Learning Together Slowly: Bayesian Learning About Political Facts*

Journal of Politics, 79 (4), 2017

Seth J. Hill
University of California, San Diego†

August 7, 2017

Abstract: Although many studies suggest that voters learn about political facts with prejudice towards their preexisting beliefs, none have fully characterized all inputs to Bayes' Rule, leaving uncertainty about the magnitude of bias. This paper evaluates political learning by first highlighting the importance of careful measures of each input and then presenting a statistical model and experiment that measure the magnitude of departure from Bayesian learning. Subjects learn as cautious Bayesians, updating their beliefs at about 73 percent of perfect application of Bayes' Rule. They are also modestly biased. For information consistent with prior beliefs, subject learning is not statistically distinguishable from perfect Bayesian. Inconsistent information, however, corresponds to learning less than perfect. Despite bias, beliefs do not polarize. With small monetary incentives for accuracy, aggregate beliefs converge towards common truth. Cautious Bayesian learning appears to be a reasonable model of how citizens process political information.

Keywords: Bayesian learning; perceptual bias; political information; crossover scoring method.

*The experiments presented here were approved by the UCSD Human Research Protections Program. I thank the anonymous reviewers and editor, John Bullock, Jamie Druckman, Conor Dowling, Anthony Fowler, James Fowler, Don Green, Greg Huber, Sara Kerosky, Gabe Lenz, Skip Lupia, Neil Malhotra, Salvo Nunnari, Brendan Nyhan, Molly Roberts, Timothy Ryan, Dave Wiens, and seminar audiences at the University of Chicago Harris School, Northwestern, Stanford, and UCSD for feedback.

†Department of Political Science, 9500 Gilman Drive #0521, La Jolla, CA 92093-0521; sjhill@ucsd.edu, http://www.sethjhill.com.

Under many theories of representative democracy, citizens evaluate politicians by the economic and social outcomes obtained under the incumbent government. Fair evaluation requires objective appraisal of those outcomes. Models of accountability often assume citizens update their beliefs in response to new information via Bayes' Rule. Yet scholars have long observed that individuals appear to process new information with a bias towards their previously-held views. Classic studies in American political science echo this concern, from the "spiraling effect of political reinforcement (Berelson, Lazarsfeld, and McPhee, 1954, 223)" to the "perceptual screen (Campbell et al., 1960, 133)" of partisan identification. Another famous model of political behavior argues that political party can be used as a cue for citizens who do not wish to invest extensively in gathering new information (Downs, 1957), a shortcut which seems plausibly dissimilar from objective appraisal of new events (c.f. Fiorina, 1981).

Current evidence in political science suggests that citizens process political information with motivated skepticism, confirmation bias, and selective attention (e.g. Bartels, 2002), and that citizens update imperfectly even with objective information without partisan implications (Huber, Hill, and Lenz, 2012). Even more troubling is evidence that new information leads to *divergence* rather than convergence in political beliefs (Lord, Ross, and Lepper, 1979). Thus, a central question for the operation of democracy is how well citizens update politically-relevant beliefs. How much do biases in information processing limit accurate accumulation of political knowledge?

Research in economics and psychology evaluates learning in non-political contexts. This literature has not, however, evaluated if and by how much learning about political facts differs from learning about non-political facts. Is the role of Bayesian learning similar in the context of political facts to the context of non-political facts?

Measuring how learning of political information operates in the real political world is incredibly challenging. First, most measures of learning to date come from opinion surveys where responses to questions about political topics may be clouded by motivations other than accuracy, for example by partisan cheerleading (Bullock et al., 2015) or shirking (Prior and Lupia, 2008). Second, signals in the real world are generally ambiguous. That is, different individuals reading the same newspaper article may have different interpretations about what that article means with respect to their beliefs about the political world. Third, existing studies have not generally measured the uncertainty surrounding either respondents' beliefs about political facts nor uncertainty about the information present in the signals.

In this article, I present an experimental design that controls for these three problems to measure how individual learning about political information compares to learning via Bayes' Rule. Subjects receive noisy signals about salient political facts over the course of multiple rounds. The structure of the signals is such that there is no ambiguity about how they should be used to update beliefs with Bayes' Rule. In each round subject beliefs are elicited with incentives, creating measures of prior and posterior beliefs less clouded by cheerleading or shirking. Measuring prior and posterior beliefs along with signals of known form allows me to characterize how subjects should learn with Bayes' Rule, and measure to what extent and in what direction observed learning departs. The statements of fact subjects evaluate relate to political information thought to be important under both retrospective and prospective theories of voting.

I find that individuals consistently update political beliefs in the appropriate direction, even on facts that have clear implications for political party reputations, though they do so cautiously and with some bias. By cautious, I mean that they do not update their beliefs in response to new information as much as indicated by perfect application of Bayes' Rule. By biased, I mean that the amount of learning is not only less than Bayesian (cautious), but varies with prior beliefs in a way it should not (bias). Subjects do not, however, polarize. Though subjects were cautious in general and particularly cautious with signals opposed to their initial beliefs, on average they converged towards the same true value in response to information. Interestingly, those who identify with one of the political parties are no more biased or cautious that pure independents in their learning, conditional on initial beliefs.

I also compare learning about political facts to learning by the same subjects about their performance on an IQ quiz and, in a second experiment, to learning about an ego-irrelevant fact. Relative to political facts, I find more caution in learning about performance on the IQ quiz but less caution in learning about an abstract fact. In both cases subjects exhibit more bias in learning about political facts, though differences are small. Importantly I find departures from Bayesian learning for both the IQ and the abstract fact, which is consistent with other work and suggests the experimental setup here does not uniquely generate unusually rational learning.

This article makes contributions to the literatures on perceptual bias in politics and on information processing more generally. First, the experiment and tests of Bayesian learning relative to motivated reasoning on political facts implies that citizens may learn more rationally and closer to Bayes' Rule than the exiting literature suggests. With the design here I am able to directly measure the magnitude of bias without assumptions about prior beliefs, with incentives to be accurate, and with limited concern about errors in interpretation of signals or selective exposure. Because the design quantifies all inputs to Bayes' Rule, it allows a careful statement on the magnitude of departure, and may also provide a path forward to more measurement of the process and context by which citizens learn political information.

Second, the results suggest that learning about political facts is not notably different from learning about non-political facts and that Bayesian learning is not an unreasonable model of how individuals respond to new political information. Although subjects learn with more bias towards prior beliefs about political facts than about abstract or ego-relevant non-political facts, in neither of two experiments are differences particularly large. Political learning appears only modestly different from learning about other facts. About each type of fact, subjects learn slowly towards common truth.

Because the experimental design I introduce measures each input to Bayesian learning of political information, it may be useful for other scholars interested in evaluating political behavior and political information processing. Across many research questions, students of politics are interested in the subjective probabilistic beliefs of both experts and average citizens. Who will win the election? How likely is country Z to develop a nuclear weapon? What are the chances you will turn out to vote? With increasing evidence that survey responses about statements of fact with political implications are clouded by motivations other than accuracy (Bullock et al., 2015; Prior and Lupia, 2008; Prior, Sood, and Khanna, 2015; Taber and Lodge, 2006), the design presented here may be of wide value to scholarly inquiry and builds on recent other efforts to use incentivized experiments to elucidate issues of political accountability (Huber, Hill, and Lenz, 2012; Woon, 2012).

These results suggest that formal models of accountability, which usually assume citizens update as perfect Bayesians, may benefit by considering the implications of cautious or modestly biased Bayesian processing of signals about incumbent performance. The estimates of magnitude of departure from Bayesian learning also suggest the need to evaluate how much learning is sufficient for good accountability.[1] A remaining empirical question for further research is to what degree selective exposure or information environments more complicated than that in this experiment drive perceptual bias outside the laboratory.

The essay proceeds as follows. I first highlight the importance of measuring all inputs to Bayes' Rule to evaluate bias in political information processing, then present the crossover scoring method design to elicit probabilistic beliefs and the statistical tests used to evaluate learning relative to the Bayesian ideal. I next present the experimental design and results from two experiments, consider robustness to alternative models of learning and alternative mechanisms, and finally offer concluding thoughts on implications for understanding of how citizens process political information.

