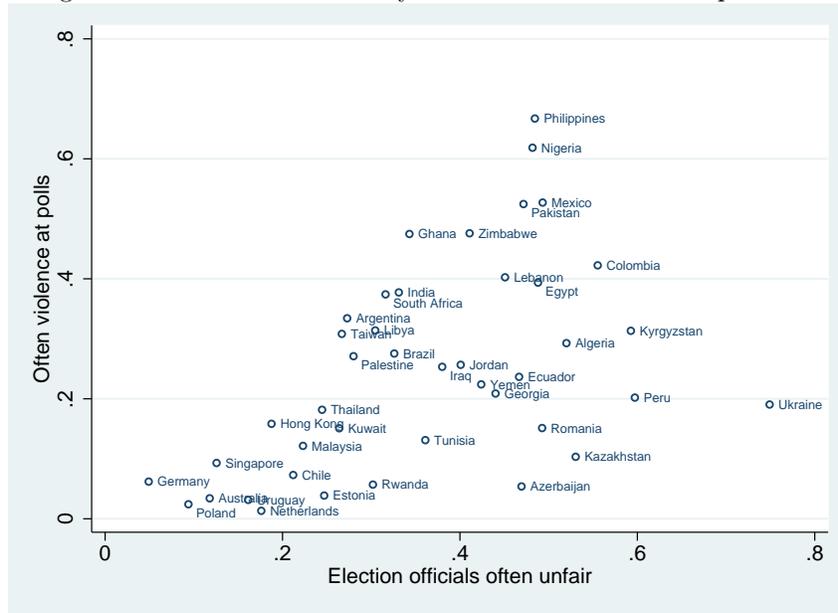# "Enfranchising Your Own? Experimental Evidence on Bureaucrat Diversity and Election Bias in India", Yusuf Neggers - Online Appendix

Appendix A: additional figures and tables

Figure A1: World Values Survey – election administration problems

Notes: Measures computed using World Values Survey Wave 6 (2010-2014). "Election officials often unfair" is the weighted percentage of respondents in each country, when asked "In your view, how often do the following things occur in this country's elections?", answering "Not at all often" or "Not often" to "Election officials are fair", against the alternatives of "Very often", "Fairly often", or "Don't know/Not answer". "Often violence at polls" is the percentage answering "Very often" or "Fairly often" to "Voters are threatened with violence at the polls."

Figure A2: Example of polling officer team during election day proceedings

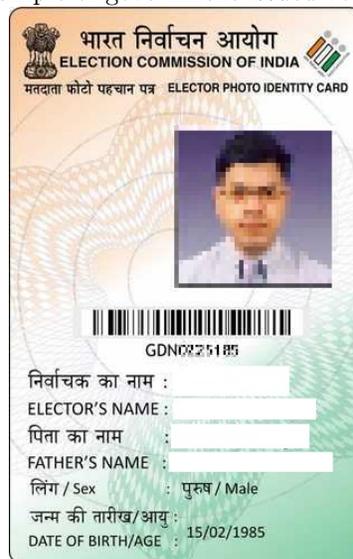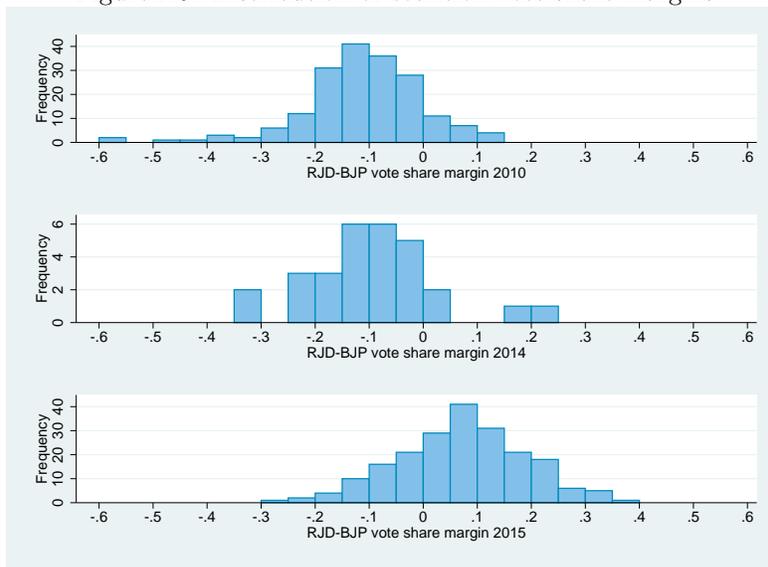Figure A3: Example of government-issued voter identity card

