

## **Part 3**

# **Bayesian Updating of Political Beliefs: Normative and Descriptive Properties**

Social scientists increasingly use Bayes' Theorem as a normative standard of rational political thinking (Bartels 2002; Gerber and Green 1998, 1999; Tetlock 2005; Steenbergen 2002) and as a tool to describe how people actually think (Bartels 1993; Achen 1992, 2002; Husted, Kenny, and Morton 1995; Grynaviski 2006; Lohmann 1994; Neill 2005). Models based on the Theorem are attractive because they offer a way to account for the weight that people place on old beliefs and new influences when revising their political ideas. They are also formal models, and as such, they bring the benefits of mathematical exposition to topics that have usually lacked it (Lupia 2002; see also Luce 1995). But uncertainty remains about the basis of their normative appeal and about whether they can accommodate everyday features of political cognition.

This essay clarifies those matters. After explaining the Theorem's appeal as a standard of rationality, I show that four important features of political thought—increased uncertainty in response to surprising information, selective

perception, attitude polarization, and enduring disagreement—are inconsistent with the most widely-used Bayesian updating model but quite consistent with other Bayesian models. Bayes' Theorem proves capable of capturing many features of real-world political thinking. But this flexibility is the Theorem's downside: precisely because it is consistent with so many different ways of thinking about politics, it is inadequate as a standard of rationality.

## Bayes' Theorem and Bayesian Updating Models

Most of the matters that interest political scientists—public opinion toward candidates, implications of new policies, the probability of terrorist attacks—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. A politician's ability to manage the economy, for example, may oscillate over time around a fixed but unknown mean. 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:

$$p(S|E) = \frac{p(E|S)p(S)}{p(E)} = \frac{p(E|S)p(S)}{\int p(E|S)p(S) dS}, \quad (3.1)$$

where  $S$  and  $E$  are events in a sample space and  $p(\cdot)$  is a probability distribution function. In words, the Theorem indicates that  $p(S|E)$ , the probability that  $S$  occurs conditional on  $E$  having occurred, is a function of the conditional probability  $p(E|S)$  and the marginal probabilities  $p(E)$  and  $p(S)$ . 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  $p(S)$  be a *belief* about  $S$ , i.e., a probability

distribution indicating someone's estimate of the extent to which  $S$  is true.  $E$  is evidence bearing on the belief. In this version of Bayes' Theorem, the estimated probability of  $S$  before observing  $E$  is given by  $p(S)$ ; it is often called the *prior probability of  $S$* , or simply the "prior." The estimated probability that  $S$  is true after observing  $E$  is  $p(S|E)$ , often called the *posterior probability of  $S$* . And  $p(E|S)$  is the likelihood function that one assigns to the evidence; it reflects a person's guess about the probability distribution from which the data are drawn. Understood in this way, the Theorem tells us how to revise any belief after receiving relevant evidence and subjectively estimating its likelihood. It is most often applied to beliefs about future events (Tetlock 2005), but it is fundamentally a tool for calculating probabilities, and it applies with equal force to all ideas that can be described in probabilistic terms. See Figure 3.1; for applications of Bayesian models to political attitudes and evaluations, see Bartels (1993, 2002) and Gerber and Green (1999).<sup>1</sup>

Bayes' Theorem is attractive as a normative standard of belief updating because it can be derived from two fundamental axioms of probability:

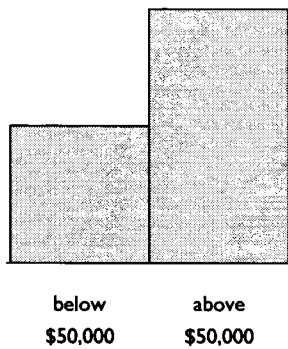
$$p(S \cap E) = p(E \cap S) \quad (3.2a)$$

