Year: 2020

Learning from polls

Leemann, Lucas ; Stoetzer, Lukas ; Traunmueller, Richard

Originally published at:
Leemann, Lucas; Stoetzer, Lukas; Traunmueller, Richard (2020). Learning from polls. In: EPSA Annual Conference, Virtual, 18 June 2020 - 19 June 2020, EPSA.

Learning From Polls*

Lucas Leemann[†], Lukas F. Stoetzer[§], Richard Traunmueller[‡]

June 17, 2019

Abstract

Voters' expectations of party strengths are a central part of many foundational political science theories that posit a strategic act by the voter. But how do voters develop these beliefs and how is this belief formation affected by polling reports? In this article, we present a dynamic Bayesian learning model that serves as a baseline for how beliefs are formed. This also allows us to infer how and when belief formation deviates from the theoretical ideal. We validate the model and illustrate its potential based on a number of experiments conducted on MTurk. We find encouraging results in this pilot study that validate the baseline model.

*We thank Marco Steenbergen for valuable comments on the paper. An earlier version was presented at the workshop of the politial methdology chair of the University of Zurich at Villa Garbald, 2019. We thank Lucien Baumgartner for impeccable research assistance. Authors in alphabetical order.

[†]Department of Political Science, University of Zurich, Switzerland. URL: http://www.lucasleemann.ch, Email: leemann@ipz.uzh.ch.

[§]Department of Political Science, University of Zurich, Switzerland. URL: http://lukas-stoetzer.org//, Email: lukas.stoetzer@uzh.ch.

[‡]Department of Political Science, University of Frankfurt, Germany. URL: http://www.richardtraunmueller.com, Email: traunmueller@soz.uni-frankfurt.de.

1 Introduction

Polls are important sources of political information that shape voters' beliefs about uncertain political outcomes: the support for parties, candidates, and policies or the likely results of elections and referendums. Given the fundamental role of voters' expectations in many theories of political science, it is important to understand how citizens process polling information and if they do so in an optimal way. The recurring ritual of criticizing polls for failing to predict political outcomes, even if polls perform as expected (Jennings and Wlezien, 2018), points to the possibility of biased or motivated forms of information processing (e.g., Bartels, 2002; Taber and Lodge, 2006). To assess this possibility we need a benchmark of what an 'optimal' processing of polling results would look like.

In this paper we propose a Bayesian learning model of how voters (should) learn from poll results. Bayes' Rule is widely regarded as the ideal procedure for learning from political information (e.g. Gerber and Green, 1999; Bartels, 2002; Bullock, 2009; Sinclair and Plott, 2012; Hill, 2017): current beliefs are a weighted combination of prior beliefs and new evidence. This model therefore allows us to evaluate and eventually explain how voters integrate evidence provided by polls and their previously held beliefs. If voters' behavior indeed matches the model, they can be said to process the political information contained in polls in a rational way and the model at hand provides good explanation of how they do it. Any divergence from the model would still yield important information because this would indicate that voters' either have cognitive limitations or process the polling results in a politically motivated way. Next to a principled way of forming expectations from poll results and previously held beliefs, the model also provides a set of key quantities which can be used to further characterize this learning process: the rate of adaptation to new evidence, the subjective credibility of the poll, and the discount rate or 'stickiness' of prior beliefs.

Using a crowd-sourced online experiment, we randomly instill different prior beliefs and then sequentially present respondents with three sets of changing polling results from a hypothetical two-party electoral race. The evidence from this first experiment suggests that Bayesian learning provides a reasonable description of how voters adjust their expectations about the electoral race to new polling information. We find that while different prior beliefs matter, respondents very quickly adapt to new evidence and tend to converge in their expectations (yet not completely). In fact, we find that they give more weight to new evidence than optimal updating based on Bayes' Rule would warrant. This suggests that voters are indeed 'hasty' Bayesians that move too quickly and not 'cautious' Bayesians as suggested by Hill (2017).

2 Bayesian Learning from Poll Results

Voters who are interested in matters of politics need to learn from available information. During electoral campaigns, polls are the most readily available source to form beliefs about the support of political candidates. To understand how polls influence voters' beliefs, they can be represented by a probability distribution. From a research perspective those probability distribution encode both the expectation a voter has about the race in terms of the mean and how certain she is about this conclusion in terms of the variance.

A candidate's support at a particular point during the campaign, denoted as θ_t, can theoretically range from 0% to 100%. Voter i's belief about it is represented by the probability distribution $p_i(\theta_t)$. As the belief and thereby the distribution differs between individuals, the distribution is denoted with i. The question we set out to answer is how a particular poll y_t that publishes a vote share for the candidate influences those beliefs.

Rational voters can use Bayes rule to update their beliefs about a candidate's support. After observing the poll each voter can change her belief by conditioning the probability distribution on the evidence from the poll $p_i(\theta_t|y_t)$. Bayes rule states this learning process as a function of marginal and conditional distributions:

$$p_i(\theta_t|y_t) = \frac{p_i(y_t|\theta_t)p_i(\theta_t)}{p_i(y_t)}. \tag{1}$$

In essence, the rule specifies how a voter can revise her belief about the candidate's vote share before observing the poll $p_i(\theta_t)$ - which is commonly refereed to as the *prior belief*. For this, each voter estimates a subjective *likelihood* of observing the poll's result $p_i(y_t|\theta_t)$. It reflects a distribution which – in the eye of the voter – is likely to have generated the poll. A poll that sees a candidate e.g. at 55% is more likely to result from a population where 55% instead of 40% support the candidate. $p(y_t)$ is the probability of the poll that normalizes the product of the likelihood and the prior.[1] The updated conditional distribution $p_i(\theta_t|y_t)$ is also refereed to as the *posterior belief*.

