APPENDIX FOR "WHY SO LITTLE STRATEGIC VOTING IN INDIA?"

Contents

Appendix A. Supplemental discussion of different strategies to identify voters' party preferences

To classify voters as strategic or not, we require three types of information for each respondent:

1. Vote choice;
2. Beliefs about the likely outcome of the race in which the respondent is voting; and
3. Party preference.

To identify vote choice, we directly ask respondents for whom they voted. To identify beliefs about the likely outcome of the race, we ask respondents three separate questions in which we ask them to identify the party of the likely winner in their constituency, the party of the likely second place candidate, and the party of the likely third place candidate.

To identify party preference, we adopt several different strategies, presenting seven different ways of identifying a respondent's preferred party. All of these, however, are rooted in two broad strategies for identifying a voter's preferred party. The first relies on a survey item that asks respondents whether they feel close to a party or not. If they respond that they do, respondents were asked which they felt close to. We take that response as their party preference. The second approach uses a series of questions in which respondents were asked to rate six parties in the state (BJP, BSP, INC, SP, RLD, and Peace Party), indicating whether they like a lot, somewhat like, neither like nor dislike, dislike somewhat, or dislike a lot a particular party.[1] Respondents were also given the option to say that they had not heard of a party. We treat those who evaluated one party more positively than all others as preferring that party.

These two methods have different pros and cons. On the one hand, because the closeness measure directly asks respondents to identify a party, some voters might be more likely to let their vote choice shape their response to this question, seeking to ensure consistency between their responses or simply treating their vote choice as evidence of closeness without actually reflecting on whether they truly feel an affinity for the party. In contrast, the measure based on party ratings does not require respondents to directly state a preference. Any inconsistency in their responses would be far less obvious if, for example, they reported voting for one party but then gave a more positive rating to another party, particularly if they also gave a fairly positive rating to the party they voted for. Consciously ensuring that one rates the party they voted for more favorably than all other parties—not because that is their sincere preference but out of a desire to ensure consistency between responses—would take considerable cognitive effort on the part of respondents. In this way, this rating-based measure is less likely to be contaminated by vote choice. On the other hand, because the closeness measure allows respondents to say that they are not close to a party, this measure is arguably getting at the notion of a strong underlying preference. Indeed, the number of respondents coded as "unattached"—for whom we cannot identify a party preference—is much higher using the closeness measure than the ratings measure. The party rating measure potentially risks attributing a clear preference when one does not actually exist.

Using these two types of measures as our bases, we subsequently impose further requirements for identifying party preferences to see if our results are robust if we further restrict who counts as having a party preference. To that end, we employ measures asking about prior vote choice in the 2014 national election and the 2012 Uttar Pradesh state election, treating a respondent as having a party preference only if the preference derived from the closeness or rating measure matches prior vote choice. Though respondents may misremember prior vote choice or misstate it to match current vote choice, responses to these prior vote questions either indicate 1) an enduring preference for the party over time if responses are accurate or 2) a sufficiently strong preference at the time of the survey that the respondent was unwilling, consciously or not, to admit to having voted for another party in prior elections. Either way, we would argue that these questions should weed out those who have only a weak preference for a party. Based on these various survey items, we arrive at seven different methods of identifying a party preference:

1. Identify as close to a party;
2. Identify as close to a party that the respondent voted for in the 2014 national (Lok Sabha or LS) election;
3. Identify as close to a party that the respondent voted for in both the 2014 national and 2012 Uttar Pradesh (Vidhan Sabha or VS) elections;

[1] In the open-ended question about which party a respondent is close to, 99.7% of respondents who provided an answer cited one of these seven parties.

4. Rate a party more favorably than all others;
5. Rate a party more favorably than all others and voted for it in the 2014 LS election;
6. Rate a party more favorably than all others and voted for in both the 2014 LS and 2012 VS elections;
7. Identify as close to a party and rate it more favorably than others.

Figure A1 presents the share of strategic voters (out of all voters) using these seven different methods for identifying a voter's preferences. The definition we use in the main text ("Close") yields the greatest share of strategic voters. Measures that are different (using party ratings) or more restrictive produce even smaller shares of strategic voters, typically well under 1%.

Figure A1. Share of strategic voters using different measures of party preference

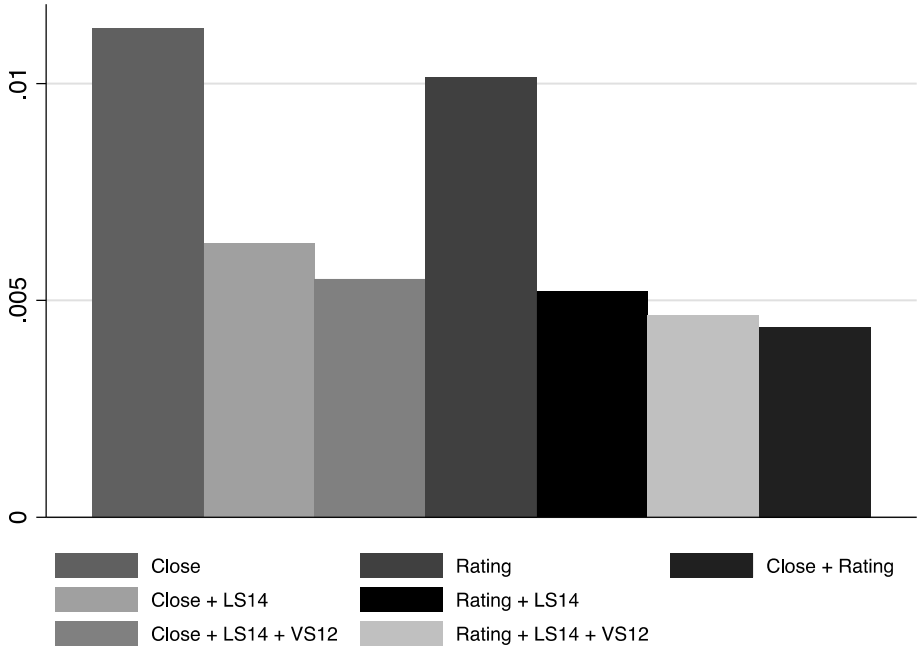


Table A1 elaborates on Figure A1. It lists in the far-left column the seven methods for identifying a respondent's party preference. The next column indicates the share of respondents with identifiable party preferences using this method. As is clear, the party rating measure (when used alone) generates the greatest share of respondents with identifiable party preferences. The middle column indicates the share of respondents who are classified as strategic voters, taking all respondents used in the analysis into account. The next column, second from the right, presents the share of respondents who are strategic voters when excluding unattached and constrained voters who cannot—whether because of their lack of party attachment or the constituency in which they are voting—possibly be strategic voters. The share of strategic voters is somewhat larger, but still quite small. Finally, the column on the right is the share of respondents who are strategic voters from among those whose beliefs suggest they are potential strategic voters. In other words, this is the number of strategic voters divided by the number of strategic and expressive voters combined. Strategic and expressive voters are the two groups who possess party preferences, can vote for their preferred party, and believe that their preferred party is not likely to be in contention to win. As this column shows, for six of seven measures, the majority of voters who see their preferred parties as unlikely to win ultimately cast strategic votes. In the final column, the number in brackets is the number of voters who believe that their preferred party is uncompetitive. As these numbers indicate, few respondents believe that their preferred party is uncompetitive.

Table A1. Share of strategic voters with different methods for identifying preferences

Method for identifying party preference	*% of respondents with a preference*	*% strategic voters (of all respondents)*	*% strategic voters (excluding unattached and constrained voters)*	*% strategic voters (of those preferring uncompetitive party) [N]*
Close to a party	76.0%	1.1%	1.6%	74.6% [55]
+ LS 2014 vote	66.8%	0.6%	1.0%	63.9% [36]
+ VS 2012 vote	59.6%	0.6%	1.0%	60.6% [33]
Preferred party	93.3%	1.0%	1.1%	61.7% [60]
+ LS 2014 vote	78.2%	0.5%	0.7%	50.0% [38]
+ VS 2012 vote	67.8%	0.5%	0.7%	47.2% [36]
Close + preferred	68.7%	0.4%	0.7%	57.1% [28]

Appendix B. Muslims and strategic voting

Muslims are commonly believed to behave more strategically in India because they strongly oppose the BJP and are therefore willing to abandon their preferred party in an effort to prevent the BJP from winning. With this in mind, Table A2 categorizes voters into the six types described in the main text, but broken down by religion. Columns 1 and 2 include all Muslim and Hindu respondents; columns 3 and 4 exclude constrained and unattached respondents, thereby limiting the pool of respondents to those who could conceivably vote strategically. Table A2 confirms that, among Muslims, strategic voters are much more common than among Hindus. Interestingly, there are no expressive voters among Muslims, meaning that *all* Muslim respondents who believe that their preferred party is unlikely to win vote strategically. In contrast, only a slight majority of Hindus do the same.