$$p(S|E) = \frac{P(S \cap E)}{p(E)}. \quad (3.2b)$$

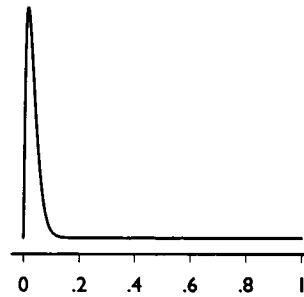
---

<sup>1</sup> If it seems confusing to think of attitudes in probabilistic terms, consider that attitudes are merely beliefs that objects are good or bad in some way, often accompanied by affective responses to those objects (Zanna and Rempel 1988; Abelson 1986). If I like John McCain, I have assigned a high probability to the hypothesis that he belongs to a category of objects that I like. And if reviewing new evidence causes me to like McCain less, I assign a lower probability to that hypothesis. This definition of attitude is in keeping with the view that much mental categorization is probabilistic (e.g., Smith and Medin 1981; Smith 1990).

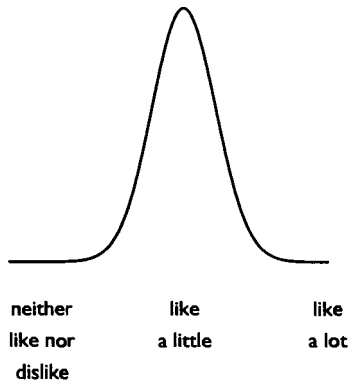
**Median Household Income in America**



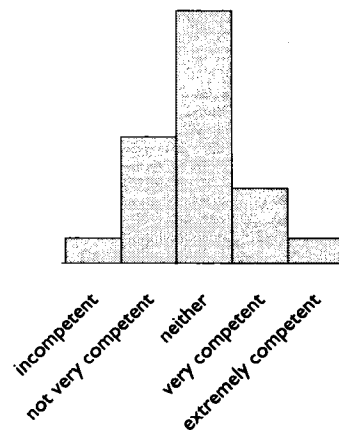
**Probability that Gingrich Wins the Presidency**



**Attitude toward John Edwards**



**Bill Clinton as Steward of the Economy**



**Figure 3.1: Attitudes, Evaluations, and Factual Beliefs Are Probability Distributions.** All of political cognition can be conceived in terms of probability distributions, and in doing so we win for political science the vast body of knowledge about subjective probability theory, the chief element of which is Bayes' Theorem. The upper left-hand panel depicts a factual belief about household income in the U.S.: the person holding this belief estimates that there is a 35% chance that the median household income is below \$50,000 and a 65% chance that it is greater than that. This is just a two-category discrete probability distribution. The upper right-hand panel depicts a belief about a future matter, Newt Gingrich's chance of winning the Presidency in 2008. The belief is a beta distribution: continuous, asymmetric, and bounded between 0 and 1 (Paolino 2001; Jackman 2008, ch. 2). The lower left-hand panel depicts a positive but somewhat ambivalent attitude about John Edwards: it is a normal distribution. The lower right-hand panel is an ambivalent voter's evaluation of Bill Clinton as a manager of the national economy—a discrete probability distribution with five categories.

Equation 3.2a says that the probability that  $S$  and  $E$  both occur equals the probability that  $E$  and  $S$  both occur. Equation 3.2b is a definition of conditional probability.<sup>2</sup>

No one consciously rejects either axiom, and following both of them requires that beliefs be updated according to Bayes' Theorem:  $p(S|E) = \frac{p(E \cap S)}{p(E)}$ , and  $p(E \cap S) = p(E|S)p(S)$ , so  $p(S|E) = \frac{p(E|S)p(S)}{p(E)}$ . By contrast, updating that does not correspond to Bayes' Theorem constitutes an implicit rejection of either or both of the axioms.<sup>3</sup>

Because the denominator of Equation 3.1 only serves to ensure that the posterior density integrates to one (and is therefore a proper probability density), Bayes' Theorem is more commonly expressed as

$$p(S|E) \propto p(E|S)p(S).$$

In words, "the posterior belief is proportional to the prior belief times the likelihood."

People are Bayesian if their posterior beliefs are determined in this fashion.

Importantly, Bayes' Theorem says nothing about what one's priors should be, what evidence one should use to update, or how one should interpret the evidence that one does use—a point to which we shall return.