Figure A4: Example of neighboring polling stations in close proximity

Figure A5: Distribution of coalition vote share margins

Notes: Figure plots the distribution of the vote share margin between the RJD and BJP coalitions in Bihar, for the 185 of 243 races where these two coalitions fielded the top two candidates in the 2010 state assembly elections, the 29 of 40 races in the 2014 national parliamentary elections, and the 206 of 243 races in the 2015 state assembly elections.

Figure A6: Election day experiences – orderly environment
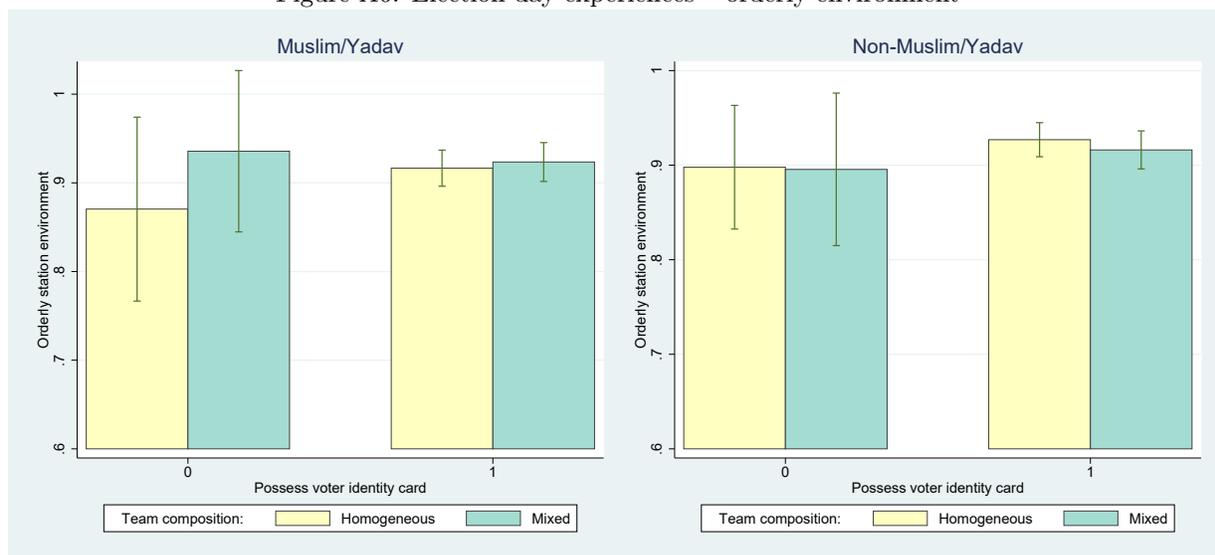
Notes: The left and right panels depict the estimated likelihood of the listed election day outcome, separately by type of potential voter (Muslim/Yadav, voter identity card possession) and polling station officer team composition. Estimates are based on corresponding regressions in Table 8, assuming mean values of all control variables. Error bars signify 95 percent confidence intervals.

Table A1: Cross-position balance

	Presiding officer (1)	Polling officer 1 (2)	Polling officer 2 (3)	Polling officer 3 (4)	Polling officer 4 (5)
Muslim/Yadav presiding officer		-0.007 (0.014)	0.006 (0.014)	-0.007 (0.014)	-0.018 (0.029)
Muslim/Yadav polling officer 1	-0.007 (0.013)		-0.001 (0.013)	-0.019 (0.013)	-0.015 (0.027)
Muslim/Yadav polling officer 2	0.006 (0.014)	-0.001 (0.014)		0.013 (0.015)	-0.009 (0.026)
Muslim/Yadav polling officer 3	-0.006 (0.013)	-0.019 (0.012)	0.012 (0.014)		-0.020 (0.029)
Muslim/Yadav polling officer 4	-0.016 (0.026)	-0.017 (0.030)	-0.011 (0.030)	-0.017 (0.025)	
Observations	5,561	5,561	5,561	5,523	1,178

Notes: Each column reports coefficients from an OLS regression of the listed outcome (Muslim/Yadav assignment to the specified position) on dummies for Muslim/Yadav assignment to the other polling officer team positions specified in table. Additionally included are sub-constituency-by-team-size fixed effects. Robust standard errors in parentheses.