Learning political information

Most scholars of political information processing agree that the ideal procedure for learning is Bayes' Rule. For example, "Every opinion is a marriage of information and predisposition (Zaller, 1992, p. 6)" (see also, Bartels, 2002; Bullock, 2009; Gerber and Green, 1999; Taber and Lodge, 2006). Bayes' Rule provides a coherent path to transform the two inputs of prior beliefs and new information into posterior beliefs. When confronted with new information, citizens should evaluate the information and update their beliefs by a weighted combination of prior beliefs and the meaning of that information. In this article, I consider beliefs about binary factual statements, i.e. the subject has a probabilistic belief that the statement is true, with uncertainty reflected by

[1] See Ashworth and Bueno de Mesquita (2014) for a model where citizens are better off *not* learning perfectly about incumbent performance.

the magnitude of the probability. New information is a *signal*, and along with the prior belief is updated to a posterior belief through Bayes' Rule,

$$\Pr(T|S) = \Pr(T)\frac{\Pr(S|T)}{\Pr(S|T)\Pr(T) + \Pr(S|F)\Pr(F)} \tag{1}$$

with T a belief that the statement is true, F a belief that the statement is not true, $\Pr(T) = 1 - \Pr(F)$, S a signal about the statement, and $\Pr(\cdot)$ returning the probabilistic belief about its argument.

The difficulty of evaluating how well political citizens follow this Bayesian model of learning is highlighted by the terms in Equation 1. Testing for divergence from Bayesian learning requires observing or making assumptions about each quantity in (1) for each individual in the population: posterior beliefs $\Pr_i(T|S)$, prior beliefs $\Pr_i(T)$, and beliefs about the likelihood of observing the signal S if the fact is true versus false, $\Pr_i(S|T)$ and $\Pr_i(S|F)$. I highlight that these quantities might each vary across the population by subscripting each probability for individual i. Existing research concludes that citizens process political information with perceptual bias such as motivated skepticism or confirmation bias, meaning that the amount of learning from new information varies with prior beliefs more than would be indicated by objective application of Bayes' Rule.

Research studies always make assumptions to simplify analysis. In fact, without assumption about or measurement of $\Pr_i(S|T)$ and $\Pr_i(S|F)$, any pattern of learning can be made consistent with Bayes' Rule. Much of the work on political information processing has made an assumption about the signals delivered or received in order to evaluate learning. However, this means most evidence currently cited on perceptual bias rests on these assumptions.

While recent research appreciates that individuals may vary in their prior beliefs, the problem of potential heterogeneity in beliefs about signals remains central. Designs that use panel data to observe prior and posterior beliefs over time along with changes in the state of the world [such as changes in the economy (Bartels, 2002) or outcomes in the war in Iraq (Gaines et al., 2007)] assume that objective changes in the state of the world are received as consistent signals to respondents to the panel survey with respect to the outcome measure, e.g. that for respondent i $\Pr_i(S|T) =$

$\mathrm{Pr_j(S|T)}\ \forall\ i \neq j$ and $\mathrm{Pr_i(S|F)} = \mathrm{Pr_j(S|F)}\ \forall\ i \neq j$.[2] Without this assumption, the appearance of motivated reasoning, i.e. variation in Pr(T|S) in response to the same signal, could be due to heterogeneous interpretation of signals instead of biased processing. Knowing how individuals interpret signals is central to making inferences about how individuals learn.

Empirical research has moved from early work that identified cross-sectional differences in beliefs (e.g., Berelson, Lazarsfeld, and McPhee, 1954, ch. 10) to more recent work that measures priors and posteriors and considers assumptions about signals (e.g., Bartels, 2002; Bolsen, Druckman, and Cook, 2014; Gaines et al., 2007; Jerit and Barabas, 2012; Lauderdale, 2015; Rahn, 1993; Taber and Lodge, 2006; Zaller, 1992). Recent work has commonly found partisan divergence in the evaluation of new political information, and strongly suggests that partisans are biased in the way they evaluate political signals (for evidence closer to Bayesian learning like that here, see Guess and Coppock, N.d.). In fact, some work finds the bias to be inconsistent with Bayesian learning, with subjects moving in the direction opposite of the signal through "biased assimilation" (Lord, Ross, and Lepper, 1979) leading to polarization and hardening of views (see also, Bartels, 2002; Nyhan and Reifler, 2010; Taber and Lodge, 2006).[3]

Eliciting probabilistic beliefs

The experimental design here measures all quantities required to compute beliefs via Bayes' Rule and allows characterization of the magnitude and direction of departure from Bayesian learning. The experiment delivers noisy signals about political facts over multiple rounds. Subjects are informed that the signals are noisy but informative: signals are correct on average three out of four times. Thus, the signals are simple, clear, stochastic, consistently-delivered, and of known form common to all participants in the study.

Prior to the delivery of the first signal and after the delivery of each signal, subjects' beliefs are elicited using monetary incentives with the crossover scoring method. The crossover method

[2] Gaines et al. (2007, Figure 2) do find monotonically consistent updating of factual beliefs over the panel rounds with respect to casualties and weapons of mass destruction during the Iraq War, but with a good bit of noise. Bartels (2002, Equation 5) assumes constant meaning to signals in reduced form regression specifications.

[3] Research following Lord, Ross, and Lepper (1979) did not always replicate biased assimilation, finding it contingent on various factors of the individual and the study. See the presentation in Gerber and Green (1999, p. 195–7).

asks participants for what probability p they would be indifferent between receiving a payment with probability p and receiving a payment if their answer is correct. With these incentives, the subject maximizes their probability of payment by accurately reporting their subjective belief about the factual statement.[4] This method of eliciting beliefs was proposed formally by Allen (1987), Karni (2009), and Möbius et al. (2011). My experimental design is similar to that in Möbius et al. (2011). Holt and Smith (2016) show that this method outperforms the Quadratic Scoring Rule in an experimental comparison.

This experimental design has three key features. First, it measures prior and posterior in quantitative terms and with incentives for accuracy. Second, it delivers randomized signals of known form over multiple rounds. Third, signals are unambiguous and presented without other information, lessening the likelihood of differential interpretation of signals. In the context of correcting misperceptions as in Nyhan and Reifler (2010), the design tests how individuals respond when the correction is of unambiguous likelihood.

The experiment here is similar to economic experiments on non-political learning, which compare observed choices to choices that would be made under perfect application of Bayes' Rule, such as Möbius et al. (2011). For example, Anderson and Holt (1997) run lab experiments to learn how respondents behave in the context of information cascades where observing the behavior of others should lead subjects applying Bayes' Rule to, at some points, depart from their own private signals. Anderson and Holt (1997) use regression models to account for random decision errors, similar to those I estimate below, to show that subjects are mostly Bayesian and do not suffer from a handful of proposed behavioral biases (status quo bias, representativeness bias, or counting heuristic). Instead, subjects behave about 73 percent that of perfect Bayesian (see p. 858), a number nearly the same as that I estimate below.

As with any simplified experiment, the advantages of internal validity come with drawbacks surrounding external validity. A signal of a specific likelihood as delivered in this experiment can

[4] I present the scoring rule and prove incentives are maximized at true beliefs in Online Appendix Section A. In the actual experiment, I presented the details of the mechanism in simple terms and highlighted at multiple points that participants would maximize payment conditional on beliefs by accurately reporting their beliefs.

approximate a variety of information flows citizens may observe in the real world. For example, observing a news story about rising gas prices is a signal with a specific likelihood about the state of inflation in the world (probability of observing the story given rising inflation versus the probability of observing the story given declining inflation). But a simple, unambiguous signal from a computer is not likely to fully reflect more complicated and, particularly, competitive information environments outside of the laboratory. This experiment uses binary statements of fact and delivers accurate signals, which may not always be the case outside of the lab. However, this abstraction from complication buys the ability to directly measure quantitative departures from Bayesian learning and the magnitude of any bias.

Testing for departure from Bayesian learning

To measure how the responses of experimental subjects compare to ideal Bayesian learning, I use the log-odds specification of Bayes' Rule, which transforms Equation 1 to

$$\text{logit}[\Pr(T|S = s)] = \text{logit}[\Pr(T)] + \log[\Pr(S = s|T)/\Pr(S = s|F)] \tag{2}$$

(see Online Appendix Section B for the derivation), which can then be specified as the regression model

$$\begin{aligned}\text{logit}[\Pr_{it}(T|S_{it} = s)] = \quad & \delta \, \text{logit}[\Pr_{it-1}(T)] \\ + \quad & \beta \, \mathbf{1}[S_{it} = t] \times \log[\Pr(S = t|T)/\Pr(S = t|F)] \\ + \quad & \beta \, \mathbf{1}[S_{it} = f] \times \log[\Pr(S = f|T)/\Pr(S = f|F)] + \varepsilon_{it}\end{aligned} \tag{3}$$

where i indexes subjects and t indexes rounds, $\mathbf{1}[\cdot]$ returns a 1 when its argument is true, and 0 otherwise, and ε is a random disturbance to the updating.