In political science, one Bayesian updating model is far more common than others: the "normal-normal" model, so-called because it assumes both that people's priors are normally distributed and that they perceive new information to be normally distributed (e.g., Achen 1992; Bartels 1993, 2002; Gerber and Green 1999; Husted, Kenny, and Morton 1995; Gerber and Jackson 1993; Zechman 1979). Suppose that a voter is trying to learn about  $\mu$ , a politician's level of honesty. Initially, her belief

---

<sup>2</sup> Although conditional probability is usually presented as an axiom, Bernardo and Smith (1994, ch. 2) show that it can be derived from simpler axioms.

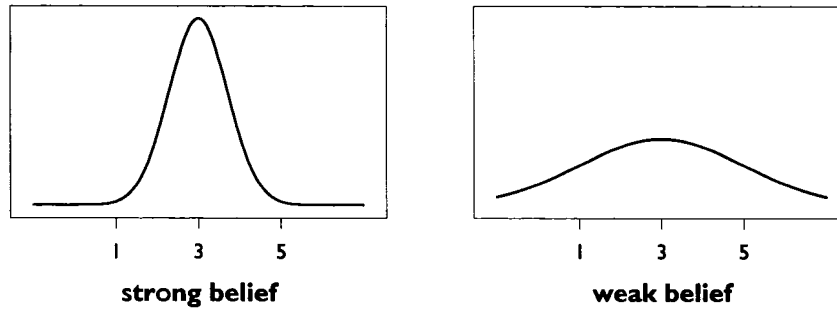
<sup>3</sup> This is the simplest valid treatment of a complex topic. For elaborate efforts to root Bayesian statistics in an axiomatic framework, see Savage (1954) and Pratt, Raiffa, and Schlaifer (1964).

## 94 Bayesian Updating of Political Beliefs

about his honesty is normally distributed:  $\mu \sim N(\mu_0, \sigma_0^2)$ . Later, she encounters a new message,  $x$ , that contains information about his level of honesty. She assumes that the message is a draw from a distribution with a mean equal to the parameter of interest.<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 normal. We write  $x \sim N(\mu, \sigma_x^2)$ . The variance of this distribution,  $\sigma_x^2$ , captures how definitive the new information is. If it is communicated directly from a highly credible source, the signal it sends is clear and the variance is quite small. But if it is merely a rumor that the voter spots in a tabloid, the signal is only slightly informative and its variance will be high. Similarly, the variance of a prior or posterior belief is a measure of the confidence with which it is held: the higher the variance, the less confidence one places in one’s estimate of  $\mu$ . (See Figure 3.2.)

---

<sup>4</sup> Assuming that the mean of the message distribution is  $\mu$  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 needs to adjust for that bias before updating. This is no obstacle to Bayesian updating (e.g., Jackman 2005), but it is not part of the normal-normal model.



**Figure 3.2: The Variance of a Belief Indicates Its Strength.** Both panels depict beliefs that are normal distributions with means of 3. The distribution in the left-hand panel has a variance of .25: the person who holds this belief is quite confident that the parameter of interest is about 3. By contrast, the distribution in the right-hand panel has a variance of 4. The person who holds this belief is not confident that the parameter of interest is close to 3; to him, it could easily be around 1 or 5 or some value even more distant from 3.

By a common result (e.g., Box and Tiao 1973), a voter with a normal prior belief who updates according to Bayes' Theorem in response to  $x$  will have posterior belief  $\mu|x \sim N(\mu_1, \sigma_1^2)$ , where

$$\begin{aligned} \mu_1 &= \frac{\sigma_0^2 x + \sigma_x^2 \mu_0}{\sigma_0^2 + \sigma_x^2} \\ &= \mu_0 \left( \frac{1/\sigma_0^2}{1/\sigma_0^2 + 1/\sigma_x^2} \right) + x \left( \frac{1/\sigma_x^2}{1/\sigma_0^2 + 1/\sigma_x^2} \right), \quad \text{and} \end{aligned} \quad (3.3a)$$

$$\sigma_1^2 = \frac{\sigma_0^2 \sigma_x^2}{\sigma_0^2 + \sigma_x^2} = \frac{1}{1/\sigma_0^2 + 1/\sigma_x^2}. \quad (3.3b)$$