Table A2. Voter type by major religious groups in UP

	All respondents		*Excluding constrained and unattached voters*	
Voter Type	***Muslim***	***Hindu***	***Muslim***	***Hindu***
	(1)	**(2)**	**(3)**	**(4)**
Sincere	59.6%	68.5%	91.6%	92.8%
Strategic	3.2%	0.7%	4.9%	1.0%
Expressive	0.0%	0.5%	0.0%	0.6%
Insincere	2.2%	4.1%	3.5%	5.6%
Constrained	8.5%	2.7%	--	--
Unattached	26.4%	23.6%	--	--
N	624	2,984	406	2,201

Like Table A2, Table A3 similarly addresses the question of whether Muslims are more likely to behave strategically than Hindus. However, rather than look at the distribution of voter types within the two major religious groups in UP as Table A2 does, Table A3 looks at each voter type, broken down by religion. (Smaller religious groups, which account for less than 1% of respondents, are excluded). Muslims comprise about 17% of our total sample. Nevertheless, they account for nearly half of all strategic voters. They are also somewhat overrepresented among constrained voters, likely because many are Samajwadi Party supporters, and the Samajwadi Party was in an alliance with Congress. Consequently, it did not contest in a number of seats, leading to more constrained voters. Notably as well, there are no expressive Muslims voters. Within the other categories (sincere, insincere, and unattached), the share of Muslims largely mirrors their share in the overall survey. Table A3 therefore corroborates, like Table A2, the intuition that Muslims are more likely than Hindus to vote strategically.

Table A3. Share of voter types belonging to major religious groups in UP

Vote Type	*Muslim*	*Hindu*
Sincere	15.2%	83.7%
Strategic	48.8%	51.2%
Expressive	0%	100.0%
Insincere	10.2%	89.8%
Constrained	39.9%	60.2%
Unattached	18.9%	80.5%
Total	17.2%	82.0%

Table A4. Replication of analysis of *Winner* by Religious Group

	Muslims	Hindus	Muslims	Hindus
	(1)	(2)	(3)	(4)
Distance from contention 2017	-0.02	-0.01**	-0.02	-0.02***
	(0.01)	(0.01)	(0.01)	(0.01)
Close to party	3.75***	3.86***	3.76***	3.94***
	(0.36)	(0.12)	(0.37)	(0.13)
Distance from contention 2017			0.03	0.04***
X Close to party			(0.03)	(0.01)
Distance from contention 2012	-0.02	0.01***	-0.02	0.01***
	(0.01)	(0.00)	(0.01)	(0.00)
Contact	3.77***	1.31***	3.75***	1.32***
	(1.04)	(0.22)	(1.04)	(0.22)
BSP	1.45***	-0.32**	1.42***	-0.33**
	(0.44)	(0.16)	(0.44)	(0.16)
SP	1.74***	-0.37**	1.70***	-0.37**
	(0.51)	(0.15)	(0.50)	(0.16)
N	1,319	7,139	1,319	7,139
Pseudo-R^2	0.72	0.66	0.72	0.66

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Table A4 replicates the analysis of the *Winner* dependent variable that is the basis for Figures 2 and 3 in the main text. However, Table A4 analyzes Hindus and Muslims separately. All models are conditional logits. Interestingly, the results for both religious groups are substantively similar. The coefficients on *Distance from contention 2017*, *Close to a party*, and the interaction between the two are in the same direction for both subsamples and of roughly similar magnitude. Several coefficients in the Muslim subsample do not reach conventional levels of statistical significance. But, this subsample is considerably smaller than the Hindu subsample. Some of the results for control variables differ between the two subsamples. Notably, the coefficients on the parties differ because few Muslims are BJP voters, whereas many Hindus are. We suspect that the *Distance from contention 2012* variable differs between these subsamples because the BJP did not do particularly well in 2012 and was, in many seats, far from contention. But, since many Hindus voted for the BJP in 2017, *Distance from contention 2012* may simply capture the fact that the BJP was often the party further from contention in 2012. The correlation between *Distance from contention 2012* and *Winner* among Hindus may therefore reflect the fact that many Hindus (who voted for the BJP) expected it to be the winner and that the party was, in 2012, frequently far from contention. Interestingly, the contact variable has a much larger substantive effect on Muslims than Hindus.

As a general take-away, although Muslims are more likely to vote strategically than Hindus, we find little difference between the two groups in what predicts beliefs about who will win in a respondent's constituency. The objective indicator of performance only weakly predicts beliefs, primarily for parties that a respondent is not close to. Partisan leanings strongly predict believing that a party is going to win. The seeming disconnect between the results in Table A4 and the patterns in Tables A2 and A3 may reflect the tiny number of strategic voters, even among Muslims. Their small number may not be sufficient to produce different patterns with respect to the predictors of beliefs about election competitiveness.

Appendix C. Additional figures and tables associated with Figures 2 and 3

Table A5. Analysis for Figure 2: Predictors of election forecasts

	Winner	Winner	Likelihood of victory	Likelihood of victory	Likelihood of victory	Likelihood of victory
	(1)	(2)	(3)	(4)	(5)	(6)
Distance from contention 2017	-0.01**	-0.01**	-0.01***	-0.01***	-0.02***	-0.02***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Close to party	3.94***	3.88***	2.21***	2.18***	4.21***	4.16***
	(0.12)	(0.12)	(0.02)	(0.02)	(0.13)	(0.13)
Distance from contention 2012		0.00		-0.00		0.00
		(0.00)		(0.00)		(0.00)
Contact		1.57***		0.27***		0.79***
		(0.21)		(0.05)		(0.13)
BSP	-0.29**	-0.25*	0.02	-0.02	0.04	0.02
	(0.12)	(0.14)	(0.04)	(0.04)	(0.09)	(0.11)
SP	-0.26**	-0.18	-0.17***	-0.21***	-0.51***	-0.52***
	(0.10)	(0.14)	(0.03)	(0.04)	(0.08)	(0.11)
Constant			1.93***	1.75***		
			(0.02)	(0.05)		
N	8,538	8,538	9,284	9,284	8,941	8,941
Pseudo-R^2 / R^2	0.64	0.65	0.59	0.60	0.59	0.60

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Table A5 provides the results that are the basis for Figure 2, which graphically presents the results of models 2 and 4 in this table. Models 1, 3, and 5 include just the main predictors (*Distance from contention 2017* and *Close to party*) along with party dummies, whereas Models 2, 4, and 6 add in controls for previous distance from contention and party contact. Models 1 and 2 are estimated using conditional logits, with *Winner*, a binary variable, as the dependent variable. Models 3 and 4 are estimated using OLS with respondent fixed effects, with *Likelihood of victory*, which is a five-point scale, as the dependent variable. Given the ordered nature of the dependent variable in models 3 and 4, OLS is not necessarily the optimal estimation strategy, though it has the advantage of being easily interpretable. Models 5 and 6 therefore replicate models 3 and 4 using ordered logits with respondent fixed effects. The results are substantively the same.

Table A6. Analysis for Figure 3: Interacting party closeness and distance from contention

	Winner	Winner	Likelihood of victory	Likelihood of victory	Likelihood of victory	Likelihood of victory
	(1)	(2)	(3)	(4)	(5)	(6)
Distance from contention 2017	-0.01***	-0.01***	-0.01***	-0.01***	-0.02***	-0.02***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Close to party	4.01***	3.95***	2.23***	2.20***	4.23***	4.17***
	(0.12)	(0.12)	(0.03)	(0.03)	(0.13)	(0.13)
Distance from contention 2017	0.04***	0.04***	0.01***	0.01***	0.02*	0.02*
X Close to party	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
Distance from contention 2012		0.00		-0.00		-0.00
		(0.00)		(0.00)		(0.00)
Contact		1.58***		0.28***		0.80***
		(0.21)		(0.05)		(0.13)
BSP	-0.29**	-0.26*	0.01	-0.02	0.05	0.03
	(0.12)	(0.14)	(0.04)	(0.04)	(0.09)	(0.11)
SP	-0.25**	-0.18	-0.17***	-0.21***	-0.50***	-0.52***
	(0.10)	(0.14)	(0.03)	(0.04)	(0.08)	(0.11)
Constant			1.93***	1.74***		
			(0.02)	(0.05)		
N	8,538	8,538	9,284	9,284	8,941	8,941
Pseudo-R^2 / R^2	0.64	0.65	0.60	0.60	0.59	0.60

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Table A6 provides the results that are the basis for Figure 3, which graphically presents the results of model 2 in this table. Models 1, 3, and 5 include just the main predictors (*Distance from contention 2017, Close to party*, and their interaction) along with party dummies, whereas Models 2, 4, and 6 add in controls for previous distance from contention and party contact. Models 1 and 2 are estimated using conditional logits, with *Winner*, a binary variable, as the dependent variable. Models 3 and 4 are estimated using OLS with respondent fixed effects, with *Likelihood of victory*, which is a five-point scale, as the dependent variable. Given the ordered nature of the dependent variable in models 3 and 4, OLS is not necessarily the optimal estimation strategy, though it has the advantage of being easily interpretable. Models 5 and 6 therefore replicate models 3 and 4 using ordered logits with respondent fixed effects. The results are substantively the same.

