

# Partisan Bias and the Bayesian Ideal in the Study of Public Opinion

**John G. Bullock** Yale University

*Bayes' Theorem is increasingly used as a benchmark against which to judge the quality of citizens' thinking, but some of its implications are not well understood. A common claim is that Bayesians must agree more as they learn and that the failure of partisans to do the same is evidence of bias in their responses to new information. Formal inspection of Bayesian learning models shows that this is a misunderstanding. Learning need not create agreement among Bayesians. Disagreement among partisans is never clear evidence of bias. And although most partisans are not Bayesians, their reactions to new information are surprisingly consistent with the ideal of Bayesian rationality.*

In February 2004, 11 months after the American invasion of Iraq, emerging evidence suggested that Iraq had not possessed weapons of mass destruction at the time of the invasion. Partisans were divided on the matter: 40% of Democrats reported believing that Iraq had WMD at the time of the invasion, against 83% of Republicans who said the same (ABC News/The Washington Post 2004). In the following year, the evidence became clearer. The failure of the Iraq Survey Group to locate WMD was widely publicized, and the chief U.S. weapons inspector declared that WMD “were not there” at the time of the invasion (Duelfer 2004, 6). But the revelations made little difference: by March 2005, 35% of Democrats said that Iraq had WMD at the time of the invasion, and 78% of Republicans said the same (ABC News/The Washington Post 2005). The gap between partisans’ beliefs was as wide as ever.

Public opinion scholars will recognize this as a new instance of an old pattern. Enduring disagreements between Republicans and Democrats are a hallmark of American politics, and they have long been taken as evidence that identifying with a party biases partisans’ subsequent political views (e.g., Berelson, Lazarsfeld, and McPhee 1954, Chap. 10). Of course, some disagreements may be due to differences in values that are causally prior to party ID (Downs 1957; Fiorina 1981), but value differences cannot explain partisan disagreements over factual questions and other matters to which values are irrelevant. Such

disagreements seem to be especially strong evidence of partisan bias (Bartels 2002).

The most direct challenge to this view comes from Gerber and Green (1999), who maintain that even these “hard cases” are not evidence of partisan bias. They use a Bayesian model of belief revision to argue that even when Republicans and Democrats share the same values and are free from partisan bias, disagreements between them can endure. The argument is unusual, but the model is not: Bayesian models like theirs are increasingly used as frameworks through which changes in public opinion are studied (e.g., Achen 1992; Bartels 1993; Gerber and Jackson 1993; Grynaviski 2006; Husted, Kenny, and Morton 1995; Lohmann 1994; Zechman 1979) and as normative benchmarks against which to judge the quality of citizens’ thinking about politics (e.g., Bartels 2002; Gerber and Green 1999; Steenbergen 2002; Tetlock 2005).

Bartels (2002, 119–23) counters that Gerber and Green misunderstand their model: unbiased application of Bayes’ Theorem, he writes, implies not enduring disagreement but increasing agreement whenever people revise their beliefs in response to new information. Achen (2005, 334), Grynaviski (2006, 331), and Goodin (2002) agree in a series of thoughtful articles, with Grynaviski declaring that unbiased Bayesians “inexorably come to see the world in the same way” and Goodin finding something sinister: “Bayesian rationality requires minorities to cave in

too completely, too quickly.” But none of these articles prove any properties of the Bayesian model that lies at the heart of the dispute, and as this article shows, each of them elides vital points about the connection between Bayesian learning and partisan disagreement. In the very long run, partisans who revise their beliefs according to Bayes’ Theorem will agree if they receive the same evidence and interpret it in the same way. But in the shorter term—the one that we care about when we study public opinion trends over months, years, and election campaigns—Bayesian learning is compatible with sustained disagreement. And contrary to Bartels (2002, 121–23), Tetlock (2005, 128), and others, Bayesian learning is compatible with the polarization of attitudes and beliefs—even when people receive the same information and interpret it in the same way.

These findings show that enduring disagreement and polarization can never be *ipso facto* proof of partisan bias, because they can occur even among unbiased Bayesians. They also matter because of the expanding role that Bayesian models play in analyses of public opinion. Partisans are not Bayesians, but Bayesian updating matters because it is increasingly the normative standard against which partisan updating is judged. We have an interest in sorting out what it entails.

I begin by reviewing Bayes’ Theorem and considering objections to its use in the study of political beliefs. The next section proves properties of the dominant Bayesian model of public opinion, establishing conditions under which learning will create agreement and conditions under which it will promote disagreement. This model pertains only to learning about political conditions that never change; in the following section, I generalize the model and the results to account for learning about changing political circumstances. I proceed by arguing that the conditions for “Bayesian agreement” established in the previous sections are rarely met in practice. The last section reviews and concludes.

## The Utility of Bayesian Models of Public Opinion

Many matters that interest political scientists—popular beliefs about politics, politicians’ true preferences, implications of new policies—can be thought of as probability distributions. Like most distributions, they have means and variances, and the task that we set for ourselves is to learn about these parameters.

Messages about a politician’s ability to manage the economy, for example, may oscillate around a fixed but unknown mean, which is the politician’s true ability level. Learning about politics becomes a matter of learning about probability distributions—a task to which Bayesian statistics is especially well-suited.

The bedrock of Bayesian statistics is Bayes’ Theorem, an equation that relates conditional and marginal probabilities:

$$p(S|E) = \frac{p(E|S)p(S)}{p(E)},$$

where  $S$  and  $E$  are events in a sample space and  $p(\cdot)$  is a probability distribution function. Stated thus, the Theorem is merely an accounting identity, but a change in terminology draws out its significance. This time, let  $S$  be a statement about politics and  $E$  be evidence bearing on the statement. Before observing  $E$ , a person’s *belief* about  $S$ —that is, a probability distribution indicating his estimate of the extent to which  $S$  is true—is given by  $p(S)$ . Because it is his belief about  $S$  prior to observing  $E$ , it is often called his *prior probability* of  $S$ , or simply his “prior.” After he observes  $E$ , his estimate of the probability that  $S$  is true is  $p(S|E)$ , often called his *posterior probability* of  $S$ . And  $p(E|S)$  is the likelihood function that he assigns to the evidence; it reflects his guess about the probability distribution from which the evidence is drawn. Understood in this way, Bayes’ Theorem tells us how to revise any belief after we have received relevant evidence and subjectively estimated its likelihood. It is most often applied to beliefs about future events (e.g., Tetlock 2005, Chap. 4), but it is fundamentally a tool for calculating probabilities, and it applies with equal force to attitudes about politicians, evaluations of their abilities, and all other ideas that can be described in probabilistic terms.<sup>1</sup>

The charge most frequently leveled against Bayesian models of political belief revision is that people do not update their beliefs as Bayes’ Theorem demands. The charge has merit,<sup>2</sup> but it misses two important points. One is that use of Bayesian models in public opinion research does not always entail an

<sup>1</sup>See Bartels (1993, 2002) and Gerber and Green (1998, 1999) for applications of Bayesian models to political attitudes and evaluations.

<sup>2</sup>The most straightforward and best-documented violation of Bayesian updating is “cognitive conservatism,” the overweighting of prior beliefs when updating (e.g., Edwards 1982). Tetlock (2005), Peffley, Feldman, and Sigelman (1987), and Steenbergen (2002) argue that the effect holds generally for political beliefs.

assumption that people use Bayes' Theorem to revise their beliefs. Often, the models instead provide a normative benchmark against which to judge real belief revision (e.g., Bartels 2002; Steenbergen 2002; Tetlock 2005). To know how far people fall from a normative ideal of belief revision, we need to specify a normative ideal, and the increasing use of Bayesian public opinion models means that unbiased Bayesian updating is that ideal.

The other important point is that Bayesian models of public opinion can be heuristically useful even if we wrongly assume that people are Bayesians, because they offer a systematic way to account for the relative influences of old beliefs and new information (Fischhoff and Lichtenstein 1978, 242). Bartels' 1993 study of media effects in the 1980 presidential campaign is a good example. Bartels assumes Bayesian updating among respondents in the American National Election Studies and concludes that voters had strong prior beliefs about the candidates at the outset of the campaign.<sup>3</sup> His data are also consistent with the assumption that ANES subjects had weak priors but overweighted them, i.e., that they acted not as Bayesians but as "cognitive conservatives" (Edwards 1982). But from a political standpoint, there is little difference between weighting strong priors according to Bayes' Theorem and overweighting weak priors. The substantive upshot is usually the same, and so long as it is, the assumption of Bayesian updating can be useful even when it is wrong.

### Disagreement When Learning about an Unchanging Condition

In studies of public opinion, one Bayesian learning model is far more common than the rest. It assumes that prior beliefs are normally distributed and that people perceive new information to be normally distributed too (e.g., Achen 1992; Bartels 1993, 2002; Gerber and Green 1999; Husted, Kenny, and Morton 1995; Lohmann 1994; Steenbergen 2002; Zechman 1979). To see how it works, suppose that a voter is trying to learn about a politician's level of honesty. She conceives of honesty as a continuum—there are infinitely many degrees of honesty and dishonesty—and the politician's level of honesty is an unknown point on this continuum. We call it  $\mu$ . Initially, the voter's belief about  $\mu$  is normally distributed:

<sup>3</sup>The chief moral of Bartels' article is that media effects must also have been strong—stronger than previously believed—to change prior beliefs as much as they did.

$\mu \sim N(\hat{\mu}_0, \sigma_0^2)$ . The mean of this distribution,  $\hat{\mu}_0$ , is the voter's best guess at time 0 about the politician's honesty. The variance of this distribution,  $\sigma_0^2$ , indicates the confidence that she reposes in this guess. If  $\sigma_0^2$  is low, she is quite sure that the politician's true level of honesty is around  $\hat{\mu}_0$ ; if  $\sigma_0^2$  is high, she allows that his true level of honesty might well be far from  $\hat{\mu}_0$ .