The posterior mean,  $\mu_1$ , is a weighted average of the mean of the prior belief and the new message. The weights are determined by the *precisions*, i.e., the reciprocals of the variances of the prior belief and the new message. This is a fantastically convenient result, as it permits us to compute posterior means without multiplying the prior probability distribution by the likelihood. And this convenience helps to account for

the popularity of the normal-normal model. But the model has shortcomings that make it normatively unattractive and descriptively unrealistic:

1. *Surprise is impossible.* The model implies that people always become more certain of their beliefs over time.
2. *Polarization is unthinkable.* Under widely assumed conditions, the model implies that people who initially disagree are literally incapable of holding posterior beliefs that are less alike than their priors.
3. *Agreement is inevitable.* The model implies that people will always disagree less as they learn more. Furthermore, learning enough will always cause their beliefs to converge to agreement.<sup>5</sup>

All of these shortcomings are peculiar to the normal-normal model; they are not inherent properties of Bayesian updating. In the remainder of this article, I elaborate each of these shortcomings and define other Bayesian updating models that surmount them.

---

<sup>5</sup> These two statements are not equivalent: it is possible for an updating model to imply ever-diminishing disagreement without implying eventual agreement. But the normal-normal model implies both.



## The Normal-Normal Model Implies that People Always Become More Certain, But Other Bayesian Updating Models Do Not

A desirable property of any learning model is that surprising new evidence can cause people to become less sure of their beliefs. A distressing property of the normal-normal model is that surprising new evidence always makes people *more* sure of their beliefs. Formally, note that the variance of a posterior belief in the normal-normal model can be expressed as

$$\sigma_1^2 = \sigma_0^2 \left( \frac{\sigma_x^2}{\sigma_0^2 + \sigma_x^2} \right).$$

If  $\sigma_x^2$  is finite,  $\frac{\sigma_x^2}{\sigma_0^2 + \sigma_x^2} < 1$ , and therefore  $\sigma_1^2 < \sigma_0^2$ : updating will always make people more certain of their belief. Moreover, the extent to which new information is surprising has no bearing on the extent to which it changes the certainty of one's beliefs. This shortcoming of the normal-normal model is often noted (Leamer 1978; Gerber and Green 1998; Bartels 1993, 2002; Grynaviski 2006), even though it has not dented the model's popularity. Fortunately, the problem is anything but endemic to Bayesian updating. It is an artifact of an unrealistic assumption of the normal-normal model: that we know the variance of the distribution that we are trying to learn about, and need only estimate its mean. This situation is as rare in politics as it is in any other domain. A model in which we simultaneously learn about the mean and variance of the unknown distribution is both more realistic on its face and capable of accommodating cases in which people become less certain over time.

## Updating with Unknown Mean and Unknown Variance

Instead of trying to learn only about the mean of a distribution, we are now trying to learn about both its mean,  $\mu$ , and its variance,  $\sigma^2$ . Our prior belief about these parameters is a bivariate joint distribution,  $p(\mu, \sigma^2)$ . It can be expressed as the product of a conditional prior belief about the mean,  $p(\mu|\sigma^2)$ , and a marginal prior belief about the variance,  $p(\sigma^2)$ . Many different densities might be used to model these priors, but two of the most obvious choices are a normal distribution for the mean and an inverse-Gamma distribution for the variance. (The inverse-Gamma distribution is attractive for modeling variances because it is continuous, flexible, and has a lower bound of 0 but no upper bound.) Specifically, the normal/inverse-Gamma prior distribution can be expressed as the product of two densities,

$$\mu|\sigma^2 \sim N\left(\mu_0, \frac{\sigma^2}{n_0}\right)$$

$$\sigma^2 \sim \text{inverse-Gamma}\left(\frac{\nu_0}{2}, \frac{\nu_0\sigma_0^2}{2}\right),$$

where

- $\mu_0$  is the mean of the prior belief about  $\mu$
- $\sigma^2/n_0$  is the variance of the prior distribution for  $\mu$ , conditional on  $\sigma^2$ .  $n_0$  is interpretable as “prior sample size”: the smaller it is, the larger the variance of the prior belief, reflecting the fact that a prior based on less information (or fewer “prior observations”) is less precise than a prior based on more information
- $\nu_0 > 0$  is a prior shape parameter, i.e., a prior “degrees of freedom” parameter