Because by experimental design signals reveal the truth three out of four times, the likelihood ratios of each signal are known: $\Pr(S = t|T)/\Pr(S = t|F) = (3/4)/(1/4) = 3$ for a true signal and $\Pr(S = f|T)/\Pr(S = f|F) = (1/4)/(3/4) = 1/3$ for false. With signals of this nature, posterior beliefs (in log odds) should increase by log(3) when the subject receives a true signal and by log(1/3) when false. Because log(3) is positive and log(1/3) negative, this application of Bayes' Rule is also intuitive: beliefs move towards true with a true signal and away from true with a false signal.[5] The values log(3) and log(1/3) enter the regression model as "data" so that the coefficients β and δ measure the magnitude of departure from Bayesian. Perfect application of Bayes' Rule leads to β and δ values of one.

Existing research suggests a set of specific departures from Bayesian learning. First is confirmation bias or motivated reasoning, where the amount of learning relative to Bayesian learning depends upon the consistency of the new information with the individual's pre-existing beliefs. To account for this possibility, I extend regression specification (3) to specification (4) to allow β to vary by whether the signal is consistent with the individual's initial beliefs as measured in the first round prior to any signals,

$$\begin{aligned}
\text{logit}[\Pr_{it}(T|S_{it} = s)] = \quad & \delta \text{ logit}[\Pr_{it-1}(T)] \\
+ \quad & \delta_2 \, \mathbf{1}[S_{it} = c_i] \times \text{logit}[\Pr_{it-1}(T)] \\
+ \quad & \beta \, \mathbf{1}[S_{it} = t] \times \log[\Pr(S = t|T)/\Pr(S = t|F)] \\
+ \quad & \beta_2 \, \mathbf{1}[S_{it} = c_i] \times \mathbf{1}[S_{it} = t] \times \log[\Pr(S = t|T)/\Pr(S = t|F)] \\
+ \quad & \beta \, \mathbf{1}[S_{it} = f] \times \log[\Pr(S = f|T)/\Pr(S = f|F)] \\
+ \quad & \beta_2 \, \mathbf{1}[S_{it} = c_i] \times \mathbf{1}[S_{it} = f] \times \log[\Pr(S = f|T)/\Pr(S = f|F)] + \varepsilon_{it}, \quad (4)
\end{aligned}$$

where $\mathbf{1}[S_{it} = c_i]$ is an indicator function that takes the value of one when a signal is consistent with subject i's initial belief and zero otherwise, and δ_2 and β_2 allow differential fealty to prior and differential response to signals as a function of consistency of signal. Statistical tests on β_2 evaluate differential learning of political information as a function of initial beliefs. I define consistency as cases where the subject's initial probabilistic belief (before any signals) match the signal in that round. That is, if the subject initially believed the statement to be true and the signal was true, the signal is consistent. Likewise, an initial belief of false is consistent with a false signal. Note that this experimental design allows a direct measure of consistency rather than by assumed relationship

[5] These two likelihood ratios are also symmetric (log(3) = –log(1/3)), which also arises intuitively from the signals reflecting the truth symmetrically three out of four times.

to self-reported partisanship.[6]

A second, stronger form of perceptual bias is the theory of biased assimilation, which argues that individuals polarize (beliefs move in opposite directions) in response to information inconsistent with initial beliefs (e.g. Lord, Ross, and Lepper, 1979). This suggests that confirming evidence would have an especially positive influence on learning ($[\beta + \beta_2] \gg 1$) and/or that discordant evidence would have a *negative* influence on learning ($\beta \ll 0$). The expanded specification (4) also evaluates this theory with statistical tests on β and β_2.

Implementation

Between September 17 and 23, 2015, I recruited 990 participants aged 18 and older and U.S. citizens from Amazon.com's Mechanical Turk (MTurk) worker platform to participate in the experiment. Participants were paid a $0.50 flat fee and offered the opportunity to earn bonuses of up to $4.50 depending upon their performance in the experiment, which was advertised to and did take about 15 minutes. The study did not deceive, which was advertised prominently on the consent screen. (I present details of a second experiment below.)

One concern about MTurk is that its sample is overly young, educated, and Democratic, although recent work shows that MTurk samples yield experimental treatment effects highly similar to other samples (Berinsky, Huber, and Lenz, 2012; Mullinix et al., 2015). To mitigate unrepresentativeness, I asked survey questions exactly as they were asked on the 2014 Pew Polarization Survey that allow me to construct post-stratification weights raked to the marginal distributions of respondents to the Pew Survey, which was a nationally-representative telephone-based sample of 10,013 respondents surveyed January to March 2014. I rake to questions related to political confirmation bias and the MTurk sample composition making the weighted distribution of subjects more representative. See Online Appendix Section C for details of the weighting procedure. Results are quite similar with unweighted analysis (I reproduce the main tables unweighted in Online Appendix Section H).

Upon consenting to participate, subjects first took an IQ-like quiz. They had three minutes to

[6] I present results by partisanship below.

answer up to 30 logic and reasoning questions. They were paid \$0.10 for each point of their total score on the quiz, which was the number answered correctly less the number answered incorrectly, skipped questions not counted. The average quiz score was 4.3 (4.9 unweighted), with a minimum of -14 and a maximum of 13.[7]

After the IQ-like quiz, subjects were taught about the main section of the experiment. They were told that they would participate in a contest consisting of 15 rounds. For each round won, they would be paid a \$0.10 bonus, \$0.00 otherwise. In each round, they would be asked to evaluate a difficult factual statement with a number from 0 to 100 that described how likely they believed the statement to be true.[8] The instructions presented the response as a probability in terms designed to be accessible to those not trained in statistics. The instructions then explained how participants would win each round, which was a function of their probabilistic belief through the crossover design. The experiment presented the crossover design in simple terms and highlighted at multiple points that the subject's chances of winning would be highest if they accurately reported their probabilistic belief.

After presenting the overview of the contest and the mechanism of payment, subjects were instructed that they would evaluate the same factual statement in multiple rounds, and that in some rounds they would receive a signal from the computer about whether or not the statement was true. They were told that the signal from the computer would indicate that the correct answer was true or false, and that this signal would be correct three out of four times on average. They were told that they might want to change their beliefs in response to the signal, and that the set of signals given would be stored and presented for them throughout the contest.

After the instructions for the contest, the subjects played three practice rounds evaluating the factual statement "It rained (more than 0.00 inches of precipitation) in Santa Fe, New Mexico on July 7, 2004." Mimicking the contest they would play, in the first round they evaluated the

[7] Subjects were told that money would not be deducted from the show-up fee for scores less than zero.

[8] The prompt in each round was "Please tell us how likely you believe this statement is true: [Statement presented]. How likely you believe that the statement is true (for example, 1 if you believe it almost certainly false, 99 if you believe it almost certainly true, 50 if totally unsure): [textbox entry]." Full instructions as presented are in Online Appendix Section J.

statement without any signal from the computer. In the second and third rounds, they received signals from the computer and again evaluated the statement. After the third round, the instructions explained how they would be paid as a function of their response.[9]

Once the practice contest was complete, participants then proceeded to the main contest for which they would be paid based upon their performance. For each of three statements, beliefs were elicited for five rounds. Beliefs were elicited in the first round for each statement prior to the delivery of any signal, measuring the subject's initial belief. In each round subsequent to the first, their previous response was presented for their reference.[10] In rounds two through five for each statement, they received one new (independent) signal in each round from the computer about the statement and reported their (potentially-updated) belief. In each round with a signal, the subject was reminded that the signal would be correct three out of four times. Additionally, in rounds three, four, and five, the signals from the previous round(s) were presented so that the subject would not have to keep track. With this design, I observe how subjective beliefs about the statement change over time in response to the noisy signals received.