Figure A2. Replication of Figure 3 with *Likely to win* as the dependent variable

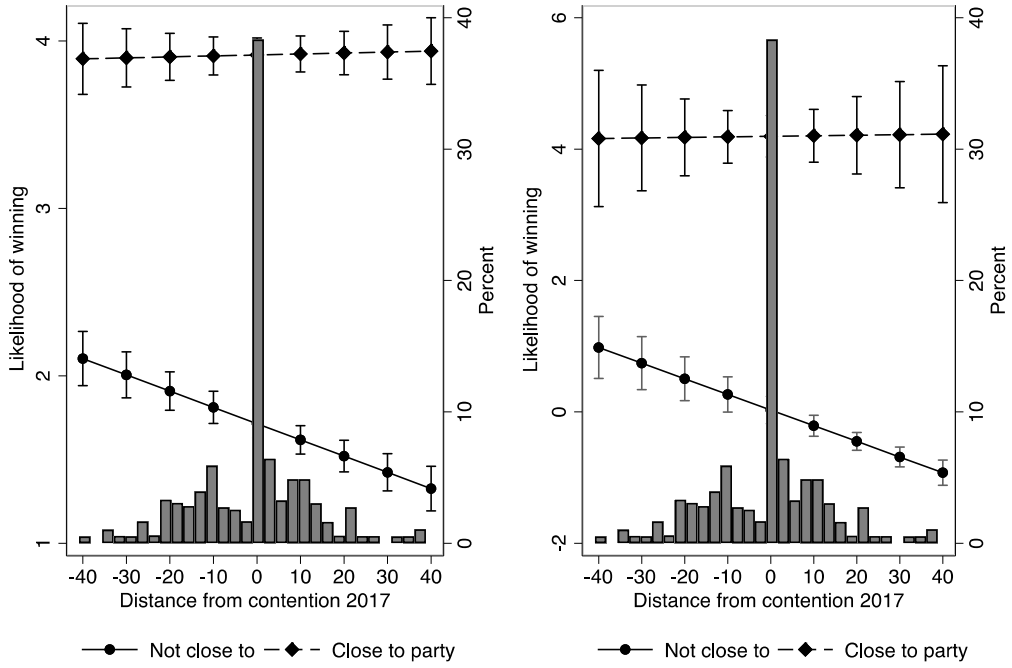


Figure A2 graphically presents the results of models 4 and 6 in Table A6 above. In other words, the two panels in Figure A2 are comparable to Figure 3 in the main text, but with *Likely to win* as the dependent variable, instead of *Winner*. The left panel in Figure A2 presents the results from model 4, estimated using OLS with respondent fixed effects; the right panel presents the results from model 6, estimated using an ordered logit with respondent fixed effects. Just as with Figure 3, the histogram plots the distribution of observations for *Distance from contention 2017.* The dashed line with triangles represents the predicted response to the question about a party's likelihood of winning for respondents who are close to the party in question. The solid line with circles represents the predicted responses for those who are not close to the party in question. A "4" on the scale represents "extremely likely to win," while 0 represents "not at all likely to win." Just as in Figure 3, the line for those close to a party is essentially flat, indicating near universal belief that the party is extremely likely to win. Similarly too, the line plotting predicted likelihood of winning for those not close to a party is negative, indicating that respondents' beliefs about a party's likelihood of winning decrease as the party gets farther from contention.

Appendix D. Robustness checks

Table A7. Voter types by respondent uncertainty about likely winner and runner-up

	(1) All respondents	(2) Has beliefs about likely winner and runner-up	(3) "Don't know" response for winner and/or runner-up.
Sincere	2,441 (67.1%)	2,234 (67.7%)	207 (61.4%)
Strategic	41 (1.1%)	41 (1.2%)	--
Expressive	14 (0.4%)	14 (0.4%)	--
Insincere	137 (3.8%)	122 (3.7%)	15 (4.5%)
Constrained	133 (3.7%)	96 (2.9%)	37 (11.0%)
Unattached	873 (24.0%)	795 (24.1%)	78 (23.2%)
Total	3,647	3,302	335

Table A7 addresses whether the breakdown of voter types depends on how we count respondents who answer "don't know" to the questions about the likely winner in their constituency and/or the likely runner-up. As noted in footnote 4 in the main text, by definition voters can only be classified as strategic or expressive if they express beliefs about who the winner and runner-up will be. Voters are only strategic if they abandon their preferred party and believe that it is not in contention, which requires beliefs about which parties are in contention. Voters are only expressive if they stick with their preferred party fully believing it to be likely to lose, which also requires beliefs about which parties are in contention. Nevertheless, we examine how our breakdown of respondents changes when we separate out those who offer at least one "don't know" response. One obvious difference between those who articulate beliefs about the likely winner and runner-up—and who constitute the overwhelming majority of our respondents—and those who do not is that there are no strategic or expressive voters among those who offer "don't know" responses. Given that those types of respondents account for a small share of all respondents, the distribution of voter types among those who offer "don't know" responses and those who do not are relatively muted. We see a somewhat lower share of sincere voters among those who offer "don't know" responses and a slightly higher share of constrained voters. The elevated share of constrained voters may be somewhat idiosyncratic, given the small number of respondents. Overall, however, those who express "don't know" responses are not all that different from those who offer responses to both questions about who the likely winner and runner-up in their assembly constituency are.

Table A8. Replication of analysis of *Winner* with party ratings in place of *Close to Party*

	1	2	3	4
Distance from contention 2017	0.01**	-0.00	0.01	-0.01
	(0.01)	(0.02)	(0.01)	(0.02)
Party Rating	2.22***	2.23***		
	(0.08)	(0.08)		
X Distance from contention		0.01		
		(0.01)		
Like party a lot			4.81***	4.98***
			(0.37)	(0.39)
X Distance from contention				0.04*
				(0.02)
Somewhat like party			1.94***	2.04***
			(0.38)	(0.40)
X Distance from contention				0.01
				(0.02)
Neither like nor dislike party			0.22	0.31
			(0.40)	(0.42)
X Distance from contention				0.01
				(0.03)
Somewhat dislike party			-0.45	-0.35
			(0.47)	(0.49)
X Distance from contention				0.06*
				(0.03)
Distance from contention 2012	-0.01	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Contact	1.80***	1.78***	1.59***	1.50***
	(0.29)	(0.29)	(0.30)	(0.30)
BSP	-0.65***	-0.64***	-0.53**	-0.53**
	(0.22)	(0.23)	(0.25)	(0.25)
SP	-0.18	-0.19	-0.07	-0.08
	(0.21)	(0.21)	(0.23)	(0.24)
N	7,618	7,618	7,618	7,618
Pseudo-R^2	0.81	0.81	0.83	0.83

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

To test the robustness of the findings in Tables A5 and A6, which are the basis for Figures 2 and 3 in the main text, we use an alternative to *Close to party* in Table A8. Instead of examining whether feeling close to a party predicts the dependent variable *Winner,* we use respondents' answers to questions in which they are asked to rate parties on a five-point scale ranging from "dislike a lot" to "like a lot." All models in Table A8 are estimated using conditional logits. In models 1 and 2, we include *Party rating* as a continuous independent variable that ranges from 0 to 4. In models 3 and 4 we enter dummy variables for each response to this question, with "dislike a lot" (0) as the reference category. Models 1 and 3 are identical to model 2 in Table A5 except for the inclusion of *Party rating* in place of *Close to party*, while models 2 and 4 are identical to Model 2 in Table A6 except for the inclusion of *Party rating* in place of *Close to party.* The results in Table A8 are broadly similar to those in Tables A5 and A6, though *Distance from contention* does not always reach conventional levels of statistical significance. The magnitude and direction of the coefficients are, however, quite similar. Figure A3 graphically presents the results from models 2 and 4 of Table A8.