Table A2: Position- and number-specific impacts on voting outcomes

	Vote share margin RJD-BJP	Ln votes RJD	Ln votes BJP	Ln total votes
	(1)	(2)	(3)	(4)
Panel A. Presiding vs polling				
Muslim/Yadav presiding officer	0.009	-0.005	-0.017	-0.010
	(0.020)	(0.052)	(0.043)	(0.018)
Muslim/Yadav polling officer	0.034	0.075	-0.057	0.007
	(0.012)	(0.031)	(0.024)	(0.009)
Observations	5,293	5,276	5,290	5,293
F-test p-value: equality of coeffs.	0.257	0.159	0.386	0.369
Panel B. Position				
Muslim/Yadav presiding officer	0.009	-0.005	-0.017	-0.010
	(0.020)	(0.052)	(0.043)	(0.018)
Muslim/Yadav polling officer 1	0.034	0.094	-0.019	0.017
	(0.019)	(0.050)	(0.037)	(0.011)
Muslim/Yadav polling officer 2	0.022	0.055	-0.065	0.001
	(0.020)	(0.052)	(0.044)	(0.014)
Muslim/Yadav polling officer 3	0.038	0.055	-0.086	0.001
	(0.019)	(0.050)	(0.040)	(0.019)
Muslim/Yadav polling officer 4	0.068	0.173	-0.069	0.013
	(0.041)	(0.090)	(0.081)	(0.019)
Observations	5,293	5,276	5,290	5,293
F-test p-value: equality of coeffs.	0.657	0.439	0.649	0.632
Panel C. Number				
Any Muslim/Yadav officer	0.028	0.058	-0.047	0.003
	(0.011)	(0.028)	(0.022)	(0.008)
Multiple Muslim/Yadav officers	-0.022	-0.059	0.037	-0.009
	(0.025)	(0.062)	(0.054)	(0.018)
Observations	5,552	5,535	5,549	5,552

Notes: All columns in Panel A and B report OLS estimates from polling-station-level regressions of the listed variable on indicators for Muslim/Yadav presence in the indicated polling party position(s), conditional on there being 1 or fewer total MY officers at the station. All columns in Panel C report OLS estimates from regressions at the polling station level of the listed variable on indicators for the degree of Muslim/Yadav presence. Additionally included in all regressions are the same controls as in Table 2. Robust standard errors in parentheses.

Table A3: Cross-station externalities – extended range

	Vote share margin RJD-BJP (3)	Ln votes RJD (1)	Ln votes BJP (2)	Ln total votes (4)
Panel A. Buffer radius				
Number mixed team neighbor	0.028	0.039	-0.047	-0.002
stations	(0.010)	(0.025)	(0.022)	(0.008)
Number mixed team stations	-0.005	0.015	0.021	0.007
within 0.25km	(0.008)	(0.023)	(0.015)	(0.007)
Number mixed team stations	-0.001	-0.005	-0.002	-0.002
within 0.25-0.75km	(0.004)	(0.008)	(0.007)	(0.002)
Mixed team composition	0.031	0.059	-0.056	-0.001
	(0.011)	(0.028)	(0.022)	(0.008)
Observations	5,190	5,174	5,187	5,190
Number locations	3,378	3,368	3,376	3,378
Panel B. Village boundaries				
Number mixed team neighbor	0.053	0.107	-0.092	0.011
stations	(0.019)	(0.053)	(0.044)	(0.013)
Number mixed team stations	0.016	-0.003	-0.051	-0.001
within village	(0.015)	(0.039)	(0.029)	(0.007)
Number mixed team stations	0.008	0.009	-0.014	0.004
in neighboring villages	(0.006)	(0.014)	(0.011)	(0.003)
Mixed team composition	0.039	0.089	-0.070	0.011
	(0.015)	(0.039)	(0.032)	(0.011)
Observations	3,209	3,193	3,207	3,209
Number villages	1,246	1,243	1,246	1,246

