Gauging a Horse Race: Experimental Evidence Towards a Theory of Political Poll Comprehension

Nicholas K. Neuteufel

Abstract

Research into elections and voting patterns shows that citizens can be strategic actors who alter voting intentions and contributing patterns to achieve ends. There is sparse research, however, on how individuals interpret and react to pre-election polls and their uncertainty. This paper presents a survey experiment with a fake horse race poll of a South Sudan run-off election. Respondents (N=250) were randomly assigned to a text-only, text and bar plot, or text and dot plot with margins of error presentation of the poll and then asked to evaluate the probability that the poll winner would win an imminent election. An ordered probit model was used to evaluate the impact of the treatment and pretreatment effects. Statistics education was found to have no effect on poll comprehension, even for those taking the class in the past year. Distrusting polls' predictive power had a negative effect on assigned likelihood, and the bar plot treatment seems to have led respondents to overstate the likelihood of the poll winner winning compared to the control and dot plot. These findings are robust in several model specifications, including a Bayesian approach. These results suggest that even seemingly trivial choices by media—including graphics—play a direct role in how individuals perceive the likelihood of electoral outcomes and thus may influence election dynamics.

Keywords Polling, horse race journalism, strategy frame, survey research, framing, public opinion, data visualization.

Reproducibility: All files—data and scripts—are in the "HorseRace" repository: `https://www.github.com/NicholasNeuteufel`.

Introduction

Gallup's decision to renounce trial heat polling for the 2016 Republican nomination for President of the United States took many in the political media by surprise. Politico's headline ("Gallup gives up the horse race") got straight to the point: Gallup's move signals a loss to a focus on which candidate is ahead of the other in the "horse race" of an upcoming election (Shepard 2015). Other names for the phenomenon include the "game schema" and "strategy frame" (Nisbet 2008, 314), though some argue that there is a "conceptual difference between a game frame and strategy frame" (Aalberg, Strömbäck, and de Vreese 2012, 167). Such journalism is "characterized by a focus on questions related to who is winning and losing, the performances of politicians and parties, and on campaign strategies and tactics" (Aalberg, Strömbäck, and de Vreese 2012, 163). Horse race stories are contrasted with pieces "foregrounding issue positions, candidate qualifications, or policy proposals" (Nisbet 2008, 314).

Journalists as Railbirds

The rise of horse race journalism has been documented for decades. Patterson (1977) analyzed television coverage of the 1976 U.S. presidential election, reporting that stories on "candidates' comings and goings on the campaign trail, their strategies for winning votes, and their prospects for victory or defeat" accounted for three-fifths of networks' election coverage (73). A mere 28 percent of television coverage discussed the issues or the platforms, records, or backgrounds of the candidates involved (Patterson 1977, 73). The trend did not cease in 1976. An increase in horse race reporting has created a large media demand for polls and "expanded the reliance on polls as news, including polls of a sort once considered not reliable for publication" (Rosenstiel 2005, 698). Traugott (2005) traces an "explosion" of U.S. presidential trial heat polling beginning in the 1980s, finding an increase of more than 900 from 1984 to 2000 (Traugott 2005, 644). U.S. election coverage using phrases like "polls say" and "polls show" increased dramatically from 1996, which witnessed 4,489 mentions during the election year, to 2004, which saw 11,327 such mentions (Frankovic 2005, 684-5).

The horse race-polling trend is a global phenomenon. Patterson (2005) documents an extensive literature analyzing similar trends in European electoral and political journalism. The number of poll reports in four top German newspapers during the run-up to German federal elections increased from 33 in 1987 to more than 650 in 2002 (Brettschneider 2008, 486). Swiss newspapers also focus on political poll reports, according to a content analysis of 31 outlets by Hardmeier (1999). Bhatti and Pedersen (2015) analyzed more than 1,070 articles discussing polls forecasting the 2011 parliamentary elections in Denmark. Outside of Europe, the Canadian media has extensively used polls to frame election coverage since at least the 1987 election (Andersen 2000). Australian evening news discussions of political opinion polls during the 1980 federal election cycle were called "extensive, superficial, and inaccurate" by Smith III and Verrall (1985, 76). Weimann (1990) went so far as to say that the Israeli press during six elections from 1969 to 1988 had an "obsession to forecast" the winners and losers beforehand.

Impacting the Derby

The strategic frame has changed the media's coverage of presidential elections, and studies have shown demonstrable impacts on voters' behavior as a result. A survey experiment conducted on 23,421 Dutch voters showed a "subtle... but societally substantive" effect of polling framing on vote intention (van der Meer, Hakhverdian, and Aaldering 2015,

22). The treatment was measured to affect a change of "two to three additional seats in the 150-seat Dutch parliament" [1] (van der Meer, Hakhverdian, and Aaldering 2015, 22). Most importantly, van der Meer, Hakhverdian, and Aaldering (2015) found that it was not actual poll results that impacted voting intention, but rather framing on latest events in the horse race. The presentation of poll results with text stating that a party is gaining in the race—that is, its support increased compared to an earlier poll—led the party to "subsequently obtain more votes than if the party had not been framed as a winner" (van der Meer, Hakhverdian, and Aaldering 2015, 21-2).

Bandwagon voting is not the only potential impact of horse race journalism on voters' intentions in upcoming elections. Hall and Snyder (2015) analyze primary races for statewide executive offices as well as for the U.S. Senate and House between 1990 and 2010 and find that voters act strategically in both campaign contributing and voting. Information on the horse—who is likely to win and who is not—is isolated by media markets in select races and is shown to be important in voters' decisions. This supports Myatt (2007)'s theory of strategic voting, which includes a component for information in which "Voter i knows her own preferences, but not those of others" (259). Polls may act as a signal of the electorate's preferences, however such signals can be "noisy" and can convey uncertainty (264). This is especially true when a poll is even somewhat close or "noisy" about the electorate's belief—such a public signal can remain "a step away from complete knowledge of the electoral situation" (264).