Figure A3. Graphical representation of results from Table A8

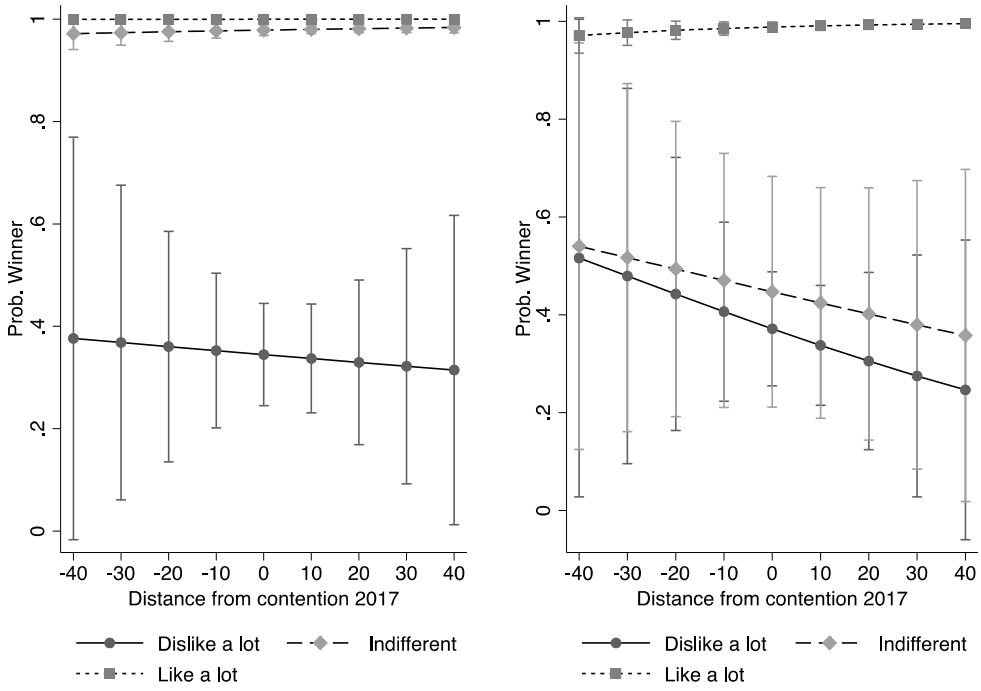


Figure A3 graphically presents the results from models 2 and 4 of Table A8. Both panels present the predicted probability of citing a party as the likely winner as *Distance from contention 2017* varies. The left panel presents the results from model 2, with *Party rating* entered as a continuous variable, while the right panel presents the results from model 4, with *Party rating* entered as a series of dummy variables. Rather than present predicted probabilities for all five response options, both panels present the predicted probabilities for "Like a lot" (4), Neither like nor dislike (2) (which appears in the panels as "Indifferent"), and "Dislike a lot" (0). Consistent with Figure 3 in the main text, for those who like a party a lot the probability of citing it as the likely winner in the constituency is nearly 100% and does not vary based on distance from contention. The two panels indicate that the results for other responses differ somewhat depending on the modeling strategy. The analysis in Table A8 certainly suggests that the effects of moving from one response category to another are not uniform across the 5-point scale, which means that treating *Party rating* as a continuous measure may not be justifiable. Entering each response category as a dummy variable (right panel) produces results very similar to Figure 3 in the main text and Figure A2. For respondents who really like a party, that party's actual competitiveness does not predict a respondent's likelihood of citing that party as the likely winner. However, for respondents who are indifferent to or dislike the party, their probability of citing it as the likely winner decreases as the party gets further from contention in the constituency.

Table A9. Replication of analysis of *Likelihood of Victory* with party ratings in place of *Close to Party*

	1	2	3	4
Distance from contention 2017	-0.00***	-0.00**	-0.00***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.00)
Party Rating	0.84***	0.84***		
	(0.01)	(0.01)		
X Distance from contention		-0.00		
		(0.00)		
Like party a lot			2.97***	2.97***
			(0.04)	(0.04)
X Distance from contention				0.00
				(0.00)
Somewhat like party			1.80***	1.80***
			(0.05)	(0.05)
X Distance from contention				0.00
				(0.00)
Neither like nor dislike party			1.00***	1.01***
			(0.04)	(0.04)
X Distance from contention				0.01**
				(0.00)
Somewhat dislike party			0.43***	0.44***
			(0.05)	(0.05)
X Distance from contention				0.00
				(0.00)
Distance from contention 2012	-0.00	-0.00	-0.00**	-0.00**
	(0.00)	(0.00)	(0.00)	(0.00)
Contact	0.24***	0.24***	0.18***	0.18***
	(0.04)	(0.04)	(0.04)	(0.04)
BSP	0.04	0.04	0.04	0.04
	(0.04)	(0.04)	(0.03)	(0.03)
SP	-0.08**	-0.08**	-0.10***	-0.10***
	(0.04)	(0.04)	(0.03)	(0.03)
Constant	0.21***	0.22***	0.76***	0.75***
	(0.04)	(0.04)	(0.05)	(0.05)
N	8,533	8,533	8,533	8,533
R^2	0.76	0.76	0.77	0.77

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

To test the robustness of the findings in Tables A5 and A6, which are the basis for Figures 2 and 3 in the main text, we use an alternative to *Close to party* in Table A9. Instead of examining whether feeling close to a party predicts the dependent variable *Likelihood of victory*, we use respondents' answers to questions in which they are asked to rate parties on a five-point scale ranging from "dislike a lot" to "like a lot." All models in Table A9 are estimated using OLS with respondent fixed effects. In models 1 and 2, we include *Party rating* as a continuous independent variable that ranges from 0 to 4. In models 3 and 4 we enter dummy variables for each response to this question, with "dislike a lot" (0) as the reference category. Models 1 and 3 are identical to model 4 in Table A5 except for the inclusion of *Party rating* in place of *Close to party*, while models 2 and 4 are identical to Model 4 in Table A6 except for the inclusion of *Party rating* in place of *Close to party*. The results in Table A9 are similar to those in Tables A5 and A6 in that, first, evaluations of a party *strongly* predict a respondent's beliefs about how likely a party is to win and, second, a party's actual distance from contention weakly predicts a respondent's beliefs about a party's likelihood of victory. The key difference between Table A9 and the results in Tables A5 and A6 is that the interactions between *Party rating* and *Distance from contention* are generally not statistically significant.

Figure A4. Graphical representation of results from Table A9

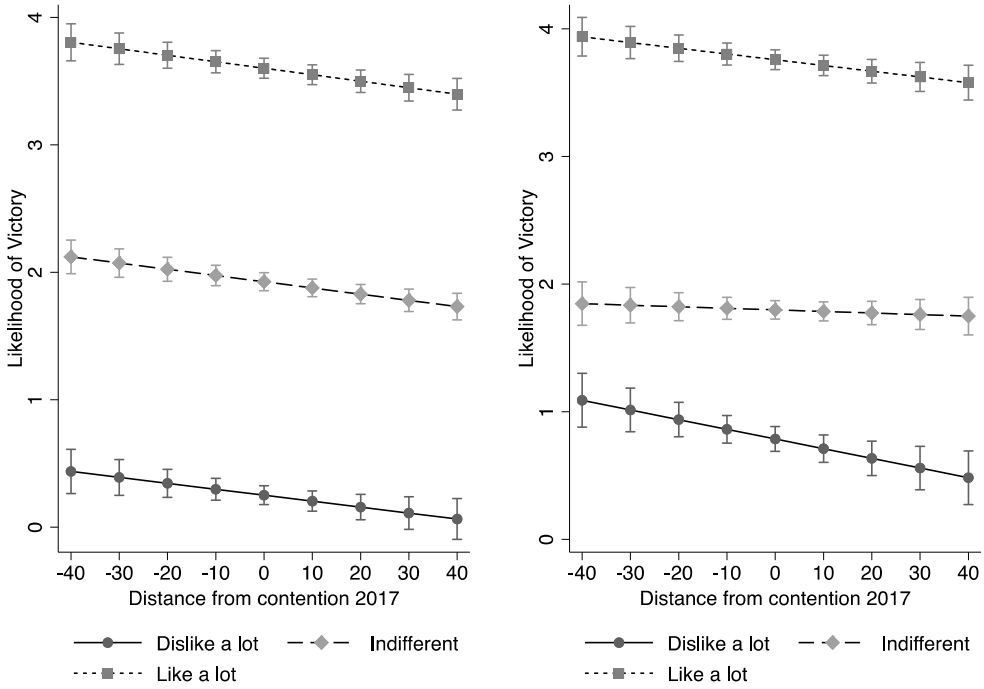


Figure A4 graphically presents the results from models 2 and 4 of Table A9. Both panels present the predicted probability of citing a party as the likely winner as *Distance from contention 2017* varies. The left panel presents the results from model 2, with *Party rating* entered as a continuous variable, while the right panel presents the results from model 4, with *Party rating* entered as a series of dummy variables. Rather than present predicted probabilities for all five response options, both panels present the predicted likelihood of victory for "Like a lot" (4), Neither like nor dislike (2) (which appears in the panels as "Indifferent"), and "Dislike a lot" (0). Consistent with Figure A2, those who like a party a lot generally predict that the party is extremely likely to win the constituency. As suggested by Table A9, Figure A4 demonstrates that how a respondent rates a party very strongly predicts her beliefs about its likelihood of victory. Further, Table A9 suggests a weak association between *Distance from contention 2017* and respondents' beliefs about a party's likelihood of victory. The main difference between Figure A2 and Figure A4 is that Figure A4 does not provide strong evidence that the relationship between *Distance from contention 2017* and *Likelihood of victory* varies much by how a respondent feels toward a party. In the left panel, the modest downward slopes of the various lines are similar for those who like, dislike and are indifferent toward a party. In the right panel, we see some variation. Most notably—and somewhat consistent with our other findings—we see the steepest slope, indicating the strongest association between a party's objective performance and respondents' beliefs about the party's competitiveness, among those who dislike a party a lot. That being said, for those who like the party a lot, there is still a modest negative slope indicating some, weak, association between beliefs about a party's competitiveness and the party's objective performance. This aspect of the findings in Figure A4 contrasts with the findings presented in the main text and in other portions of the appendix. While we would certainly not dismiss the findings presented in Figure A4, many elements of the findings are entirely consistent with our main results, and we do not view the inconsistent results in Figure A4 (and Table A9) as so strong that they ought to cast serious doubt on the otherwise consistent results across the other specifications. Nevertheless, we acknowledge that these results suggest that among those who like a party a lot, beliefs about competitiveness may be weakly tied to a party's objective performance. It is still worth recalling how modest the substantive effect of objective performance is relative to a respondent's feelings toward a party. In that regard, the findings from Figure A4 are consistent with our broader story and present a far different account from prior research, which suggests beliefs that are firmly rooted in objective party performance.