Later, the voter encounters a new message that contains information about the politician's level of honesty. She interprets the message as having value  $x_1$ , where higher values suggest greater honesty on the politician's part. She assumes that the message is a draw from a distribution with a mean of  $\mu$ .<sup>4</sup> The normal distribution is usually a sensible assumption: if the message can theoretically assume any real value, and if error or "noise" is likely to be contributed to it by many minor causes, the central limit theorem suggests that it is likely to be normally distributed. We write  $x_1 \sim N(\mu, \sigma_x^2)$ . The variance of this distribution,  $\sigma_x^2$ , captures how definitive the new information seems to the voter. If it comes directly from a source that she trusts, the signal it sends is clear and the variance is low; but if it is a rumor, the signal is weak and the variance is high.

By a common result (e.g., Lee 2004, 34–36), a voter with a normal prior who updates according to Bayes' Theorem in response to  $x_1$  will have posterior belief  $\mu|x_1 \sim N(\hat{\mu}_1, \sigma_1^2)$ , where

$$\hat{\mu}_1 = \hat{\mu}_0 \left( \frac{\tau_0}{\tau_0 + \tau_x} \right) + x_1 \left( \frac{\tau_x}{\tau_0 + \tau_x} \right), \quad \text{and} \quad (1a)$$

$$\sigma_1^2 = \frac{1}{\tau_0 + \tau_x}, \quad (1b)$$

and where  $\tau_0 = 1/\sigma_0^2$  and  $\tau_x = 1/\sigma_x^2$  are the *precisions* of the prior belief and the new message. The mean of the posterior belief,  $\hat{\mu}_1$ , is a weighted average of the mean of the prior belief and the new message; the weights are determined by the precisions. Note that although  $\hat{\mu}$  and  $x$  are indexed by time—we can have  $\hat{\mu}_2$ ,  $x_3$ , and so on— $\mu$  is not indexed. The implicit assumption of this model is that people are learning about a political condition that is not changing.

A useful aspect of this model is that it permits direct comparison of the strength of voters' prior

<sup>4</sup>This is tantamount to assuming that the message comes from an unbiased source. If the voter believes that the message comes from a biased source, she is likely to adjust for that bias before revising her belief. This is no obstacle to Bayesian updating (e.g., Jackman 2005), but it is not part of the model that is typically considered in political science.

beliefs to the strength of the new information that they receive. For example, consider a Bayesian voter whose belief about a politician's honesty is  $\mu \sim N(2, 2)$ . This belief can be thought of as reflecting all of the relevant information about the politician's honesty that the voter has received throughout her life. Imagine that she then receives message  $x = 4$  with variance  $\sigma_x^2 = 2$ , suggesting that she had underrated the politician's honesty. Because the variances of the new message and the prior belief are equal, the new message seems as "strong" to her as all of the information that she had previously received. By equations (1a) and (1b), she accords equal weight to her prior belief and the new information, and her new belief is thus  $\mu \sim N(3, 1)$ . Comparability of beliefs and information is a major virtue of Bayesian learning models, and we will make use of it throughout this article.

Debate about the possibility of increasing disagreement under unbiased Bayesian updating revolves around the properties of this model. Of course, some disagreements cannot be resolved under any circumstances. They reflect value differences that will not be resolved by more learning.<sup>5</sup> This is widely understood in the debate over Bayesian updating of public opinion. The case in which people receive the same evidence, interpret it in the same way regardless of value differences, and follow the same updating rule should be the hardest case for which to prove the possibility of increasing disagreement; it is the case that I take up here.

When disagreement does not last—when people disagree before updating and agree after updating—their views have *converged to agreement*. Before embarking on a series of proofs, it will help to define this term and several others:

- Let  $\hat{\mu}_{Dt}$  be the mean of voter  $D$ 's belief about  $\mu$  at time  $t$ . Let  $\hat{\mu}_{Rt}$  be the mean of voter  $R$ 's belief about  $\mu$  at time  $t$ . If  $\hat{\mu}_{Dt} = \hat{\mu}_{Rt}$ ,  $D$  and  $R$  agree at time  $t$ . If  $\hat{\mu}_{Dt} \neq \hat{\mu}_{Rt}$ , they disagree at time  $t$ .
- Beliefs *converge* between time  $t$  and later time  $u$  if and only if  $|\hat{\mu}_{Dt} - \hat{\mu}_{Rt}| > |\hat{\mu}_{Du} - \hat{\mu}_{Ru}|$ . They *diverge* if and only if  $|\hat{\mu}_{Dt} - \hat{\mu}_{Rt}| < |\hat{\mu}_{Du} - \hat{\mu}_{Ru}|$ .
- Beliefs *polarize* between time  $t$  and time  $u$  if and only if they diverge between  $t$  and  $u$  and  $(\hat{\mu}_{Du} - \hat{\mu}_{Dt})/(\hat{\mu}_{Ru} - \hat{\mu}_{Rt}) < 0$ .

<sup>5</sup>A related point is made in the economics literature on "agreeing to disagree" (e.g., Aumann 1976), which assumes people who share the same prior belief and whose posterior beliefs are common knowledge. Neither assumption is appropriate to the study of public opinion, which is why I do not take up the "agreeing to disagree" literature here.

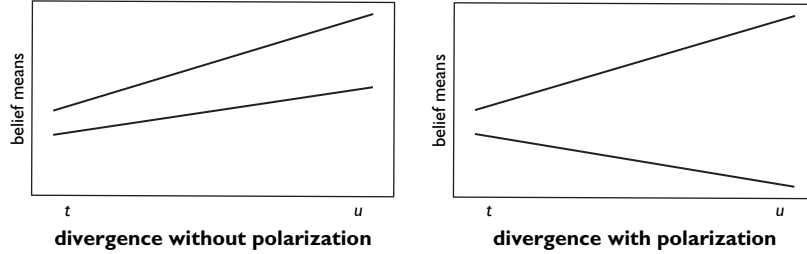
- Suppose that  $D$  and  $R$  update their beliefs in response to a sequence of messages  $x_1, \dots, x_n$ . If and only if  $\text{plim}(\hat{\mu}_{Dn} - \hat{\mu}_{Rn}) = 0$ , their beliefs will *converge to agreement* as  $n \rightarrow \infty$ .
- A person *learns* when his belief changes in response to a message.
- People are *unbiased* if they correctly perceive the value and distribution of the messages that they encounter. For example, an unbiased person will perceive message  $x$  from distribution  $N(\mu, \sigma_x^2)$  as having value  $x$  and as drawn from  $N(\mu, \sigma_x^2)$ . A biased person encountering the same message will perceive it as having a value other than  $x$  or as drawn from a distribution other than  $N(\mu, \sigma_x^2)$ . Note that unbiased partisans necessarily interpret messages in the same way: they agree on the "locations" of the messages that they receive and perceive those messages to be drawn from the same distributions.

The distinction between divergence and polarization is crucial. Divergence implies that the gap between updaters' beliefs is widening between time  $t$  and time  $u$ . Polarization implies both that this gap is widening and that updaters' beliefs are moving in different directions. Divergence can occur even as all beliefs move in the direction of new evidence, but polarization requires that at least one person's belief move away from the new evidence, a strange phenomenon that Tetlock claims is not just "contra-Bayesian" but also "incompatible with all normative theories of belief adjustment" (2005, 128). As we shall see, these claims are too strong.

Figure 1 illustrates the distinction between divergence and polarization, which is not drawn in psychology, political science, or the popular press. In all three fields, "polarization" is used to describe both phenomena; for example, when political scientists maintain that Bayesian updating is incompatible with "polarization," they generally mean that it cannot even accommodate divergence (e.g., Gerber and Green 1999, 199). Maintaining the distinction is important to understanding Bayesian models of public opinion: such models accommodate divergence more easily than polarization.

With these definitions in hand, we can establish conditions under which updating will produce convergence, divergence, and polarization of beliefs. In all cases, we consider two unbiased Bayesians,  $D$  and  $R$ , who have prior beliefs  $\mu \sim N(\hat{\mu}_{D0}, \sigma_{D0}^2)$  and  $\mu \sim N(\hat{\mu}_{R0}, \sigma_{R0}^2)$ , respectively. We assume  $0 < \sigma_{D0}^2, \sigma_{R0}^2$ : neither  $D$  nor  $R$  is absolutely certain of

**FIGURE 1 Divergence with and without Polarization.** Each panel depicts the means of two people’s beliefs at times  $t$  and  $u$ . In the left-hand panel, the beliefs are diverging but not polarizing: although the gap between the means is growing, they are moving in the same direction. In the right-hand panel, the beliefs are diverging and polarizing: the gap between the means is growing and they are moving in opposite directions. By definition (see previous page), polarization requires divergence.



his prior belief. All proofs appear in the online appendix.