2.1 Quantities of interest

Different quantities of the learning process are of interest for researchers. Central is the question how closely voters will adapt their expectation to the polls. We define this as the *rate of adaption*. In particular we are interested how strongly the prior expectation is shifted in direction of the new evidence. This is sometimes referred to as the delta rule (see e.g. Nassar et al., 2010). If we denote the expectation of the *posterior belief* as $\mu_{it} = \mathbb{E}[p_i(\theta_t|y_t)]$ we can write the adaption equation as

$$\mu_{it} = \mu_{it-1} + \delta_{it}(y_t - \mu_{it-1}). \tag{2}$$

In this, the posterior expectation is equal to the prior expectation plus the adaptation rate δ_t times the difference between the new evidence and the prior expectation. The adaptation rate is defined between 0 and 1. 0 means no adaption to the new poll and 1 implies perfect adaption.

In the Bayesian learning model the rate of adaption depends on a number of aspects. First, how informative a voter perceives the poll to be about the vote share. Relating survey shares to the support in the population comes with several sources of uncertainty. Most relevant is the fact that only a limited amount of people are asked about their vote

[1] It can be calculated by integrating over the possible vote shares for the candidate $p_i(y_t) = \int_0^1 p_i(y_t|\theta_t)p_i(\theta_t)d\theta_t$.

intention, leaving room for sampling error. The true support can fall within a margin of error of the poll result. Second, recipients of the poll might consider alternative factors that lower or increase the information value. For example, they consider the poll as a forecasting device which means that next to sampling error they expect a forecast error for the final support. But they could also judge the poll to be more precise in estimating the true support and mistake the survey as a nearly complete census. All those aspects can be encoded in the *subjective standard deviation of the poll* - how voters think that the true support deviates from the poll result. We can calculate this as the standard deviation of the individual likelihood:

$$SD_i[y_t] = \sqrt{\mathbb{VAR}[p_i(y_t|\theta_t)]}. \tag{3}$$

The subjective notation can be compared to the objective standard error of poll. Sampling theory states that under random sampling the standard error of a proportion is equal to: $\sqrt{\frac{y_t(1-y_t)}{N}}$. In the latter analysis, it is insightful to compare this value to the *subjective standard deviation of the poll.*

Second, the rate of adaption depends on how strongly voters carry over their beliefs. Over an electoral campaign multiple polls will be published, a situation a Bayesian learning model can accommodate. In essence the *posterior belief* from the past period can form the new *prior belief* that are to be revised in light of the evidence. With this the question arises how beliefs evolve:

$$p_i(\theta_t|\theta_{t-1}) \tag{4}$$

Different specification of the process can be used to analyze the evolution process. In the application here we employ a power discount model (Smith, 1979), which we describe below. An alternative process could be a random-walk as employed in dynamic linear learning models.

2.2 Parametric model of Bayesian Learning from Polls

The analysis of the learning process requires certain assumptions about the probability distributions that represent voters' beliefs. We choose a set of flexible distributions that permit for a conjugate learning process: A Beta-Binomial model.[2] This model is well-suited for our purpose as it permits us to analyze the rate of adaption, the subjective standard deviation of the poll and the evolution of beliefs.

Again y_t defines the share of support for candidate A reported in poll at time t. The poll is based on a survey of N_t respondents. We assume that the Likelihood function voters have in mind when evaluating the poll is a binomial distribution $p(y_t N_t|\theta_t)$, where the count of respondents that support that candidate is the product of the poll's share and the number of survey respondents. In the current form we assumes that all voters employ the same likelihood when evaluating the poll. Hence, no i subscript.

$$p(y_t N_t|\theta_t) = \binom{N_t}{y_t N_t} \theta_t^{y_t N_t} (1 - \theta_t)^{N_t - y_t N_t} \tag{5}$$

We further assume that beliefs are beta-distributed. The posterior belief from the last update $p(\theta_{t-1}|y_{t-1})$ are distributed with $\alpha_{it-1}, \beta_{it-1}$. Those shape parameters of the beta distribution guide an individual's belief about the support for the candidate:

$$p_i(\theta_{t-1}|y_{t-1}) = \frac{\Gamma(\alpha_{it-1} + \beta_{it-1})}{\Gamma(\alpha_{it-1})\Gamma(\beta_{it-1})} \theta_{t-1}^{\alpha_{it-1}-1} (1 - \theta_{t-1})^{\beta_{it-1}-1} \tag{6}$$

The beliefs are carried over to form the new prior beliefs power discount model (Smith, 1979):

$$p_i(\theta_t) = p_i(\theta_{t-1}|y_{t-1})^d, \tag{7}$$

where the discount parameter d specifies how much of the new prior depends on the

[2] An alternative would be a dynamic linear learning model. However, the bounds of the shares are better suited to a beta model. In addition, the beta is more flexible in that it permits for non-symmetric beliefs. A more general form might be found when employing a dynamic generalized linear model, at the cost of conjugacy.

past posterior belief. It ranges from 0 to 1, where one means that the beliefs carry over and 0 that the posterior is not taken into consideration for the new period. Again we assume a common discount parameter for all learners.

