the order and placement of races on ballots [see, e.g., 20, 16, 12, 13, 14, 19, 1].

In this paper, we study the effects of a ballot design in which candidates' party designations (or labels) are included for certain *partisan* contests and are omitted for other *nonpartisan* contests. Because voters' decisions are often strongly influenced by party identification in both such contests [3], the absence of party labels increases the chance that voters abstain voting in those contests [also known as *roll off*, cf. 9, 24, 1] or resort to ballot cues [8, 9, 17, 20, 16, 18]. In our analysis, we focus on ballot cues and examine whether *partisan voters*—those who vote by party even in nonpartisan contests—use the *party order* from a contest with party designations (i.e., the sequence in which parties in that contest appear on the ballot) to inform their voting in nonpartisan contests. If partisan voters cast their ballots assuming that the party orders of partisan and nonpartisan contests are the same, then their votes would not represent their intent when the two party orders actually differ. We say that a *(party-order) flip* occurs when the party order of a race differs from that of the U.S. presidential race in the same election. We are interested in measuring the impact of such a flip on nonpartisan vote shares, and we postulate that such impact is *heterogeneous* with respect to the presidential vote share. We are interested in two different estimands. The first estimand is the *flip effect*, which measures the conditional average impact of a party-order flip on the vote share of a party's candidate in a contest without party designations, compared to the vote share of the presidential candidate of the same party. The second estimand is the share of *partisan-voting mistakes*, which measures the proportion of partisan voters whose vote is cast incorrectly due to the

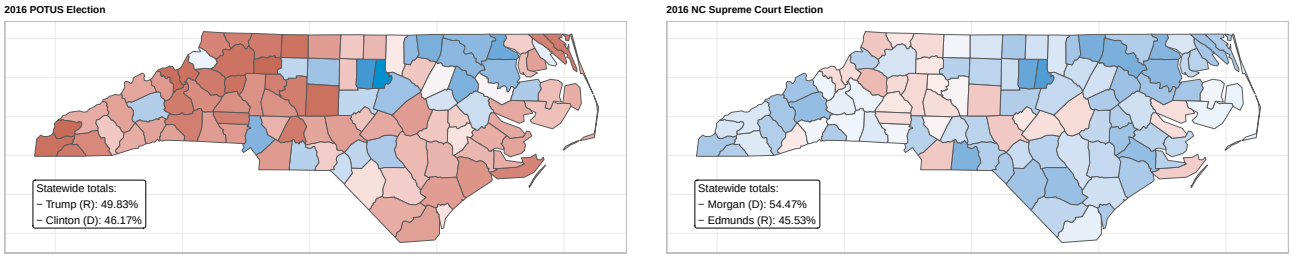

Fig. 1. 2016 North Carolina electoral maps by county. The left chart displays the outcome of the 2016 presidential race in each county in North Carolina. The republican (democratic) presidential candidate won the counties colored in red (blue), with darker shades indicating larger wins. The right chart shows the analogous county outcomes of the 2016 NC Supreme Court race, using the same color coding. The 2016 NC Supreme Court race was on the same ballot as the 2016 presidential election. The presidential candidates were listed on the ballot *with* their party labels: Donald J. Trump (R) was listed first, followed by Hillary Clinton (D). In contrast, the NC Supreme Court candidates were listed *without* party labels. Michael R. Morgan—endorsed by the democrats—was listed first, while Robert H. Edmunds—endorsed by the republicans—was listed second.

flip. In essence, partisan-voting mistakes represent the overall proportion of votes that do not reflect voter intent, while the flip effect measures the net impact on vote shares, considering that mistakes by opposing partisan voters cancel out.

To measure the flip effect, we analyze 13 North Carolina (NC) statewide judicial races *without* party designations during the general elections of 2004, 2008, 2012, and 2016. We leverage the NC General Statute to identify a quasi-randomized natural experiment with contest-level treatment assignment. By attributing party affiliations through contemporaneous endorsements, we find that 7 of these contests exhibit a party-order flip relative to the presidential race, while 6 do not. Using double machine learning [cf. 5, 21] on a dataset containing precinct-level election and demographic data for these 13 contests (34, 345 observations), we estimate the flip effect. The use of double machine learning allows us to adaptively achieve a more flexible specification (e.g., functional form and model selection), increasing robustness and efficiency and enabling us to construct valid confidence intervals and p-values. We use cluster-robust standard errors to account for the dependence of election outcomes within the same contest. Additionally, due to the heterogeneity of the effect across different levels of the presidential vote share, we use bootstrap to construct a simultaneously valid confidence band for the flip effect for all levels of the presidential vote share.

Our analysis reveals that a ballot design with party-order flips and without party designations can affect election outcomes in the following ways:

- We estimate that 12.08% (p-value 0.00086) of democratic partisan voters and 13.63% (p-value 0.00172) of republican partisan voters cast their votes incorrectly due only to the flip.
- Allowing for heterogeneity is crucial for estimating the impact of a party-order flip on election outcomes of contests without party designations. Our analysis rejects the hypothesis that the flip effect is homogeneous (p-values $<$ 0.01 for both parties). In turn, the flip effect is positive when the corresponding presidential vote share is low and negative when it is high (p-value 0.00280 for democrats and 0.00724 for republicans).
- The overall average treatment effect of a party-order flip in the NC judicial races is not statistically significant (p-value 0.95842 for democratic candidates and 0.99615 for republican candidates). However, this result is an artifact of partisan-voting mistakes from both parties canceling each other out and North Carolina having an approximately even party split among presidential voters.