Notes: Each column within a panel reports OLS estimates from a regression at the polling station level of the listed variable on an indicator for mixed team composition and variables for the numbers of polling stations within the various distances indicated that are mixed team composition and of each team size between three and five (not shown). Each regression includes sub-constituency-by-team-size fixed effects and controls for log total registered voters and share Muslim/Yadav registered voters. Neighbor stations are those within the same building/compound of a given polling station. Stations within 0.25 and 0.25-0.75km are non-neighbor stations within the stated distance of a given polling station. Numbers of stations within a village and in neighboring villages are the numbers of non-neighbor polling stations within the same village as a given station and in villages adjacent to a given station's village. Panel A is restricted to stations matched to the dataset of station GPS locations. Panel B further excludes stations in the top 1 percent of villages in terms of number of stations contained within, or their neighboring villages. Standard errors clustered at the location and village level in parentheses for Panel A and B, respectively.

6

Table A4: Heterogeneity by voter identity card coverage – outlier robustness checks

| | Top and bottom percentile | | | | |
| | 1 | 0.5 | 0.75 | 1.5 | None |
	(1)	(2)	(3)	(4)	(5)
Panel A. Trim					
Mixed team * Voter ID coverage percent	-0.007	-0.006	-0.007	-0.005	-0.006
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Mixed team	0.556	0.488	0.568	0.459	0.486
	(0.315)	(0.321)	(0.318)	(0.312)	(0.322)
Observations	5,442	5,498	5,470	5,386	5,552
Interacted sub-constituency controls	X	X	X	X	X
Panel B. Winsorize					
Mixed team * Voter ID coverage percent	-0.006	-0.006	-0.006	-0.006	
	(0.003)	(0.003)	(0.003)	(0.003)	
Mixed team	0.490	0.487	0.488	0.490	
	(0.321)	(0.322)	(0.322)	(0.320)	
Observations	5,552	5,552	5,552	5,552	
Interacted sub-constituency controls	X	X	X	X	

Notes: All columns report OLS estimates from regressions at the polling station level of the RJD-BJP coalition vote share margin on an indicator for mixed team composition interacted with the sub-constituency-level percentage of registered voters with a voter ID card. Also included are the same controls as in the even-numbered columns of Table 4. Panel A the trims and Panel B winsorizes the top and bottom percentiles of observations in terms of coalition vote share margin as specified in each column heading. Robust standard errors in parentheses.

Table A5: Cross-station externalities and heterogeneity by voter identity card coverage

| | Vote share margin RJD-BJP | |
	(1)	(2)
Mixed team * Voter ID coverage percent	-0.004	-0.006
	(0.002)	(0.003)
Number mixed team neighbor stations * Voter ID coverage percent	-0.003	-0.005
	(0.002)	(0.003)
Mixed team	0.356	0.506
	(0.174)	(0.312)
Number mixed team neighbor stations	0.266	0.488
	(0.178)	(0.259)
Observations	5,442	5,442
Interacted sub-constituency controls		X

Notes: All columns report OLS estimates from regressions at the polling station level of the listed variable on an indicator for mixed team composition and variables for the numbers of mixed team and total (not shown) neighboring polling stations, and their interactions with the sub-constituency-level percentage of registered voters with a voter ID card. Also included in the first and second columns are the same controls as in the odd- and even-numbered columns of Table 4, respectively. Sample trims top and bottom one percent of observations in terms of coalition vote share margin. Standard errors clustered at the location level in parentheses.

Table A6: Vignette experiment – heterogeneity by officer type

	Ability to cast vote	
	4-point scale	0-1 indicator
	(1)	(2)
Muslim officer/potential voter type match	0.088	0.022
	(0.175)	(0.086)
Yadav officer/potential voter type match	0.271	0.160
	(0.197)	(0.099)
Brahmin officer/potential voter type match	1.099	0.339
	(0.541)	(0.213)
Observations	818	818
F-test p-value: equality of coeffs.	0.200	0.309

Notes: Column (1) reports OLS estimates from an officer-level regression of a variable taking the 1-4 scale value ("Very unlikely (1)", "Unlikely (2)", "Likely (3)", "Very likely (4)") of the assessed likelihood of the hypothetical individual being allowed to vote on indicator variables for whether the officer's own type matches that of the randomly assigned voter name (Muslim, Yadav, Brahmin), separately by each officer type for which a voter name match was possible. Column (2) reports OLS estimates from a regression with the outcome as an indicator variable taking value 1 if the respondent answers "Very likely" or "Likely" as opposed to "Unlikely" or "Very unlikely" and 0 otherwise. Controls are the same as in Table 6. Standard errors clustered at the polling station level in parentheses.