The first two statements each subject evaluated were drawn at random from a set of six political factual statements, the full text of which is presented below in Table 1. The statements are about economic and social outcomes under various presidential administrations and the vote shares received by presidential candidates. There were three statements where a true signal favored the Democratic Party/president, and three statements where a true signal favored the Republican Party/president. Each participant was assigned to receive one of the three statements favoring the Democratic Party and one of the three statements favoring the Republican party, and assigned at random which question would be presented in the first contest.

The third and final statement was not political. Instead, all respondents evaluated one of two factual statements about their score on the IQ quiz,

50 United States citizens aged 18 and over recruited from Mechanical Turk on June

[9] See Online Appendix Section J for the feedback and instructions.

[10] In all rounds, subjects had 20 seconds to evaluate the statement, to limit the option of searching for the truth on the web. After 20 seconds, responses were recorded and they were automatically forwarded to the next round.

> 11, 2015 completed the same 3-minute IQ-like quiz as you. They were also paid a $0.50 show up fee and $0.10 per point of their quiz score, the number marked correct minus the number marked incorrect in the 3 minutes. Your score is in the [top half (above the 25th out of these 50 scores) /OR/ bottom half (at or below the 25th out of these 50 scores)] on the quiz.

I had recruited 50 participants to take the same IQ-like quiz on June 11, and did use their median quiz score (5) to score the validity of the statement for each participant. Subjects were assigned at random whether the statement that they would evaluate described them in the top or bottom half of the distribution, as indicated within the brackets. Participants evaluated this statement over five rounds exactly as they did the two political statements.

Finally, after completing the three contests, participants answered a series of survey questions about their demographics, political attitudes, and political behaviors. This includes standard demographics and political questions such as partisanship and ideology, and the set of questions taken from the 2014 Pew Polarization Survey to construct stratification weights.[11] On the final screen, a code was presented to the subjects for them to submit on Mechanical Turk in order to collect any bonuses from the IQ-like quiz responses and the three contests.

Results

In this section, I present an overview of the questions of the experiment and the amount of learning. I show that on average participants did respond to the signals and that partisans did diverge in their prior and posterior beliefs about partisan facts yet learned in common direction towards the truth.

In Table 1, I present the full set of factual statements evaluated along with average prior and posterior beliefs. Prior beliefs are the subjective probabilities that each statement is true in the first round of each contest before any signals are received. Posterior beliefs are the subjective probabilities in the fifth round of each contest after four signals have been received. I present the prior and posterior for all respondents, as well as separately for self-identified Democrats and Republicans, sorted in ascending order of prior beliefs among Democrats.[12]

[11] I find below no evidence of post-treatment bias in these responses.

[12] For the statements about score on the IQ quiz, I tabulate responses separately for those subjects who were above and below the top half. I pool respondents regardless of whether they evaluated a statement about being in the top or

Table 1: Prior and posterior beliefs by question and partisanship

Question	All respondent means		Democrat means		Republican means	
	Prior	Posterior	Prior	Posterior	Prior	Posterior
From 2009, when President Obama took office, to 2012, median household income adjusted for inflation in the United States fell by more than 4 percent. (TRUE)	57.8	73.5	49.9	70.4	63.0	73.8
The rate at which American women aged 15-44 had legal abortions fell more between 1980 and 1988, while Ronald Reagan was president, than between 1992 and 2000, while Bill Clinton was president. (FALSE)	56.8	36.1	55.8	35.2	60.7	37.0
In the 2004 Presidential Election, John Kerry was defeated by George W. Bush. In the nation as a whole, of all the votes cast for Kerry and Bush, Kerry won less than 48 percent. (FALSE)	62.3	36.9	61.1	34.6	68.5	40.0
The total public debt of the United States federal government more than doubled from quarter 2 in 1981 to quarter 1 in 1989 while Ronald Reagan was president. (TRUE)	60.5	67.5	63.8	66.3	57.5	68.9
From January 2001, when President Bush first took office, to January 2005, when President Bush started his second term in office, the civilian unemployment rate increased by more than 1 percentage point. (TRUE)	73.5	77.7	75.2	81.6	68.9	78.5
In the 2012 Presidential Election, Barack Obama defeated the Republican Mitt Romney. In the nation as a whole, of all the votes cast for Obama and Romney, Romney won less than 48 percent. (FALSE)	74.7	47.6	80.0	50.0	67.6	50.6
Your IQ quiz score is in the top half (respondents for which TRUE).	59.2	75.5	56.0	71.3	61.8	78.3
Your IQ quiz score is in the top half (respondents for which FALSE).	53.4	39.8	56.4	40.1	51.3	41.4

Note: Each cell presents the average elicited probabilistic belief that the statement is true among the participants who evaluated that fact. Prior beliefs are the beliefs offered in the first round prior to any signals. Posterior beliefs are the beliefs offered in the fifth round after receiving four signals. All subjects evaluated their score on the IQ quiz and evaluated two of the partisan statements drawn at random.

There are three notable observations from Table 1. First, participants learn from signals. In all cases, average posterior beliefs are closer to the truth than average prior beliefs. Note that this learning occurs even though subjects received noisy signals in an abstract environment, one out of four of which were inaccurate. For example, the first row presents results for a statement about change in median household income under Democratic President Barack Obama. The statement is true, and thus participants on average received three out of four true signals. The average beliefs for all participants move from a prior probability true of 57.8 to a posterior probability true of 73.5. For reference, three true and one false signal transforms a prior belief of 57.8 to posterior 92.5 with perfect application of Bayes' Rule. The observed average posterior of 73.5 is evidence of caution in learning, that subjects learned less than perfect Bayes. In the second row, a statement that is false, beliefs move from an average prior of 56.8 to an average posterior of 36.1.

A second observation from Table 1 is partisan differences in prior and posterior beliefs. For the partisan questions, prior beliefs for Democrats were more favorable to the Democratic president or candidate and less favorable to the Republican president or candidate than the prior beliefs of Republican subjects. This reproduces the longstanding result of partisan differences in factual beliefs (Berelson, Lazarsfeld, and McPhee, 1954; Campbell et al., 1960).

Despite divergent prior and posterior beliefs, a third observation from Table 1 is the absence of polarization. Democratic and Republican beliefs always move in the same correct direction in response to the signals. In the aggregate, identifiers from both parties learn together in the same direction about politically-relevant facts.

Table 1 presents an aggregate overview of the results, but a particular value of this experimental design is the observation of individual-level learning over five rounds in response to signals. To provide intuition for the experiment, I present examples of four subjects' responses and signals in Online Appendix Figure A1. The figure shows how subject beliefs evolve in response to specific signals, and highlights individual examples of caution, bias, and inattentiveness. The fourth subject plotted, for example, did not update beliefs in response to any signal in any experiment. A small set

bottom half by differencing from 100 the responses of subjects assigned to evaluate the bottom half statement.

of participants appear not to have engaged in the game. Nonetheless, I include all subject responses in analysis below, whether or not they appear to have ignored the signals or actively participated. While some might drop subjects whose beliefs never change, they remain in the sample here to represent citizens who may not revise their beliefs in response to political information. Non-changing beliefs are post-treatment.[13]

Bayesian learning about political facts

In this section, I evaluate how well the Bayesian model of learning captures political learning in the experiment. I estimate regressions consistent with specification (3) in Table 2. Each observation is one round of one of the two partisan contests, with the first coefficient estimate that of δ, the influence of the (prior) belief from the most recent round of the contest. The second coefficient estimates β as an evaluation of Bayesian learning. That is, the variable Signal takes the value of log(3) when the signal was true and log(1/3) when the signal was false. These values are symmetric (log(3) $\approx$ 1.098 $\approx$ –log(1/3)), so pooling the two together makes for straightforward inference about the parameters of learning, though one could estimate separate coefficients for true and false signals if desired (as in Eq. 3). As noted above, the regression model is derived directly from Bayes' Rule, but also has the intuitive interpretation that β measures how much beliefs move towards true in response to a true signal and away from true in response to a false signal, on average.[14]

Table 2 pools multiple responses of the same subject consistent with Bayes' Rule being memoryless. Standard errors are clustered at the subject-game level to account for potential within-game correlation in subject responses.[15]

The results in the basic specification of column one show that subjects are not perfect Bayesians,

[13] About 24 percent of partisan contests exhibit no change to beliefs during the five rounds of that contest. Considering all three contests for each individual, 58 subjects (5.9 percent) never revised initial beliefs.