The effects of the strategy frame go beyond voter persuasion. Frankovic (2005) adds to a sizable literature on the importance of polls and horse race perceptions in primary elections. One major concern of primary voters is "electability," the perception of which is very much controlled by polling data (Frankovic 2005). Frankovic (2005) cites perceptions of electability and general-election-focused polls as a reason for John Kerry winning the 2004 Democratic nomination. Mutz (1995) demonstrates via times-series analysis that "horse race spin," or "the extent of media coverage suggesting a candidate is gaining or losing political support," may have played a key role in campaign contributing by primary activists during the 1988 Democratic presidential primary. Media stories focused on a candidate's relative position in the horse race have been shown to play a role in strategic contributing, raising questions of how poll presentation in media stories affect not just future polls but resources like funds as well.

Misleading Methods

Even Oxford-educated media magnates misunderstand interpretation of polling data or use such misconceptions to their advantage (Murdoch 2015). Loewen, Lupia, and Rubenson (2015) argue that "members of the media misread the polls and misled their audiences" in the weeks before the Canadian and British parliamentary elections of 2015 because such analysts failed to appreciate the late-changing nature of elections and the propensity of strategic voting. Patterson (2005) demonstrates that media sources regularly misrepresent and misinterpret margin of error and overstate statistically insignificant changes in polls in short time periods. Rather than attribute most small shifts to sampling error or differences in poll organization methods or questionnaires, many pundits seek alternative explanations about the nature of the horse race in question (Patterson 2005). This is perfectly illustrated by a November 2015 tweet by Politico national political reporter Gabriel Debenedetti: "Movement in the Democratic primary since early October (WSJ/NBC poll) [...] Clinton 58% $\rightarrow$ 62% Sanders 33% $\rightarrow$ 31%." The tweet implies the different point estimates of two polls signify "movement" in the race, even though the shifts are both less than the general margin of error (reported in the piece linked to as 5 percentage points). The Sanders shift (-2) was less than half of that margin of error.

Graphic presentation of polls are another way citizens may be misled about the nature of a specific poll or the race as a whole, impacting those of lower educational attainment most (Hollander 1993). Conventional journalistic style, by representing one margin of error (see Franklin (2002)) or using a simple bar chart without margins of error, might be consistently misleading readers about the nature of an ongoing race (Franklin 2002). Bar plots without margins of error may be passing off polls as more certain than they are especially when presented in isolation without a link to the poll press release or accompanying information. Journalists not trained in polling methods may be making errors passed to the public by a simple retweet or graphic inclusion. Such media miscues may be impacting races where voters are shown to vote strategically, such as in the races analyzed in Hall and Snyder (2015). Studying the graphic presentation of polling data may have become more important as voters have come to engage with and process stories in digital formats and spread information and pictures more quickly.

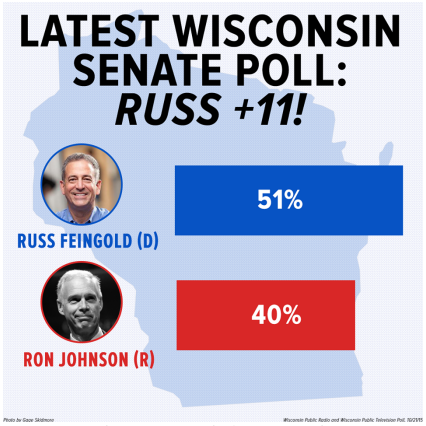


Figure 1. Democratic Senatorial Campaign Committee (DSCC) Facebook post (October 21, 2015).

Instances of misleading or unhelpful communication do

[1] The equivalent of a five-to-nine-member swing in the U.S. House, though differing electoral systems make the comparison weak at best.

not originate solely from media or those claiming objectivity. Misleading poll presentations occur within political communications from party and campaign offices. One recent example is Figure 1, taken from the Facebook page of the Democratic Senatorial Campaign Committee (DSCC). Mostly notably, the DSCC post had no mention of the poll's margin of error, sample size, sponsor/institution, or methods. There is a tactical element to these communications as communications actors understand the strategic elements of volunteering, contributing, and voting noted earlier. The graphic even goes so far as to emphasize the point estimates' difference (+11 for Democrat Russ Feingold) with the margin of error noticeably absent from the graphic.

Hypotheses

Based on Hollander (1993) and evidence from shifting vote intentions in the Dutch election experiment by van der Meer, Hakhverdian, and Aaldering (2015), one should expect, *ceteris paribus*:

Hypothesis 1: There will be a significant difference in probabilities assigned to the poll leader winning an upcoming election based on graphic stimulus.

Hypothesis 1a: The bar plot with margin of error information transmitted via caption text will result in higher likelihoods assigned.

Hypothesis 1b: The dot plot with margins of error embedded in the graph will result in more modest likelihoods assigned to the poll leader compared to the bar plot.

Hypothesis 2: If there is a difference in treatment effect, it will be mediated by whether the person has reported they have taken a statistics course.

(These imply a null hypothesis: There will be no significant difference in probabilities assigned.)

Methods

This section details a survey experiment designed to test how media presentation and framing of poll results influence perceptions of a horse race. First, I address issues that some argue may limit the external validity of the design, namely: university student sampling and a Web-based survey, then I detail the questionnaire design.

Addressing Potential Validity Shortfalls

There are two key hurdles an otherwise rigorous Internet-based student-sample survey experiment must pass to claim strong external validity. They are:

(1) showing validity beyond a student sample and

(2) accounting for Internet coverage bias.

(1) A common refrain in assessing student sampling is that "a student sample lacks external generalizability" and acts as a hurdle to drawing an inference from an experimental study (Druckman and Kam 2009, 3). This trend, which began in social psychology with Sears (1986), has made its way to political science experimentation. Gerber and Green (2008) specify studies in political communication and social cues as situations where "the external validity of lab studies of undergraduates has inspired skepticism" (358).

However, Druckman and Kam (2009) give evidence that the validity concern of a student sample is often exaggerated. Drawing from Monte Carlo experiments on three samples differing wildly in their distributions, Druckman and Kam (2009) show that "if the treatment effect is the same across populations, the nature of a particular sample is largely irrelevant for establishing that effect" (12). There is, regrettably, no guarantee in this scenario as survey research on presenting polling results is sparse at best. However, previous research can guide us on the impact of a student sample.