**Proposition 1 (convergence to agreement).** *If  $D$  and  $R$  receive a sequence of messages  $x_1, \dots, x_n \stackrel{iid}{\sim} N(\mu, \sigma_x^2)$  where  $\sigma_x^2$  is finite, their beliefs will converge to agreement as  $n \rightarrow \infty$ .*

Proposition 1 shows that unbiased Bayesian learning will indeed create agreement among partisans—eventually. The result is familiar to statisticians (see the online appendix for discussion), but it speaks

$$a = (\hat{\mu}_{D0} - \hat{\mu}_{R0})[(\tau_{D0} + \tau_{x^*})(\tau_{R0} + \tau_{x^*}) - \tau_{D0}\tau_{R0}] \text{ and}$$

$$b = (\hat{\mu}_{R0} - \hat{\mu}_{D0})[(\tau_{D0} + \tau_{x^*})(\tau_{R0} + \tau_{x^*}) + \tau_{D0}\tau_{R0}].$$

This condition is somewhat complicated, but it has a straightforward explanation. Just as any voter’s posterior belief mean is a weighted average of his prior belief mean and the new information that he receives, the distance between any two voters’ posterior belief means is a weighted average of the distance between their prior belief means and the distance between each voter’s prior belief mean and the new information:

$$|\hat{\mu}_{Dn} - \hat{\mu}_{Rn}| = \left| \frac{(\hat{\mu}_{D0} - \hat{\mu}_{R0})\tau_{D0}\tau_{R0} + (\bar{x} - \hat{\mu}_{R0})\tau_{R0}\tau_{x^*} + (\hat{\mu}_{D0} - \bar{x})\tau_{D0}\tau_{x^*}}{(\tau_{D0} + \tau_{x^*})(\tau_{R0} + \tau_{x^*})} \right|.$$

only to belief trends as partisans receive and process an infinite amount of information. Because most people update their political views infrequently and subsist on a diet that is nearly devoid of political information—points that I elaborate later—it is more important to learn what happens before they attain political-information Nirvana. And in the shorter run, Proposition 2 shows that even unbiased Bayesians may disagree more after updating than they did before. Indeed, they *will* disagree more after updating whenever the condition of Proposition 2 is satisfied:

**Proposition 2 (divergence).** *Assume that  $D$  and  $R$  update in response to a sequence of messages  $x_1, \dots, x_t, \dots, x_n \stackrel{iid}{\sim} N(\mu, \sigma_{xt}^2)$ , where  $\sigma_{xt}^2$  is the variance of the distribution from which message  $x_t$  is drawn. Let  $\bar{x} = \sum_{t=1}^n x_t/n$ ,  $\tau_{xt} = 1/\sigma_{xt}^2$ , and  $\tau_{x^*} = \sum_{t=1}^n \tau_{xt}$ . Divergence occurs between time 0 and time  $n$  if and only if  $(\bar{x} - \hat{\mu}_{R0})\tau_{R0}\tau_{x^*} + (\hat{\mu}_{D0} - \bar{x})\tau_{D0}\tau_{x^*} \notin [\min\{a, b\}, \max\{a, b\}]$ , where*

The left-hand side of this equation,  $|\hat{\mu}_{Dn} - \hat{\mu}_{Rn}|$ , is the distance between the voters’ posterior belief means. The right-hand side is the absolute value of a weighted average with three components. The first component,  $\frac{(\hat{\mu}_{D0} - \hat{\mu}_{R0})\tau_{D0}\tau_{R0}}{(\tau_{D0} + \tau_{x^*})(\tau_{R0} + \tau_{x^*})}$ , is the difference between the voters’ prior belief means, weighted by the strength of those prior beliefs. The second component,  $\frac{(\bar{x} - \hat{\mu}_{R0})\tau_{R0}\tau_{x^*}}{(\tau_{D0} + \tau_{x^*})(\tau_{R0} + \tau_{x^*})}$ , is the difference between  $R$ ’s prior belief mean and the new information, weighted by the strength of  $R$ ’s prior belief and the new information. The third component,  $\frac{(\hat{\mu}_{D0} - \bar{x})\tau_{D0}\tau_{x^*}}{(\tau_{D0} + \tau_{x^*})(\tau_{R0} + \tau_{x^*})}$ , is the difference between  $D$ ’s prior belief mean and the new information, weighted by the strength of  $D$ ’s prior belief and the new information. Proposition 2 says that when the second and third components are sufficiently great, beliefs will diverge.

The simplest way for unbiased Bayesian updating to cause divergence is best conveyed by an example.

Consider two voters,  $D$  and  $R$ , who believe that the United States has the same high probability of winning a war ( $\hat{\mu}_{D0} = \hat{\mu}_{R0}$ ) but are not equally confident of this belief:  $D$  is more confident than  $R$  of U.S. victory ( $\sigma_{D0}^2 < \sigma_{R0}^2$ ). Both then receive a message suggesting that the chances of victory are lower than they believed ( $x_1 < \hat{\mu}_{D0} = \hat{\mu}_{R0}$ ). Both respond by lowering their estimates of the chance of victory—but because  $R$  has the weaker prior belief, he is swayed more by the evidence, and he revises his belief more than  $D$  does. Before updating,  $D$  and  $R$  agreed; after updating, they disagree: their beliefs have diverged. Divergence can also occur when  $D$  and  $R$  do not have the same prior belief means; for example, if  $R$  initially thinks the chances of victory slightly higher but is much less confident of his belief, divergence is still likely to occur.

Although the necessary-and-sufficient condition for divergence stated in Proposition 2 is somewhat complicated, two necessary conditions are much simpler:

**Corollary to Proposition 2 (necessary conditions for divergence).** *Under the assumptions of Proposition 2, divergence can occur between time 0 and time  $n$  only if  $\sigma_{D0}^2 \neq \sigma_{R0}^2$  and  $\bar{x} \notin [\min\{\hat{\mu}_{D0}, \hat{\mu}_{R0}\}, \max\{\hat{\mu}_{D0}, \hat{\mu}_{R0}\}]$ .*

Previous explorations of this model (e.g., Bartels 2002) focused on the special case in which members of different parties have equally strong prior beliefs ( $\sigma_{D0}^2 = \sigma_{R0}^2$ ) and respond only to “moderate” messages that fall within the range of the prior belief means ( $x_1, \dots, x_t, \dots, x_n \in [\min\{\hat{\mu}_{D0}, \hat{\mu}_{R0}\}, \max\{\hat{\mu}_{D0}, \hat{\mu}_{R0}\}]$ ). The Corollary to Proposition 2 shows that these are conditions under which divergence cannot occur. And by extension, they are conditions under which convergence must occur every time that people revise their beliefs. If a Democrat believes that the probability of U.S. victory in a war is 10%, a Republican believes that the probability is 90%, and the news always suggests 50%, beliefs must converge every time they are revised. Similarly, if the Democrat puts the probability at 80% and the Republican at 90%, and their beliefs are equally strong, beliefs must converge every time that they are revised—no matter what the news indicates about the probability of victory. The focus on these two conditions in previous research may account for the erroneous conviction that unbiased Bayesian learning must *always* produce convergence when people revise their beliefs.

The impossibility of divergence when updaters have equally strong prior beliefs deserves special attention. This condition does not make divergence

between individuals unlikely in practice, because any two people are unlikely to have priors of exactly equal strength. But is there a systematic partisan difference in belief strength—not between any given Democrat and Republican, but between Democrats and Republicans in general? Data on this point are scarce, but findings that liberals tend to be more ambivalent than conservatives (Feldman and Zaller 1992; see also Tetlock 1986) and that white men tend to hold exceptionally strong beliefs about candidates (Alvarez 1998, Chap. 6) suggest that there may be a systematic partisan difference in belief strength, too.

What of the possibility that beliefs will polarize as well as diverge in response to new information? It simply cannot happen under the model considered in this section:

**Proposition 3 (polarization).** *Under the assumptions of Proposition 2, polarization cannot occur.*

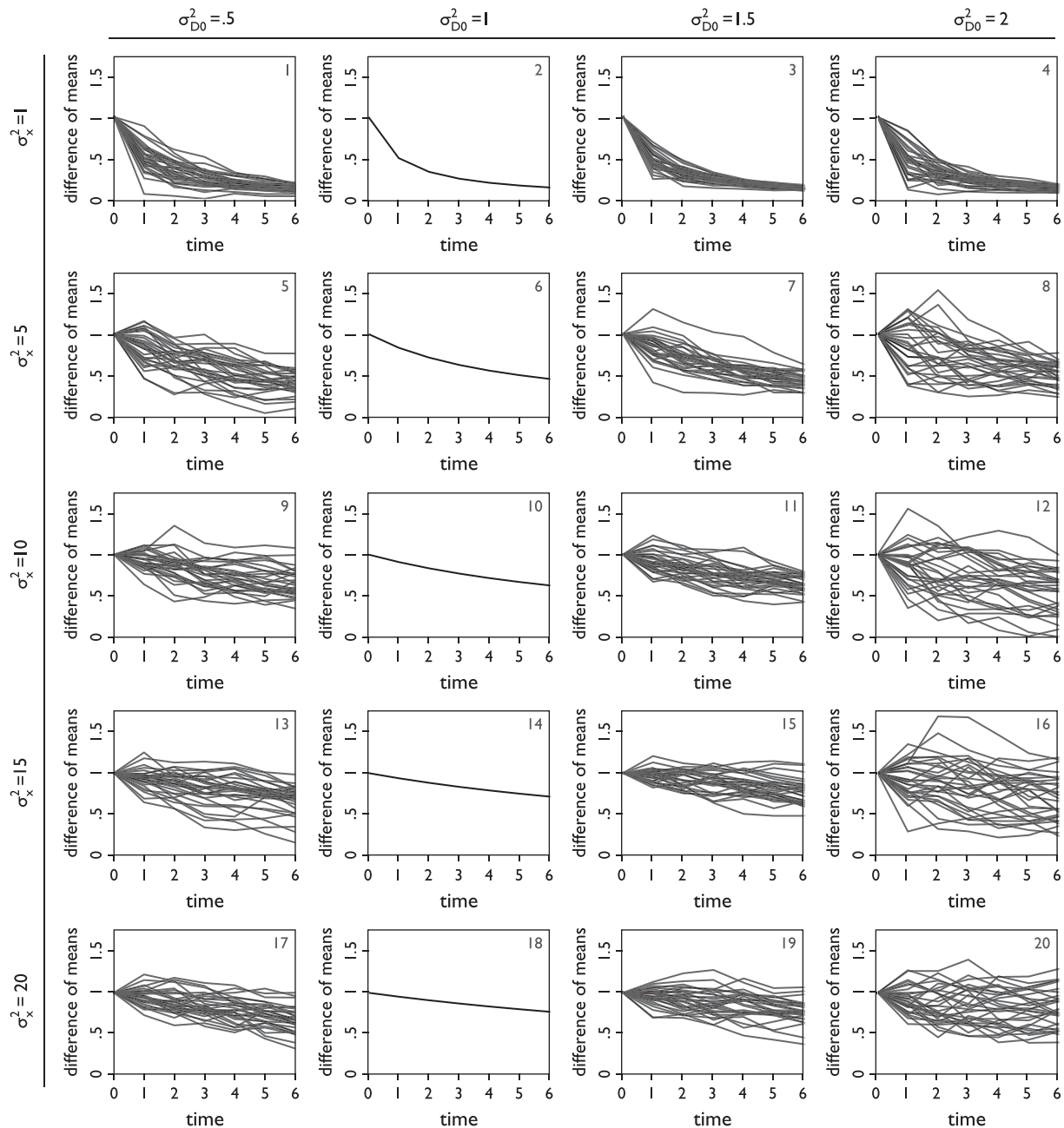
Recall that by equation (1a), posterior beliefs are weighted averages of prior beliefs and new information. When people with different prior beliefs receive the same information, interpret it in the same way, and revise their beliefs according to equation (1a), their beliefs must shift in the direction of the new information. This is all that is required to preclude polarization.