By conducting a placebo test on 7 flipped and 6 non-flipped NC judicial contests *with* party designations during the 2016 and 2020 general elections, our analysis also finds that the flip effect disappears when party designations are included on the ballot (p-value 0.96513 for the democratic partisan voters and 0.99386 for the republican partisan voters).

Taken together, our results from judicial races with and without party designations suggest that including party designations for all contests on the ballot helps correctly capture the intent of the electorate.

Motivating example

The 2016 general election in North Carolina featured a key judicial contest for a seat on the NC Supreme Court as Associate Justice. Heading into the election, the republican justices on the court held a 4-3 majority, with the republican incumbent Robert H. Edmunds running to remain in his seat and retain that majority. The challenger, democratic Judge Michael R. Morgan, was running for the same seat seeking to secure a democratic majority on the court. In accordance with the NC General Statute (see Section 3.1), the race was nonpartisan. This meant that the names of the two candidates would appear on the ballot in a random order and without the corresponding party labels: Judge Morgan (D) was listed first, and Justice Edmunds (R) was listed second. In contrast, at the time, the law prescribed that candidates from the party of the sitting governor (republican) would be listed first in all partisan statewide races, including the U.S. presidential race. In the language of this paper, the *party order* of the 2016 NC Supreme Court contest was *flipped*, i.e., the party order of this nonpartisan judicial contest was the opposite of the party order of the presidential race.

Figure 1 presents the North Carolina, county-level, electoral maps of the 2016 U.S. presidential race (left panel) and of the 2016 NC Supreme Court race (right panel). The two maps show that, in most counties, the victory of a given party's presidential candidate corresponded to a loss or a weaker victory for the candidate of the same party running for the NC Supreme Court. The extent of this phenomenon suggests that the presence of a *flip effect*: the party-order flip of the 2016 NC Supreme Court race might have played a role in how voters cast their ballots.

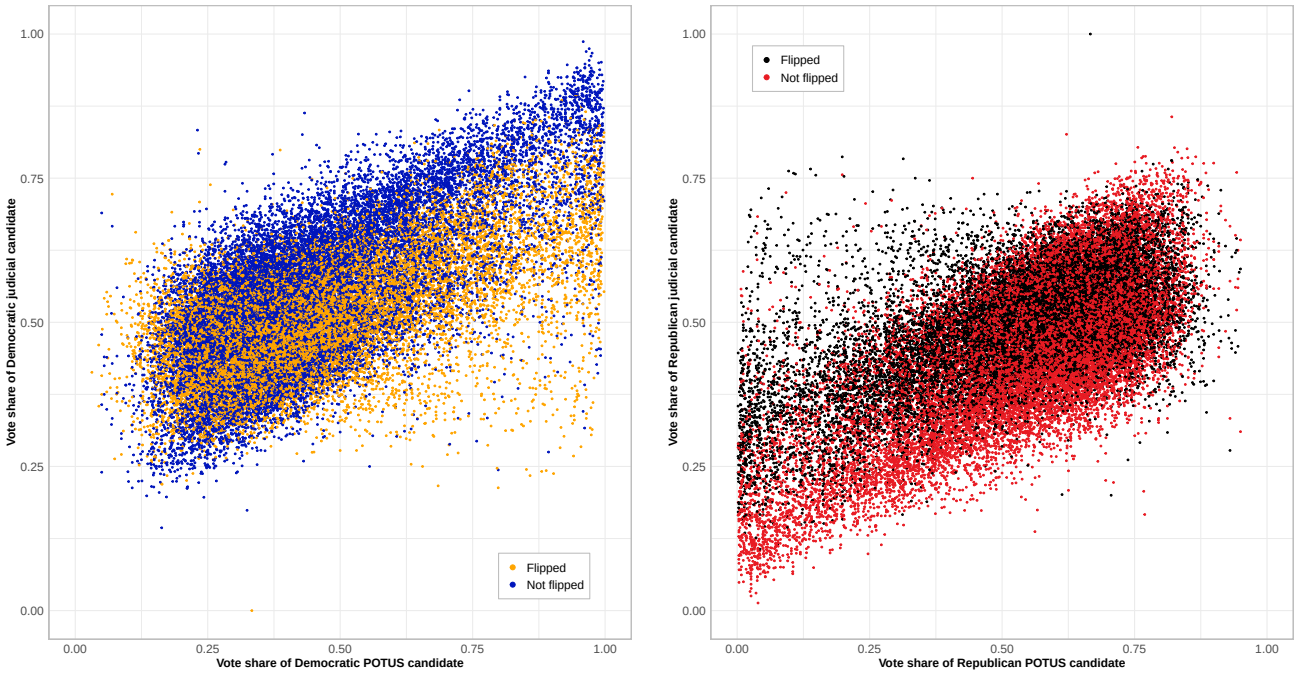

Fig. 2. NC precinct-level vote share by party. The left chart displays the precinct-level vote share of democratic judicial candidates (y-axis) against the democratic presidential vote share in the same election (x-axis) in North Carolina. Judicial races with the same party order as the presidential race are in blue. Judicial races in which the party order is flipped are in orange. The right chart displays the corresponding plot for the republican party with non-flipped judicial races in red and flipped judicial races in black.