Table A7: Impacts on attempting to vote

	Went to station on election day	
	Muslim/ Yadav	non- Muslim/ Yadav
	(1)	(2)
Mixed team	0.017	-0.007
	(0.011)	(0.011)
Observations	1,946	2,620

Notes: Each column reports OLS estimates from an individual-level regression of the listed variable on the mixed team polling station indicator, for the sample of individuals indicated in each column. Additionally included are the same set of controls as in Table 8, with the exception of the individual- and household-level characteristics, which are not available for individuals who indicated they did not go to the polling station on election day. Standard errors clustered at the polling station level in parentheses.

Table A8: Type-specific impacts of officer identity on voting outcomes

	Vote share margin RJD-BJP (1)	Ln votes RJD (2)	Ln votes BJP (3)	Ln total votes (4)
Any Muslim officer	0.024	0.055	-0.035	0.012
	(0.012)	(0.030)	(0.024)	(0.008)
Any Yadav officer	0.045	0.065	-0.104	-0.034
	(0.022)	(0.057)	(0.044)	(0.025)
Observations	5,293	5,276	5,290	5,293
F-test p-value: equality of coeffs.	0.378	0.872	0.146	0.065

Notes: All columns report OLS estimates from polling-station-level regressions of the listed variable on indicators for Muslim and Yadav presence, conditional on there being 1 or fewer total Muslim/Yadav officers at the polling station. Additionally included are the same controls as in Table 2. Robust standard errors in parentheses.

Appendix B: data collection and experimental details

Vignette experiment

Experiment prompt

"Please consider the following situation: A voter named [RANDOMLY ASSIGNED] arrives at the polling station without an EPIC card but has a government voter's slip without a photograph. He can recite his name and other particulars. On a scale of 1 to 4, how likely do you think it is that he would be allowed to cast a vote based on this information?", where the potential responses are "Very unlikely (1)", "Unlikely (2)", "Likely (3)", "Very likely (4)".

Randomly assigned names

- Muslim: Najam Uddin, Mustak Ansari, Mohammed Alam
- Yadav: Ajay Yadav, Kailesh Yadav, Surendra Yadav
- Brahmin: Arjun Tripathi, Rohit Mishra, Alok Chaturvedi

List randomizations

Introductory prompt

"I'm going to read you a list of various statements, and I would like for you to tell me how many of them occurred during the previous 2014 Lok Sabha election. Please, count to yourself. Do not tell me which ones, only HOW MANY IN TOTAL. For example, it might be that none of them occurred, all of them occurred, or any number in between."

Non-sensitive statements

Experiment 1
- Your polling station was located in a government school building.
- There were other polling stations at the same building/location as your polling station.
- You saw campaign posters in your neighborhood before the election.
- One or more of the election officers at your polling station was female.

Experiment 2
- You heard political party advertisements on television or radio before the election.
- You stood in line while waiting to vote at your polling station.

- A police/security officer was at your polling station while you were there.
- One or more of the election officers at your polling station was from your village.

Survey sampling

Potential voters survey

Polling stations in urban areas, where locating specific individuals based on the information available in the electoral roll would not have been feasible, were excluded from the sample (8.3 percent of total). Also excluded were polling stations with only three election officers (0.7 percent of total), as were polling stations that were split across a main polling station and an extension station (9.8 percent of total). The publicly available lists of registered voters were at the (main+extension) level, so it was not possible to determine to which of the main station or extension individuals were assigned. The only difference between having a main and extension station versus two polling stations in the same location is whether the threshold for maximum registered voters at a single station was reached after the formal yearly deadline to split polling stations. Administration is otherwise identical.

In some locations, fewer than three Muslims or two Yadavs were identified in the list. If too few Muslims were available, Yadavs were randomly drawn to fill the positions when possible, and vice versa. If fewer than five Muslims and Yadavs in total were identified, individuals that were neither Muslim nor Yadav were randomly drawn to fill the position.