The major drawback of the student sample is limiting heterogeneity of responses to questions of education level, which may be a useful predictor of polling comprehension. College students, by definition, are limited to responding to question of highest education level with a response like "some college" or "graduated college" depending on credit status. Hollander (1993) gives evidence that graphic presentation of polling differentially impacts perception of public opinion. The difference is based on level of education—those with lower levels of education were more persuaded and compelled by a bar plot of poll results than contradictory text information within the same article. The homogeneity of education level within this study's student sample is a hindrance to generalizability and by itself merits replication in a larger, more heterogeneous sample.

Within that constraint, a student sample may actually *underestimate* any effect size this experiment may demonstrate for two reasons. First, even university students focusing on non-math subjects are much more likely to have recent mathematical training than non-mathematically-focused adults. Second, students undergo selection bias. Students at exclusive universities are chosen by admissions committees to some extent based on performance during high school, including grades and test scores in math curricula. Thus even non-mathematically-inclined students surveyed are more likely to appreciate the mathematical ambiguity of the prompt than similar adults.

(2) An additional potential hurdle to generalizing the results of this survey is Internet coverage error. The issue is not generalizing from the students accessing an online form of the survey to the general population of students, because students are required to have e-mail access at UNC and are given campus Wi-Fi access. The generalizability problem here is applying the results of online students to a general population of American adults—including persons not using the Internet. 11% of U.S. adults do not regularly access the Internet (Pew Research Center 2015). Those 11% tend to be less knowledgeable about current events and political issues (Pew Research Center 2015).

Representativeness problems in political polling have been corrected by adjustment and stratification strategies. One interesting case is Wang et al. (2014), who use very unrepresentative polling data from Xbox Live users in the United States

(who skew very much male and 18-29) to project the 2012 U.S. presidential election quite well at the state level for all 50 states. These strategies, however, depend upon outside data to regress onto or to analyze alongside. These data do not currently exist for the experiment or concept at hand, thus adjustment is not a feasible strategy.

Rather than statistically adjust, one can consider the results of the experiment with a theoretical basis in mind. The expectation of Internet coverage bias is that the sample used in this experiment may be more informed politically than non-Internet users, as Pew's surveys show is generally the case (Pew Research Center 2015). This may be another factor underestimating any treatment effect found in the experiment. The only way to be sure is to replicate this study with a mail or random-digit-dialing (RDD) design.

Experimental Design

This section summarizes the questionnaire used to collect data analyzed later. The first section of the survey asked for basic information about the respondent: gender identity, age, education level, and primary college major in terms of categories (humanities, natural sciences, social sciences, etc.). These demographic elements were collected to be analyzed as potential predictor variables. The last question of the section read:

> South Sudan became a country in 2011 following a referendum with more than 90% of South Sudan's voters voting for independence from Sudan. Since 2011, the country has seen tension and conflict, including instances of election violence.
> Do you think democracy in South Sudan is a good idea?

This question was included not for the response, but to prime participants about South Sudanese elections and make sure they knew South Sudan was a real country[2] and that it held elections.

The next section was the treatment. The control group was exposed to the following text designed to emulate journalistic style of stories regarding new polls:

> A reputable non-ideological polling firm specializing in African elections wanted to predict the outcome of South Sudan's upcoming run-off election between the Movement Party and the Opposition Party. Using dozens of trained volunteers, the firm randomly sampled more than 370 South Sudanese citizens representative of the country's demographic and ethnic diversity.
> The firm sent out a press release detailing its results, including the following: "Our recent poll in South Sudan showed that 54% of respondents supported the Movement Party while 46% will vote for the Opposition in the upcoming run-off election. The poll's margin of error is ± (plus or minus) 5 percentage points."

The treatment groups were exposed to the text and either a traditional bar plot (Figure 2) or a dot plot with margin of error bands around the point estimates (Figure 3).[3]

Figure 2. Bar plot treatment

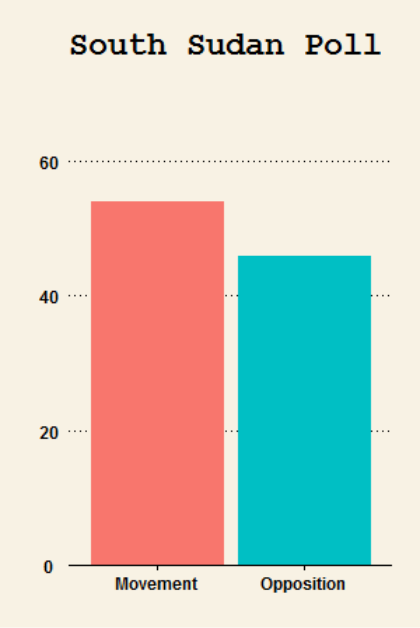
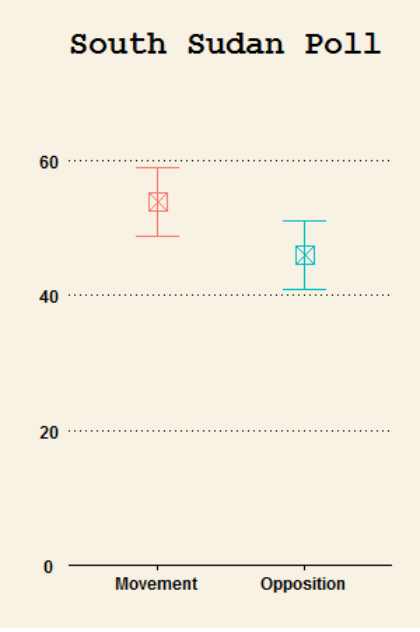


Figure 3. Dot plot treatment



[2] A few respondents were confidently skeptical and told me after the survey that South Sudan had to be a fabrication. They were corrected.

[3] Note that both plots have perfectly equal axes. The theme is "wsj," designed to emulate the *Wall Street Journal* from the packages "ggthemes."

Each group was then asked: "Based on this poll and this poll alone, what probability would you give to the Movement party winning the election, which takes place this week?" Respondents in all three groups were then asked to identify a probability level they would assign to the Movement Party winning the (fake) upcoming election given the (fake) poll. They could select on an ordinal scale from "Very Unlikely" to "Very Likely" with an option for "Neither Likely Nor Unlikely" as well as "I Do Not Know."