While the 2016 NC Supreme Court race provided us with anecdotal evidence of a flip effect, a richer dataset with all of the judicial contests without party labels during the general elections of 2004, 2008, 2012, and 2016 supports the same conclusion. Figure 2 plots the precinct-level vote share of a given party's judicial candidate against the precinct-level vote share of that party's presidential candidate. By using different colors to distinguish judicial contests that are flipped from those that are not, we again see evidence of a flip effect. If the presidential candidate of a given party wins in a precinct, then the win of the judicial candidate of the same party tends to be weaker (and possibly a loss) if the party order of that judicial contest is flipped. Similarly, if the presidential candidate loses in a precinct, then the judicial candidate of the same party tends to lose by a smaller margin (or even wins) if the party order is flipped. As such, the effect of a party-order flip on the vote share of a judicial candidate would be positive when the vote share of the corresponding presidential candidate is low and negative when it is high. In other words, the flip effect varies (is *heterogeneous*) relative to the vote shares of the presidential contest.

The flip effect shows how a party-order flip affects a judicial candidate's vote share, but it does not break down the components of this effect, obscuring the magnitude of partisan voter mistakes, which may be canceling out. To measure the magnitude of partisan-voter mistakes due to a flip, we can estimate the flip effect in hypothetical precincts where the presidential vote share is 100%, resulting in mistakes from voters of only one party.

Election laws, data, and methods

In this section, we briefly review the evolution of the NC General Statute regulating party designations and candidate orders (Section 3.1), discuss the data used for our analysis (Section 3.2), and establish our causal inference framework (Section 3.3).

Party affiliation and order of candidates on official ballots

The NC General Statute, Chapter 163, §165.5, determines whether a ballot should feature the party designations of the candidates in a specific contest, while §165.6 dictates the order in which these candidates must be listed.

For instance, during the general elections of 2004, 2008, and 2012, presidential candidates along with their party affiliations were arranged alphabetically by party (for the parties with a minimum 5% statewide voter registration). Consequently, ballots featured the democratic presidential candidate first, the republican presidential second, followed by third-party candidates (in alphabetical order) and write-ins. In contrast, in those same elections, statewide judicial candidates were arranged alphabetically by last name, without party designation included on the ballot.

Several amendments to the law were implemented leading up to the 2016 general election. Notably, party designation became mandatory for the candidates running for the NC Court of Appeals (Session Law, SL, 2015-292), and candidates of the party that received the most votes in the latest gubernatorial race would be listed first (SL 2013-381 and 2016-109). Consequently, in 2016, the republican presidential candidate appeared first on the ballot, and so did all of the republican candidates for the NC Court of Appeals. In contrast, the candidates for Associate Justice of the NC

Supreme Court lacked any party designation and were arranged alphabetically by last name, starting from the letter H. This ballot arrangement was the result of two random drawings: the first drawing selected the letter H from a bowl, and the second drawing determined the order as alphabetical rather than reverse alphabetical.

Following the 2016 election, the NC General Assembly further amended the NC General Statute (SL 2016-125 and SL 2018-99) by (i) characterizing all statewide judicial elections as partisan, (ii) mandating their party designations to be printed on the ballot, (iii) and requiring all candidates in any election ballot to be listed in either alphabetical or reverse-alphabetical order, according to a drawing conducted by the State Board of Elections. In the 2020 general election, two random drawings determined candidates to be listed alphabetically, commencing with the letter O.

By regulating the order of candidates for each contest, the NC General Statute establishes a corresponding party order, either apparent with party labels or implied by party endorsements and affiliations. As we discuss in Section 3.3, the difference in the arrangement rules for partisan and judicial contests provides us with quasi-randomized party-order flips and the exogenous treatment variation that is at the core of the identification strategy for our causal analysis.

Data

Our dataset consists of precinct-level data for 26 statewide judicial races (NC Supreme Court and NC Court of Appeals) that took place during the general elections of 2004, 2008, 2012, 2016, and 2020. Each of these races included two candidates from opposing parties (a republican and a democrat), so we have a total of 69,161 observations of judicial vote shares for each party. During these elections, each ballot included between 15 and 20 statewide races, with the judicial races placed after all the other statewide races.

The dataset was built by combining precinct-level election results of statewide races with precinct-level information about NC voters who participated or were registered to vote in each election. This information is broken down by precinct, voting method, and voter demographics such as party affiliation, race, ethnicity, gender, and age. The election and voter data we used is maintained by the NC State Board of Elections, and it is publicly available at `https://www.ncsbe.gov/`. Additionally, we include election turnout statistics, as well as the gender of each judicial candidate, whether they are listed first in their contest, if they are running as an incumbent, and their party affiliation.