- $\nu_0\sigma_0^2$  is a prior scale parameter, equivalent to the sum of squared residuals one would obtain from a previously observed dataset of size  $\nu_0$  in which each observation came from a  $N(\mu_0, \sigma_0^2)$  dataset.

The marginal inverse-Gamma prior distribution has a  $\chi^2$  shape, and indeed, the inverse- $\chi^2$  distribution is a special case of the inverse-Gamma distribution. (Gelman et al. 2004 discuss updating with normal/inverse- $\chi^2$  priors; see Grynaviski 2006 for an application.)

Suppose that we encounter messages  $\mathbf{x} = (x_1, \dots, x_n)$  that we believe bear directly on the parameters about which we are trying to learn; i.e., we believe that  $x_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2) \forall i \in (1, \dots, n)$ . If our prior belief about  $\mu$  and  $\sigma^2$  is a normal/inverse-Gamma distribution with parameters  $\mu_0, n_0, \nu_0$ , and  $\sigma_0^2$ , our posterior belief will also be a normal/inverse-Gamma distribution:

$$\mu|\sigma^2, \mathbf{x} \sim N\left(\mu_1, \frac{\sigma^2}{n_0 + n}\right) \text{ and} \tag{3.4a}$$

$$\sigma^2|\mathbf{x} \sim \text{inverse-Gamma}\left(\frac{\nu_1}{2}, \frac{\nu_1\sigma_1^2}{2}\right), \tag{3.4b}$$

where  $\mu_1 = \frac{n_0\mu_0 + n\bar{x}}{n_0 + n}$ ,  $\nu_1 = \nu_0 + n$ , and  $\nu_1\sigma_1^2 = \nu_0\sigma_0^2 + \sum_{i=1}^n (x_i - \bar{x})^2 + \frac{n_0n}{n_0 + n} (\mu_0 - \bar{x})^2$ .

Usually, interest focuses on the marginal posterior distribution of  $\mu$ , which is a  $t$  distribution:

$$\mu|\mathbf{x} = t_{\nu_1}\left(\mu_1, \sqrt{\sigma_1^2 / (n_0 + n)}\right).$$

Now, the marginal prior distribution of  $\mu$  is  $t_{\nu_0}(\mu_0, \sqrt{\sigma_0^2/n_0})$ . If  $\sigma_0^2/n_0 < \sigma_1^2/(n_0 + n)$ , the posterior belief about  $\mu$  has a higher variance than the prior. This occurs when

$$\begin{aligned} \frac{\sigma_0^2}{n_0} &< \frac{\sigma_1^2}{n_0 + n} \\ \sigma_0^2 \left( \frac{n_0 + n}{n_0} \right) &< \frac{\nu_0 \sigma_0^2 + \sum_{i=1}^n (x_i - \bar{x})^2 + \frac{n_0 n}{n_1} (\mu_0 - \bar{x})^2}{\nu_0 + n} \\ \sigma_0^2 \left( \frac{n_0 + n}{n_0} \right) (\nu_0 + n) - \nu_0 \sigma_0^2 &< \sum_{i=1}^n (x_i - \bar{x})^2 + \frac{n_0 n}{n_0 + n} (\mu_0 - \bar{x})^2 \\ \frac{\sigma_0^2}{n_0} (n \nu_0 + n n_0 + n^2) + \frac{\sigma_0^2}{n_0} (n_0 \nu_0) - \nu_0 \sigma_0^2 &< \sum_{i=1}^n (x_i - \bar{x})^2 + \frac{n_0 n}{n_0 + n} (\mu_0 - \bar{x})^2 \\ \frac{\sigma_0^2}{n_0} [n (\nu_0 + n_0 + n)] &< \sum_{i=1}^n (x_i - \bar{x})^2 + \frac{n_0 n}{n_0 + n} (\mu_0 - \bar{x})^2. \quad (3.5) \end{aligned}$$