Posterior beliefs $p_i(\theta_t|y_t)$ are formed according to Bayes rule and result in conjugate beta-distributed beliefs:

$$p_i(\theta_t|y_t) = \frac{\Gamma(\alpha_{it} + \beta_{it})}{\Gamma(\alpha_{it})\Gamma(\beta_{it})}\theta_t^{\alpha_{it}-1}(1-\theta_t)^{\beta_{it}-1} \tag{8}$$

where $\alpha_{it} = d\alpha_{it-1} + y_t N_t$ and $\beta_{it} = d\beta_{it-1} + (1 - y_t)N_t$ are both a function of the discount parameter, the poll and the prior beliefs. Hence all those aspects shape a voter's learning process. To see this more clearly, we can formulate the posterior expectation.

$$\mu_{it} = \frac{d\alpha_{it-1} + y_t N_t}{d(\alpha_{it-1} + \beta_{it-1}) + N_t} \tag{9}$$

If the beliefs do not carry over, and $d = 0$, than the mean expectations are equal to the poll's reported share y_t. If the survey is relatively small $\lim_{N\to 0}$, the posterior expectation is dominated by the prior belief expectation as the discount factor cancels out: $\frac{\alpha_{t-1}}{\alpha_{t-1}+\beta_{t-1}}$. We can also study the rate of adaption, which yields similar results. Substituting the the expectations in the equation 2 and solving for δ_t gives:

$$\delta_{it} = \frac{(y_t\alpha_{it-1} + y_t\beta_{it-1} - \alpha_{it-1})N_t}{(y_t\alpha_{it-1} + y_t\beta_{it-1} - \alpha_{it-1})(d(\alpha_{it-1} + \beta_{it-1}) + N_t)}. \tag{10}$$

Again, if $d = 0$ the rate of adaption is 1 and the expectations are perfectly adjusted to the evidence. If the poll is small the rate of adaption tends to towards zero $\lim_{N\to 0} \delta_{it} = 0$.[3]

[3] At the present state the model does not allow for varying subjective standard deviation of the poll. As the number of respondents if fixed, the variance of the likelihood is the same for all recipients of the poll $\mathbb{VAR}[p(y_t N_t|\theta_t)] = N_t\theta_t(1-\theta_t)$. An alternative could be to allow for individual perceptions of the sample size by adding a positive constant to the equation: $y_t N_t c_i$. It could be interpreted as scaling the true sample size up (if above 1) and down if between 0 and 1. This addition would end up in the calculation of the standard deviation of the poll $\mathbb{VAR}[p_i(y_t|\theta_t)] = c_i\theta_t(1-\theta_t)$

3 Experimental Set-up

For the purpose of data collection we relied on the crowdsourcing platform Amazon Mechanical Turk (MTurk). The main advantages of crowd-sourced experiments are the relative low cost, the short time needed to arrive at the required responses and the overall easy handling. Early on researchers asked whether MTurk can be used to produces adequate samples and found that it performs quite well when compared to more established internet surveys (Berinsky et al., 2012; Mason and Suri, 2012; Thomas and Clifford, 2017).

Recent research shows that treatment effects and treatment heterogeneity can be estimated adequately in MTurk samples. Coppock (2018) replicates fifteen survey experiments and compares the estimates based on random samples with estimates based on a MTurk sample. In general, the two sets of estimates overlap. MTurk samples promise to provide a fast and cost-efficient way to cary out online experiments.

We carried out an initial pilot survey with 300 participants. We follow standard practice and recruit workers on MTurk with an approval rating of more than 97% and more than 5000 HIT submissions (but see also Robinson et al., 2019). The median time for taking the survey was less than 6 minutes and we payed participants 1.10$ to pay the California minimum wage. We also include an attention check at the end asking for the color combinations of the plots. In the analyses presented here, we rely on the raw data and do not apply any weights following the recommendation of Miratrix et al. (2018).

Before describing the structure of the survey in detail, we turn to a fundamental element of this experiment. How can we elicit respondents's beliefs? It does not suffice to ask people what they think is the most likely vote share since that would only provide the expected value of the distribution (Winkler, 1967) . Throughout the experiment we rely on a set of questions from Manski (2009). This choice is based on a recent evaluation study that compares six different prior elicitation methods and shows that Manski's set of questions is best suited to elicit prior beliefs (Leemann et al., 2019). The

Table 1: Manski Prior Elicitation

Question	Response
What is the most likely vote share of party A? Please give your response in percentage points.	b_i^m
What do you think is a likely range of the vote share that party A will receive? Please indicate the lower bound in percentage points.	b_i^l
Now, please indicate the upper bound in percentage points.	b_i^u
What is the probability that party A will get a vote share of less than b_i^l percent?	p_i^l
What is the probability that party A will get a vote share of more than b_i^u percent?	p_i^u

Manski questions, shown below in Table 1, is the only elicitation technique that across all conditions was able to uncover the objective distribution.

The survey is set in a hypothetical election where party A and party B are competing with each other. We ask after information updates the same question in the same format. The first question elicits the expected value, while questions 2 to 5 allow to infer the variance in an individual belief (Table 1).