Because the law regulating the content and arrangement of official ballots has varied substantially over time (see Section 3.1), our dataset includes 13 judicial races without party designations and 13 races with party designations. For the 13 races *without* party designations, we determine the party affiliation of each judicial candidate. We find 7 judicial races with a party-order flip relative to the presidential race and 6 without (totaling to 34, 345 per party). Among the 13 races *with* party designations, there are 7 that feature a party-order flip relative to the presidential race and 6 that do not (totaling to 34, 816 per party). Thus, for each judicial race, we also include the binary treatment variable *flip* to track whether the party order of the judicial race differs from the party order of the presidential race in the same election.

Causal inference framework, identification, and estimation

For a given nonpartisan contest c and a given electoral precinct p,[1] we let Y_{cp} denote the vote share of a given party's candidate in contest c and in precinct p. We let $Y_{cp}(0)$ denote the potential vote share when the party order of contest c is the same as the party order of the presidential race on the same ballot, whereas $Y_{cp}(1)$ represents the potential vote share when the party order of contest c is flipped. We use the binary variable $T_c \in \{0, 1\}$ to denote the treatment status of contest c. We set $T_c = 1$ if the party order of contest c is flipped, and $T_c = 0$ otherwise. In our analysis, we make the following assumptions for identification [25].

Assumption 1 (No interference among units) The precinct's potential vote share of each given contest depend only on the treatment assigned to that contest and do not depend on the treatment assigned to other contests.

In the context of NC judicial elections without party labels, this assumption implies that the potential vote shares of a judicial candidate in one precinct do not depend on the party order of other judicial elections. This assumption is likely to hold because voters either (i) have little or no visibility into the mechanisms that lead to different party orders on a ballot, or (ii) are so well-informed about the election that they are able to vote for the candidates they actually support regardless of the party order of their contests. Such an assumption is common in many causal studies, and it has been successfully applied in related causal analyses of election ballot design [cf. 12, 1].

The NC General Statute that regulates the content and arrangement of official ballots motivates us to make the following assumption about the treatment (flip) assignment.

Assumption 2 (Known random assignment) The treatment is randomly assigned by a known mechanism.

In our application, the mechanism that determines the party order of a judicial contest has varied over time (see Section 3.1). However, the within-ballot order of candidates that such mechanisms induced is essentially randomized. To see why this is true, consider a given judicial contest and take any two last names at random. Then the arrangement rules of Section 3.1 give us a 50% chance that either of those last names would be listed first on that contest's ballot. Because the party labels are absent and because each candidate is endorsed by or affiliated with one party, this results in a 50% chance of a flip.

We are interested in two (related) estimands. The first estimand is the *flip effect*, which is the conditional average effect of a party-order flip on the vote share of a judicial candidate relative to the corresponding presidential vote share. We expect the flip effect to be *heterogeneous* with the presidential vote share. Well-informed voters are able to vote for their intended candidate, while less-informed partisan voters might use party order as a cue to identify the candidate affiliated with their party. As such, they would vote for the candidate in the same within-contest position as the presidential candidate of their party, and their vote would not represent their intent if the party-order of that contest is flipped. The fraction of votes that a nonpartisan candidate would receive from voters of their party would likely be higher

[1] The contest and precincts subscripts are both election year specific.

in the absence of a party-order flip, and lower otherwise. Because all voters are exposed to the same nonpartisan contests, the ultimate share of votes that a candidate receives when a party-order flip occurs comes from the votes cast correctly by their partisan voters (which is likely lower than the corresponding vote share when no flip occurs), as well as from the votes cast incorrectly by the partisan voters of the opposing party (which likely increase the candidate's vote share). Thus, allowing for such heterogeneity is essential for accurately modeling precinct-level, down-ballot voting without party labels.

The second estimand is the share of *partisan-voting mistakes* that can be attributed to such ballot design. Because the vote share of a given judicial candidate includes the votes cast correctly by the candidate's partisan voters and the votes cast incorrectly by the partisan voters of the opposing party, the first estimand is directly associated with the outcome (actual and potential). In contrast, the second estimand focuses on capturing the intent of partisan voters. This estimand could face an identification problem since we only observe the net effect on each outcome. However, it can be identified by examining (hypothetical) precincts where the presidential candidate of a given party receives 100% of the votes. In these precincts, there are no voters from the opposing party, eliminating the votes incorrectly cast by them. Identifying the heterogeneity of the flip effect (first estimand) then allows us to identify the share of partisan-voting mistakes (second estimand), which we assume to be homogeneous across the presidential election vote share. (This homogeneity holds if the propensity for partisan voters to make mistakes is independent of the precinct in which they vote.)