Only two polled entities (the parties) were selected so as not to introduce confusion between the common use of the term margin of error and the margin of error of proportions in polls with more than two choices. (See Franklin (2002) on this common error.) South Sudan was selected as the scenario because American media rarely, if ever, covers South Sudan in depth, making it extremely unlikely that a U.S. college student has a strong opinion on a false election in which two almost-fake political parties are participating. Thus there is little risk of a pretreatment effect specific to South Sudan, its parties, or an upcoming election. The lack of information also helps to reduce the risk of partisanship bias inherent in polling. Similar survey design asking about Democratic or Republican candidates' chances of winning given just one poll may induce motivated reasoning or wishful thinking that clouds mathematical considerations of the polls' presentation and framing.

It is important to note that the *most correct* answer is "Neither Likely Nor Unlikely." This is because the 95% confidence interval for the difference of proportions in a single question response is

$$CI_{(p_1-p_2)} = 1.96 \cdot \sqrt{\frac{(p_1+p_2)-(p_1-p_2)^2}{n-1}},$$

which in this case has $p_1 = 0.54$ (Movement Party support), $p_2 = 0.46$, and $n \approx 371$, resulting in a ≈ 10.1 percentage point margin of error (Franklin 2002, 3). This margin of error is greater than the lead the Movement Party has (8 points), so it is difficult to ascertain a likelihood that the Movement Party is actually winning the race. Another nod in this direction is that confidence intervals are to be interpreted as examples of the procedure of repeated random sampling, not as intervals where an analyst is 95% confident the true parameter is encompassed. Thus it strains credulity to interpret one random sample as evidence strong enough to move from a prior of equal probability to a belief of likelihood. This is especially true because we have no evidence of turn-out or voting probability within the electorate.

A third section asked respondents about their experience with polling. Participants were asked if they have ever taken a statistics class, how long ago that was if they have, how often they read about polls or political campaigns, what their opinion of horse race media stories is, and whether they agree or disagree that "polls of randomly selected voters can be trusted to predict upcoming elections" on a scale ranging from "Strongly Disagree" to "Strongly Agree." These questions were included to evaluate possible preatreatment effects.

The final portion of the questionnaire posited a communications strategy evaluation by the Movement Party. The first question asked participants to rank, in order, "the most persuasive arguments that the poll [from the control/treatment section] means the Movement Party will win" from most to least convincing. The four arguments were: "This poll shows that the Movement Party is gaining a bigger lead compared to a poll last week," "The Movement Party's lead (8 points) is larger than the margin of error (5 points)," "Even if the Opposition gained most of the margin of error, the Movement Party would still be ahead," and "The Movement Party is at 54% and the Opposition is at 46%."

The survey was fielded from October 20, 2015 to October 25, 2015, mostly at the University of North Carolina at Chapel Hill. The primary recruiting tool for respondents was social media advertising and e-mail requests. The final number of completed responses was 250. The survey software used was Qualtrics. A full spreadsheet of the data can be found in the Github referenced below the abstract on page 1.

Results

Summary Statistics

Summary statistics for the sample are found in Tables 1 (gender and education level) and 2 (age). Summary statistics for pretreatment measures are found in Table 3. Treatment and outcome measures are found in Table 4.

Table 1. Summary statistics: Gender & Educational attainment

Identity	N	Percentage
Female	131	52.4%
Male	114	45.6%
Other identities	5	2.0%
Highest education level		
Some HS/less	30	12.0%
Completed HS/GED	6	2.4%
Vocational training	6	2.4%
Some college	129	51.6%
Graduated college	46	18.4%
Post-grad/more	33	13.2%

N=250.

Table 2. Summary statistics: Age

Min	1Q	Median	Mean	3Q	Max
14	20	21	28.93	36.75	88

N=250.

Table 3. Pretreatment measures summary statistics

Frequency of horse race exposure	N	Percentage
Daily	35	14.0%
2-3 times a week	60	24.0%
Once a week	55	22.0%
Once a month	40	16.0%
Less than once a month	44	17.6%
Never	16	6.4%
Utility of horse race journalism		
Not useful	55	22.0%
Somewhat useful	177	70.8%
Very useful	18	7.2%
Trust in polls' predictive power		
Strongly disagree	23	9.2%
Somewhat disagree	67	26.8%
Neither agree nor disagree	44	17.6%
Somewhat agree	99	39.6%
Strongly agree	17	6.8%
Stats education		
Less than a year ago	45	18.0%
1-3 years ago	59	23.6%
3-5 years ago	23	9.2%
6+ years ago	46	18.4%
No stats class	77	30.8%

N=250.

Table 4. Summary statistics: Treatment & Outcome

Treatment	N	Neither (Don't Know)	Somewhat (Likely) Very	Somewhat (Unlikely) Very
Control	90	22	37	5
		(4)	(17)	(1)
			4	-
Bar	80	7	40	4
		(3)	(18)	(1)
			5	2
Dot	80	17	37	7
		(5)	(8)	(1)
			3	2

N=250.

Note that in there is no significant difference in "Don't Know" responses based on treatment assignment in Table 4—there is no evidence that assignment confused respondents' answers, even considering the non-traditional dot plot presentation.

Ordered Probit Model

In analyzing the survey data presented earlier, an ordered probit with the following predictor variables (X_i) was modeled. (For a formal model of the ordered probit, please see the Appendix.)

$$X_1 = \text{level of trust person puts in polls for prediction,}$$

$$X_2 = \text{frequency of exposure to horse race journalism,}$$

$$X_3 = \text{whether person prefers text or graphic presentation,}$$

$$X_4 = \text{whether person has taken a stats class, and}$$

$$X_5 = \text{treatment assignment.}^4$$

The estimates of the probit analysis are found in Table 5, which reports the average marginal effect of changes in each regressor along with its associated White/robust standard error (Fernihough 2013).[5] The probit excluded all responses on the unlikely spectrum as well as all "I Do Not Know" responses for a final model sample size (N) of 215. 80 of those 215 respondents were assigned to the control group by Qualtrics, 70 the bar plot, and 65 the dot plot.

It is important to note that the $\partial Y/\partial X$ for trust in polls (grouped into "Agree," "Neither agree nor disagree," and "Disagree") is based on the change from "Agree." Horse race exposure—grouped into "Never" (which includes all respondents indicating they read horse race journalism less than once a month), "Occasional" (one time a month to one time a week), and "Often" (multiple times a week, including daily)—is based on change from the "Never" group. The treatment effects are changes from the control group.