Respondents are asked to assess the winning chances of party A in district D. To start, we need to instill a prior belief before giving respondents the first election poll. To do so, we present respondents with 100 election outcomes of districts that are similar to district D (Goldstein and Rothschild, 2014). We do so by relying on a GIF that quickly shows 100 outcomes – Figure 1 shows four still frames from the GIF.

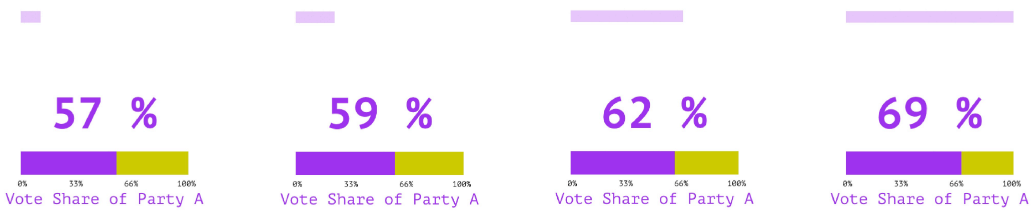

Figure 1: Four still frames from the GIF. Each still frame is shown for about half a second. The purple bar at the top is a progress bar.

The GIF exists in two versions. A first version is symmetric and follows a Beta distribution with mean at 50% ($\mathcal{B}(60, 60)$). The second version is an asymmetric Beta distribution with mean at 67% ($\mathcal{B}(60, 30)$). The examples above are from the asymmetric distribution.

Following the display of of these 100 hundred election outcomes respondents are asked what their expectation is for a similar district. They are then asked to asked what they tHink party A's vote share will be and presented with the five Manski questions. This allows us to determine their prior before showing them the first polling results. Figure 2 shows how polling results are communicated during the experiment.

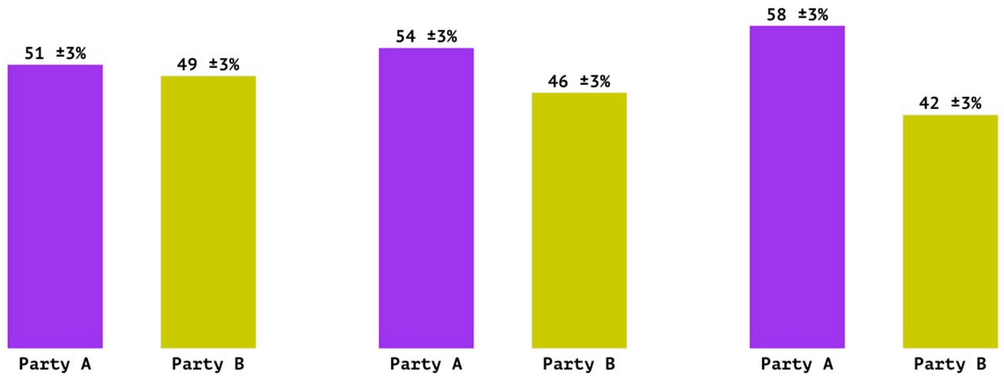

Figure 2: Screenshot the three possible information treatments. Respondents are presented with one plot with polling information in each round.

After showing the respondents a plot based on a poll they are again presented with the question of how large party A's vote will be and provided the five Manski questions. After this they receive two more times a new poll and after each poll they are asked to respond to the five Manski questions.

The following illustration shows the treatments and the sequence during the experiment.

T_1: Respondents are shown GIF with 100 election results.

M_1: Respondents' beliefs are measured.

T_2: Respondents are shown an election poll.

M_2: Respondents' beliefs are measured.

T_3: Respondents are shown an election poll.

M_3: Respondents' beliefs are measured.

T_4: Respondents are shown an election poll.

M_4: Respondents' beliefs are measured.

T_1: Respondents are shown the symmetric distribution showing a close race (50:50) or an asymmetric distribution showing a one-sided race (67:33). Randomization 1.

T_2: Half the respondents receive a poll showing 58:42, the other half receives 51:49. Randomization 2.

T_3: All respondents receive a poll showing 54:46.

T_4: Half the respondents receive a poll showing 51:49, the other half receives 58:42.

The polls are hence either starting high and falling or they start low and increase. Randomization 1 and 2 are independent of each other and this ensures that some people receive the first polling information that is close to their prior while others receive information that is further away. In Table 2 we show the randomization check across a number of variables and conditions.

Table 2: Balance/Randomization Check

	symmetric	reverse	N	female	age	university	polint
1	Non-symmetric Priors	Increasing Polls	91	0.50	41.40	0.66	2.72
2	Non-symmetric Priors	Decreasing Polls	70	0.40	43.80	0.46	2.82
3	Symmetric Priors	Increasing Polls	81	0.55	39.64	0.57	2.75
4	Symmetric Priors	Decreasing Polls	100	0.45	41.97	0.64	2.75

In the next section we describe how we estimate the parameters of interest. We then discuss the empirical results of the pilot study and show how close respondents' behavior is to the model.

4 Statistical model using elicited beliefs

In this section we describe a method to estimate the parameters from the parametric model above from our experiment. In the experimental set-up we measure a set of respondents' $i \in (1, \ldots, R)$ belief about the vote share of a candidate, every time $t \in (1, \ldots, T)$ we present them with a new poll result. We initial the prior beliefs using sequences of election results (see Figure 1). The prior beliefs are denoted by $t = 0$. The beliefs are measured using Manski's questions. We use the notation described in Table 1: b^m_{it} denotes the expected value, b^l_{it} the lower bound, b^u_{it} the upper bound, p^l_{it} the probability to observe a value below the lower bound, and p^u_{it} the probability for the upper bound. We use the data to estimate the parameters of the model to approximate how the respondents update their beliefs.