To conduct our analysis, we let X_p be the vote share in precinct p of the presidential candidate on the same ballot and of the same party as the candidate in contest c, and we let W_{cp} and Z_c respectively be the (same-party or party independent) precinct- and contest-level controls for contest c and precinct p. We then consider the following specification that allows for contest-level treatment assignment and heterogeneous treatment effect:

$$Y_{cp} = f(X_p)T_c + g(X_p, W_{cp}, Z_c) + \epsilon_{cp}, \tag{1}$$

where f captures the *flip effect*, the conditional average treatment effect of a party-order flip relative to the presidential vote share X_p. In (1), the function g captures the contribution of X_p and of other controls, and the error terms ϵ_{cp} are assumed to be independent across contests and centered, i.e., $\mathbb{E}\left[\epsilon_{cp}|T_c, X_p, W_{cp}, Z_c\right] = 0$.

To estimate the heterogeneity of the flip effect with respect to the presidential vote share, we consider the following third-degree specification for the flip-effect:[2]

$$f(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3, \tag{2}$$

where x is the presidential vote share and $\theta_0, \theta_1, \theta_2$ and θ_3 are coefficients to be estimated.

Based on the specifications in (1) and (2), we apply double machine learning to estimate the flip effect. This framework allows us to incorporate modern machine learning methods to obtain a flexible specification for the impact of the covariates (nuisance function) without affecting the validity of our confidence regions and p-values for the estimands of interest. Even with randomized assignment, incorporating these additional controls is important to obtain tighter confidence bands (higher efficiency) for the conditional average treatment effect given each value of the presidential vote share. Because of the correlation of the precinct-level election outcomes for the same contest, we need to account for cluster-robust standard errors where each judicial contest corresponds to a cluster. Moreover, we also account for such dependence in the construction of the bootstrap which is needed to build simultaneously valid confidence bands across different presidential vote shares.

[2] We also separately study a first-degree model and obtain findings that are similar the ones we discuss here.

Results from North Carolina's Judicial Elections

In this section, we utilize our dataset of NC judicial elections to estimate the effect of a flip in the party order of judicial candidates relative to the party order of presidential candidates. We perform our analysis by slicing our dataset by party because of the dependence that arises due to the vote shares of democratic and republican candidates essentially adding up to one. Whenever judicial contests are listed on the ballot *without* party designations, this allows us to separately estimate the flip effect for the democratic and the republican party (see Section 4.1). We also use the outcome of judicial races listed on the ballot *with* party designations to conduct a placebo test (see Section 4.2).

Main result: conditional average flip effect without party labels

In this section, we estimate the heterogeneous flip effect in (1) and (2) for the democratic and the republican party, respectively denoted by $\hat{f}_d$ and $\hat{f}_r$. Our estimates use 34,345 precinct-level observations corresponding to 7 flipped and 6 non-flipped statewide judicial races *without* party designations in North Carolina (see Section 3.2).

Figure 3 plots the estimated flip effect for judicial candidates without party designations as a function of the vote share of the corresponding presidential candidate of the same party. In both panels, the solid lines represent the conditional average effect of a party-order flip as a function of the presidential vote share x, the shaded regions provide 95% cluster-robust pointwise confidence intervals, and the outer dashed lines correspond to the 95% uniform confidence band. Our analysis finds a statistically significant flip effect for both parties (p-value 0.00280 for the democratic party and 0.00724 for the republican party). Moreover the analysis also supports the heterogeneity of the flip effect relative to the presidential vote share. That is, it rejects the null hypothesis of a homogeneous flip effect with p-values smaller than 0.01 for both parties.

The left panel depicts the flip effect for democratic judicial candidates. The x-axis plots the vote share of the democratic presidential candidate in the same precinct and in the same election, and the y-axis plots the estimated conditional average flip effect $\hat{f}_d(x)$ (solid blue line). For a given democratic presidential vote share x, the value $\hat{f}_d(x)$ estimates the causal impact of a party-order flip on the vote share of a democratic judicial candidate. *Low* vote shares (small x) for the democratic presidential candidate correspond to a *positive* conditional average flip effect ($\hat{f}_d(x) > 0$), while *high* presidential vote shares (large x) correspond to a *negative* conditional average flip effect ($\hat{f}_d(x) < 0$). In other words, in a precinct with a smaller democratic base, the effect of a party-order flip