Holding a skeptical view of polls' capability to predict upcoming elections reduces the probability the respondent reports a higher likelihood of the Movement Party winning the election compared to those who favorably evaluate polls' capabilities, while being neutral does not have a statistically significant difference. Compared to never or rarely reading horse race stories, those who occasionally read such media have a higher probability of evaluating the Movement Party's chances higher by 0.126 (robust standard error of 0.062).

There is no significant difference between those reporting never/rarely reading and those reading horse race stories multiple times a week. There was no significant difference between those preferring graphics to text and those who had taken a stats class before and those who had not when taking into account the pretreatment effects. The only treatment with a significant difference from the control group is the bar plot treatment group with a 0.193 probability increase (robust standard error of 0.059.)

[4] Gender and college major were excluded so as to include as many respondents as possible without restricting the model's power.

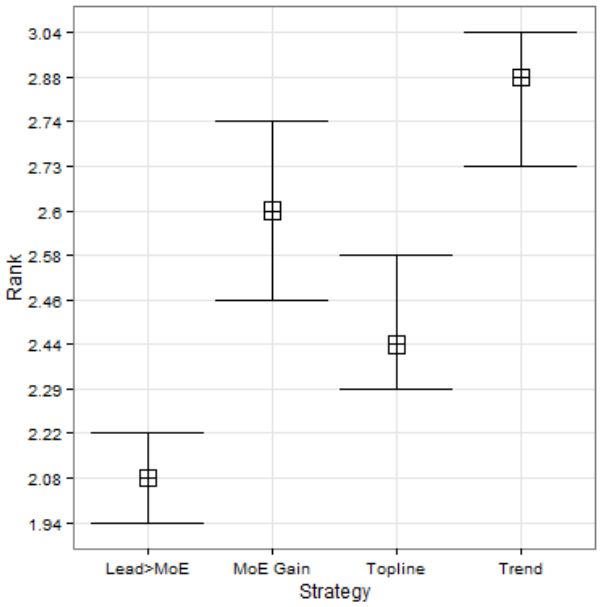
[5] The difference between robust and normal standard errors was trivial. The mean difference has an absolute value of 0.001, and the median difference 0.0006. These indicate good model fit.

Table 5. Average marginal effects

	∂Y/∂X
Disagrees that polls can be trusted	−0.202*
	(0.065)
Neutral about polls' predictive power	−0.029
	(0.084)
Occasional horse race exposure	0.126*
	(0.062)
Often exposed to horse race journalism	0.044
	(0.067)
Text preference	0.033
	(0.059)
Took stats class	−0.044
	(0.058)
Bar treatment (N=70)	0.193*
	(0.051)
Dot treatment (N=65)	0.025
	(0.059)

N=215. Robust standard errors. * = p-value < 0.05.

Figure 4. Ranking communications strategies (95% confidence intervals). Lower = more highly ranked.



Communicating Poll Results

The following section details the ranking of communication strategies as evaluated by the 212 respondents who completed this optional section. The mode order of the strategies' rankings was:

1. The Movement Party's lead (8 points) is larger than the margin of error (5 points).
2. Even if the Opposition gained most of the margin of error, the Movement Party would still be ahead.
3. The Movement Party is at 54%, and the Opposition is at 46%.
4. This poll shows that the Movement Party is gaining a bigger lead compared to a poll last week.

Table 6 summarizes the ranking statistics of the arguments with the mean and standard deviation of the rankings.

Table 6. Summary statistics: Communications Strategies

	Mean	SD
Topline	2.44	1.07
Lead > MoE	2.08	1.05
MoE Gain	2.60	1.05
Trend	2.88	1.16

N=212.

Figure 4 shows the confidence intervals for the rankings of communication strategies.

Discussion

Robustness Checks

The following is a demonstration of the robustness of the specification of the prior ordered probit model found in Table 5 via alternative specifications dropping poll trust, horse race exposure, and both from the models. The results are found in Tables 7, 8, and 9 along with White/robust standard errors. The bar plot treatment remains significant, and the average marginal effect remains positive in all four with an average marginal effect ranging from 0.172 to 0.196. The occasional horse race group remains significant and positively-signed, and disagreeing in the poll trust question remains significant and negatively-signed.

Table 7. Average marginal effects without poll trust

	∂Y/∂X
Occasional horse race exposure	0.144*
	(0.064)
Often exposed to horse race journalism	0.090
	(0.065)
Text preference	0.038
	(0.058)
Took stats class	−0.039
	(0.058)
Bar treatment (N=70)	0.176*
	(0.054)
Dot treatment (N=65)	0.010
	(0.061)

N=215. Robust standard errors. * = p-value < 0.05.

A further check of the surprising finding that having taken a stats class is not significant is to limit the model space to those reporting themselves as having taken a stats class and

Table 8. Average marginal effects without horse race exposure

	$\partial Y/\partial X$
Disagrees that polls can be trusted	−0.207*
	(0.066)
Neutral about polls' predictive power	−0.037
	(0.083)
Text preference	0.024
	(0.058)
Took stats class	−0.039
	(0.058)
Bar treatment (N=70)	0.196*
	(0.051)
Dot treatment (N=65)	0.029
	(0.059)

Table 9. Average marginal effects with neither

	$\partial Y/\partial X$
Text preference	0.037
	(0.060)
Took stats class	−0.026
	(0.059)
Bar treatment (N=70)	0.172*
	(0.055)
Dot treatment (N=65)	0.012
	(0.062)

analyze if there is a time limit to a stats class's impact. In this modeling category, N=153, and respondents were grouped into 3 categories: stats class was less than a year ago (N=39), 1-3 years ago (N=55), and 3+ years ago (N=59). The results are found in Table 10. Note that stats class changes are from the base of the group having taken a stats class in the past year.

The statistical significance of the treatments disappears in this more constrained model, but it is important to remember that this is, to some extent, a result of the sample sizes being more limited in this modeling space. These results ask for more inquiry into the effect of specific skills education in political socialization and understanding of elections by citizens.