Table A10. Replication of analysis of *Winner* with alternative measures of party performance

	1	2	3	4
Vote 2017	0.01***	0.02***		
	(0.00)	(0.00)		
Distance from contention 2017 (alt)			-0.01	-0.02
			(0.01)	(0.01)
Close to party	3.91***	5.23***	3.89***	4.07***
	(0.11)	(0.36)	(0.12)	(0.16)
X Vote		-0.04***		
		(0.01)		
X Distance from contention 2017 (alt)				0.07*
				(0.04)
Vote 2012	-0.01*	-0.01		
	(0.00)	(0.00)		
Distance from contention 2012 (alt)			-0.01	-0.02
			(0.01)	(0.01)
Contact	1.64***	1.69***	1.57***	1.56***
	(0.19)	(0.19)	(0.21)	(0.21)
BSP	-0.10	-0.12	-0.50***	-0.50***
	(0.14)	(0.14)	(0.12)	(0.11)
SP	-0.02	-0.01	-0.43***	-0.42***
	(0.14)	(0.14)	(0.11)	(0.11)
N	8,989	8,989	8,538	8,538
Pseudo-R^2	0.65	0.65	0.65	0.65

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

To test the robustness of the findings in Tables A5 and A6, which are the basis for Figures 2 and 3 in the main text, we use alternatives to *Distance from contention 2017* in Table A10. The dependent variable in all models is *Winner*, and all are conditional logit models. In models 1 and 2, we include *Vote 2017* as a measure of objective party performance in lieu of *Distance from contention 2017. Vote 2017* is a party's vote share in the 2017 election. We expect that as a party's vote share increases, respondents will be more likely to cite it as the likely winner. *Vote 2017* differs from *Distance from contention 2017* in that it does not take account of a party's relative placement compared to other parties in the constituency. In models 3 and 4, we include a modified *Distance from contention 2017* variable that aims to distinguish winners and runners-up, on the one hand, from those of out contention, on the other hand. To modify the variable, we recode all winners and runners-up as 0s. However, because the data are then skewed, we use a log transformation of the variable (which requires adding a small constant to all the 0 values). The modified variable therefore captures the logged distance from contention for parties in third place or worse. In doing so, it starkly differentiates between candidates in contention and those that are not in contention. *Vote 2012* and the modified *Distance from contention 2012* are calculated in the same way but based on the 2012 election results. Models 1 and 3 are identical to model 2 in Table A5 except for the inclusion of the alternative measures of party performance in place of the original *Distance from contention 2017* measure, while models 2 and 4 are identical to model 2 in Table A6 except for the inclusion of the alternative measures of party performance in place of the original *Distance from contention 2017* measure. The results in models 1 and 2, with *Vote 2017* as the measure of party performance, are substantively identical to those in Tables A5 and A6. (Note that the coefficients are in the opposite direction because we expect that as vote share increase, a respondent is *more* likely to cite a party as the winner, whereas when a party gets farther from contention—positive values of the variable—we expect a respondent to be *less* likely to cite that party as the winner). The results in models 3 and 4, with the modified *Distance from contention 2017* variable as the measure of party performance, are largely consistent with those in Tables A9 and A10 except that several variables are less precisely estimated. Specifically, the coefficient on *Distance from contention 2017 (alt)* is not statistically significant (p-value = 0.18) and its interaction with *Close to party* is statistically significant only at the 10% level. Thus, Table A10 suggests that the results in the main text are not an artifact of how we measure a party's performance. The results broadly hold with alternative measures of a party's objective performance.

Table A11. Replication of analysis of *Likelihood of Victory* with alternative measures of party performance

	1	2	3	4
Vote 2017	0.01***	0.01***		
	(0.00)	(0.00)		
Distance from contention 2017 (alt)			-0.01*	-0.01**
			(0.00)	(0.00)
Close to party	2.15***	2.66***	2.19***	2.23***
	(0.02)	(0.08)	(0.03)	(0.04)
X Vote 2017		-0.02***		
		(0.00)		
X Distance from contention 2017 (alt)				0.02*
				(0.01)
Vote 2012	0.00	0.00		
	(0.00)	(0.00)		
Distance from contention 2012 (alt)			-0.01***	-0.01***
			(0.00)	(0.00)
Contact	0.30***	0.32***	0.27***	0.28***
	(0.04)	(0.04)	(0.05)	(0.05)
BSP	0.02	0.01	-0.16***	-0.16***
	(0.04)	(0.04)	(0.03)	(0.03)
SP	-0.18***	-0.17***	-0.33***	-0.33***
	(0.04)	(0.04)	(0.03)	(0.03)
Constant	1.45***	1.32***	1.79***	1.78***
	(0.06)	(0.07)	(0.05)	(0.05)
N	9,791	9,791	9,284	9,284
R^2	0.59	0.59	0.60	0.60

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

To test the robustness of the findings in Tables A5 and A6, which are the basis for Figures 2 and 3 in the main text, we use alternatives to *Distance from contention 2017* in Table A11. The dependent variable in all models is *Likelihood of victory*, and all are OLS models with respondent fixed effects. In models 1 and 2, we include *Vote 2017* as a measure of objective party performance in lieu of *Distance from contention 2017*. *Vote 2017* is a party's vote share in the 2017 election. The expectation is that as a party's vote share increases, respondents will be more likely to cite it as the likely winner. *Vote 2017* differs from *Distance from contention 2017* in that it does not take account of a party's relative placement compared to other parties in the constituency. In models 3 and 4, we include a modified *Distance from contention 2017* variable that aims to distinguish winners and runners-up, on the one hand, from those of out contention, on the other hand. To modify the variable, we recode all winners and runners-up as 0s. However, because the data are then skewed, we use a log transformation of the variable (which requires adding a small constant to all the 0 values). The modified variable therefore captures the logged distance from contention for parties in third place or worse. In doing so, it starkly differentiates between candidates in contention and those that are not in contention. *Vote 2012* and the modified *Distance from contention 2012* are calculated in the same way but based on the 2012 election results. Models 1 and 3 are identical to model 4 in Table A5 except for the inclusion of the alternative measures of party performance in place of the original *Distance from contention 2017* measure, while models 2 and 4 are identical to model 4 in Table A6 except for the inclusion of the alternative measures of party performance in place of the original *Distance from contention 2017* measure. The results in models 1 and 2, with *Vote 2017* as the measure of party performance, are substantively identical to those in Tables A5 and A6. (Note that the coefficients are in the opposite direction because we expect that as vote share increase, a respondent is *more* likely to believe that the party will win, whereas when a party gets farther from contention—positive values of the variable—we expect a respondent to be *less* likely to believe that the party will win). The results in models 3 and 4, with the modified *Distance from contention 2017* variable as the measure of party performance, are also consistent with those in Tables A5 and A6.

Table A12. Analysis of whether respondents correctly predict the constituency winner

	1	2	3	4	5
Close to winner	4.33***			4.36***	
	(0.36)			(0.38)	
Winning party rating					
Like a lot					6.11***
					(0.51)
Somewhat like					1.43**
					(0.57)
Neither like nor dislike					-0.24
					(0.50)
Somewhat dislike					-0.02
					(1.01)
Contacted by winner	1.07***			1.14***	1.15***
	(0.27)			(0.27)	(0.37)
Interest		-0.10		-0.27*	-0.25*
		(0.07)		(0.15)	(0.15)
Media		0.06		0.10	-0.09
		(0.05)		(0.11)	(0.23)
Knowledge		-0.15			
		(0.27)			
Education level					
Graduate +		0.25		0.09	0.24
		(0.16)		(0.20)	(0.55)
Matric + Intermediate		0.32**		0.34	-0.07
		(0.14)		(0.23)	(0.32)
Primary school		0.23		-0.00	-0.17
		(0.15)		(0.32)	(0.27)
Margin of victory			-0.89		
			(0.89)		
Same winner			0.04		
			(0.19)		
ENP			-0.38**	-0.10	0.20
			(0.17)	(0.24)	(0.60)
Constant	-2.40***	-0.44	0.81	-1.90**	-4.34**
	(0.23)	(0.29)	(0.60)	(0.83)	(2.13)
N	3,647	3,647	3,647	3,647	3,234
R^2	0.43	0.01	0.00	0.43	0.69

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Because the modeling strategy adopted in the main text exploits variation across party-respondent observations, we can only include independent variables that vary at the level of the party (e.g., whether a respondent reports having been contacted by a party) and we cannot control for individual- or constituency-level factors. In Table A12 we test the robustness of our findings by adopting a very different modeling strategy. Instead of using stacked data, each respondent in the analysis in Table A12 appears once in the data. In this analysis, the dependent variable is whether the respondent correctly forecasts the winning party in her constituency when asked which party is likely to win. Those who correctly forecast the party are coded as 1, all others (including those who did not provide a response) are coded as 0. Because our dependent variable is binary, all models are logistic regressions with standard errors clustered by constituency. In models 1, 2, and 3 we include different sets of controls. Model 4 then includes those independent variables that are (close to) statistically significant. Model 5 replicates model 4 but replaces *Close to winner* with responses to the *Winning party rating* variable (derived from the same variables used Tables A8 and A9), which captures the respondent's rating of the winning party. A description of all of the variables used in Table A12 can be found below in Table A12a.