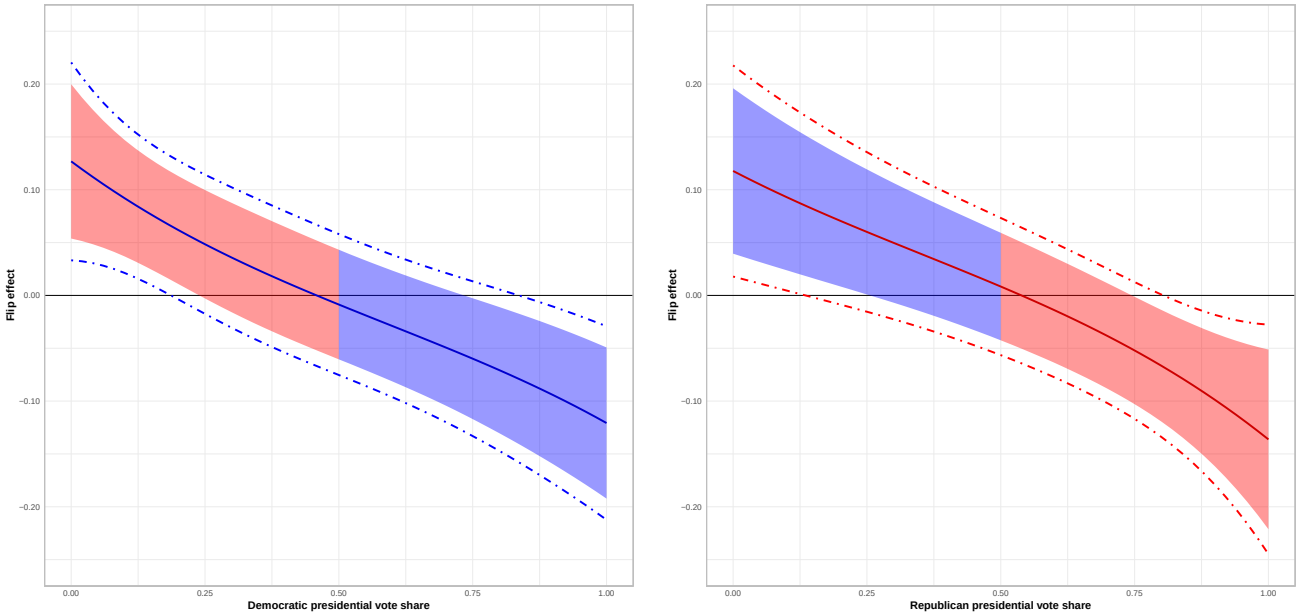

Fig. 3. Flip effect without party labels. The left chart shows the estimated flip effect $\hat{f}_d(x)$ for democratic judicial candidates without party designation as a function of x, the vote share of the democratic presidential candidate in the same election and precinct (solid blue line). Shaded region: 95% pointwise confidence intervals. Outer blue dashed lines: 95% uniform confidence band. The right chart shows the analogous flip effect for republican judicial candidates (lines in red).

and the absence of party labels help the democratic judicial candidate whose vote share increases, on average, because the larger share of votes cast incorrectly is by republican partisan voters. In contrast, in a precinct with a larger democratic base, the biggest share of incorrectly cast votes comes from democratic partisan voters thereby reducing the vote share of the democratic judicial candidate on average.

In a precinct where the votes in the presidential race are almost evenly split between two parties ($x \approx 0.5$), the conditional average flip effect $\hat{f}_d(x)$ is not statistically significant. That is, votes cast incorrectly by democratic partisan voters and votes cast incorrectly by republican partisan voters cancel each other out, making the effect of a party-order flip negligible. This negligibility, however, should not be interpreted as reassuring. The fact that partisan-voting mistakes from both parties cancel out does not mean that voters are effectively casting their votes for their intended candidate. Our model also allows us to estimate the fraction of *partisan-voting mistakes* by setting the presidential vote share $x = 1$ and the estimate $\hat{f}_d(1)$ (where a negative estimate corresponds to partisan-voting mistakes of democratic partisan voters). In a hypothetical precinct with $x = 1$ there are only presidential voters of one party, so partisan-voting mistakes cannot cancel out. We find that the average proportion of partisan-voting mistakes made by democrats is 0.12077 (p-value 0.00086), and the 95% confidence interval is given by $(0.04951, 0.19202)$.

In the right panel of Figure 3, we find the conditional average flip effect for republican judicial candidates and obtain similar findings. All of the earlier observations about the left chart apply here as well with the caveat that, in the right chart, small values of x correspond to precincts in which the republican presidential candidate has few votes, while large values of x correspond to precincts in which the republican presidential candidate receives a large fraction of votes. Just like above, when $x = 1$ in the right chart, we find that the average proportion of partisan-voting mistakes made by republicans is 0.13625 (p-value 0.00172), and the 95% confidence interval $(0.05141, 0.22109)$.

Our analysis also allows us to test whether partisan-voting mistakes vary by party. In our dataset, we find that $\hat{f}_d(1) - \hat{f}_r(1)$ is not statistically different from zero (p-value > 0.05). We can also test if partisan-voting mistakes cancel out when the presidential vote share is evenly split. Our data show that both $\hat{f}_r(1/2)$ and $\hat{f}_d(1/2)$ are not statistically significant (p-values > 0.05). Taken together, these results support the homogeneity of partisan-voting mistakes across parties.

We conclude this section by computing the average treatment effect (ATE) of a flip for both parties, disregarding the heterogeneity with respect to the presidential vote share. In such a model, the estimate for the ATE for the democratic party is 0.00123 (p-value 0.95842). Similarly, the ATE for the republican party is 0.00012 (p-value 0.99615). These estimates suggest that the (unconditional) average flip effect is not significant, though such a conclusion is an artifact of the approximately even party mix of voters in the state, and of the reciprocal transfer of voting mistakes from the candidate of one party to the candidate of the opposing party. Heterogeneity with respect to the presidential vote share again appears to be crucial to quantify the causal effect of a party-order flip. When heterogeneity is taken into account, our model suggests that average share of partisan-voting mistakes of the democratic and the republican party are respectively 12.08% and 13.63%. As a reference point, we note that in the 2000 U.S. Presidential Election less than 1% of democratic voters in Florida's Palm Beach County were misled by the infamous butterfly ballot [22, 27, 23].