[14] For example, one could alternatively code true signals 1 and false signals -1, in which case β would be a more standard regression coefficient. Using log(3) and log(1/3) rescales this standard regression coefficient so that learning may be compared to Bayes' Rule.

[15] Because responses of 0 or 100 are undefined in logits, in all data analysis I recode responses of 0 and 100 to 1 and 99. In practice, many subjects did revise their beliefs in response to signals after stating beliefs of 0 and 100. In Online Appendix Table A1, I present estimates for each round of the contest, and show no apparent trend in learning by round or contest. All regressions use weighted least-squares with Pew post-stratification weights.

Table 2: Bayesian learning about political facts

VARIABLES	(1) Pooled	(2) Signal consistent	(3) Not consistent	(4) Pooled	(5) Dems/Reps only
Logit prior (δ)	0.61**	0.52**	0.60**	0.60**	0.60**
	(0.02)	(0.04)	(0.03)	(0.03)	(0.03)
Signal (β)	0.73**	1.06**	0.59**	0.59**	0.59**
	(0.05)	(0.10)	(0.06)	(0.06)	(0.07)
Signal*Signal consistent (β_2)				0.47**	0.45**
				(0.12)	(0.14)
Logit prior*Signal consistent (δ_2)				-0.085	-0.078
				(0.05)	(0.06)
Observations	7,664	3,294	4,227	7,521	6,138
R-squared	0.445	0.596	0.304	0.444	0.439
Std. error of regression	2.36	2.11	2.55	2.37	2.38
N subjects	990	902	958	988	804
Wald test on null $\delta = 1$	0	0	0	0	0
Wald test on null $\beta = 1$	8.2e-09	1	0	0	2.7e-09

Robust standard errors in parentheses
** $p<0.01$, * $p<0.05$

Note: Dependent variable is logit-beliefs that the statement is correct in that round for rounds 2 through 5. A perfect Bayesian would have coefficients of 1 on both variables. Standard errors clustered on the subject-game.

but do respond to signals similar to Bayes' Rule. The coefficient on the prior is 0.61 and on the signal 0.73. While the hypothesis of perfect Bayesian learning can be rejected at standard levels for both coefficients (see final two rows for p-values on Wald tests for perfect Bayesian learning), subject posterior beliefs are a weighted average of prior belief and the likelihood ratio of the signal received, consistent with Bayes' Rule. Subjects update their beliefs in response to signals 73 percent as much as they would with perfect application of Bayes' Rule. This provides initial evidence that citizens process political information in a manner close to Bayesian.

Columns two and three evaluate whether subjects learn differently in response to signals that are consistent or inconsistent with their initial beliefs, i.e. motivated reasoning. I separate the sample into cases where the signal received in the round was consistent or inconsistent with the subject's first round belief. That is, for those subjects whose initial belief was less than 50, False signals are consistent with their prior beliefs and True signals are inconsistent. For subjects whose initial belief was greater than 50, True signals are consistent and False signals inconsistent.[16] Previous work has separated samples by self-reported partisan identification on the assumption that this separates individuals into different types of bias in processing. I am able to directly match initial beliefs to subsequent signals on consistency, regardless of partisanship.[17]

The two models separating rounds by consistent and inconsistent signals estimate coefficient δ at similar magnitude (0.52 and 0.60), though do indicate that subjects held on slightly more to their previous beliefs when the signal was inconsistent with their initial belief. The coefficients on signals more strongly suggest motivated bias. For signals that are consistent with first round beliefs, subjects update beliefs at 106 percent of the rate they should have as perfect Bayesians, though the difference from perfect application of Bayes rule ($\beta = 1$) is not statistically significant. However, for signals inconsistent with first round beliefs, subjects update beliefs at 59 percent of the rate of perfect Bayesian. The model in the fourth column pools the two sets of observations together and adds interaction terms to test the difference, consistent with specification (4), finding a difference in updating of 0.47 with a standard error of 0.12 (significant at $p < .05$). Difference

[16] I exclude subjects whose initial beliefs were exactly 50.
[17] Results by partisanship are presented in Online Appendix Table A2 and discussed in the text below.

in fealty to prior is not statistically significant, with a coefficient of -0.085.

In the final column, I consider only subjects who identify as Democrats or Republicans, including leaners. This evaluates if partisan identifiers are more biased than independents conditional on their initial round beliefs. I find no evidence to this effect, and in fact the coefficient on consistent signals for partisans is of smaller magnitude than for all subjects (difference not statistically significant). This suggests that, conditional on initial beliefs, partisanship is not related to additional bias in processing of consistent or inconsistent signals.[18]

The regression coefficient estimates represent average learning across all subjects and rounds. To understand how cautious subjects are on average at the individual level, I pool each subject's responses to the two partisan contests in which they participated and run the regression specification from Table 2 on these eight observations, estimating the parameters of learning for each individual separately.[19] Of 911 subjects with enough variation in responses to estimate the model, 574 (63.0 percent) had a coefficient on the signal less than 1. Of those, 235 were statistically significant from 1 in a one-tailed test. Cautious learning is common at the individual level.

The overall result from Table 2 is that subjects learn from signals about political facts less than they should with perfect application of Bayes' Rule. Additionally, they learn more from signals that are consistent with their initial beliefs than signals inconsistent with those beliefs. However, subjects learn in the appropriate direction from both consistent and inconsistent signals. Importantly, even limiting the samples to party identifiers does not change this result, with partisans learning in the appropriate direction about political facts that are inconsistent with their initial beliefs. These results are inconsistent with biased assimilation, the strong form of motivated bias in information processing, even for partisans.

[18] Partisanship was measured after the experiment, which could lead to post-treatment bias: an influence of treatment assignment on the survey response. Partisanship is generally thought to be a highly stable trait, particularly the three-value version collapsing leaning partisans as partisan. I present models in Online Appendix Table A4 that indicate no influence of assigned facts or signals on these responses.

[19] Some subjects must be dropped from this analysis because they did not update their beliefs or did not enter enough responses. Dropping those who never update does change the sample on which these numbers are calculated. However, those who never update are the most cautious of all subjects, so this analysis understates caution in the total sample.

Political learning relative to other learning

Table 2 shows that subjects update their beliefs about political facts as cautious Bayesians. In this section, I benchmark this result against these same subjects learning about their relative performance on the IQ quiz. The results show that subjects learn more, on average, about political facts than about their performance on the quiz, that motivated bias appears larger on partisan facts, but that the differences on average are not particularly large.

Table 3 presents results similar to Table 2 for the rounds from the IQ contests for each subject. The first column presents the main specification from Eq. (3), with estimates of δ and β of 0.63 and 0.64. These compare to estimates from partisan facts of 0.61 and 0.73, suggesting similar weighting to prior beliefs but more learning from signals about partisan facts. Column two adds interactions with whether the signal was consistent with the subject's initial beliefs, showing that subjects do learn closer to perfect application of Bayes' Rule when the signal is consistent. The coefficient of 0.22 is not statistically significant from zero, however, and is half the size of the coefficient on the consistent interaction for partisan facts (column four, Table 2). Column three estimates the same model for partisans, finding minor differences with the results in column two.

Column four pools IQ and partisan contests together and adds interactions to test for differences. The interaction of signal and partisan fact indicates subjects learned by 5.9 percentage points closer to perfect application of Bayes' Rule on partisan facts relative to IQ facts when the signal was inconsistent with initial beliefs. When the signal was consistent with initial beliefs, subjects learned 31 points closer to perfect application of Bayes' Rule on partisan facts (0.059 + 0.25). These differences are not statistically significant, but do suggest that in this experiment subjects were more responsive to signals about partisan facts than to their performance on the IQ quiz.[20]

In sum, Table 3 compares learning about relative performance on an IQ quiz to learning about politically-relevant statements of fact. Subjects appear to learn more from signals about political statements, though differences are not generally statistically distinguishable. One interesting ob-

[20] There is a potential confound to these differences, which is that all subjects evaluated the IQ fact after evaluating the two partisan facts. It may be that order in the experiment could change the observed learning, e.g. through fatigue. To mitigate order effects, the second experiment (discussed below) randomized order.