For this, we assume that the observed values for the mean, the lower and upper bound are measured with normal distributed measurement error for each respondent at each time point. To keep the measurement model simple, we assume that the same error variance guides the measured values, and assume that there are no covariances between the errors:

$$b^m_{it} \sim \mathcal{N}(\mu^m_t, \sigma^2) \tag{11}$$

$$b^l_{it} \sim \mathcal{N}(\mu^l_t, \sigma^2) \tag{12}$$

$$b^u_{it} \sim \mathcal{N}(\mu^u_t, \sigma^2) \tag{13}$$

The expectations are generated from a common belief that follows the description in subsection 2.2. This belief is assumed to be beta distributed with α_t and β_t as the shape parameters. One central difference to the theoretical model, here, is that we assume that all participants hold the same belief, which we consider sensible as we are interested in learning about the average learning process for different sub-groups. We define the expectation for the mean to be equal to the expectation of the beta distributed beliefs:

$$\mu^m_t = \frac{\alpha_t}{(\alpha_t + \beta_t)} \tag{14}$$

The expectations for the lower and upper bound are calculated from the CDF of the beta beliefs $F(x;\alpha_t,\beta_t)$ using the respective measures of the probability.

$$\mu_t^l = F(p_t^l;\alpha_t,\beta_t) \tag{15}$$

$$\mu_t^u = 1 - F(p_t^u;\alpha_t,\beta_t) \tag{16}$$

The analytical results from the parametric model identify how the beliefs are revised and evolve as function of the polls. In this we do not attempt to estimate the beliefs of each respondent at every time-point instead we are interested in the parameters of the model. The central parameter is d the discount-rate. And the general prior belief before the experiment (α_0,β_0). The other shape parameters of the respondents beliefs are defined according to the equations from the learning process:

$$\begin{aligned} \alpha_{it} &= d\alpha_{it-1} + y_t N_t \\ \beta_{it} &= d\beta_{it-1} + (1-y_t)N_t \end{aligned} \tag{17}$$

With this in mind, we define the log-Likelihood function of the observed measures for each respondent at each time-point to estimate the average discount parameter and the prior beliefs. From all respondents' updating processes we calculate quantities of interest like the rate of adaption. We define $\alpha_0 = [\alpha_{i0},\ldots,\alpha_{R0}]$ and $\beta_0 = [\beta_{i0},\ldots,\beta_{R0}]$ as the vector of all beliefs before the experiment, $\mathbf{B}$ as a matrix containing the belief measures of all respondents at each time-point, and $\mathbf{P}$ as the poll results presented to the respondents at the different time points y_t, the sample size N_t. We assume that the observations for each respondent are independent, which permits us to write the joint log-likelihood:

$$\ell(\mathbf{Y},\mathbf{P} \mid d,\alpha_0,\beta_0) = \sum_{t=0}^{T} -\frac{3R}{2} log(2\pi\sigma^2) - \frac{3}{2\sigma^2}\sum_{i=1}^{R}\left((b_{it}^m - \mu_t^m)^2 + (b_{it}^l - \mu_t^l)^2 + (b_{it}^u - \mu_t^u)^2\right) \tag{18}$$

The log-likelihood is maximized numerically and results in a estimate of the d and the prior beliefs.The evolution of beliefs can be calculated from the results and compared to the measured quantiles. Of course, this is the most parsimonious model. The only parameter that defines the learning process is the discount rate. We can make the model more flexible by assuming individual specific discount rates, adding different perceptions of the poll's information value (see Footnote 4), or even imposing a hierarchical structure.

5 Results

In this section we present some preliminary results. The first part shows the descriptive results, i.e. the survey responses. The second part illustrates the learning model described in subsection 2.2.

5.1 Descriptive Results

Before estimating the joint statistical learning model described in the section above, we estimate the average respondent beliefs in the different conditions separately for each time point. The resulting belief distribution give a descriptive representation of the learning process. Figure 3 shows the intervals and expectation from the resulting beliefs.[4]

The Figure highlights three aspects: First, participants clearly update their belief in direction of the polls. In the case of increasing poll results, the expected support for party A and B goes up, and vise versa for decreasing support. Second, there can be strong jumps towards the poll results from the prior beliefs. This is particular pronounced in the cases, where the prior (about a general race) is not in line with the first poll results. As an example, in the case of non-symmetric priors and increasing polls, the expectations form a prior at 66%, the first poll is at 51%. Respondents adopt to the information from the first poll directly. This also holds for the the scenario with symmetric beliefs and decreasing polls. Third, participants take the polls at their face-value and update closely to the poll results. The expectation always aligns closely with the poll results, as