Model 1 includes predictors related to the parties—much like those in our original analysis. Both *Close to winner* and *Contacted by winner* strongly predict whether a respondent correctly forecasts the winning candidate in her race. These findings echo the findings in the main text and appendix, suggesting that those who prefer the winning party are very likely to correctly predict the winner, compared to those who are not close to the winning party. If all respondents objectively evaluated the state of the race, then we should not expect to see *Close to winner* correlate with correctly forecasting the winner. But, consistent with our findings elsewhere that those close to a party expect it to win, being close the party that wins strongly predicts correctly forecasting the eventual winner.

Model 2 includes four variables related to political sophistication: *Interest*, *Media*, *Knowledge*, and *Education* (which is entered as a series of dummy variables). Theoretically, the expectation is that those who are more interested in the election campaign, consume more political news, have more political knowledge, and are more highly educated will be more likely to correctly predict the winner in their constituency. For the most part, however, these variables are not statistically significant, though several have p-values that are not exceedingly large (i.e., < 0.2).

Model 3 includes variables related to the constituency that should make it easier or harder to identify the winner: *Margin of victory*, *Same winner*, and *ENP*. Larger margins of victory and the same winning party in consecutive elections should make it easier for voters to correctly forecast the winner; higher ENPs should make that task more difficult because there are more parties. Of these, only ENP is statistically significant.

Model 4 includes statistically significant variables or those with relatively small p-values (i.e., *Media*, *Interest*, and *Education*). *Close to winner* and *Contacted by winner* remain statistically significant predictors of respondents' correctly forecasting the winner in their constituency. *Interest* is negatively correlated with correct forecasts, and the other variables (including ENP) are very imprecisely estimated. Finally, model 5 replicates model 4, replacing *Close to winner* with dummy variables for each response option for *Winning party rating.* The results are very similar to model 4.

All told, the results from this exercise, which adopts a very different modeling strategy, are remarkably similar to the analysis presented in the main text. Respondents accurately forecast the winner mainly when they feel close to the party that ultimate wins or rate the eventual winner very favorably. Other variables related to how politically sophisticated respondents are or how difficult it should be, based on the actual race, to predict the outcome do not actually correlate with correct predictions about the winner. The coefficient on *Interest* is in the incorrect direction, which may be capturing the fact that those who are politically uninterested are more likely to be politically unaligned. Being politically unaligned may, in spite of being uninterested in politics, help respondents to correctly forecast the winner because their views are not colored by their political preferences. Though the results about political sophistication and characteristics of the race may be surprising, it is worth noting just how strongly feelings toward the winning party predict a correct forecast. Such a tight correlation between feelings toward a party and correct forecast likely overshadow other plausible predictors of correct forecasts. In sum, the results of this analysis, which allows us to control for individual- and constituency-level factors, suggest nearly identical conclusions to those presented in the main text.

Table A12a. Variables used in Table A12

Variable name	Description
Close to winner	Dummy variable indicating that the respondent feels close to the party that is the eventual winner in the respondent's constituency.
Feelings toward winning party	Response on a five-point scale indicating the respondent's feelings toward the party that ultimately won in the respondent's constituency. Responses are entered as a series of dummy variables for each response, with "dislikes the party a lot" as the reference category.
Like a lot	Dummy variable indicating that the respondent likes a lot the party that won in the constituency; reference category is dislikes the party a lot.
Somewhat like	Dummy variable indicating that the respondent somewhat likes the party that won in the constituency; reference category is dislikes the party a lot.
Neither like nor dislike	Dummy variable indicating that the respondent neither likes nor dislikes the party that won in the constituency; reference category is dislikes the party a lot.
Somewhat dislike	Dummy variable indicating that the respondent somewhat dislikes the party that won in the constituency; reference category is dislikes the party a lot.
Contacted by winner	Dummy variable indicating the respondent was contacted by the party whose candidate eventually won the constituency.
Interest	Response to a survey item asking about a respondent's interest in the campaign. 0 = not at all interested, 1 = a little interested, 2 = somewhat interested, 3 = very interested.
Media	Average of three variables asking about the frequency with which respondents obtained news from various media sources (TV news, newspapers, WhatsApp). Responses for each of the three components are: daily (4), regularly (3), occasionally (2), rarely (1), never (0).
Knowledge	Dummy variable indicating whether the respondent correctly said that one of the major parties in the election was in an election alliance and correctly identified such an alliance.
Education	Response from a variable asking respondents about their highest level of education. The original variable has 9 values. For the purposes of creating a smaller number of categories with more respondents in each, we create a four-point education variable. Each value is entered separately, with non-literate as the reference category.
Graduate +	Dummy variable indicating that the respondent has at least a college education or more; reference category is non-literate.
Matric + intermediate	Dummy variable indicating that the respondent's highest level of education was class 10, 11, or 12; reference category is non-literate.
Primary school	Dummy variable indicating that the respondent's highest level of education was between class 1 and class 9; reference category is non-literate.
Margin of victory	Vote share separating the winner and runner-up in the respondent's constituency in 2017.
Same winner	Dummy variable indicating that the same party that won in the respondent's constituency in 2017 also won the constituency in 2012.
ENP	Effective number of parties in the respondent's constituency in 2017.

Table A13. Analysis of *Winner* by respondents' expressed uncertainty in survey items

	All	Certain	Uncertain	All	Certain	Uncertain
	(1)	(2)	(3)	(4)	(5)	(6)
Distance from contention 2017	-0.01**	-0.01	-0.00	-0.01***	-0.01**	-0.01
	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
Close to party	3.88***	3.89***	4.02***	3.95***	4.01***	3.99***
	(0.12)	(0.14)	(0.23)	(0.12)	(0.15)	(0.22)
Distance from contention 2017				0.04***	0.04***	0.03
X Close to party				(0.01)	(0.01)	(0.02)
Distance from contention 2012	0.00	-0.00	0.02**	0.00	-0.00	0.02*
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
Contact	1.57***	1.87***	1.23***	1.58***	1.86***	1.25***
	(0.21)	(0.30)	(0.31)	(0.21)	(0.30)	(0.31)
BSP	-0.25*	-0.18	-0.99***	-0.26*	-0.18	-0.98***
	(0.14)	(0.16)	(0.37)	(0.14)	(0.16)	(0.37)
SP	-0.18	-0.24	-0.31	-0.18	-0.24	-0.28
	(0.14)	(0.16)	(0.32)	(0.14)	(0.17)	(0.32)
N	8,538	6,287	2,251	8,538	6,287	2,251
Pseudo-R^2 / R^2	0.65	0.65	0.67	0.65	0.65	0.67

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

One concern is that although respondents are providing answers to questions about who the likely winners and runners-up are, these may not reflect genuine beliefs so much as a desire to satisfy a survey enumerator by providing some sort of answer. However, we find that in response to other survey items, significant shares of respondents say that they do not know. Respondents who say they do not know the answers to other survey items may be especially likely to volunteer what are, in effect, non-opinions about the likely winner in their race. We therefore replicate model 2 in Tables A5 and A6, but break down our analysis by those who express uncertainty on multiple survey items (models 3 and 6) and those who do not (models 2 and 5). (The original results from Tables A5 and A6 appear in models 1 and 4 in Table A13). We classify respondents as "uncertain" if they answer "don't know" to any one of three questions about 1) how close the race will be (very, somewhat, not at all), 2) the expected margin separating the 1st and 2nd place candidates (large or small), and 3) the expected margin separating the 2nd and 3rd place candidates (large or small). The results for both groups are virtually identical. Some coefficients among the "uncertain" group do not reach statistical significance, but this is a much smaller group. Since the results are stronger among those whom we classify as "certain," it is unlikely that our results are being driven by respondents offering responses simply to satisfy survey enumerators.