Figure 2 illustrates possible patterns of belief revision among unbiased Bayesians. Each panel depicts 30 simulations in which  $D$  and  $R$  revise their beliefs six times in response to new information. Different lines in each panel represent different simulations: when the lines slope down, disagreement between  $D$  and  $R$  is diminishing; when the lines slope up, disagreement is increasing. In each simulation,  $D$  and  $R$  are trying to learn about a feature of politics,  $\mu$ , that has a value of 2.  $D$ 's prior belief about  $\mu$  is always accurate but uncertain:  $\mu \sim N(2, \sigma_{D0}^2)$ , where  $\sigma_{D0}^2$ , the strength of his prior belief, varies from column to column.  $R$ 's prior belief is always  $\mu \sim N(1, 1)$ . At each time  $t \in 1, 2, \dots, 6$ ,  $D$  and  $R$  update their beliefs in response to a message  $x_t \stackrel{iid}{\sim} N(2, \sigma_x^2)$ , where the precise value of  $x_t$  varies from simulation to simulation and where  $\sigma_x^2$ , the strength of the messages, varies from row to row.

The simulation results differ from panel to panel because  $\sigma_{D0}^2$  and  $\sigma_x^2$  differ from panel to panel, so it is vital to consider which values of these variances are plausible. Previous research offers almost no guidance: measurement of prior belief strength is rare (Alvarez and Franklin 1994), and measurement of the quality of news sources is rarer still. But for our purposes, what is most relevant is not the individual

**FIGURE 2 Bayesian Learning about an Unchanging Condition Can Increase Disagreement Between Partisans.** Each panel depicts 30 simulations of learning by two people who update according to Equations 1a and 1b. In each simulation, they receive a new message at each of six times. A single line is plotted for each simulation, showing how the absolute difference between the means of their beliefs changes as they learn from the new messages. Most of these lines trend toward 0 in every panel, indicating that updaters' beliefs are converging to agreement. But there are upward-sloping line segments in every panel save those in the second column, indicating that disagreement between updaters can increase from one time to the next even as their beliefs trend gradually toward agreement.

In each panel,  $\mu = 2$ ,  $D$ 's belief at time 0 is  $\mu \sim N(2, \sigma_{D0}^2)$ , and  $R$ 's belief at time 0 is  $\mu \sim N(1, 1)$ . Lines are darker when they overlap.



variances but their ratio, which indicates the strength of new information relative to prior beliefs. And we can assess the plausibility of different ratios. For example, if  $\sigma_x^2 = 1$  and  $\sigma_{D0}^2 = 1$ , the 1:1 ratio implies that each new message carries as much weight as the Democrat's entire lifetime of prior experience and that six new messages collectively carry as much weight as six lifetimes of prior experience. If campaign messages were as strong as such ratios imply, we would have reason to expect colossal campaign effects, which manifestly do not occur (Bartels 1993; Shaw 1999). Higher ratios are thus much more plausible. For example, if  $\sigma_x^2 = 20$  and  $\sigma_{D0}^2 = 1$ , the 20:1 ratio implies that each message is weighted 5% as heavily as the Democrat's lifetime of prior experience and that six messages are collectively weighted 30% as heavily. These messages are still influential, but they do not demolish prior beliefs as they would if  $\sigma_x^2 = 1$ . In the absence of strong evidence, the reader may disagree about the relative plausibility of different ratios, which is why Figure 2 reports results for a range of message variances and prior belief variances.

Consider Panel 1. In this panel, the Democrat's prior belief is twice as strong as the Republican's ( $\sigma_{D0}^2 = .5$  versus  $\sigma_{R0}^2 = 1$ ). Each of the six messages received by these partisans is drawn from a distribution with variance  $\sigma_x^2 = 1$ , implying that the messages collectively carry six times as much information as the Republican has previously received throughout his life, and three times as much information as the Democrat has previously received. When new information is so much stronger than both updaters' prior beliefs, we should see rapid convergence. And we do: the 30 simulation lines trend sharply to 0; and at time 6, they are all below 1, indicating that the partisans disagree less at time 6 than at time 0 in all 30 simulations.

The same trend toward agreement appears in all other panels. This is unsurprising: even in Panel 17, where new information is weakest relative to updaters' priors ( $\sigma_{D0}^2 = .5$ ,  $\sigma_{R0}^2 = 1$ ,  $\sigma_x^2 = 20$ ), the cumulative weight of the information is 15% as great as the weight of the Democrat's lifetime of prior experience, and 30% as great as the weight of the Republican's prior experience. Moreover, all new messages are drawn from a distribution with  $\mu = 2$ , and 2 is within the range of the updaters' prior belief means. In expectation, then, new messages will not fall outside the range of prior belief means, and a necessary condition for divergence (set forth in the Corollary to Proposition 2) will not be satisfied.

And yet, divergence from one time to the next occurs in 12 of the 16 panels, as indicated by the

upward-sloping line segments within each panel. (It is absent only from the panels of the second column, where it cannot occur because  $D$  and  $R$  have equally strong prior beliefs.) It is rarest when new information overwhelms updaters' priors; for example, it occurs in expectation only 10% of the time in Panel 1. But it grows more common as information grows weaker and as the difference between the strengths of updaters' priors increases. In Panel 20, where information is weakest and the gap between updaters' prior belief strengths is greatest, divergence occurs in expectation 40% of the time that people update their beliefs. In all of these panels, save those in the second column, a general principle is at work: gradual convergence to agreement does not preclude increased disagreement in response to specific new messages. Indeed, increased disagreement from one time to the next can be quite common, as it is in the panels of the rightmost column. And disagreement is increasing in Figure 2 even under conditions that sharply favor convergence. If  $\mu$  were outside the range of updaters' prior belief means, if the updaters' prior belief strengths were less similar, or if new information were less overpowering, divergence would be sharper and even more frequent.

The distinctive panels of the second column deserve explanation.  $D$  and  $R$  have equally strong prior beliefs in these panels, which ensures that the gap between their beliefs is reduced by the same amount whenever they learn from new messages, regardless of the values of those messages. Within each panel of the column, updating trends are therefore identical for all 30 simulations. This accounts for the panels' appearance: although it appears that each panel contains but one line, there are really 30 lines in each panel, each line tracing the same path.

## Disagreement When Learning about a Changing Condition

Almost all Bayesian analyses of public opinion assume that people are trying to learn about political conditions that never change. The assumption is apt when we are trying to learn about the past (did the New Deal prolong the Great Depression?) and perhaps when we update our beliefs over the short term (which party is better for me *right now?*), but it is inappropriate when the feature of politics that we want to learn about is a moving target. The state of the economy changes because of demographic trends and unforeseen events. My preferences over policies change as new proposals are placed on the table or



taken off of it. And candidates may improve during their time in office or fall under the sway of constituents whose views I oppose. In these cases, the unchanging-condition assumption is a poor approximation of reality. The assumption is embodied in the model described in the previous section, which is the basis for almost all Bayesian analyses of public opinion (e.g., Achen 1992; Bartels 1993, 2002; Gerber and Green 1999; Lohmann 1994; Steenbergen 2002; Zechman 1979). But it is easy to generalize the model so that it accommodates political change.

Suppose that a politically interesting condition changes according to the rule

$$\mu_t = \gamma\mu_{t-1} + \epsilon_{\mu,t}, \quad \epsilon_{\mu,t} \sim N(0, \sigma_\mu^2) \quad (2)$$

where  $t$  is the current time and  $\gamma \geq 0$ . In this equation,  $\gamma$  is an autoregressive parameter that dictates systematic change in the condition of interest: when  $\gamma > 1$ , the absolute expected value of  $\mu$  increases with  $t$ ; when  $\gamma < 1$ , the absolute expected value of  $\mu$  decreases with  $t$ . For tractability, assume that the people whose beliefs we are studying know the true value of  $\gamma$ .<sup>6</sup>  $\epsilon_{\mu,t}$  is a disturbance with known and finite variance  $\sigma_\mu^2$ ; it determines random variation in  $\mu$  over time. When  $\gamma = 1$  and  $\sigma_\mu^2 = 0$ ,  $\mu$  is unchanging, just as in the model from the previous section.

Let  $x_1, \dots, x_n$  be messages containing information about  $\mu$  at times  $1, \dots, n$ . At any time  $t \in 1, \dots, n$ , the relationship between a new message and  $\mu$  is  $x_t \sim N(\mu, \sigma_x^2)$ , where  $\sigma_x^2$  is a known and finite variance.<sup>7</sup> The voter's belief after updating in response to message  $x_{t-1}$  is  $\mu_{t-1} | \gamma, \sigma_\mu^2, \sigma_x^2, \mathbf{x}_{t-1} \sim N(\hat{\mu}_{t-1}, \sigma_{t-1}^2)$ , where  $\mathbf{x}_{t-1}$  is the vector of messages  $x_1, \dots, x_{t-1}$ . It can be viewed either as a posterior belief (because it is the voter's belief after receiving message  $x_{t-1}$ ) or as a prior belief (because it is the voter's belief before receiving message  $x_t$ ).