Perhaps the statistics finding is not surprising because a statistics course may change a person's view on the validity of polling or a person's propensity to consume horse race journalism. The data in this sample do not support that view. See Tables 11 and 12 for two-way tables illustrating the absence of these phenomena—both have χ^2 p-values very much exceeding the standard α of 0.05. On the other hand, there does seem to be a relationship between poll trust and horse race exposure frequency (see Table 13).

Table 7 (the model without poll trust) and Table 10 (the model limited to those having taken a stats class) suggest robustness of the "Occasional" horse race exposure group result. In both specifications as well as the original model, "Occasional" is both significant and positively-signed. The average marginal effect is at a much greater magnitude than "Often" and "Never" in all three was well. *However*, these

Table 10. Average marginal effects; model limited to stats class participants

	$\partial Y/\partial X$
Horse race exposure (Occasional)	0.209*
	(0.070)
Horse race exposure (Often)	0.152*
	(0.075)
Poll trust (Disagree)	−0.213*
	(0.074)
Poll trust (Neither agree nor disagree)	0.154*
	(0.073)
Text preference	0.081
	(0.066)
Stats class (1-3 years ago)	−0.041
	(0.086)
Stats class (3+ years ago)	−0.007
	(0.089)
Bar Plot Treatment (N=48)	0.096
	(0.069)
Dot Plot Treatment (N=52)	−0.040
	(0.074)

N=153. Robust standard errors. * = p-value < 0.05.

Table 11. Stats class and poll trust

Stats class	Poll trust		
	Agree	Disagree	Neither
No	27	26	11
Yes	77	50	24

χ^2 p-value = 0.476.

Table 12. Stats class and horse race exposure

Stats class	Horse race exposure		
	Never	Occasionally	Often
No	19	23	22
Yes	31	61	59

χ^2 p-value = 0.348.

Table 13. Poll trust and horse race exposure

Poll trust	Horse race exposure		
	Never	Occasionally	Often
Agree	13	39	52
Disagree	23	31	22
Neither	14	14	7

χ^2 p-value = 0.001.

results are not found in a Bayesian approach to the ordered probit (see Appendix Table 1), though the other significant effect sizes are. Given the failure of the "Occasional" effect size to replicate in the Bayesian ordered probit, one cannot be confident that the effect is robust, though the traditional ordered probits raise the question of how exposure to horse race journalism affects interpretation and action upon poll results.

Insight into Poll Comprehension

The experimental evidence analyzed in this paper give the field insight into how individuals interpret polls in a near-vacuum of information relating to the race at hand. Therefore, inferences from this paper may fit better into a Myatt-style individual analysis of the singular, somewhat isolated voter rather than a social network or group analysis of reaction to polls as individuals may influence others' perceptions of either the mathematical validity of an interpretation of a poll or the implication the poll has outside of its statistical validity. The experiment also, by design, fails to allow for motivated or partisan reasoning, which can play a role in mediating or altering the inferences below.

The main inference from the data is that the major focus for readers of poll results seems to be the relationship between the difference in point estimates between candidates and the margin of error presented. Thus a typical citizen's perception or interpretation of a poll result with regard to some choice or candidate (φ) is some estimation function, Ξ, based on the difference between the point estimates ($p_1 - p_2 = d$), the margin of error (ζ), and the person's prior belief about the strength of the poll or polls generally (ω):

$$\text{Likelihood}_{\varphi} \sim \Xi(d, \zeta, \omega).^{6}$$

This line of thought is supported by the statistically significant difference between the ranking of the argument that "[t]he Movement Party's lead (8 points) is larger than the margin of error (5 points)" and the ranking of all three of the other communications strategies as visualized by confidence intervals in Figure 4. More evidence to this interpretation is the general ranking of the arguments presented in Table 6. The margin of error argument ("Even if the Opposition gained most of the margin of error, the Movement Party would still be ahead") had a mode rank of second, and the argument reiterating the poll result (point estimates) had a mode rank of third. The latter argument ("The Movement Party is at 54%, and the Opposition is at 46%") was ranked as more persuasive than the trend argument (that this poll's difference between parties was larger than a poll's last week, echoing van der Meer, Hakhverdian, and Aaldering (2015)) to a statistically significant degree (see Figure 4).

[6] One may even expect a stronger form where Likelihood $\sim \Xi(d|\zeta, \omega)$. Note that these are functions for the difference between two point estimates. It may be a special case where d is a vector of estimate differences. It is almost certain that Ξ changes depending on the number of choices in the poll and the electoral system. Further study is needed.

This explanation provides solid analytical grounding for understanding the impact of the bar plot on assigned likelihood (supporting Hypothesis 1a). The bar plot may be hindering cognitive comparisons of the point estimate difference to the margin of error. Imposing the need to perform mathematical calculations on top of the bar plot, even with the margin of error given in text literally adjacent to the bar plot, seems to allow, encourage, or lead respondents to overstate the likelihood of the poll winner winning an upcoming election. In a similar vein, the dot plot with the embedded margins of error had a negligible impact on assigned likelihood when compared to the control, giving support for Hypothesis 1b. The exact pathway cannot be determined from these data, calling for more political psychology research on the cognitive processing of mathematical or polling data. Likewise, the exact nature of Ξ cannot be determined without further research.

Keeping in mind the limitations present in creating the theoretical framework around Ξ, this study shows a potential need to re-evaluate media approaches to graphical representations of polling data. Even seemingly trivial notions such as the type of graph may alter a voter's Ξ function in ways that are hard to perceive. The simple presence of a bar plot—even controlling for poll trust, horse race exposure frequency, presentation preference, and statistics education—seems to be responsible for a large upwards shift in the respondents' likelihood evaluation in favor of the poll leader. This was true even though a contextualizing story was present right next to the graph. Bar graphs could be changing of the dynamics of an election if the media source disseminating these graphs has a large circulation within the voting electorate. A more cautious and helpful approach may be to use a dot plot similar to the one used in the experiment (see Figure 3.) However, replication of this study is needed before this is a substantive and well-supported recommendation.