Placebo test: conditional average flip effect with party labels

To test the specifications of our heterogeneous treatment effect framework in (1), we conduct a placebo test using 34,816

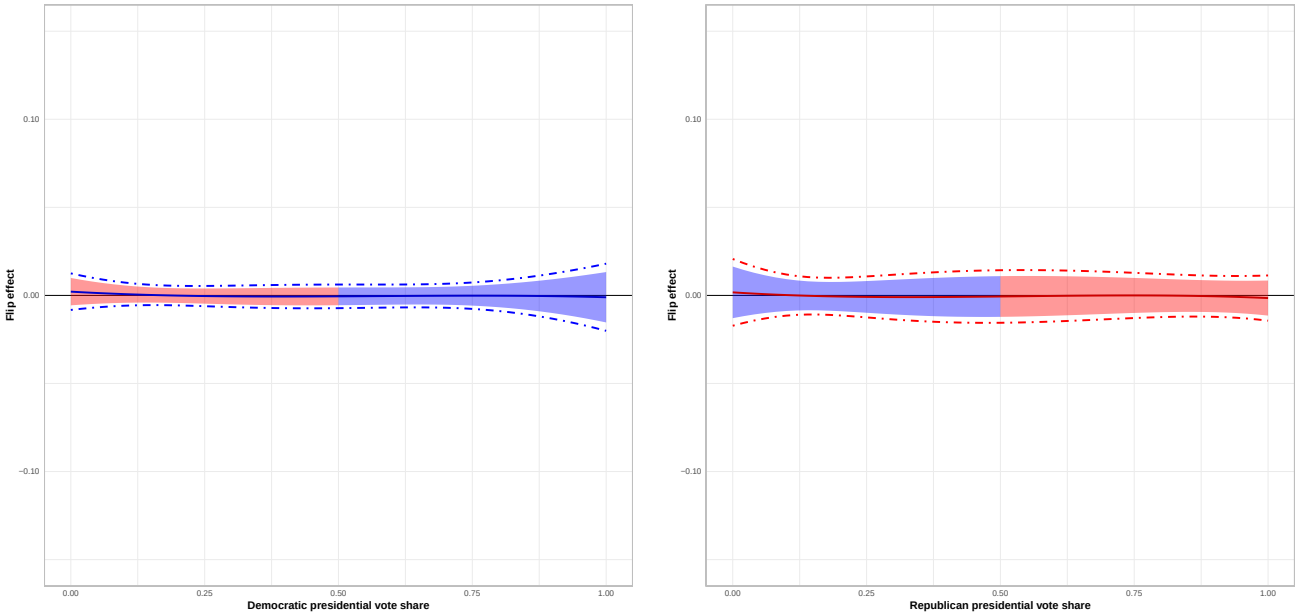

Fig. 4. Flip effect with party labels. The left chart shows the estimated flip effect $\hat{f}_d(x)$ for democratic judicial candidates with party designation as a function of x, the vote share of the democratic presidential candidate in the same election and precinct (solid blue line). Pointwise confidence intervals are given by the shaded region, while uniform confidence intervals are depicted by the outer blue dashed lines. The right chart shows the analogous flip effect for republican judicial candidates (lines in red).

precinct-level observations from 7 flipped and 6 non-flipped statewide judicial races with party designations in North Carolina (see Section 3.2). Our placebo analysis mirrors Section 4.1. Importantly, we establish that the presence of party designations nullifies the effect that party-order flips have on judicial vote shares. This finding supports the robustness of our original research design.

Figure 4 plots the estimated democratic (left) and republican (right) flip effects for judicial contests with party designations conditional on the vote share of the corresponding presidential candidate of the same party. In other words, Figure 4 is the analog of Figure 3 when party labels are included on the ballot. The figure shows that the presence of party labels neutralizes the causal effect of a party-order flip. For any given presidential vote share x, the estimated flip effects are close to zero and not statistically significant (p-value 0.96513 for the democrats and 0.99386 for the republicans).

Conclusion

We show that electoral contests without party designations, appearing on ballots alongside races with party designations, can mislead a significant number of voters. Partisan voters might incorrectly infer the party affiliation of candidates in races without party labels by looking at the (party) order of the candidates in races with party labels. Consequently, the outcomes of such races without party labels might not reflect the true preferences of the electorate.

In a close contest, the mistakes of partisan voters from opposing parties could cancel each other out, concealing the number of votes incorrectly cast due to incorrect inferences about candidates' party affiliations (flip).

Leveraging changes in the NC General Statute that affected the arrangement and order of candidates on official ballots over the last 20 years, we use a quasi-randomized natural experiment with state-level treatment assignment pertaining to NC judicial elections. This allows us to identify both the net flip effect and the total partisan-voting mistakes. Using precinct-level election and demographic data, we use double machine learning to accurately capture the heterogeneous effect of a party-order flip for voters of each party. Specifically, we estimate that 12.08% (95% confidence interval: 4.95%, 19.20%) of democratic partisan voters and 13.63% (95% confidence interval: 5.14%, 22.10%) of republican partisan voters cast their votes incorrectly.