When a political condition is changing over time, Bayesian belief revision is a two-step process. In the first step, the voter forecasts the value that the condition,  $\mu$ , will have at time  $t$ . She does this by multiplying the mean of her prior belief,  $\hat{\mu}_{t-1}$ , by the

<sup>6</sup>The same assumption is made in other analyses of learning about changing conditions, notably Gerber and Green (1998). Divergence and polarization are less likely under this assumption. The formal results in this section are thus especially noteworthy: they show that divergence and polarization can occur even under an unfavorable assumption.

<sup>7</sup>As in Gerber and Green (1998),  $\gamma$ ,  $\sigma_\mu^2$ , and  $\sigma_x^2$  are held constant for simplicity of exposition. But the model described in this section can accommodate the case in which they change over time. See Meinhold and Singpurwalla (1983) for an example.

rate of change in  $\mu$ , which is  $\gamma$ . Her forecast is thus  $\gamma\hat{\mu}_{t-1}$ . In the second step, the voter receives message  $x_t$  and adjusts her forecast to account for it:

$$\mu_t | \gamma, \sigma_\mu^2, \sigma_x^2, \mathbf{x}_{t-1}, x_t \sim N(\hat{\mu}_t, \sigma_t^2), \text{ where}$$

$$\hat{\mu}_t = \gamma\hat{\mu}_{t-1} + K_t(x_t - \gamma\hat{\mu}_{t-1}), \quad (3a)$$

$$\sigma_t^2 = \frac{1}{1/(\gamma^2\sigma_{t-1}^2 + \sigma_\mu^2) + 1/\sigma_x^2}, \quad (3b)$$

and  $K_t = \sigma_t^2 / \sigma_x^2$ . This second step is an "error correction" in which the voter modifies her forecast to account for the distance between it and the new information that she had received ( $x_t - \gamma\hat{\mu}_{t-1}$ ). Its effect is always to draw beliefs toward the new information. The weight that it assumes in the updating process is  $K_t$ , which is standardized to lie between 0 and 1. The stronger the new information (i.e., the lower the value of  $\sigma_x^2$ ), the higher the value of  $K_t$ , and the more that beliefs will be drawn toward the new information.

Equations (3a) and (3b) are known as the *Kalman filter algorithm* after Rudolf Kalman (1960), who shows that they yield the expected value of  $\mu_t$  under the assumption of normal errors.<sup>8</sup> For our purpose, the Kalman filter is useful because it helps to establish results analogous to those in the previous section, but under the assumption that people are learning about changing features of politics. Again, let  $D$  and  $R$  be unbiased Bayesians whose beliefs at time  $t = 0$  are  $\mu_0 \sim N(\hat{\mu}_{D0}, \sigma_{D0}^2)$  and  $\mu_0 \sim N(\hat{\mu}_{R0}, \sigma_{R0}^2)$ , respectively. For all finite  $t$ , assume  $0 < \sigma_{Dt}^2, \sigma_{Rt}^2$ : neither  $D$  nor  $R$  is absolutely certain of his belief. For all times  $t$ , assume that  $\mu_t$  is determined by Equation 2 and that  $x_t \sim N(\mu, \sigma_x^2)$ . Under these assumptions, the initial results resemble those from the previous section:

**Proposition 4 (convergence to agreement).** *If  $D$  and  $R$  receive a sequence of messages  $x_1, \dots, x_n$ , their beliefs will converge to agreement as  $n \rightarrow \infty$ .*

**Proposition 5 (divergence).** *Divergence occurs between time  $t$  and later time  $u$  if and only if  $K_{Du}(x_u - \gamma\hat{\mu}_{Dt}) - K_{Ru}(x_u - \gamma\hat{\mu}_{Rt}) \notin [\min\{a, b\}, \max\{a, b\}]$ , where  $a = (\gamma - 1)(\hat{\mu}_{Rt} - \hat{\mu}_{Dt})$  and  $b = (\gamma + 1)(\hat{\mu}_{Rt} - \hat{\mu}_{Dt})$ .*

As with the model from the previous section, receiving infinite information will cause partisans to agree,

<sup>8</sup>For an overview of the Kalman filter, see Beck (1990). Meinhold and Singpurwalla (1983) provide a lucid introduction from a Bayesian point of view.

but receiving finite amounts of information may increase disagreement even if both partisans interpret the information in the same way. And as with the model from the previous section, the condition for divergence is somewhat complicated, but its explanation is straightforward. Recall that when learning about a changing condition, the transition from prior to posterior beliefs takes place in two steps: first, one forecasts the condition's future value,  $\mu_u$ ; second, one receives new information and adjusts the forecast to account for it. When  $\gamma > 1$ , the forecasting process necessarily pushes beliefs apart, because  $\gamma|\hat{\mu}_{Dt} - \hat{\mu}_{Rt}| > |\hat{\mu}_{Dt} - \hat{\mu}_{Rt}|$ . In this case, Proposition 5 says that divergence occurs when the adjustments for new information also push beliefs apart, or when they pull beliefs together by less than the forecasting process pushes them apart. On the other hand, if  $\gamma < 1$ , the forecasting process pulls beliefs together, because  $\gamma|\hat{\mu}_{Dt} - \hat{\mu}_{Rt}| < |\hat{\mu}_{Dt} - \hat{\mu}_{Rt}|$ . In this case, Proposition 5 says that divergence occurs when the adjustments for new information push beliefs apart more than the forecasting process pulls them together.

To see how divergence about a changing condition might occur in practice, consider Adam Berinsky's recent analysis of Congressional support for U.S. intervention in World War II (Berinsky 2007, 987–88). In June 1940, Republican and Democratic Congressmen sharply opposed U.S. intervention, as indicated by the antiwar bent of their statements in the *Congressional Record*. But they were not equally opposed: approximately 25% of Democratic statements about the war expressed support for war, against only 12.5% of Republican statements. In the next year and a half, prowar sentiment increased among Congressmen of both parties. But it increased more quickly among Democrats, and the partisan gap in support for war therefore quadrupled: by the first week of December 1941, about 90% of Democratic statements about the war were prowar, against only 40% of Republican statements (Berinsky 2007, 988).

There is no sure way to draw conclusions about legislators' beliefs from their statements in the *Record*, but those statements suggest diverging beliefs about the benefits of war, and in a way that illustrates the logic of Proposition 5. Let  $t$  be June 1, 1940 and  $t + 1, \dots, t + 18$  be the first days of the next 18 months.  $D$  and  $R$  are Congressmen trying to learn about  $\mu_t$ , the net benefit to Americans of a declaration of war. Like a politician's honesty, this net benefit has no natural scale; for convenience, suppose that  $\hat{\mu}_{Dt} = 1.2$  and  $\hat{\mu}_{Rt} = 1.0$ . These numbers suggest that  $D$  perceives a greater benefit of going to war than  $R$ , just as the *Record* suggests that Democrats were

more supportive of war than Republicans in June 1940. Let  $\mu_t = 2.5$ , indicating that both Congressmen underestimate the true benefit of a declaration of war. Moreover,  $\sigma_{Dt}^2 = .5$  and  $\sigma_{Rt}^2 = .2$ , indicating that  $R$ 's belief is more than twice as strong as  $D$ 's. There is little random variation in the true benefit of a declaration over time:  $\sigma_\mu^2 = .01$ , suggesting that the benefit of going to war is not, at this early stage, much dependent on chance contingencies. But  $\gamma = 1.05$ , indicating that the benefit is increasing systematically as fascist successes mount in France, China, East Africa, and the Soviet Union. The news received in any particular month is relatively weak:  $\sigma_x^2 = 10$ , implying that each new month's information matters only 2% as much as  $R$ 's prior learning about the merits of going to war, and 4% as much as  $D$ 's prior learning on the same subject. But over 18 months,  $\sigma_x^2 = 10$  implies substantial learning. The cumulative weight of the news received in these months is 36% as great as the weight attached to  $R$ 's lifetime of prior experience, and fully 90% as great as the weight attached to  $D$ 's lifetime of prior experience.

This is all that is needed to produce belief divergence. In June 1940,  $D$  and  $R$  will forecast the benefit that a declaration of war will have on July 1:  $\gamma\hat{\mu}_{Dt} = 1.26$ , and  $\gamma\hat{\mu}_{Rt} = 1.05$ . The distance between these forecasts ( $1.26 - 1.05 = .21$ ) will exceed the distance between the Congressmen's prior beliefs ( $1.2 - 1.0 = .2$ ); thus, the forecasting process will push beliefs apart. Now suppose that both Congressmen receive new information about the benefit of a declaration of war on July 1,  $x_{t+1} = 2.625$ , indicating that the true benefit is greater than they anticipated. (In expectation, this is the new information that they will receive, because  $x_{t+1}$  is drawn from a normal distribution that is centered in expectation at  $\gamma\mu_t = 2.625$ .) They will adjust their forecasts to account for this new information. For  $R$ ,  $K_{R,t+1}$  is approximately .02, reflecting the low weight that he attaches to the new information. By equation (3a), his posterior belief will therefore be  $\hat{\mu}_{R,t+1} \approx (1 - .02)(1.05)(1.0) + .02(2.625) \approx 1.08$ . For  $D$ ,  $K_{D,t+1}$  is approximately .05, reflecting his weaker prior belief and correspondingly greater responsiveness to new information. His posterior belief will therefore be  $\hat{\mu}_{D,t+1} \approx (1 - .05)(1.26) + .05(2.625) \approx 1.33$ . The Congressmen's adjustments for new information will therefore push their beliefs apart, just as the forecasting process did. The combined effect of the forecasting and the adjustment for new information will be divergence: on June 1, the difference between the Congressmen's beliefs was  $1.2 - 1.0 = .2$ ; on July 1, it will be  $1.33 - 1.08 = .25$ . And this divergence will

occur even though both Congressmen receive the same information, interpret it in the same way, and follow the same updating rule.