[4]Figure A1 in the Appendix plots the distributions instead of the intervals

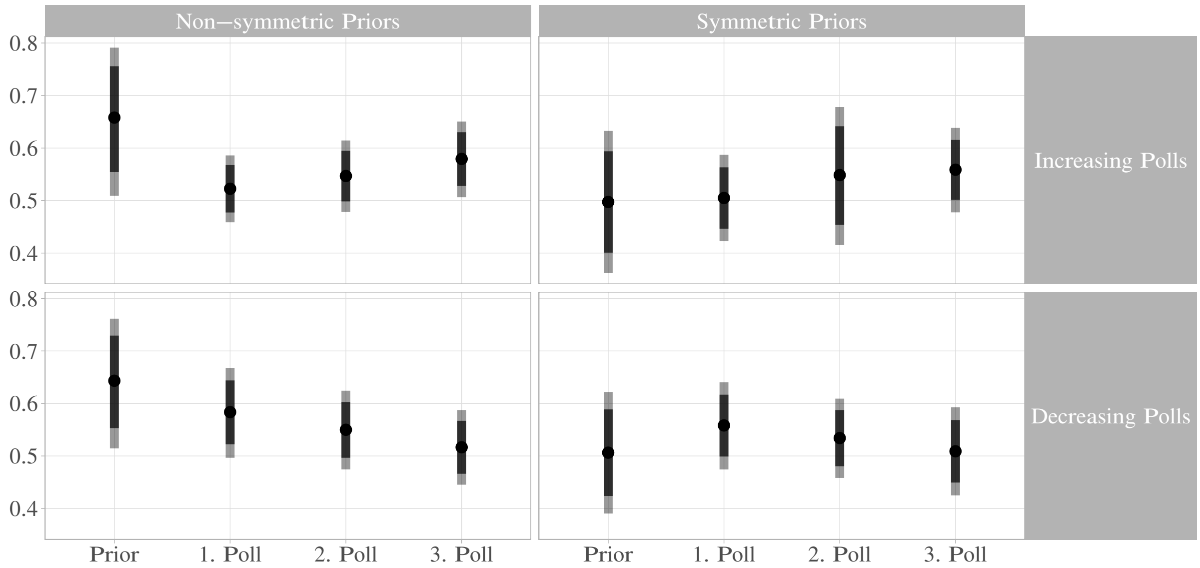

Figure 3: Respondents' beliefs about support for party A for different experiential scenarios and time-points. The parameters of the beta distributions are estimated for each time point separately. The bars indicate 99% and 95% coverage of the beliefs, the points the expectations.

the mean absolute deviation of the expectation to the poll deviation is 0.9% percentage points.

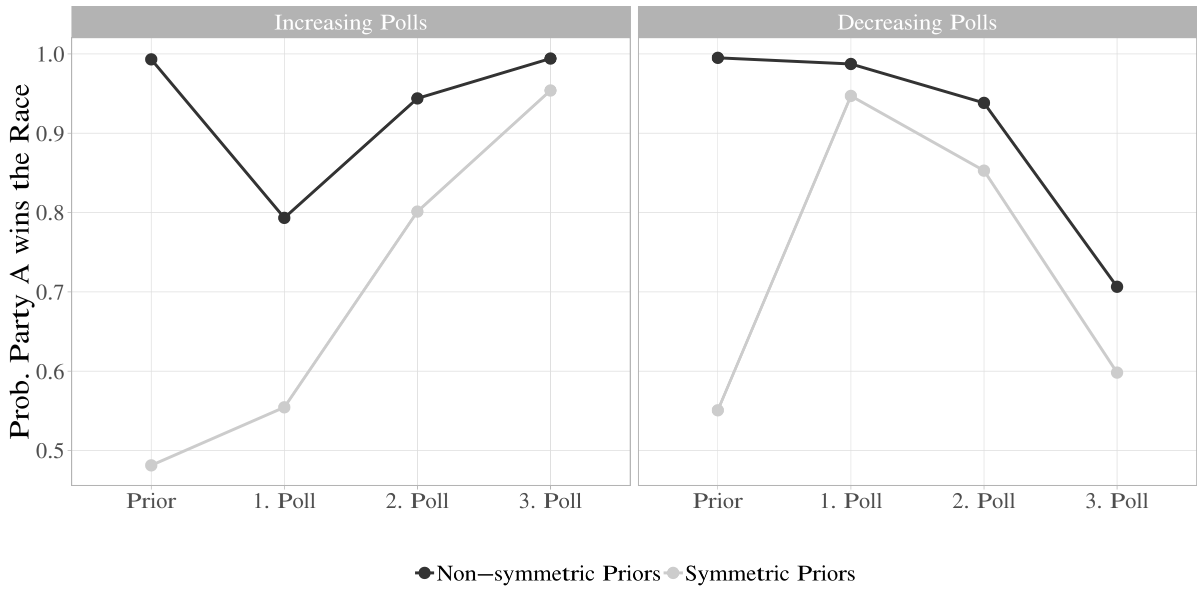

Figure 4: Respondents' beliefs about the probability that party A wins more than 50% of support, for different experiential scenarios and time-points. The parameters of the beta distributions are estimated for each time point separately.

An additional pattern is that the effect of priors does not vanish over the three poll results in our experiment. Figure 4 shows that this can have important implications how voters perceive the electoral race. It plots the probability that party A has a majority of support given the estimated beliefs in the electorate for the different scenarios. Of particular interest is the comparison within participants that received the same polls, but hold different priors. In the left panel, we plot the probability among participants who observed increasing poll results for party A. Both groups, those with symmetric and asymmetric priors, increase their chances that party A will win. But participants with with symmetric priors, who before the race believed that it's a close call, take longer to update. While the difference gets smaller over the time, it persistence even after the final poll. The same holds for the decreasing poll results, where the participants with symmetric priors are always less pessimistic that party A will win.