Table 3: Learning about relative quiz performance as benchmark

VARIABLES	(1) Pooled	(2) Pooled	(3) Dems/Reps only	(4) All contests
Logit prior (δ)	0.63**	0.59**	0.58**	0.59**
	(0.03)	(0.05)	(0.06)	(0.05)
Signal (β)	0.64**	0.53**	0.57**	0.53**
	(0.07)	(0.12)	(0.14)	(0.12)
Signal*Signal consistent (β_2)		0.22	0.17	0.22
		(0.18)	(0.21)	(0.18)
Logit prior*Signal consistent (δ_2)		0.051	0.054	0.051
		(0.07)	(0.08)	(0.07)
Logit prior*Partisan fact				0.017
				(0.06)
Signal*Partisan fact				0.059
				(0.13)
Partisan*Signal*Signal consistent				0.25
				(0.22)
Partisan*Logit prior*Signal consistent				-0.14
				(0.09)
Observations	3,863	3,808	3,104	11,329
R-squared	0.437	0.438	0.424	0.442
Std. error of regression	2.34	2.33	2.37	2.35
N subjects	988	969	791	990
Wald test on null $\delta = 1$	0	0	0	0
Wald test on null $\beta = 1$	8.5e-07	0.00020	0.0036	0.00019

Robust standard errors in parentheses
** $p<0.01$, * $p<0.05$

Note: Dependent variable is logit-beliefs that the statement is correct in that round for rounds 2 through 5. A perfect Bayesian would have coefficients of 1 on both variables. Column four pools partisan and IQ contests together. Standard errors clustered on the subject-game.

servation is that there does not appear to be much variation in the weighting of prior beliefs as a function of consistency of signal or of IQ versus partisan statements. This suggests that it is not so much that subjects hold dearly to their previous beliefs and ignore new signals. Rather, subjects are more cautious in updating beliefs in response to signals inconsistent with their initial beliefs.

Learning about political versus abstract facts

In order to provide a second benchmark against which to compare political learning and to make a connection to a literature in economics with learning about ego-irrelevant abstract facts, I fielded a second experiment. From September 8 to 12, 2016, each of 395 subjects participated in an experiment similar to the first except that each subject evaluated two statements of fact. One of the statements was an abstract ego-irrelevant fact asking about the length of the day from sunrise to sunset in Doha, Qatar on January 8, 2012. The other statement was selected at random from the Obama household income and Reagan debt questions from the first experiment. Order of fact presentation was also randomized, and as before subjects received four signals accurate at probability 0.75 and were incentivized with the crossover scoring rule. Full details of the experiment are in Online Appendix Section E.

Table 4 presents the results of this second experiment. Columns one through five analyze learning about the partisan facts only, replicating the results of the first experiment: subjects learn cautiously from signals (estimate of β in column one of 0.70), learn more from signals consistent with initial beliefs and less from inconsistent signals (estimates of β of 0.99 and 0.55, columns two and three), and Democrats and Republicans do not exhibit notably greater bias or caution than pure independents (column five versus column four).

Columns six and seven evaluate learning about the abstract fact about the length of the day in Doha. Subjects are less cautious learning about this fact, with an estimate of β of 0.85 (column six) versus 0.70 (column one) for these same subjects on the political facts. These differences are not statistically distinct, however, and a general observation is that learning about the two types of facts is not too dissimilar. Column seven tests for bias towards initial beliefs on the abstract fact, and

Table 4: Bayesian learning about political versus abstract facts (Experiment 2)

VARIABLES	(1) Partisan fact	(2) Partisan fact Signal consistent	(3) Partisan fact Not consistent	(4) Partisan fact	(5) Partisan fact Dems/Reps only	(6) Abstract fact	(7) Abstract fact	(8) All facts
Logit prior (δ)	0.58**	0.51**	0.57**	0.57**	0.58**	0.57**	0.57**	0.57**
	(0.03)	(0.05)	(0.04)	(0.04)	(0.05)	(0.03)	(0.04)	(0.04)
Signal (β)	0.70**	0.99**	0.55**	0.55**	0.55**	0.85**	0.76**	0.76**
	(0.06)	(0.12)	(0.08)	(0.08)	(0.09)	(0.06)	(0.07)	(0.07)
Signal*Signal consistent ($\beta 2$)				0.44**	0.45**		0.31*	0.31*
				(0.14)	(0.16)		(0.14)	(0.14)
Logit prior*Signal consistent ($\delta 2$)				-0.061	-0.065		-0.060	-0.060
				(0.06)	(0.07)		(0.06)	(0.06)
Logit prior*Partisan fact								0.0019
								(0.06)
Signal*Partisan fact								-0.20
								(0.11)
Partisan*Signal*Signal consistent								0.13
								(0.20)
Partisan*Logit prior*Signal consistent								-0.0017
								(0.08)
Observations	1,580	702	878	1,580	1,292	1,580	1,580	3,160
R-squared	0.428	0.567	0.292	0.433	0.437	0.422	0.425	0.429
Std. error of regression	2.23	2.08	2.33	2.22	2.22	2.28	2.28	2.25
N subjects	395	278	332	395	323	395	395	395
Wald test on null $\delta = 1$	0	0	0	0	0	0	0	0
Wald test on null $\beta = 1$	3.4e-06	1	1.1e-07	9.6e-08	4.1e-07	0.028	0.0012	0.0012

Robust standard errors in parentheses
** $p<0.01$, * $p<0.05$

Note: Dependent variable is logit-beliefs that the statement is correct in that round for rounds 2 through 5. A perfect Bayesian would have coefficients of 1 on both variables. Standard errors clustered on the subject-game.

finds subjects are biased towards their initial beliefs with an interaction estimate of 0.31. This bias interaction point estimate is smaller than the bias estimate for partisan facts of 0.44 (column 4). Subjects learn more from inconsistent signals on this abstract fact than from inconsistent signals about the political fact, 0.76 (column seven) versus 0.55 (column four).

Column eight pools partisan facts with the abstract fact and tests for differences in learning. Differences in fealty to prior beliefs (δ) are of very small magnitude. In general the differences in learning from signals measured by the interactions are not of large substantive magnitude and are not statistically significant. The interaction of signal and partisan coefficient suggests subjects learn 20 percent less than perfect application of Bayes' Rule (coefficient of -0.20) from inconsistent signals on political facts relative to inconsistent signals on abstract facts.

The overall results from the second experiment are that learning about political facts is roughly similar to learning about abstract facts. The finding of imperfect Bayesian learning about the abstract facts also shows that this experimental setting does not necessarily generate overly rational behavior among subjects. This amount of learning is similar to learning observed in other experiments about non-political ego-irrelevant facts, with Bayes' Rule being a fair but imperfect descriptor of individual behavior.

Partisanship and other moderators to learning

The experimental design here allows a different analysis than the standard approach in the political science literature on motivated bias, which looks for differences in response to signals by partisanship regardless of initial beliefs. In Online Appendix Table A2, I present results separately by question and partisanship, setting aside initial round beliefs as a definition of consistency. The conclusions from Table 2 hold in this analysis. Subjects update beliefs in the appropriate direction yet cautiously relative to perfect application of Bayes' Rule. Subjects exhibit some bias in this cautious updating, with signals consistent with their partisan identity leading to more learning than inconsistent signals (larger coefficients for Democratic-favored facts for True relative to False and the opposite for Republicans). Finally, there is little evidence of biased assimilation. While a few coefficients are estimated greater than one, their magnitude is not dramatic. The largest coefficient estimate is 1.28, 28 percent greater than Bayesian learning. Strangely, this coefficient is estimated for Republicans responding to a True signal when True *opposes* their Republican president (Reagan debt question) – in the direction opposite of biased assimilation. There is only one coefficient estimated less than zero, not statistically significant, for Republicans learning in the wrong direction to signals that household income *did not* fall at a fast rate under Democratic President Obama. This is consistent with biased assimilation ($\beta \ll 0$). While this may be consistent with an argument that only some issues generate biased assimilation, it is also consistent with sampling variability and multiple testing. Questions on abortion and Ronald Reagan do not exhibit biased assimilation. The overall pattern is one of cautious and direction-appropriate learning.

Variation in learning by political behaviors and attitudes

I present in Online Appendix Table A3 variation in learning about political facts by individual characteristics: primary voters, ideology, interest in political compromise, and political activity, a set of questions from the 2014 Pew Polarization Survey related to political polarization.[21] Two key results from Table A3 are that primary voters exhibit more bias in learning than non-primary voters and that self-described liberals and conservatives exhibit more bias in learning than moderates. These results suggest that among specific subsets of the population, learning may depart more from the Bayesian ideal. Even so, these subsets learn in the appropriate direction with only the magnitude of caution varying with the consistency of signal.