In expectation,  $D$  and  $R$  will continue to diverge through December 1, 1941. The importance of a declaration of war will grow as it appears more likely that the Allies will lose without American intervention, and this will be reflected by change in  $\mu$ , which will grow to 6.0. A week before the attack on Pearl Harbor,  $D$ 's posterior belief will be  $\hat{\mu}_{D,t+18} = 5.2$ , and  $R$ 's will be  $\hat{\mu}_{R,t+18} = 4.5$ . That is, they will both perceive the benefits of war as greater than they did in June 1940—but even so, the difference between their beliefs will almost quadruple over the 18 months (from .2 to  $5.2 - 4.5 = .7$ ), much as Berinsky observes a quadrupling of the gap between the proportions of prowar statements made by Democratic and Republican Congressmen over the same time span.

Partisan divergence in Congressional support for American intervention ended with Pearl Harbor. In formal terms, the attack on Pearl Harbor can be seen as a new message with extremely low variance ( $\sigma_x^2 \approx 0$ ). The effect of such a message is to cause people to discount all of the information that they had previously received: in effect, posterior beliefs are determined almost exclusively by the new message. In the context of deliberations about war,  $\sigma_x^2 \approx 0$  implies that when legislators heard the news about Pearl Harbor, their beliefs about the net benefit of a declaration of war were determined almost exclusively by that news, and not by the different beliefs that they held before they heard the news. In the context of Bayesian disagreement, Pearl Harbor is instructive because it shows how unusually clear new information can produce convergence to agreement even among people whose beliefs had been sharply diverging.

In this example,  $D$  and  $R$  start with beliefs that are not equally strong, and they learn from new information that is “extreme” in the sense that it is outside the range of their prior belief means. The Corollary to Proposition 2 shows that these are necessary conditions for divergence *if* people are learning about something that is not changing over time. But if people are learning about something that is changing, the Corollary to Proposition 5 shows that divergence can occur even when they have equally strong prior beliefs and respond to “moderate” information:

**Corollary to Proposition 5.** *Divergence can occur between  $t$  and  $u$  if  $\sigma_{Dt}^2 = \sigma_{Rt}^2$  or  $x_u \in [\min\{\hat{\mu}_{Dt}, \hat{\mu}_{Rt}\}, \max\{\hat{\mu}_{Dt}, \hat{\mu}_{Rt}\}]$ .*

This follows almost immediately from Proposition 5. If  $D$  and  $R$  have equally strong prior beliefs ( $\sigma_{Dt}^2 = \sigma_{Rt}^2$ ), they attach the same weight to the new information that they receive ( $K_{Du} = K_{Ru} = K_u$ ). In this case, the condition for divergence boils down to  $(1 - K_u)\gamma > 1$ . This condition can be met whether or not new information is “extreme.” The Corollary to Proposition 5 is thus a retort to the Corollary to Proposition 2.

Another striking difference between learning about fixed and changing conditions concerns polarization:

**Proposition 6 (polarization).** *Polarization occurs between  $t$  and  $u$  if and only if divergence occurs between  $t$  and  $u$  and either (a)  $K_{Du}(x_u - \gamma\hat{\mu}_{Dt}) > (1 - \gamma)\hat{\mu}_{Dt}$  and  $K_{Ru}(x_u - \gamma\hat{\mu}_{Rt}) < (1 - \gamma)\hat{\mu}_{Rt}$  or (b)  $K_{Du}(x_u - \gamma\hat{\mu}_{Dt}) < (1 - \gamma)\hat{\mu}_{Dt}$  and  $K_{Ru}(x_u - \gamma\hat{\mu}_{Rt}) > (1 - \gamma)\hat{\mu}_{Rt}$ .*

When unbiased Bayesians with normally distributed prior beliefs learn about fixed conditions from normally distributed messages, polarization cannot occur. But when they learn about *changing* conditions from normally distributed messages, polarization can occur. To see why polarization is only possible for such people when conditions are changing, recall that learning about changing conditions takes place in two steps, and that in the second step, people always shift their beliefs toward the new information that they receive. If this is all that happens—if beliefs do not also change in the first step—polarization cannot occur. And when people learn about an unchanging condition, the second step *is* all that happens. The first step, in which people forecast the new value of  $\mu$  by accounting for its rate of change, does not occur because  $\mu$  is not changing.

Similar reasoning shows why polarization among unbiased Bayesians will be relatively rare. Unlike divergence, polarization requires new information to contradict expectations: at least one person must forecast an increase in  $\mu$  while information indicates that it has decreased or stayed the same, or forecast a decrease in  $\mu$  while information indicates that it has increased or stayed the same. For example, imagine that  $M$  is Martin Luther King, Jr., and that  $H$  is Harris Wofford, a member of the Kennedy administration and the main liaison between Kennedy and civil rights groups. Both men are trying to learn about Kennedy's support for new civil rights legislation in the summer of 1963, just after the brutal suppression of protesters in Birmingham, Alabama. They share an impression that Kennedy supported new legislation in June ( $\hat{\mu}_{Mt}, \hat{\mu}_{Ht} > 0$ ) and that his support is increasing over time ( $\gamma > 1$ ). Before receiving information

about his level of support later in the summer, they forecast this level. Then they receive new information about it ( $x_u$ ). By equation (3a), they adjust their forecasts by moving in the direction of this information. We know that both men forecast increased support from Kennedy, so if the new information also indicates increased support ( $x_u > \hat{\mu}_{Mt}, \hat{\mu}_{Ht}$ ), both will conclude that Kennedy's support for new legislation has increased ( $\hat{\mu}_{Mu} > \hat{\mu}_{Mt}$  and  $\hat{\mu}_{Hu} > \hat{\mu}_{Ht}$ ). Because their beliefs will move in the same direction, polarization will not occur.

In this example, polarization will occur only if the new information suggests that both King and Wofford were wrong about the direction of Kennedy's support for civil rights, i.e., only if  $x_u < \hat{\mu}_{Mt}, \hat{\mu}_{Ht}$ . And this is only a necessary condition, not a sufficient one. If the new, negative information is too weak, King and Wofford will ignore it; both will conclude that Kennedy's support for civil rights increased, and polarization will not occur. On the other hand, if the information is too strong, it will overwhelm the forecasts of both King and Wofford; both will conclude that Kennedy's support declined, and polarization will not occur. To polarize King and Wofford, the new information must be negative and strong enough to overwhelm one man's optimistic forecast, but not so strong that it overwhelms both men's forecasts. Formally, it must fit within a range of values that is determined by the weight that each man places on it ( $K_{Mu}$  and  $K_{Hu}$ ), the mean of each man's prior belief about Kennedy's position ( $\hat{\mu}_{Mt}$  and  $\hat{\mu}_{Ht}$ ), and the rate at which Kennedy's position is changing ( $\gamma$ ). This is the range that is specified in Proposition 6.

Something like this may explain what happened in 1963. In late June of that year, King and Wofford believed that Kennedy's support for civil rights was increasing, and they therefore expected stronger support from Kennedy in the near future (Garrow 1986, 269; Wofford 1980, Chap. 5). Both men subsequently noticed Kennedy's apparent wavering, which included pressuring liberal Congressmen not to demand stronger legislation and an "empty show of federal response" to the Birmingham church bombing (Branch 1998, 250; see also Garrow 1986, 296, 302; Wofford 1980, 172–74). But Wofford, who had stronger prior beliefs about Kennedy, was largely unmoved; late in 1963, he thought Kennedy more committed than ever to civil rights. King, who had long been unsure of Kennedy's true level of support for civil rights, was more affected by July's events, and he revised downward his view of Kennedy (Branch 1998, 246–50). The result was that King and Wofford polarized over Kennedy's commitment to civil rights.

Figure 3 illustrates possible patterns of belief revision among unbiased Bayesians who are learning about a changing condition. Each panel depicts 30 simulations in which  $D$  and  $R$  revise their beliefs six times in response to new information. Different lines in each panel represent different simulations: when the lines slope up, disagreement is increasing; when they slope down, it is decreasing. The strength of the messages received varies from row to row: they are strongest in the top row ( $\sigma_x^2 = 1$ ) and weakest in the bottom row ( $\sigma_x^2 = 20$ ). In all panels,  $\sigma_\mu^2 = 0$ , meaning that  $\mu$ , the feature of politics that  $D$  and  $R$  are learning about, does not vary stochastically over time.<sup>9</sup> In these respects, Figure 3 is like Figure 2. But in Figure 3,  $\mu$  varies systematically from one time to the next, because  $\gamma > 1$  for every panel in the figure—and this is an important difference.  $D$  and  $R$  are now learning about a feature of politics that is changing over time. In every panel,  $\mu$  starts at 2 ( $\mu_0 = 2$ ) and increases as the simulation progresses; the rate of increase is determined by  $\gamma$ , which varies from column to column. As in Figure 2, the prior belief of  $R$  is  $\mu_0 \sim N(1, 1)$  in every panel, but the prior belief of  $D$  is now fixed at  $\mu_0 \sim N(2, 1.25)$ .<sup>10</sup>