Equation 3.5 establishes that people who update according to Equations 3.4a and 3.4b will become less sure if they learn from data that are sufficiently surprising ( $(\bar{x} - \mu_0)^2$  is large enough) or sufficiently vague ( $\sum_{i=1}^n (x_i - \bar{x})^2$  is large enough). Thus, in contrast to the normal-normal model, a model that presumes that both a mean and a variance are unknown permits new information to shake the confidence that people repose in their beliefs.

## Bayesian Updating Models Can Accommodate Biased Interpretation of Political Information

It is common to hear or read that “Bayesian updating requires independence between priors and new evidence” (Taber and Lodge 2006, 767; see also Ottati 1990, 160; Fischle 2000). The notion underpinning the claim is, presumably, that Bayes’ Theorem demands that beliefs correspond to some objective conception of reality. But there

is not even a germ of truth to such claims. Nothing in Bayes' Theorem—nothing in the other writing of Reverend Bayes—nothing in the writing of his contemporary, Laplace—nothing in the whole of Bayesian statistics past or present warrants such a claim. “Objective perception” of political information may belong in a standard of rationality—and this is a point to which I shall return—but it has no place in a framework for belief updating. Indeed, the entire history of Bayesian scholarship militates against the notion that understanding of probability requires or profits from objective perception of the world (de Finetti 1974; Savage 1964; Jeffrey 2004).<sup>6</sup> Still, there are some for whom the dream will never die, and Bayesian updating models are quite able to accommodate them.

Partisanship is often thought to influence political views through *selective perception*: the assimilation, from the evidence at hand, of only or chiefly those details that support one's prior beliefs (Campbell et al. 1960, esp. Chapter 6; Bartels 2002; Taber and Lodge 2006; Jacobson 2006; Gaines et al. 2007; see also Bullock 2006). Nothing in the normal-normal model is incompatible with selective perception; like Bayes' Theorem, it makes no prescription about the way in which evidence is to be interpreted. Still, researchers who have a standard of objective perception may prefer a model that explicitly distinguishes between selective and objective perception. The normal-normal model can be adapted to this task.

We begin by distinguishing “good” messages that comport with one's prior and “bad” messages that do not. Formally, let the former set of messages be  $\mathbf{x}_g = (x_1, \dots, x_t)$  and the latter set be  $\mathbf{x}_b = (x_{t+1}, \dots, x_T)$ .  $x_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma_x^2) \forall i \in (1, \dots, T)$ .

---

<sup>6</sup> Of course, this is not to deny that there are objective probabilities; although many contemporary “subjectivist Bayesians” (e.g., de Finetti, Savage, Diaconis) deny that there are, Bayes and Laplace did not. Note that there is a school of thought that goes by the name “Objective Bayes,” but the “objectivity” that its members favor amounts to the rejection of certain types of prior beliefs as inappropriate—not to the assumption that there are objectively correct likelihoods for data, least of all in the ambiguous world of politics. On both counts, see Press 2003.

$\mathbf{x}_g \equiv x_g \sim N(\mu, \sigma_x^2/t)$ ;  $\mathbf{x}_b \equiv x_b \sim N(\mu, \sigma_x^2/(T-t))$ . If one's prior belief is  $\mu \sim N(\mu_0, \sigma_0^2)$ , and one updates according to the normal-normal model, one's posterior is  $\mu|\mathbf{x}_g, \mathbf{x}_b \sim N(\mu_1, \sigma_1^2)$ , where

$$\mu_1 = \mu_0 \left( \frac{\sigma_1^2}{\sigma_0^2} \right) + \bar{x}_g \left( \frac{\sigma_1^2}{\sigma_x^2/t} \right) + \bar{x}_b \left( \frac{\sigma_1^2}{\sigma_x^2/(T-t)} \right),$$