Robustness and threats to inference

In this section, I consider robustness to three threats to inference. First, I show that an alternative model of learning separate from Bayes' Rule does not more effectively explain observed responses. I then show that results of the analysis are robust to levels of attention and to choice over post-stratification weight construction.

[21] See the note to Table A3 for exact question wording. Each of these characteristics was measured after assignment to treatment. To assess potential post-treatment bias, I show in Online Appendix Table A4 that randomized assignment to facts and signals do not predict any of these measurements.

Robustness to alternative model of learning

An alternative model of learning with bias is a tipping point model (e.g., Gerber and Green, 1998, p. 816). Under such a model, subjects respond cautiously to each individual signal, but once a set of signals accumulates in a consistent fashion (e.g., four true signals or four false signals), this pushes subjects beyond their biases to update their beliefs. In Online Appendix Section G, I present a nonparametric evaluation of whether a tipping point model better characterizes learning of political information than the Bayesian model. A tipping point model of learning suggests that the largest revisions of beliefs should be for subjects who receive a consistent set of signals, say TTT, FFF, TTTT, or FFFF, thus "tipping" them over into finally updating their beliefs. The Bayesian model of learning, in contrast, is memoryless: At any belief, a true or false signal has the same meaning regardless of the prior pattern of signals because previous signals are fully reflected in the prior.

For each pattern of signals received by participants I tabulate mean and median revision in beliefs to the most recent signal. Online Appendix Table A7 presents revisions for each pattern, sorted descending by largest absolute revision in belief. It shows that the largest revisions almost always occur in cases with a mixed set of true and false signals, rather than the consistent signals of a tipping point pattern. All but one of the tipping point patterns for partisan contests occur in the final rows of the table with the smallest revisions. For IQ contests, the first tipping point pattern is about one third down the table with the remaining in the bottom third of the table. In sum, a tipping point model of learning does not appear to be a more effective explanation of the observed learning behavior than the Bayesian model.

Robustness to weighting approach

A second threat to interpretation is the choices made in generating post-stratification weights to make the MTurk sample look like the 2014 Pew Polarization sample. The weights used throughout the paper are created by raking the marginal distributions of 12 variables from the MTurk sample to the marginal distribution of those same variables in the Pew sample, with trimming of weights to limit variance. To show results are robust to raking choices, I reproduce the main results (Tables 1, 2, and 3) in Online Appendix Section H without any survey weights (Online Appendix Tables A8, A9, and A10). Comparison of the results shows conclusions about cautious and modestly biased learning hold with the unweighted analysis.

Robustness to attention

A third threat to the interpretation of the results of this experiment is the attentiveness of the experimental subjects. One common concern about samples from MTurk is that the attentiveness of participants to tasks diverges from what one would expect in other settings. Some argue that MTurk workers are less attentive, trying to complete tasks as quickly as possible with minimal effort. Others argue workers are too attentive because they are paid for each task and invested in gaining the approval of employers for future opportunities. Empirical evidence, however, suggests research using MTurk samples produces similar results to other samples (Berinsky, Huber, and Lenz, 2012; Mullinix et al., 2015).

To evaluate the potential influence of either of these concerns, I reproduce Table 2 in Online Appendix Tables A5 and A6 separating subjects by their score on the intelligence quiz, top and bottom half. I assume that it takes more effort and attention to score highly on the quiz, while those with poor scores are likely not paying as much attention. Readers can use these different estimates as a benchmark relative to whatever concern they might have about the level of attention from the sample. Score on the quiz also likely partially reflects the subject's numeracy. The patterns from Table 2 do not vary dramatically by score on the quiz.[22]

Discussion: Lower or upper bound on learning?

Although citizens may learn more slowly than the Bayesian ideal, the amount of learning I document here might be interpreted as impressively large. Compared to relatively slow changes in aggregate series of public opinion, the subjects in this experiment updated beliefs in some cases to a striking degree. Beliefs moved from average probability 57 that abortions fell more under

[22] Low scorers update less consistently with Bayes' Rule than do high scorers, but both exhibit caution in updating and bias towards signals consistent with their initial beliefs.

Reagan than under Clinton when first presented with the statement to average probability 36 after four noisy signals, and from 58 percent to 73.5 percent average beliefs that income fell more than 4 percent during the first term of Obama (Table 1). These are large changes in aggregate beliefs and suggest what is possible when average citizens are presented unambiguous if noisy signals.

One important question is whether the setting of this experiment is closer to a best or worst case to see Bayesian learning. There are considerations on both sides. In support of the setting being closer to an upper bound on learning, the subjects are given single signals about challenging but clear statements of fact without the complication of countervailing information. They are provided incentives to give accurate responses and participation in a survey run by an academic researcher may lead to greater trust and attentiveness.

On the other hand, a variety of considerations suggest this may not be an upper bound on learning. First, participants were unlikely to be familiar with the technology used to elicit beliefs and compensate for accuracy. Citizens in the real world are likely more familiar with their own information environments. They have developed experience and strategies to learn what they need to know, and these strategies may not easily translate to this lab setting. The departure from Bayesian learning I document here is similar to departures measured in non-political contexts with lab experiments using undergraduates (e.g., Anderson and Holt, 1997, find their subjects behave 73 percent consistent with Bayes' Rule, nearly exactly what I estimate here).

Additionally, while it is the case that the signals from the computer were unambiguous, the noise with which they were delivered (only being accurate three out of four times) can represent a variety of the complications that confront citizen information processing outside the lab. Interpreting information as a noisy yet informative signal with respect to the political fact to be evaluated is similar to information processing tasks in a complicated world. The likelihood ratio can also represent multiple signals from difference sources from a more competitive information environment. Citizens are faced with these kinds of complicated combinations of information every day in their economic, social, and political experiences. Ultimately, however, this question can only be answered by new designs and research in alternative settings.

Conclusion: What to make of cautious Bayesian political citizens

The results of this experiment suggest that citizens do not learn political information as perfect Bayesians. They are cautious in responding to signals delivered, and are modestly biased in response to signals by consistency with their initial beliefs. Nonetheless, subjects are capable of learning in the appropriate direction about partisan-relevant facts and appear to learn in a similar fashion about political and non-political facts. I considered an alternative tipping point model of learning, which appeared to be much less consistent with the observed learning than a Bayesian model. Thus, Bayes' Rule seems a reasonable model of the processing of political information, even if learning is somewhat slower than ideal. Citizens learn together slowly about political facts.[23]

My conclusions differ from those of some existing work on political information and suggest the need for further research. First, the subjects in this experiment were delivered single signals without any choice as to content. In other contexts, in contrast, individuals get to *choose* what information to consume and process, e.g. reading only parts of newspaper articles or selecting which television programs to watch. My evidence suggests that citizens are capable of learning together slowly when presented with common but noisy information about the truth. An open question remains how large a problem selective exposure is for political facts.

Second, I provided financial incentives for correct responses. Particularly outside of the laboratory but even in the laboratory of many existing studies, no one is directly paying citizens $0.10 for each correct political answer. This implies that we consider if outside of the laboratory citizens perceive their incentive to learn political information as more or less valuable than the $0.10 offered here. There is evidence that many citizens behave as if they believe their political choices have important consequences – clearly, the influence of aggregated political behavior on policy can

[23] As a real world example, consider opinions about the guilt of OJ Simpson in the famous homicides of Nicole Brown Simpson and Ron Goldman. ABC News/Washington Post asked national samples whether Simpson was guilty of these murders in three surveys over two decades, July 1994, September 2007, and September 2015. The rate believing definitely or probably guilty among white respondents grew from 63 to 74 to 83 percent over this time period. Among black respondents, the rate believing definitely or probably guilty grew from 22 to 45 to 57 percent. Even on this racially-charged issue, this is evidence that Americans learn together, if slowly. I thank Don Green for this anecdote.

be large. Perhaps for this reason, many citizens make the costly effort to turn out to vote (even in large elections, e.g., Edlin, Gelman, and Kaplan, 2007; Feddersen and Sandroni, 2006), and many make the effort to consume political news. If voters perceive their choices as important, they would value the acquisition of political information such that the learning measured here is similar to the process outside of the laboratory.