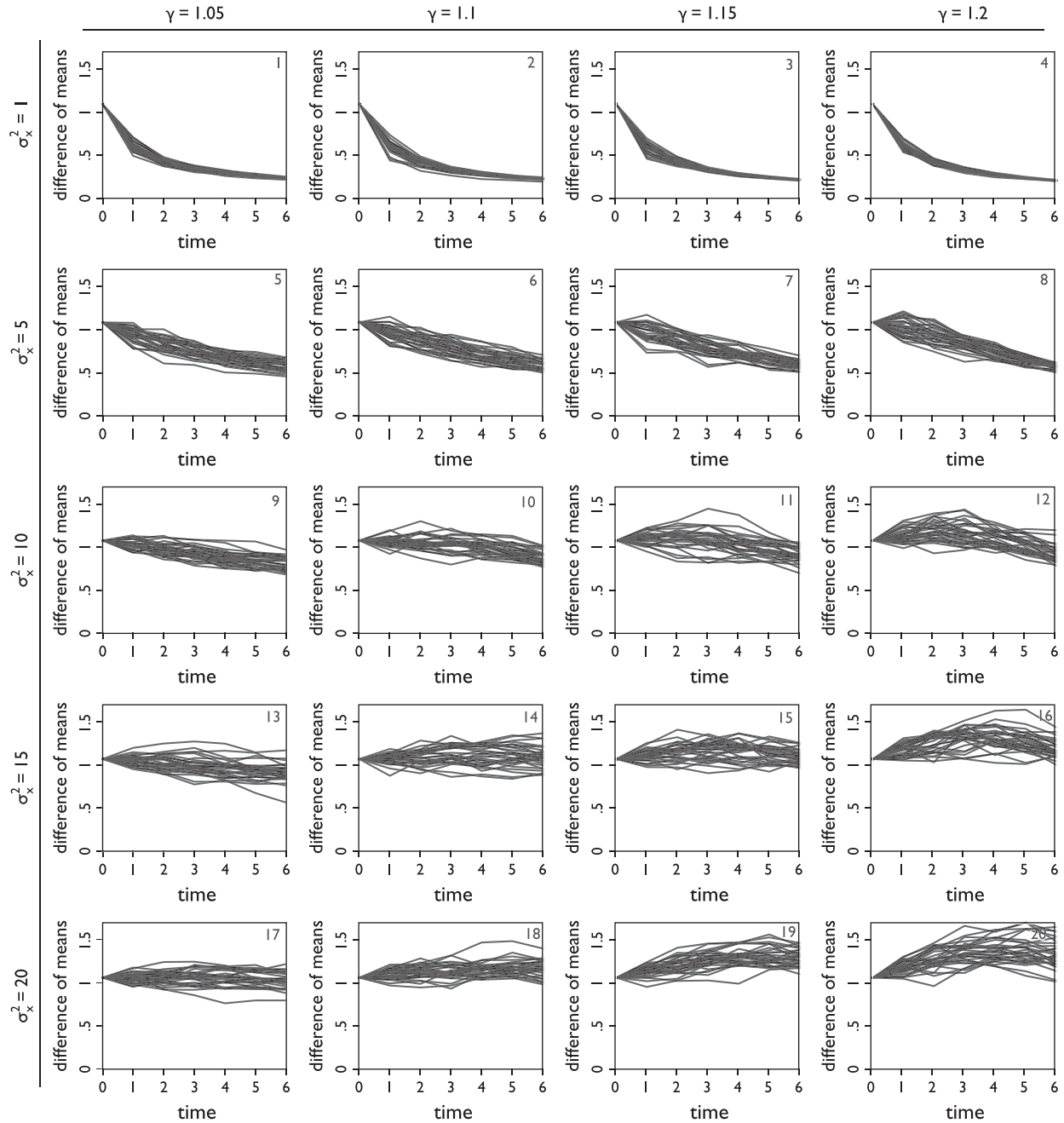
Divergence occurs throughout Figure 3. It is rare in the top row, where information is far stronger than prior beliefs and where disagreement therefore tends to diminish rapidly. But in successive rows, the trend toward agreement is less pronounced, and in the fourth and fifth rows, disagreement increases between time 0 and time 6 in almost every simulation. It increases even though the cumulative weight of the information that  $D$  and  $R$  receive is still strong. In the fourth row, where the variance of each message is  $\sigma_x^2 = 15$ , the six messages are collectively weighted 50% as heavily as the Democrat's prior experience and 40% as heavily as the Republican's. In the fifth row, where  $\sigma_x^2 = 20$ , the six messages are collectively

<sup>9</sup> $\sigma_\mu^2$  is held constant to permit the effects of changes in  $\gamma$  to emerge clearly through comparison of the panels within Figure 3, and it is set to 0 to increase the comparability of Figure 3 to Figure 2. By equation (3a), increasing  $\sigma_\mu^2$  increases the weight that  $D$  and  $R$  place on the new messages that they receive ( $K_{Dt}$  and  $K_{Rt}$ ), thereby making their beliefs converge more rapidly. This effect can be offset by increasing  $\sigma_x^2$ , which reduces the weight that  $D$  and  $R$  place on the new messages.

<sup>10</sup>The Democrat's prior belief is fixed to permit the figure to show the effect of changes in  $\gamma$  independent of changes in prior beliefs. I investigated belief updating patterns under a range of different values for  $D$ 's prior belief. Small changes (e.g., fixing  $D$ 's prior at  $\mu_0 \sim N(2, 1.5)$ ) made only small differences to the results; large changes (e.g.,  $\mu_0 \sim N(2, 10)$ ) made larger differences, typically increasing the frequency of divergence and polarization. Even in these cases, the patterns observed across panels remained the same.

**FIGURE 3 Bayesian Learning about a Changing Condition Can Increase Disagreement Between Partisans.** Each panel depicts 30 simulations of learning by two people who revise their beliefs according to Equations 3a and 3b. In each simulation, they receive a new message at each of six times. A single line is plotted for each simulation, showing how the absolute difference between the means of their beliefs changes as they learn from the new messages. In the upper rows, where the cumulative effect of new information far outweighs the effect of prior beliefs, the lines trend toward 0, indicating rapid convergence to agreement. In the lower rows, where information is weaker, upward-sloping line segments are common, indicating frequent belief divergence.

In each panel,  $\mu_0 = 2$ ,  $\sigma_{\mu}^2 = 0$ ,  $D$ 's belief at time 0 is  $\mu_0 \sim N(2, 1.25)$ , and  $R$ 's belief at time 0 is  $\mu_0 \sim N(1, 1)$ . Lines are darker when they overlap.



weighted about 38% as heavily as the Democrat's prior experience and 30% as heavily as the Republican's.

Figure 3 does not indicate when polarization occurs, but it would not look much different if it did, because polarization is uncommon in the scenarios that it depicts: it occurs twice in Panel 16, once in Panel 19, and once in Panel 20. To see why it does not occur more often, consider the properties that messages must have to polarize  $D$  and  $R$ . Because  $\gamma > 1$ , both  $D$  and  $R$  always forecast an increase in  $\mu$  from one time to the next. Polarization requires that new information contradict these forecasts: at time 1, for example, new information must suggest that  $\mu$  is *decreasing*. In addition, the new information must be extreme enough to overwhelm one updater's forecast—e.g., extreme enough to cause  $\hat{\mu}_{R1} < \hat{\mu}_{R0}$ —but not so extreme that it overwhelms both updaters' forecasts, thereby causing both  $\hat{\mu}_{R1} < \hat{\mu}_{R0}$  and  $\hat{\mu}_{D1} < \hat{\mu}_{D0}$ . The specific range of message values that will produce polarization varies from time to time and panel to panel; as an example, consider polarization between time 0 and time 1 in Panel 16. Application of Proposition 6 shows that it will occur if and only if  $-1.33 < x_1 < -1.08$ . This is a narrow range;  $x_1$  will fall within it about six times out of a thousand. Polarization under this model requires new messages to thread a needle, which is why it will not happen often.

### How Common Are the Conditions for Bayesian Agreement?

The previous sections have shown, contrary to previous claims, that unbiased Bayesian learning can sustain and even exacerbate disagreement. They have also shown, in keeping with previous claims, that unbiased Bayesian learning can produce agreement between partisans even when their initial disagreement is great. Indeed, Propositions 1 and 4 show that this will happen whenever partisans receive political messages so numerous and so credible that their prior beliefs are overwhelmed. But most partisans never receive so much information of such high quality about any political question. Bayesian updating therefore offers no expectation of convergence to agreement: the lasting differences that we observe between real-world partisans are just what we would observe if those partisans were unbiased Bayesians.

To see that partisans are unlikely to acquire enough information to bring them to agreement even if they update as unbiased Bayesians, consider

the wealth of information that we possess about their dramatic lack of political information, nicely summarized by Delli Carpini and Keeter (1996). Although partisans know more about politics than people with no partisan leanings, strength of party identification bears only a weak connection to political information levels (Delli Carpini and Keeter 1996, 144–49), and widespread political ignorance among partisans suggests that they, like nonpartisans, are not in the habit of acquiring much information about even those political questions that matter most to them. Widespread inattention to politics is also consistent with Achen and Bartels (2006), who argue that events induce political thought for most people only “once or twice in a lifetime,” and with Shaw (1999), who finds that few presidential campaign events affect support for the candidates.

Of course, it does not take much information to overwhelm weak prior beliefs. But partisans' prior beliefs—at least about political candidates, and thus about choices in most elections—are relatively strong, and much stronger than the cumulative weight of media messages received over the course of a campaign (Bartels 1993). When beliefs are this substantial, it is difficult to overwhelm them with new information.

In any ordinary human span of time, then, there is no reason to expect that partisans who disagree on political matters will acquire information sufficient to make them agree with each other. And this is so even if they are exposed to the same information, interpret it in the same way, and revise their beliefs according to Bayes' Theorem. Grynawski's (2006, 331) characterization—that mortal Bayesian learners exposed to the same information will “inexorably come to see the world in the same way”—is too strong.

### Conclusion

Writing about political belief revision, Bartels (2002, 123) tells us that “it is *failure* to converge that requires explanation within the Bayesian framework” (2002, 123). The chief contribution of this article has been just such an explanation—an explanation that does not hinge, as previous explanations have, on partisan bias or selective exposure to congenial sources of information. Even when partisans receive the same information and interpret it in the same way, Bayesian updating will lead them to agreement only if the information is of extraordinary quantity or quality. Contrary to Goodin (2002), enduring disagreement among Bayesians is no “paradox”: it is the

normal state of affairs. This result is consistent with the argument in Gerber and Green (1999) but contrary to much else that has been written on the subject, including Goodin (2002), Bartels (2002, 121–23), and Grynaviski (2006, 331).

Of course, partisans who never completely agree may yet disagree less as they learn more about the subject of their disagreement. But this article shows that Bayesian learning cannot always justify even the weaker expectation of diminished disagreement. Indeed, it may increase disagreement instead of diminishing it. Disagreement can increase through divergence, whereby all beliefs move in the same direction, but at different rates. Or it can increase through polarization, whereby different partisans' beliefs move in different directions. Contrary to Tetlock (2005, 128), Bartels (2002, 121–23), and others, both divergence and polarization are consistent with unbiased Bayesian learning.