$\bar{x}_g = \sum_{i=1}^t x_i/t$ , and  $\bar{x}_b = \sum_{i=t+1}^T x_i/(T-t)$ . As each  $x_i$  is subjectively determined, so too are  $\bar{x}_g$  and  $\bar{x}_b$ : different people interpreting the same evidence may come up with different values of these "sample means." But if we, as researchers, have a standard of objective interpretation, we can make this subjective assessment explicit by using a modified model. The formula for the mean of the posterior is given by Gerber and Green (1999):

$$\begin{aligned} \mu_1 = & \mu_0 \left( \frac{1/\sigma_0^2}{1/\sigma_0^2 + t/\sigma_x^2 + (T-t)/\sigma_x^2} \right) \\ & + \alpha_g \bar{x}_{g*} \left( \frac{t/\sigma_x^2}{1/\sigma_0^2 + t/\sigma_x^2 + (T-t)/\sigma_x^2} \right) + \alpha_b \bar{x}_{b*} \left( \frac{(T-t)/\sigma_x^2}{1/\sigma_0^2 + t/\sigma_x^2 + (T-t)/\sigma_x^2} \right). \end{aligned} \quad (3.6)$$

$\alpha_g$  and  $\alpha_b$  are the *selection weights*; they indicate the extent to which favorable and unfavorable messages are misinterpreted.  $\bar{x}_{g*}$  and  $\bar{x}_{b*}$  are the objective sample means of the favorable and unfavorable messages, i.e., the sample means in the eyes of the researcher. In the absence of selective perception,  $\alpha_g = \alpha_b = 1$ . If higher values of  $x_i$  are preferable, and selective perception consists of giving an unduly favorable interpretation to bad news,  $\alpha_b$  will be greater than 1. If selective perception consists of exaggerating the good news provided by favorable information,  $\alpha_g$  will be greater than one.

A closely related form of political bias is at work when people attribute more or less credibility to a news source than it deserves: when Communist Party officials exalt the *People's Daily*, perhaps; or when Republicans take Rush Limbaugh at face value.

This calls for a slightly different model:

$$\begin{aligned} \mu_1 = & \mu_0 \left( \frac{1/\sigma_0^2}{1/\sigma_0^2 + t/\sigma_x^2 + (T-t)/\sigma_x^2} \right) \\ & + \bar{x}_g \left( \frac{\alpha_g t/\sigma_x^2}{1/\sigma_0^2 + \alpha_g t/\sigma_x^2 + (T-t)/\sigma_x^2} \right) + \bar{x}_b \left( \frac{\alpha_b (T-t)/\sigma_x^2}{1/\sigma_0^2 + t/\sigma_x^2 + \alpha_b (T-t)/\sigma_x^2} \right). \end{aligned} \quad (3.7)$$

Here,  $\alpha_g$  and  $\alpha_b$  apply not to the content of new information but to its credibility. If one overrates the credibility of favorable messages,  $\alpha_g$  is greater than 1: it is as though he is responding to more favorable messages than he really received. If one underrates the credibility of unfavorable messages,  $\alpha_b$  is less than 1: it is as though he is responding to fewer favorable messages than he has received.

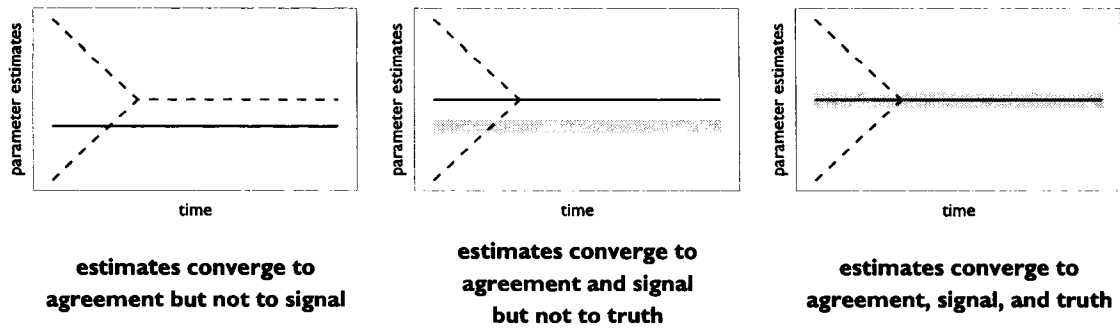
## Convergence and Polarization of Public Opinion under Bayesian Updating