The importance of general poll trust in Ξ comes from the data generated by the experiment and was not hypothesized to as large of an effect as modeled earlier. It is possible that poll trust is a secondary consideration when evaluating the likelihood—that is, likelihood estimation could be a two-step process in which a citizen evaluates the poll on its own using d and ζ and then weights the poll as a piece of evidence towards the election outcome as a part of the citizen's idea of the value of polling generally (ω). This two-stage processing may only occur for people who distrust polls. It is entirely plausible that those trusting polls to predict outcomes do not even consider this second-stage poll weighting process. If this is correct, differences in poll trust may be creating a bifurcated audience for polls and those communicating about them (media sources, campaigns, party offices, etc.). One part of the population may be primed to infer election outcomes from polling while the other is inherently skeptical even before motivated or wishful reasoning.

The data found from this experiment seem to reject Hypothesis 2, that the effect of poll presentation would be me-

diated by statistics education. Statistics classes do no seem to have an appreciable impact on likelihood *ceteris paribus*. This holds true even for those having taken the class recently. There also does not seem to be an indirect path from stats classes to change in poll trust or horse race exposure. Among those having taken a stats class in the sample, 77 agreed or strongly agreed that polls of randomly selected voters could be trusted to predict election outcomes, 50 disagreed or strongly disagreed, and 24 responded "neither." This does not significantly differ from the group of respondents not taking a stats class as seen in Table 11. The divide could be originating with exposure to horse race journalism. There was a statistically significant difference found in poll trust when considering how respondents evaluated the frequency of their reading horse race journalism (see Table 13). More study and research is needed to investigate the source of poll distrust and its role in poll evaluation.

Future Research

The most obvious direction of research would be to replicate this survey design in a truly random and representative sample. The main advantage would be to increase heterogeneity of education level, which could not be analyzed with any meaningful power in this study. A related change would be additional power in age/demographic cohort analysis, which might be helpful in analyzing generational perceptions of polls as statistical knowledge may become less salient for everyday living as one ages.

One exciting possible extension of this direction in political communications research would be to see the impact of polling visualization in the context of fundraising within political campaigns. A simple version of this experiment would randomly select a third of a large fundraising email list to receive a traditional horse race poll-version email in text, one third would receive the same with a bar plot, and the other third would receive the dot plot. The outcome variable of dollars raised would be a useful metric to analyze how political contributors process and act upon polling information contingent upon its framing. The continuous nature of the outcome would also help increase interpretability and salience of the result as well (Jackman 2000, 4-5). Another advantage of this experiment would be to drop the experimental effect context as participants would not be directly informed of the protocol and would probably respond more naturally.

A direct result of the communications portion of the survey is a need for research into how the public perceives trends in polling data and public opinion. This implies inquiry into how multiple polls are perceived in experiments similar to the design implemented and analyzed here. One possible design would involve exposing participants to time series of polls showing an increasing polling trend for some candidate with the object of trying to understand how citizens comprehend—are polls a series of snapshots or just a path to the current circumstance? The suggestion of this experiment is that citizens care more about polls as indications of current circumstance rather than as a signal of a trend or movement. This, however, may seem very susceptible to motivated reasoning as poll trends over time have been used to demonstrate changing public opinion towards same-sex marriage.

Beyond a time series trend, another key question of concern is how do citizens interpret conflicting or slightly contradictory polls. "Conflicting polls" can be interpreted as both situations in which one poll says Candidate X is winning and one says Y and as well as situations where some polls say Candidate Y is leading despite other polls putting Candidate X in first. The experiment detailed here has little to say on this question other than that there may be a strong cognitive bias to overstate confidence based on one poll. Both "trend" ideas may be exploited with a narrative understanding of the trend, as Berinsky and Kinder (2006) demonstrate with frame research surrounding news coverage of the Kosovo crises.

The culmination of these hypothetical experimental situations would be social network analysis within which actors strategically engage with polls and re-frame and disseminate the results of polls to achieve their own ends.

References

Aalberg, Toril, Jesper Strömbäck, and Claes H. de Vreese. 2012. "The Framing of Politics as Strategy and Game: A Review of Concepts, Operationalizations and Key Findings." *Journalism* 13(2): 162–78.

Andersen, Robert. 2000. "Reporting Public Opinion Polls: The Media and the 1997 Canadian Election." *International Journal of Public Opinion Research* 12(3): 285–98.

Berinsky, Adam J., and Donald R. Kinder. 2006. "Making Sense of Issues Through Media Frames: Understanding the Kosovo Crisis." *The Journal of Politics* 68: 640–56.

Bhatti, Yosef, and Rasmus Tue Pedersen. 2015. "News Reporting of Opinion Polls: Journalism and Statistical Noise." *International Journal of Public Opinion Research* Advance Access.

Brettschneider, Frank. 2008. "The News Media's Use of Opinion Polls." In *The SAGE Handbook of Public Opinion Research*. London: SAGE Publications Ltd.

Druckman, James N., and Cindy D. Kam. 2009. "Students as Experimental Participants: A Defense of the 'Narrow Data Base'." `http://www.ipr.northwestern.edu/publications/docs/workingpapers/2009/IPR-WP-09-05.pdf`.

Fernihough, Alan. 2013. *Package 'mfx'*.

Franklin, Charles H. 2002. "The 'Margin of Error' for Differences in Polls." `https://abcnews.go.com/images/PollingUnit/MOEFranklin.pdf`.

Frankovic, Kathleen A. 2005. "Reporting The Polls in 2004." *Public Opinion Quarterly* 69(5): 682–97.

Gerber, Alan S., and Donald P. Green. 2008. "Field Experiments and Natural Experiments." In *The Oxford Handbook of Political Methodology*. Oxford, U.K.: Oxford University Press.

Goodrich, Ben, and Ying Lu. 2007. "oprobit.bayes: Bayesian

Ordered Probit Regression." In *Zelig: Everyones Statistical Software*, ed. Kosuke Imai, Gary King, and Olivia Lau.

Hall, Andrew B., and James M. Snyder. 2015. "Information and Wasted Votes: A Study of U.S. Primary Elections." https://dl.dropboxusercontent.com/u/11481940/Hall_Snyder_Strategic_Voting.pdf.

Hardmeier, Sibylle. 1999. "Political Poll Reporting in Swiss Print Media: Analysis and Suggestions for Quality Improvement." *International Journal of Public Opinion Research* 11(3): 257–74.

Hollander, Barry A. 1993. "Unintended Effect: Persuasion by the Graphic Presentation of Public Opinion Poll Results.".