5.2 Results from Learning Model

In this section, we turn to the estimates form our parametric learning model. The results are reported in the Table below, containing the estimates for the different scenarios. Of particular interest is the discount factor. We estimate a value of 0.19 and 0.2 for the two cases in which the priors do not clash with the first poll result (Non-symmetric Priors and Decreasing Polls, as well as Symmetric Priors and Increasing Polls).

Symmetry	Direction	α_0	β_0	δ	σ
Non-symmetric Priors	Increasing Polls	144.96	125.99	0.00	0.10
Non-symmetric Priors	Decreasing Polls	52.41	29.51	0.19	0.09
Symmetric Priors	Increasing Polls	35.59	36.00	0.20	0.19
Symmetric Priors	Decreasing Polls	52.42	50.89	0.00	0.09

And a discount factor of 0 for the two cases where the first poll stands in contrast to the prior expectation. A discount factor of zero means that respondents do not carry over prior beliefs and completely follow the polls to form their updated beliefs. In the scenarios this might be necessary to accommodate the jump, from the general race to a particular race, at the beginning. In the other instances, we observe that participants consider their priors when updating their beliefs. If this is the case, it takes some time for the effect of priors to vanish from the belief formation about the race, which can result in the described patterns above. But even a discount factor of 0.2 implies a high rate of adaption. For the 0.2 discount factor the rate of adaption is 0.98 which means that the new expectation is to 98% at the new poll result from the previous belief expectation.

When considering the resulting beliefs from our parametric learning model (which we show in Figure A2 in the Appendix) it becomes clear that the learning model might not be flexible enough to accommodate the patterns we have described above. We observe a clear trend for the belief formation, but compared to the descriptive results, the implied beliefs after observing the polls are too precise. This is first indication that the learning model has to be adapted to understand the dynamic learning from poll results. We discuss some extensions in the Discussion.

6 Discussion

A first application of the dynamic learning model to our experimental data shows that it works reasonably well in capturing the general pattern of learning. However, certain aspects have to be adjusted to make the model fit the data better and to further be able to infer about other quantities of interest form the learning process. At the moment we only estimate the rate of adaption, but have little to say about the subjective standard deviation of a poll. In a next iteration we want to adapt the model specification to allow participants to have different perceptions of the precision of the polls. In addition we hope to accommodate the first transition from the general priors to the first poll result more accurately, as this seems to result in different estimates for our learning model.

References

Bartels, Larry. 2002. "Beyond the Running Tally: Partisan Bias in Political Perceptions." *Political Behavior* 24: 117–150.

Berinsky, Adam J, Gregory A Huber, and Gabriel S Lenz. 2012. "Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk." *Political Analysis* 20(3): 351–368.

Bullock, John G. 2009. "Partisan bias and the Bayesian ideal in the study of public opinion." *The Journal of Politics* 71(3): 1109–1124.

Coppock, Alexander. 2018. "Generalizing from survey experiments conducted on mechanical Turk: A replication approach." *Political Science Research and Methods* pp. 1–16.

Gerber, Alan, and Donald Green. 1999. "Misperceptions about perceptual bias." *Annual review of political science* 2(1): 189–210.

Goldstein, Daniel G, and David Rothschild. 2014. "Lay understanding of probability distributions." *Judgment & Decision Making* 9(1).

Hill, Seth J. 2017. "Changing votes or changing voters? How candidates and election context swing voters and mobilize the base." *Electoral Studies* 48: 131–148.

Jennings, Will, and Christopher Wlezien. 2018. "Election polling errors across time and space." *Nature Human Behaviour* 2: 276–283.

Leemann, Lucas, Lukas F. Stoetzer, and Richard Traunmueller. 2019. "Learn What Respondents Believe – Eliciting Priors in Surveys." *Working Paper* pp. 1–15.

Manski, Charles F. 2009. *Identification for prediction and decision.* Harvard University Press.

Mason, Winter, and Siddharth Suri. 2012. "Conducting behavioral research on Amazon's Mechanical Turk." *Behavior research methods* 44(1): 1–23.

Miratrix, Luke W, Jasjeet S Sekhon, Alexander G Theodoridis, and Luis F Campos. 2018. "Worth weighting? How to think about and use sample weights in survey experiments." *Political Analysis* X: 1–17.

Nassar, Matthew R, Robert C Wilson, Benjamin Heasly, and Joshua I Gold. 2010. "An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment." *Journal of Neuroscience* 30(37): 12366–12378.

Robinson, Jonathan, Cheskie Rosenzweig, Aaron J Moss, and Leib LItman. 2019. "Tapped Out or Barely Tapped? Recommendations for How to Harness the Vast and Largely Unused Potential of the Mechanical Turk Participant Pool." *PsyArXiv. June* 7.

Sinclair, Betsy, and Charles R Plott. 2012. "From uninformed to informed choices: Voters, pre-election polls and updating." *Electoral Studies* 31(1): 83–95.

Smith, JQ. 1979. "A generalization of the Bayesian steady forecasting model." *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 375–387.

Taber, Charles S, and Milton Lodge. 2006. "Motivated skepticism in the evaluation of political beliefs." *American Journal of Political Science* 50(3): 755–769.