Our results indicate that ballots mixing contests with and without party designations mislead many voters, leading to outcomes that do not reflect true voter preferences. To accurately capture voter intent, such ballot designs should be avoided.

References

1. Ned Augenblick and Scott Nicholson. Ballot position, choice fatigue, and voter behaviour. *The Review of Economic Studies*, 83(2):460–480, 12 2015.
2. Larry M. Bartels. Partisanship and voting behavior, 1952-1996. *American Journal of Political Science*, 44(1):35–50, 2000.
3. Chris W. Bonneau and Damon M. Cann. Party identification and vote choice in partisan and nonpartisan elections. *Political Behavior*, 37(1):43–66, 2015.
4. Ted Brader and Joshua A. Tucker. Following the party's lead: Party cues, policy opinion, and the power of partisanship in three multiparty systems. *Comparative Politics*, 44(4):403–420, 2012.
5. Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *Econom. J.*,

21(1):C1–C68, 2018.

6. Gary W. Cox and Jonathan N. Katz. Why did the incumbency advantage in U.S. House elections grow? *American Journal of Political Science*, 40(2):478–497, 1996.
7. Gary W. Cox and Scott Morgenstern. The increasing advantage of incumbency in the U.S. states. *Legislative Studies Quarterly*, 18(4):495–514, 1993.
8. Philip L. Dubois. Voting cues in nonpartisan trial court elections: A multivariate assessment. *Law & Society Review*, 18(3):395–436, 1984.
9. Richard L. Engstrom and Victoria M. Caridas. Voting for judges: Race and roll-off in judicial elections. *Political Participation and American Democracy*, 92:171–191, 1991.
10. Morris P Fiorina. Parties and partisanship: A 40-year retrospective. *Political Behavior*, 24:93–115, 2002.
11. Morris P Fiorina. *Unstable majorities: Polarization, party sorting, and political stalemate*. Hoover Press, 2017.
12. Daniel E Ho and Kosuke Imai. Randomization inference with natural experiments: An analysis of ballot effects in the 2003 California recall election. *Journal of the American Statistical Association*, 101(475):888–900, 2006.
13. Daniel E. Ho and Kosuke Imai. Estimating causal effects of ballot order from a randomized natural experiment: The California alphabet lottery, 1978–2002. *Public Opinion Quarterly*, 72(2):216–240, 05 2008.
14. Amy King and Andrew Leigh. Are ballot order effects heterogeneous? *Social Science Quarterly*, 90(1):71–87, 2009.
15. Gary King and Andrew Gelman. Systemic consequences of incumbency advantage in U.S. House elections. *American Journal of Political Science*, 35(1):110–138, 1991.
16. Jonathan G.S. Koppell and Jennifer A. Steen. The effects of ballot position on election outcomes. *The Journal of Politics*, 66(1):267–281, 2004.
17. Monika L. McDermott. Voting cues in low-information elections: Candidate gender as a social information variable in contemporary United States elections. *American Journal of Political Science*, 41:270–283, 1997.
18. Monika L. McDermott. Candidate occupations and voter information shortcuts. *The Journal of Politics*, 67(1):201–219, 2005.
19. Marc Meredith and Yuval Salant. On the causes and consequences of ballot order effects. *Political Behavior*, 35:175–197, 2013.
20. Joanne M. Miller and Jon A. Krosnick. The impact of candidate name order on election outcomes. *The Public Opinion Quarterly*, 62(3):291–330, 1998.
21. Miruna Oprescu, Vasilis Syrgkanis, Kieth Battocchi, Maggie Hei, and Greg Lewis. EconML: A machine learning library for estimating heterogeneous treatment effects. In *33rd Conference on Neural Information Processing Systems, Vancouver, Canada*, page 6, 2019.
22. Robert C. Sinclair, Melvin M. Mark, Sean E. Moore, Carrie A. Lavis, and Alexander S. Soldat. An electoral butterfly effect. *Nature*, 408(6813):665–666, 2000.
23. Richard L. Smith. A statistical assessment of Buchanan's vote in Palm Beach County. *Statistical Science*, 17(4):441–457, 2002.
24. James M. Vanderleeuw and Glenn H. Utter. Voter roll-off and the electoral context: A test of two theses. *Social Science Quarterly*, 74:664–673, 1993.
25. Tyler J. VanderWeele. Ignorability and stability assumptions in neighborhood effects research. *Stat. Med.*, 27(11):1934–1943, 2008.
26. Jonathan N. Wand, Michael C. Herron, and Henry E. Brady. Ballot cost Gore thousands of votes. *San Diego Union-Tribune*, pages G–3, November 19, 2000.
27. Jonathan N. Wand, Kenneth W. Shotts, Jasjeet S. Sekhon, Walter R. Mebane, Michael C. Herron, and Henry E. Brady. The butterfly did it: The aberrant vote for Buchanan in Palm Beach County, Florida. *American Political Science Review*, 95(4):793–810, 2001.