Another implication of the finding that citizens can learn in a fashion close to Bayes' Rule is that voters in the real world may indeed learn about political information as Bayesians but the challenges to measuring this learning have led many extant studies to different conclusions. That is, many individuals may feel a duty to be good democratic citizens such that they do derive utility from "getting it right" in a way analogous to the small monetary incentives provided here. Because most existing evidence captures an apparent departure from Bayesian learning but not the magnitude of this departure, it remains possible that learning of political information is not too far from Bayesian. This article provides a framework for studying learning of political information that can be extended in future experiments and in the real world. The evidence here suggests that citizens do converge towards similar beliefs, even with cautious and biased learning.

If in the real political world citizens fall more short of the Bayesian ideal than in this experiment, it is not necessarily due to their own partisan bias (Bartels, 2002) or cognitive limitations (Huber, Hill, and Lenz, 2012). The political world does not provide citizens strong incentives to invest in political learning, as Downs (1957) long ago noted. Thus, if learning does in fact fall short, we might evaluate the political institutions and elite behaviors that do not provide the incentives for citizens to learn or fully evaluate new political information. More broadly, there are a variety of models of voter behavior that conclude that citizens need not be fully informed on every political issue to enact accountability from their representatives (e.g., Lupia, 1994; Popkin, 1991), or may even benefit from being under-informed (Ashworth and Bueno de Mesquita, 2014). Theoretical treatments should consider how much is "enough" learning.

Voters are not asked to make perfect, continuous judgments about political facts. Rather, they must make categorical choices in contests weighing multiple complicated political facts. It may be that getting it close to right works almost as well in a noisy world with political conflict across multiple policy dimensions as learning as a perfect Bayesian, yet without the full costs. In fact, there may be some value in updating beliefs cautiously. There may even be value to learning with bias towards the political coalition with which you align. Note that updating beliefs about one fact may have consequences for beliefs about other facts and preferences (e.g., Andreoni and Mylovanov, 2012; Lauderdale, 2015), and that other incentives may structure both caution and bias in learning. Future theoretical and empirical research should explore more specifically the welfare implications of cautious and biased Bayesian learning.

More broadly, formal models of political accountability assume voters learn through noisy signals about incumbent performance via Bayes' Rule. The results here clarify that this assumption is plausible but does not hold perfectly. Future models may want to account for caution and bias, either in relaxing previous assumptions or in building into the models some feature of citizen decision-making that rationalizes caution and bias. It may be that joining a long coalition has implications for how citizens learn about politically-relevant facts. As such, more theory and evidence is needed to evaluate the effectiveness of cautious Bayesian learning for democratic citizens.

References

Allen, Franklin. 1987. "Discovering Personal Probabilities When Utility Functions Are Unknown." *Management Science* 33(4): 542–544.

Anderson, Lisa R., and Charles A. Holt. 1997. "Information Cascades in the Laboratory." *American Economic Review* 87(5): 847–862.

Andreoni, James, and Tymofiy Mylovanov. 2012. "Diverging Opinions." *American Economic Journal: Microeconomics* 4(1): 209–232.

Ashworth, Scott, and Ethan Bueno de Mesquita. 2014. "Is Voter Competence Good for Voters?: Information, Rationality, and Democratic Performance." *American Political Science Review* 108(3): 565–587.

Bartels, Larry M. 2002. "Beyond the Running Tally: Partisan Bias in Political Perceptions." *Political Behavior* 24(2): 117–150.

Berelson, Bernard R., Paul F. Lazarsfeld, and William N. McPhee. 1954. *Voting: A Study of Opinion Formation in a Presidential Campaign*. Chicago: University of Chicago Press.

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.

Bolsen, Toby, James N. Druckman, and Fay Lomax Cook. 2014. "The Influence of Partisan Motivated Reasoning on Public Opinion." *Political Behavior* 36: 235–262.

Bullock, John G. 2009. "Partisan Bias and the Bayesian Ideal in the Study of Public Opinion." *Journal of Politics* 71(3): 1109–1124.

Bullock, John G., Alan S. Gerber, Seth J. Hill, and Gregory A. Huber. 2015. "Partisan Bias in Factual Beliefs About Politics." *Quarterly Journal of Political Science* 10(4): 519–578.

Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. 1960. *The American Voter*. New York: Wiley.

Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper Collins.

Edlin, Aaron, Andrew Gelman, and Noah Kaplan. 2007. "Voting as a Rational Choice: Why and How People Vote to Improve the Well-Being of Others." *Rationality and Society* 19(3): 293–314.

Feddersen, Timothy J., and Alvaro Sandroni. 2006. "A Theory of Participation in Elections." *American Economic Review* 96(4): 1271–1282.

Fiorina, Morris P. 1981. *Retrospective Voting in American National Elections*. New Haven: Yale University Press.

Gaines, Brian J., James H. Kuklinski, Paul J. Quirk, Buddy Peyton, and Jay Verkuilen. 2007. "Same Facts, Different Interpretations: Partisan Motivation and Opinion on Iraq." *Journal of Politics* 69(4): 957–974.

Gerber, Alan S., and Donald P. Green. 1998. "Rational Learning and Partisan Attitudes." *American Journal of Political Science* 42(3): 794–818.

Gerber, Alan S., and Donald P. Green. 1999. "Misperceptions About Perceptual Bias." *Annual Review of Political Science* 2: 189–210.

Guess, Andrew, and Alexander Coppock. N.d. "The Exception, Not the Rule? The Rarely Polarizing Effect of Challenging Information." Working paper, Yale University, available at https://alexandercoppock.files.wordpress.com/2014/11/backlash_apsa.pdf.

Holt, Charles A., and Angela M. Smith. 2016. "Belief Elicitation with a Synchronized Lottery Choice Menu That Is Invariant to Risk Attitudes." *American Economic Journal: Microeconomics* 8(1): 110–139.

Huber, Gregory A., Seth J. Hill, and Gabriel S. Lenz. 2012. "Sources of Bias in Retrospective Decision Making: Experimental Evidence on Voters Limitations in Controlling Incumbents." *American Political Science Review* 106(4): 720–741.

Jerit, Jennifer, and Jason Barabas. 2012. "Partisan Perceptual Bias and the Information Environment." *Journal of Politics* 74(3): 672–684.

Karni, Edi. 2009. "A Mechanism for Eliciting Probabilities." *Econometrica* 77(2): 603–606.

Lauderdale, Benjamin E. 2015. "Partisan Disagreements Arising from Rationalization of Common Information*." *Political Science Research and Methods* 4(3): 477–492.

Lord, Charles G., Lee Ross, and Mark R. Lepper. 1979. "Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence." *Journal of Personality and Social Psychology* 37(11): 2098–2109.

Lupia, Arthur. 1994. "Shortcuts Versus Encyclopedias: Information and Voting Behavior in California Insurance Reform Elections." *American Political Science Review* 88(1): 63–76.

Möbius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat. 2011. "Managing Self-Confidence." NBER Working Paper 17014.

Mullinix, Kevin J., Thomas J. Leeper, James N. Druckman, and Jeremy Freese. 2015. "The Generalizability of Survey Experiments." *Journal of Experimental Political Science* 2(2): 109–138.

Nyhan, Brendan, and Jason Reifler. 2010. "When Corrections Fail: The Persistence of Political Misperceptions." *Political Behavior* 32(2): 303–330.

Popkin, Samuel L. 1991. *The Reasoning Voter*. Chicago: University of Chicago Press.

Prior, Markus, and Arthur Lupia. 2008. "Money, Time, and Political Knowledge: Distinguishing Quick Recall and Political Learning Skills." *American Journal of Political Science* 52(1): 169–183.

Prior, Markus, Gaurav Sood, and Kabir Khanna. 2015. "You Cannot be Serious: The Impact of Accuracy Incentives on Partisan Bias in Reports of Economic Perceptions." *Quarterly Journal of Political Science* 10(4): 489–518.

Rahn, Wendy M. 1993. "The Role of Partisan Stereotypes in Information Processing about Political Candidates." *American Journal of Political Science* 37(2): 472–496.

Taber, Charles S., and Milton Lodge. 2006. "Motivated Skepticism in the Evaluation of Political Beliefs." *American Journal of Political Science* 50(3): 755–769.

Woon, Jonathan. 2012. "Democratic Accountability and Retrospective Voting: A Laboratory Experiment." *American Journal of Political Science* 56(4): 913–930.

Zaller, John. 1992. *The Nature and Origins of Mass Opinion*. New York: Cambridge University Press.