This may seem to be good news. Unbiased Bayesian updating is increasingly put forth as ideal information processing, and the results here suggest that partisans are closer to that ideal than we tend to suppose, because they show that the ideal accommodates real patterns of public opinion change among partisans. But there is another way to view the results. Rather than casting citizens' thinking about politics in a rosy light, they may suggest that Bayesian updating is inadequate as a normative standard by which to judge reactions to political information. Because it permits lasting disagreement, divergence, and polarization even among people who receive the same information and interpret it in the same way, it may be insufficient for rationality in the sense of "plain reasonableness" (Bartels 2002, 125–26; Fischhoff and Lichtenstein 1978).

Readers must judge for themselves whether "plain reasonableness" precludes lasting disagreement, divergence, and polarization. What this article demonstrates is that those patterns of public opinion do not suffice to show that people are biased or to help us understand the extent of their biases. What *will* suffice is better information about partisans' beliefs and perceptions. For more than a decade, some political scientists have clamored for more research on the strength and shape—rather than just the "location"—of partisans' beliefs about policies and candidates and their perceptions of new information (Alvarez and Franklin 1994; Bartels 1986; Gerber and Green 1999; Gill and Walker 2005; Steenbergen 2002). For the study of bias in political information processing, this is exactly the right stance. To determine how well partisan belief revision

matches up against a normative baseline, we need to know not just the location of partisans' beliefs but the degree of confidence with which those beliefs are held. Coupling measures of belief location with measures of belief strength will do much to help us understand the nature and extent of political biases, and thus to help us understand just how far partisans fall from a normative ideal.

We study Bayesian updating not because we think it is what partisans do but because we think it is what they should do: in a large and increasing number of public opinion analyses, it is the ideal to which real belief updating is compared. That said, the findings reported here suggest that the gap between real and Bayesian political belief revision is not as wide as many have suggested. Real partisans do not always disagree less as they learn more. Sometimes their disagreement is undiminished by exposure to new information; sometimes, it is even increased. And all of these aspects of real-world opinion change are consistent with the Bayesian ideal.

## Acknowledgments

I thank Neal Beck, Jonathan Bendor, Adam Berinsky, Nathan Collins, Mo Fiorina, Alan Gerber, Sarah Hammond, Simon Jackman, Robert Luskin, Andrew Rutten, Paul Sniderman, and Jonathan Wand for helpful discussions. Larry Bartels, Will Bullock, Fred Cutler, Don Green, Stephen Jessee, Christopher Kam, Matt Levendusky, Nora Ng, Ben Nyblade, Angela O'Mahony, Paul Quirk, and three anonymous reviewers offered valuable comments on earlier drafts.

I wrote most of this article while I was a Killam Postdoctoral Fellow in the Department of Political Science at the University of British Columbia. I thank the Trustees of the Killam Trusts for their support.

Proofs of all formal propositions in this article are given at <http://bullock.research.yale.edu/disagreement>. All figures, simulations, and analyses can be reproduced with files available at the same URL.

*Manuscript submitted 13 March 2008*

*Manuscript accepted for publication 5 October 2008*

## References

- ABC News/The Washington Post. 2004. ABC News/Washington Post Weapons of Mass Destruction Poll, February 2004 [computer file] (Study 4034). ICPSR version. Horsham, PA: Taylor Nelson Sofres Intersearch [producer], 2004. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2004.

- ABC News/The Washington Post. 2005. ABC News/Washington Post Poll, March 2005 [computer file] (Study 4319-v1). ICPSR version. Horsham, PA: Taylor Nelson Sofres Intersearch [producer], 2005. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2005.
- Achen, Christopher H. 1992. "Social Psychology, Demographic Variables, and Linear Regression: Breaking the Iron Triangle in Voting Research." *Political Behavior* 14 (September): 195–211.
- Achen, Christopher H. 2005. "Let's Put Garbage-Can Regressions and Garbage-Can Probits Where They Belong." *Conflict Management and Peace Science* 22 (Winter): 327–39.
- Achen, Christopher H., and Larry M. Bartels. 2006. "It Feels Like We're Thinking: The Rationalizing Voter and Electoral Democracy." Presented at the Annual Meeting of the American Political Science Association, Philadelphia.
- Alvarez, R. Michael. 1998. *Information and Elections*. Revised ed. Ann Arbor: University of Michigan Press.
- Alvarez, R. Michael, and Charles H. Franklin. 1994. "Uncertainty and Political Perceptions." *Journal of Politics* 56: 671–88.
- Aumann, Robert J. 1976. "Agreeing to Disagree." *Annals of Statistics* 4 (November): 1236–39.
- Bartels, Larry M. 1986. "Issue Voting Under Uncertainty: An Empirical Test." *American Journal of Political Science* 30 (November): 709–28.
- Bartels, Larry M. 1993. "Messages Received: The Political Impact of Media Exposure." *American Political Science Review* 87 (June): 267–85.
- Bartels, Larry M. 2002. "Beyond the Running Tally: Partisan Bias in Political Perceptions." *Political Behavior* 24 (June): 117–50.
- Beck, Nathaniel. 1990. "Estimating Dynamic Models Using Kalman Filtering." *Political Analysis* 1: 121–56.
- 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. 2007. "Assuming the Costs of War: Events, Elites, and American Public Support for Military Conflict." *Journal of Politics* 69 (November): 975–97.
- Branch, Taylor. 1998. *Pillar of Fire*. New York: Simon and Schuster.
- Delli Carpini, Michael X., and Scott Keeter. 1996. *What Americans Know about Politics and Why It Matters*. New Haven, CT: Yale University Press.
- Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: HarperCollins.
- Duelfer, Charles. 2004. "Comprehensive Report of the Special Advisor to the Director of Central Intelligence on Iraq's Weapons of Mass Destruction." <http://www.gpoaccess.gov/duelfer/index.html> (October 10, 2008).
- Edwards, Ward. 1982. "Conservatism in Human Information Processing." In *Judgment under Uncertainty: Heuristics and Biases*, ed. Daniel Kahneman, Paul Slovic, and Amos Tversky. New York: Cambridge University Press.
- Feldman, Stanley, and John Zaller. 1992. "The Political Culture of Ambivalence: Ideological Responses to the Welfare State." *American Journal of Political Science* 36 (February): 268–307.
- Fiorina, Morris P. 1981. *Retrospective Voting in American National Elections*. New Haven, CT: Yale University Press.
- Fischhoff, Baruch, and Sarah Lichtenstein. 1978. "Don't Attribute This to Reverend Bayes." *Psychological Bulletin* 85 (2): 239–43.
- Garrow, David J. 1986. *Bearing the Cross*. New York: William Morrow.
- Gerber, Alan, and Donald Green. 1999. "Misperceptions about Perceptual Bias." *Annual Review of Political Science* 2: 189–210.
- Gerber, Alan, and Donald P. Green. 1998. "Rational Learning and Partisan Attitudes." *American Journal of Political Science* 42 (July): 794–818.
- Gerber, Elisabeth R., and John E. Jackson. 1993. "Endogenous Preferences and the Study of Institutions." *American Political Science Review* 87 (September): 639–56.
- Gill, Jeff, and Lee D. Walker. 2005. "Elicited Priors for Bayesian Model Specifications in Political Science Research." *Journal of Politics* 67 (August): 841–72.
- Goodin, Robert E. 2002. "The Paradox of Persisting Opposition." *Politics, Philosophy, and Economics* 1 (February): 109–46.
- Grynaviski, Jeffrey D. 2006. "A Bayesian Learning Model with Applications to Party Identification." *Journal of Theoretical Politics* 18 (July): 323–46.
- Husted, Thomas A., Lawrence W. Kenny, and Rebecca B. Morton. 1995. "Constituent Errors in Assessing Their Senators." *Public Choice* 83 (June): 251–71.
- Jackman, Simon. 2005. "Pooling the Polls Over an Election Campaign." *Australian Journal of Political Science* 40 (December): 499–517.
- Kalman, Rudolf E. 1960. "A New Approach to Linear Filtering and Prediction Problems." *Journal of Basic Engineering, Series D* 82: 35–45.
- Lee, Peter M. 2004. *Bayesian Statistics: An Introduction*. 3rd ed. London: Arnold.
- Lohmann, Susanne. 1994. "The Dynamics of Informational Cascades: The Monday Demonstrations in Leipzig, East Germany, 1989–91." *World Politics* 47 (October): 42–101.
- Meinhold, Richard J., and Nozer D. Singpurwalla. 1983. "Understanding the Kalman Filter." *American Statistician* 37 (May): 123–27.
- Peffley, Mark, Stanley Feldman, and Lee Sigelman. 1987. "Economic Conditions and Party Competence: Processes of Belief Revision." *Journal of Politics* 49 (February): 100–21.
- Shaw, Daron R. 1999. "A Study of Presidential Campaign Event Effects from 1952 to 1992." *Journal of Politics* 61 (May): 387–422.
- Steenbergen, Marco R. 2002. "Political Belief Updating: An Experimental Investigation Within a Bayesian Framework." Presented at the Annual Scientific Meeting of the International Society of Political Psychology, Berlin.
- Tetlock, Philip E. 1986. "A Value Pluralism Model of Ideological Reasoning." *Journal of Personality and Social Psychology* 50 (April): 819–27.
- Tetlock, Philip E. 2005. *Expert Political Judgment*. Princeton, NJ: Princeton University Press.
- Wofford, Harris. 1980. *Of Kennedys and Kings*. New York: Farrar Straus Giroux.
- Zechman, Martin J. 1979. "Dynamic Models of the Voter's Decision Calculus: Incorporating Retrospective Considerations into Rational-Choice Models of Individual Voting Behavior." *Public Choice* 34 (September): 297–315.

John G. Bullock is assistant professor of political science, Yale University, New Haven, CT 06520.