Jackman, Simon. 2000. "Models for Ordered Outcomes." http://web.stanford.edu/class/polisci203/ordered.pdf.

Loewen, Peter, Arthur Lupia, and Daniel Rubenson. 2015. "What the Canadian and British election polls tell us about Donald Trump." *The Washington Post*, Monkey Cage.

Murdoch, Rupert. 2015. "Tweet by Rupert Murdoch." https://twitter.com/rupertmurdoch/status/650399630463791104.

Mutz, Diana C. 1995. "Effects of Horse-Race Coverage on Campaign Coffers: Strategic Contributing in Presidential Primaries." *The Journal of Politics* 57(4): 1015–42.

Myatt, David P. 2007. "On the Theory of Strategic Voting." *The Review of Economic Studies* 74(1): 255–81.

Nisbet, Matthew C. 2008. "Horse Race Journalism." In *Encyclopedia of Survey Research Methods*, ed. Paul J. Lavrakas. Thousand Oaks, CA: Sage Publications, Inc.

O'Halloran, Sharyn. 2005. "Econometrics II." http://www.columbia.edu/~so33/SusDev/SusDev.htm.

Patterson, Thomas E. 1977. "The 1976 Horserace." *The Wilson Quarterly* 1(3): 73–9.

Patterson, Thomas E. 2005. "Of Polls, Mountains: U.S. Journalists and Their Use of Election Surveys." *Public Opinion Quarterly* 69(5): 716–24.

Pew Research Center. 2015. "Coverage Error in Internet Surveys." http://www.pewresearch.org/2015/09/22/coverage-error-in-internet-surveys/.

R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rosenstiel, Tom. 2005. "Political Polling and the New Media Culture: A Case of More Being Less." *Public Opinion Quarterly* 69(5): 698–715.

Sears, David O. 1986. "College Sophomores in the Laboratory: Influences of a Narrow Data base on Social Psychologys View of Human Nature." *Journal of Personality and Social Psychology* 51: 515–30.

Shepard, Steven. 2015. "Gallup gives up the horserace." http://www.politico.com/story/2015/10/gallup-poll-2016-pollsters-214493.

Smith III, Ted J., and Derek O. Verrall. 1985. "A Critical Analysis of Australian Television Coverage of Election Opinion Polls." *Public Opinion Quarterly* 49: 48–79.

Traugott, Michael W. 2005. "The Accuracy of the National Preelection Polls in the 2004 Presidential Election." *Public Opinion Quarterly* 69(5): 642–54.

van der Meer, Tom W.G., Armen Hakhverdian, and Loes Aaldering. 2015. "Off the Fence, Onto the Bandwagon? A Large-Scale Survey Experiment on Effect of Real-Life Poll Outcomes on Subsequent Vote Intentions." *International Journal of Public Opinion Research* Advance Access.

Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman. 2014. "Forecasting elections with non-representative polls." *International Journal of Forecasting* 31(3): 980–91.

Weimann, Gabriel. 1990. "The Obsession to Forecast: Preelection polls in the Israeli Press." *Public Opinion Quarterly* 54(3): 396–408.

Acknowledgments

I would like to thank Dr. Michael MacKuen and Brice D.L. Acree for their guidance and assistance in the survey design and quantitative analysis methods. Thomas Hunold and Sara Skutch were also helpful in survey design and copy-editing. Also of invaluable assistance was R (R Core Team 2015).

Appendix

The Ordered Probit

Borrowing terminology from O'Halloran (2005) whereby Y' is the transformation of the ordered outcome variable Y, X is the vector of observed predictor variables X_i, β is the vector of parameters to be estimated, and ε is a normally distributed stochastic error term, a general probit model is written as $Y' = X\beta + \varepsilon$.

It is important to note that Y' is a transformed latent continuous representation of an unobserved $Y*$, which would be a real number representation of the probability the participant would assign to the Movement Party winning the election. These categories are bounded by "cut-points" within the probability space of the likelihood categories with the latent real number represented by ϑ. The interval can be represented by the following inequalities where τ_i is one such cut-point:

$$Y = \text{Neither Likely Nor Unlikely, as } \tau_1 \leq \Pr(\vartheta) \leq \tau_2,$$

$$Y = \text{Somewhat Likely, as } \tau_2 \leq \Pr(\vartheta) \leq \tau_3, \text{ and}$$

$$Y = \text{Likely/Very Likely, as } \tau_3 \leq \Pr(\vartheta) \leq 1.$$

(This can be extended to include the unlikely responses, but such responses were excluded in the modeling.) The

probability of observing Y equal to the category j is represented (in terminology slightly modified from Goodrich and Lu (2007)) by Φ, the Normal cumulative distribution function with mean μ and variance 1:

$$\Pr(\text{Y = j}) = \Phi(\tau_j|\mu) - \Phi(\tau_{j-1}|\mu) \text{ for} j = 2, ..., J.$$

The ordered probit model seeks to estimate these cut-points via the maximum-likelihood method.

A Bayesian approach

A Bayesian version of the ordered probit considered earlier was fitted to test the earlier ordered probit models. Table 1 shows the mean, standard deviation, and 2.5% and 97.5% quantiles for the marginal posterior distributions using the same predictors as in Table 5. The model passes the Heidelberg (stationary and halfwidth) and Raftery diagnostic tests for convergence. Note that the marginal posterior distribution quantiles for both bar plot treatment and poll distrust are of the same sign as in the traditional model, giving more evidence to their significance. However, this does not apply to occasional exposure, which is not positive in both quantiles. One must keep in mind that these are parameter estimates, so they are not directly comparable to the tables containing the average marginal effects found in the Results section.

Table 1. Bayesian ordered probit summary

Predictor	Mean (SD)	2.5%	97.5%
(Intercept)	0.981 (0.254)	0.492	1.486
Occasional exposure	0.250 (0.210)	−0.162	0.657
Exposed often	0.107 (0.218)	−0.329	0.531
Poll distrust	−0.559 (0.179)	−0.913	−0.209
Poll neutrality	−0.365 (0.229)	−0.807	0.083
Text preference	0.157 (0.176)	−0.190	0.506
Took stats class	−0.277 (0.175)	−0.622	0.070
Bar plot	0.523 (0.189)	0.150	0.897
Dot plot	−0.789 (0.192)	−0.452	0.296