Appendix E. Supplemental discussion of the case of Uttar Pradesh

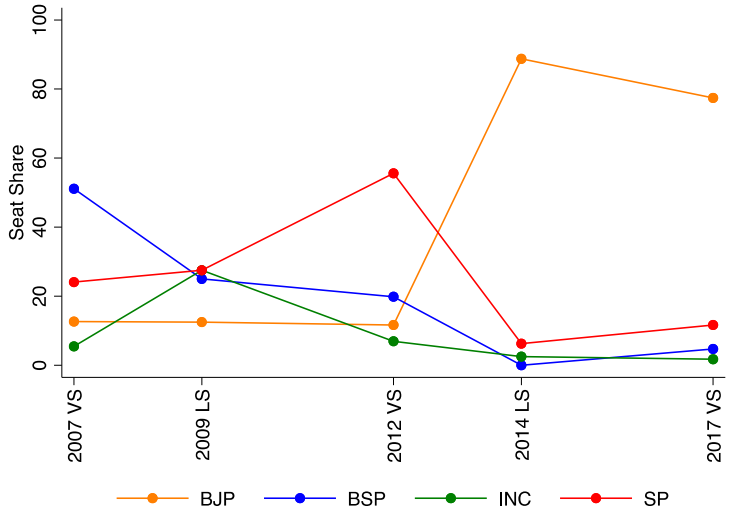
In this supplemental discussion of Uttar Pradesh, we first discuss whether our respondents' inability to correctly forecast the election makes sense given the electoral context of UP. We then explain why findings from UP should plausibly generalize to other Indian states with multiparty systems.

Why identifying constituency winners is difficult

The main text documents that most respondents do not seem to form independent judgements about their preferred party's electoral viability that correspond to objective party performance. As a result, the respondents who correctly identify the winner in their constituency are mainly those who like the party of the eventual winner (see Table A12). As we suggest in the conclusion to the article's main text, voters' inability to correctly identify the winner in their constituency likely stems from a lack of reliable information about how parties are likely to fare in an election. In particular, three elements of the electoral context in UP make the task of identifying likely winners in a constituency difficult.

First, the state has experienced major swings in seat shares from one election to the next, suggesting that a party's prior performance is not necessarily a good indicator of its likely performance in a subsequent election. In other words, elections prior to 2017 suggested that forming beliefs about the likely winner in one's own constituency based on the party that swept the prior election would not likely be particularly accurate. Obviously, seat shares and vote shares may not be strongly correlated. For instance, in recent elections, the BSP has won seat shares similar to Congress despite winning substantially larger vote shares. But, since voters may not be well versed in vote shares but have a good sense for who wins a legislative majority, seat shares are likely a better guide for shaping voters' perceptions of which parties are doing well across the state. Figure A5 presents the seat shares won by the BJP, BSP, SP, and Congress in the two national elections (2009, 2014) and two state elections (2007, 2012) held in UP preceding the 2017 state election.

Figure A5. Main parties' seat shares in UP elections, 2007-2017



State-wide, the elections before 2017 featured one clear election victory for each of the three major parties (BJP, BSP, SP) and one largely split verdict. Viewed against this backdrop, it is perhaps not surprising that voters would be optimistic about their party's chances of victory. In 2007, the BSP won the Vidhan Sabha election, capturing a (slim) majority of seats with only about 30% of the vote. In the 2009 Lok Sabha election, the BSP, SP, and Congress all won sizeable shares of the state's seats, with the SP winning a plurality, though only slightly more than Congress. In the next state election, in 2012, the Samajwadi Party won a healthy majority, though with only about 30% of the vote. However, just two years later, the BJP captured a huge majority of seats in the Lok Sabha election. As it happened, the main parties' seat shares in the 2017 Vidhan Sabha election were relatively similar to those of the preceding Lok Sabha election, but given the volatility in the previous elections, it was certainly not a foregone conclusion that this would be the case.

Second, and related, as noted in the main text, most seats in UP change party hands from one election to the next. In other words, there are few "safe seats" that voters know are likely to remain with the same party over time. The absence of safe seats is unsurprising given the dramatic changes in seat shares documented in Figure A5. Nevertheless, it is possible that a good number of seats are party strongholds and that much of the seat-share volatility is driven by a smaller set of "swing" constituencies. However, as discussed in the main text, most seats have changed hands in recent elections. In the 2007 UP state election, 64% of seats were won by a different party from the one winning in the previous election in 2002. Constituency boundaries changed between 2007 and 2012, making it difficult to say which seats remained in the hands of one party or another. Then, in 2017, nearly 80% of seats changed party hands.

Third, the vote share margins separating the winners and runners-up and the first runners-up (2nd place) and second runners-up (3rd place) are not particularly large. Table A14 lists the mean vote share difference between the first- and second-place candidates and between the second- and third-place candidates in 2017 and the preceding three Vidhan Sabha elections. Though these vote share margins were slightly larger in 2017, voters casting ballots in 2017 would have experienced multiple successive elections in which the average election was decided by a relatively modest margin, and less than 20% of the vote separated the winning candidate from the third-place candidate. Though Table A14 suggests that most races were not decided by exceedingly slim margins, it also suggests that most races were not blowouts in which one party clearly outpaced all others.

Table A14 Average vote share separating top three candidates

Election	*1st and 2nd place*	*2nd and 3rd place*
2002	8.1%	8.6%
2007	7.3%	11.4%
2012	8.0%	9.5%
2017	13.6%	9.7%

Why findings from UP could plausibly generalize across other parts of India

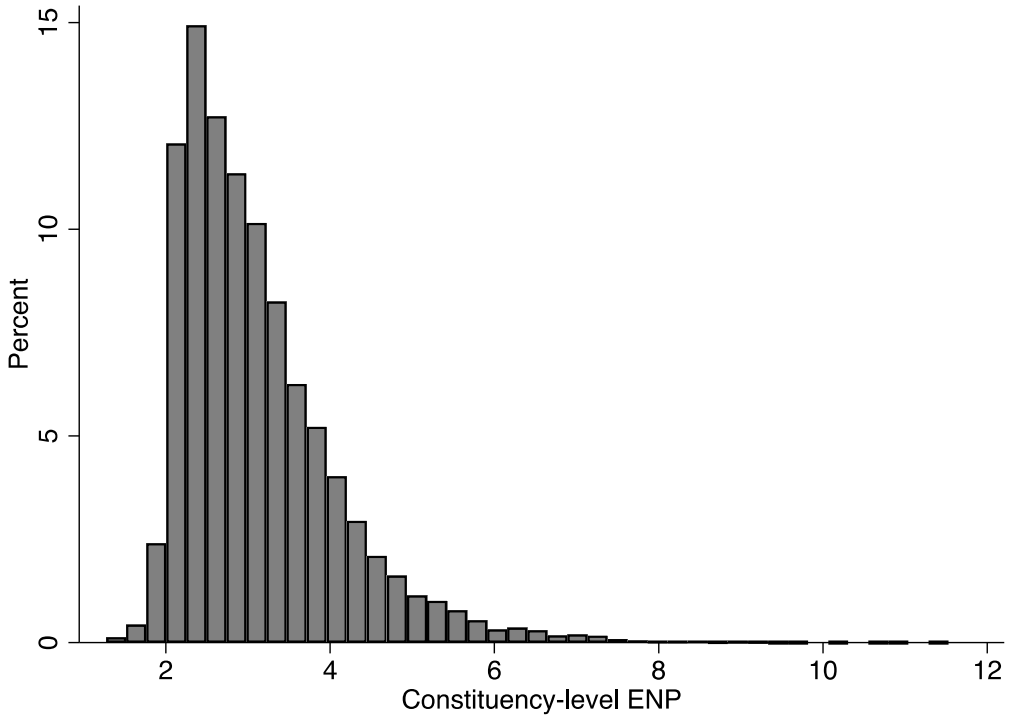
Obviously, without comparable survey data from other Indian states, we cannot say anything conclusive about the generalizability of our findings across India. Naturally, the question of strategic voting is largely irrelevant in states in which two parties dominate political competition (e.g., Chhattisgarh, Gujarat, Rajasthan, Madhya Pradesh, Uttarakhand). However, there is good reason to believe that the findings from the 2017 UP election are not anomalous.

First, one might worry that there was something about the 2017 UP election that made identifying the leading candidates especially difficult. Though the previous section suggested that the political context in UP does not make it easy for voters to gauge who the likely winners are at the constituency level, compared to other elections in UP, the 2017 election should have arguably made strategic voting easier than in other elections, not harder. Although the state had not recently seen two consecutive electoral victories by the same party, the BJP's victory in the 2014 Lok Sabha election was more emphatic—in terms of both seat and vote shares—than either the BSP or the SP's victories in the 2007 and 2012 state elections, respectively. Though the state's record of volatility quite reasonably meant that voters would hesitate to believe with a high degree of certainty that the BJP would repeat its strong performance from 2014 in 2017, the BJP was arguably more of a front-runner heading into the 2017 election than either the BSP or SP had been following their prior electoral victories. The task of forecasting was further simplified by the Samajwadi Party-Congress alliance, which diminished the number of major competitors in most seats from four to three.