Seasonal migration is common in the survey area and the electoral rolls contain errors (e.g. listed individuals may be duplicates or have moved and registered at another polling station without being deleted from the list at the previous station). Therefore, randomly drawn backup respondents were also identified for each primary respondent. In addition, if a located individual indicated that they did not go to the polling station to attempt to vote on election day, the next backup individual was then substituted. In the final sample, 36.6 percent of respondents were from the primary sample, 22.6 percent were the first backup, 14.6 percent were the second backup, 11.2 percent were the third backup, and 15 percent were fourth backup or higher. These rates of replacement are similar to those of other surveys in the region which identified respondents based on the electoral roll (Banerjee et al. 2014). Neither the rate of consent (>99 percent) among reachable individuals nor the overall proportion of primary versus backup respondents differ significantly by whether the polling station is mixed versus homogeneous team.

Election officers survey

A total of 6,251 officers served at polling stations during the 2014 election in the district in which the survey was conducted. Out of these officers, 6,045 had phone numbers listed in the administrative data which were not obviously incorrect (i.e. having the wrong number of digits or all zero numerals).

Of these 6,045 individuals, 614 officers were inferred as Muslim or Yadav. Each of these individuals was attempted to be reached by phone. One non-Muslim/Yadav officer was randomly selected for calling from each of the mixed composition teams of which the previous 614 Muslim/Yadav officers were a member. If the officer could not be reached or did not consent, another officer of the same type was selected as a replacement, if possible. An additional 600 homogeneous polling teams were randomly chosen, stratifying by sub-constituency, and an officer from within the team was randomly selected. Again, if the officer could not be reached or did not consent, another officer was selected as a replacement, if possible. A total of 2,340 officers were called. In 30 percent of instances the individual was not reachable (in large part due to the listed phone number no longer being functional/up-to-date), and initial non-consent was low (1.4 percent). From the population of mixed team polling stations with an M-Y officer and non-MY officer each confirmed as initially consenting and homogeneous polling stations with an officer confirmed as initially consenting, 305 mixed team and homogeneous polling stations each were randomly selected as described in the main text. The rate of final consent (>98 percent) among reachable individuals in this set of officers does not differ significantly by whether the polling station is mixed versus homogeneous team.

Counterfactual calculation details

The total estimated effect on the RJD-BJP coalition vote share margin of shifting to a mixed composition polling station team is approximated by the sum of the within-station effect and the cross-station spillover effect multiplied by the number of neighbor polling stations, adjusting for the sub-constituency level of voter identity card coverage, ID_c. Sub-constituency-level administrative data available for the 2010 and 2014 election cycles across the entire state of Bihar allows me to observe the voter identity card coverage and calculate the average number of polling station neighbors in each sub-constituency, N_c. The distribution of neighbors by officer team size is not available outside of the two sample districts. The results of the spatial externality regressions in Section III.B, however, are nearly identical when the overall number of neighbors is used as a control instead of separating by team size. In addition, the neighbor and identity card coverage data were not yet available for the 2015 elections, so the 2014 values are used for that election cycle.

Taking the coefficients from a modified version of equation (3) allowing for cross-station spillover effects, estimated on the sample districts for which I possess officer assignment information,

$$Y_{pc} = \mu_{co} + \beta Mixed_{pc} + \gamma T_{pc} + N_{pc} + \beta_2 \left[Mixed_{pc} * ID_c \right] + \gamma_2 \left[T_{pc} * ID_c \right] + \phi_2 \left[N_{pc} * ID_c \right] + \epsilon_{pc},$$

the impact of a change of magnitude, X, in the proportion of mixed polling stations in a sub-constituency can be estimated as $X * [(\beta + \gamma * N_c) + (\beta_2 + \gamma_2 * N_c) * ID_c]$. As the baseline shares of mixed teams outside my sample area are not observed, I assume the proportions are the same as that in the observable sample, 0.324. The value of X needed to change the outcome of the race between the RJD and BJP coalitions can then be calculated using the previous estimate together with the constituency level margins of victory. When calculating impacts at the parliamentary constituency level, I take a weighted average (based on number of polling stations) across the sub-constituencies within each parliamentary constituency.