Thomas, Kyle A, and Scott Clifford. 2017. "Validity and mechanical turk: An assessment of exclusion methods and interactive experiments." *Computers in Human Behavior* 77: 184–197.

Winkler, Robert L. 1967. “The assessment of prior distributions in Bayesian analysis.” *Journal of the American Statistical association* 62(319): 776–800.

A Additional Figures

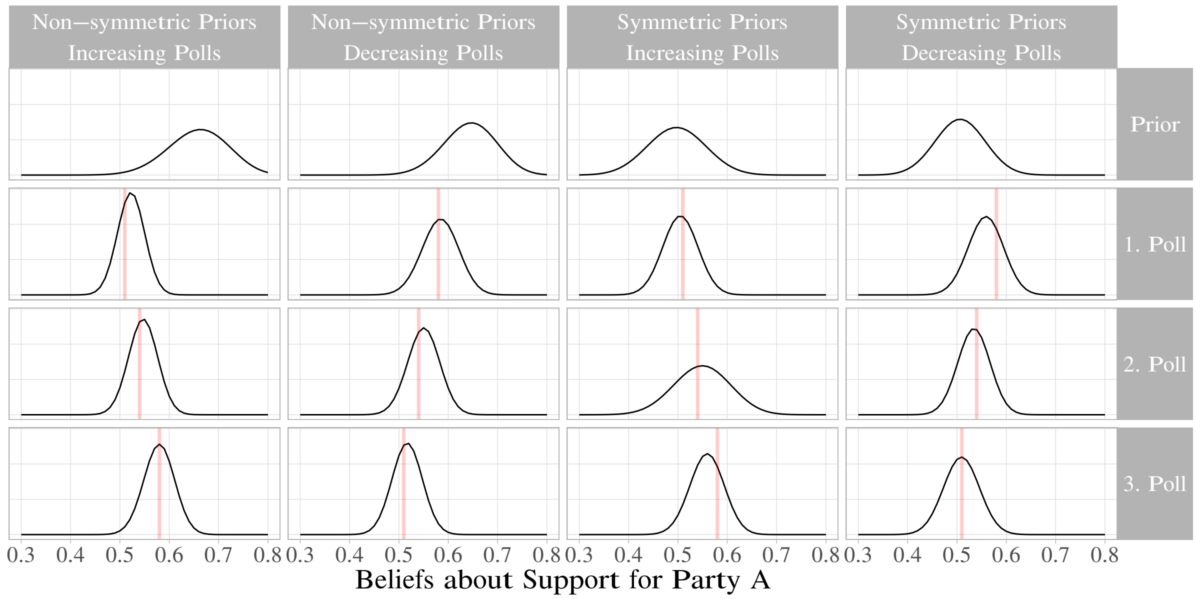

Figure A1: Respondents' beliefs about support for party A for different experiential scenarios and time-points. The parameters of the beta distributions are estimated for each time point separately. The vertical red lines indicate the poll result presented to the participants.

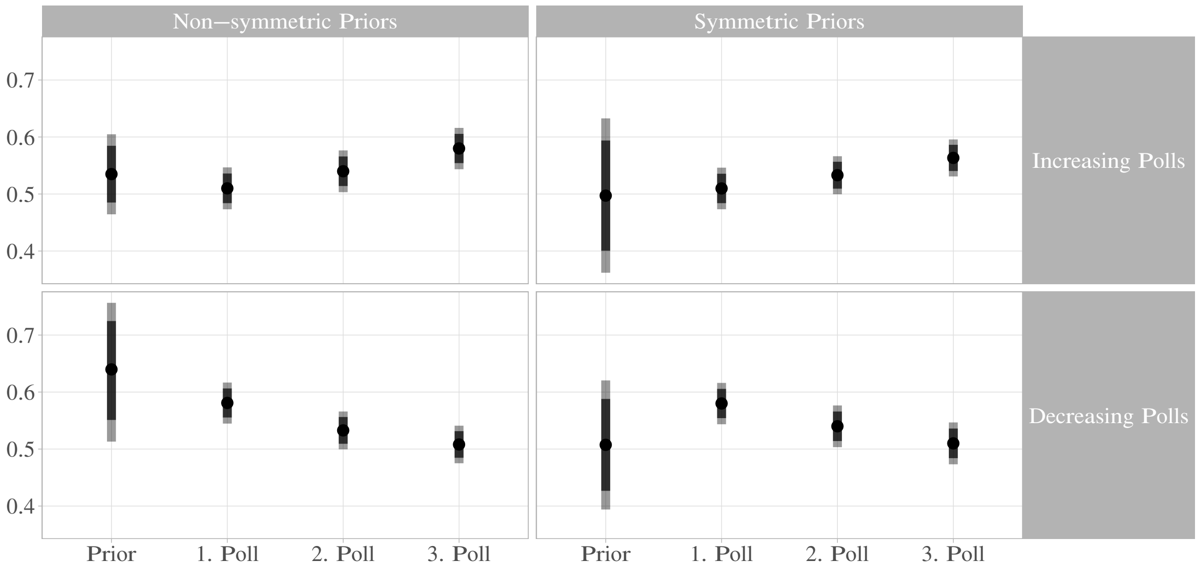

Figure A2: Respondents' beliefs about support for party A for different experiential scenarios and time-points. The parameters of the beta distributions are estimated from the learning model. The bars indicate 99% and 95% coverage of the beliefs, the points the expectations.