Second, one might worry that something specific to UP makes strategic voting especially uncommon. If this were the case, then we would expect that multiparty elections in other states would exhibit more frequent strategic voting that would, in turn, produce more frequent Duvergerian (that is, two-party) outcomes. This is not the case. In a number of Indian states, only two major parties compete, meaning that there is little opportunity for strategic voting or non-two-party outcomes. However, in many other states, multiple parties compete. In state elections in India's large states (defined as those sending ten or more legislators to the lower house of the national legislature) during the period from 1961 to 2019, there were 17,212 constituency-level races outside of Uttar Pradesh in which three or more candidates ran on party labels that won 5% or more of the statewide vote. In other

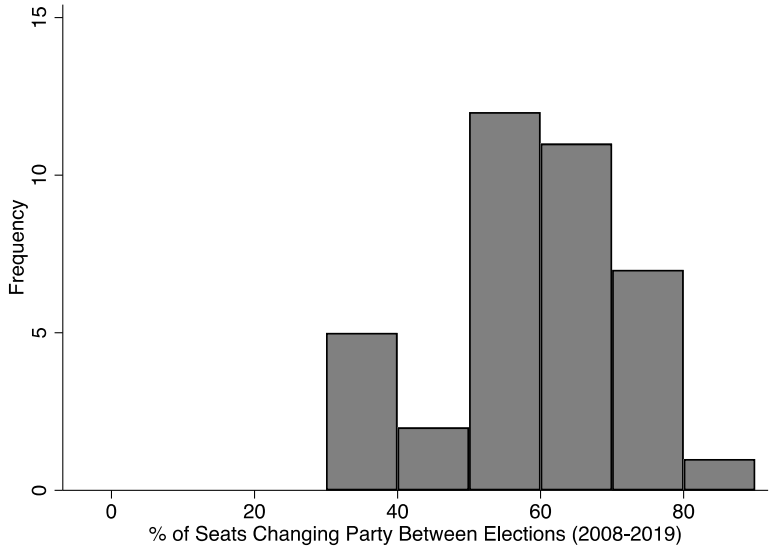
words, these should be races where there were three or more credible candidates. Figure A6 is a histogram representing the distribution of constituency-level effective number of parties (ENPs) across these multiparty races. As the figure shows, relatively few outcomes conform to the Duvergerian two-party expectation. The constituency-level ENP was less than 2.5, arguably the highest threshold one could use to classify an electoral outcome as conforming to Duverger's Law, in about 30% of races (5,233), and less than 3% (488) had a constituency-level ENP of 2 or lower. In other words, the paucity of Duvergerian outcomes in Indian state elections where multiple major-party candidates compete is not unique to Uttar Pradesh, where almost no races in 2017 conformed to the Duvergerian two-party expectation. Rather, few races witness convergence anywhere in India. Aggregate-level data suggest therefore that the individual-level findings from UP are probably not anomalous. Voters in India seem to only rarely converge on two candidates in the context of multiparty races, indicating that strategic voting is probably not particularly widespread elsewhere in India.

Figure A6. The distribution of constituency-level ENPs in state-level multiparty races outside Uttar Pradesh



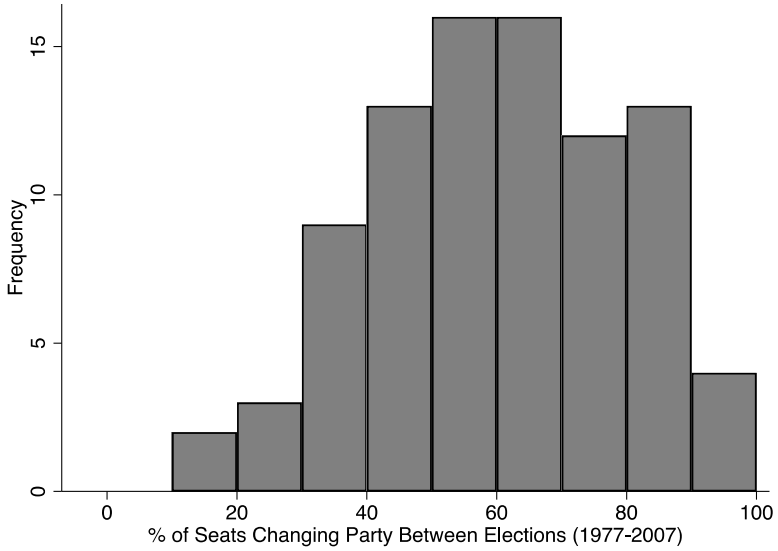
Furthermore, Uttar Pradesh is not a clear outlier in terms of the volatility of its elections or the margins of victory separating its leading candidates. Figure A7 presents a histogram depicting the share of seats changing hands in state-level elections held in India's seventeen large states between 2008 and 2019. Figure A7 includes only these recent elections because the 2008 delimitation changed the constituency boundaries in all states, making it difficult to identify whether a seat remained in the hands of the same party before and after 2008.

Figure A7. Share of seats changing party hands between elections, 2008-2019



As Figure A7 shows, the share of seats changing hands in UP in 2017 was rather high at nearly 80%; the 2007 share, at around 65%, was much closer to the mean. Figure A8 presents a comparable figure for the 1977-2007 period, during which time constituency boundaries did not change. Compared to historical figures, the share of seats changing party hands in recent UP elections is not remarkably high, even if it is above the median.

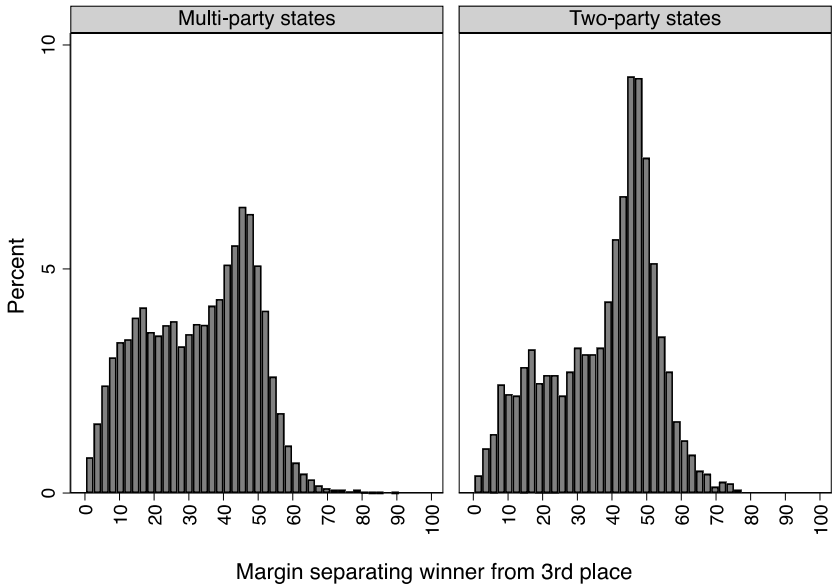
Figure A8. Share of seats changing party hands between elections, 1977-2007



Finally, Figure A9 depicts the distribution of vote shares separating first and third place candidates in races held in state legislative elections in India's seventeen large states between 2001 and 2019. The right panel includes the four large states that have had predominantly two-party competition during this period (Chhattisgarh, Gujarat, Rajasthan, and Madhya Pradesh), while the left panel includes the remaining thirteen multi-party states. Figure A9

distinguishes between two-party and multiparty states, since states with only two main parties will likely see very large differences between the first- and third-place candidates because the third-place candidates may win very small vote shares. Recall that in Uttar Pradesh, the average vote share separating first- and third-place parties in the 21st century has been slightly less than 20%. In other words, Figure A9 suggests that elections in UP are more competitive than average inasmuch as the margins separating the first- and third-place candidates are lower than average. However, the left panel of Figure A9 also includes many constituencies that did not feature multiparty competition. For instance, some elections in the left panel had largely two-party competition but are not included in the right panel because two-party competition is not necessarily the norm across all elections. (In other words, there can be elections in otherwise multiparty states that more closely approximately two-party competition). Perhaps more importantly, the left panel does not take account of the fact that many states with multiparty systems feature election alliances in which all of the main parties join one of two major alliances, effectively reducing competition to just two candidates in each electoral district. In short, though elections in UP may be slightly more closely contested than in other Indian states, they do not constitute major outliers that would lead one to expect that findings from UP should not generalize to other multiparty Indian states.

Figure A9. Vote share margins between first- and third-place candidates, 2001-2019



In sum, elections in UP are a bit more closely fought that average, and the most recent election in 2017 was somewhat more volatile in terms of seats changing hands. But, the levels of volatility and competitiveness do not stand out as particularly unusual in the Indian context. Thus, we should not expect that voters in other multiparty states in India should have appreciably better information about likely outcomes in their constituencies. Moreover, election outcomes from across India suggest that when multiple sizable parties field candidates in a constituency, voters rarely converge on just two of those candidates. Though this evidence, in Figure A6, only indirectly suggests the absence of strategic voting, we see that UP is not alone in frequently producing multiparty outcomes at the constituency level when several viable parties compete.