No issue in Bayesian analysis of public opinion is more disputed than the implications of Bayesian updating for disagreement among people with different prior beliefs. Gerber and Green (1999, 203-05) maintain that if Republicans and Democrats are Bayesian, they will agree neither more nor less as they update in response to new evidence. Empirically, Gerber and Green find just this patterning of presidential approval over time and adduce it as evidence that many people are Bayesians whose views are unaffected by partisan bias. Bartels (2002) cites the same public opinion data as evidence that people are biased Bayesians or not Bayesian at all:

unbiased application of Bayes' Theorem, he writes, implies convergence of public opinion. Achen (2005, 334) agrees, and Grynawski (2006, 331) claims that Bartels "formally proved" that Bayesian updaters will "inexorably come to see the world in the same way." (He didn't.) A closely related dispute is about polarization of public opinion: Gerber and Green (1999) write that attitude polarization (Lord, Ross, and Lepper 1979) is incompatible with Bayesian updating except under very unusual circumstances, while Steenbergen (2002, 7-8) concludes exactly the opposite. There is at least a little truth to all of these positions: Bayesian updating does imply convergence of public opinion under some conditions, but these are more numerous and more stringent than the discussions to date have acknowledged.

Before embarking on a series of proofs, it will help to distinguish between three kinds of convergence, all of which are depicted in Figure 3.3. *Convergence to agreement* occurs when prior beliefs converge to the same belief after updating. *Convergence to signal* occurs when people's beliefs converge to the mean of the distribution of messages that they are using to update; if beliefs converge to the same signal, they also converge to agreement. This kind of convergence has been widely discussed in Bayesian statistics, where it is subsumed by the broader topic of consistency of Bayes estimates. In that literature, convergence to signal has been proved to hold under general conditions for a wide variety of prior and data distributions (e.g., Diaconis and Freedman 1986; Strasser 1981); I focus here on the case of normal priors and data because of the ubiquity of these assumptions in political science. *Convergence to truth* occurs when beliefs converge to the true parameter of interest. Often, convergence to signal implies convergence to truth, but not always. If beliefs are updated in response to messages from a biased news source, convergence to signal implies that beliefs are not converging to the truth.





**Figure 3.3: Convergence to Agreement, to Signal, and to Truth.** The dashed lines depict parameter estimates by two people. They converge in each panel; this is *convergence to agreement*. The solid black line in each panel represents the mean of the distribution of information that people are using to update their estimates, e.g., the distribution from which news articles are drawn. In the leftmost panel, the parameter estimates do not converge to the mean of this distribution. In the middle and rightmost panels, they do: this is *convergence to signal* as well as convergence to agreement.

The grey line in the last two panels indicates the true parameter value. In the middle panel, it differs from the mean of the information distribution. This occurs whenever the information that people use to update their beliefs is biased on average. In the rightmost panel, the mean of the information distribution is also the true parameter value: here we have *convergence to truth* as well as to agreement and to signal.

## Convergence of Public Opinion Under the Normal-Normal

### Model

**Proposition 1.** A person’s prior belief is  $\mu \sim N(\mu_0, \sigma_0^2)$ . He updates according to Equation 3.3a in response to  $\mathbf{x}$ , a sample of  $t$  messages that he perceives to have mean  $\bar{x}$ , with

$x_i \stackrel{iid}{\sim} N(\mu, \sigma_x^2) \forall i \in (1, \dots, t)$ . If  $t$  is large enough, his belief will converge to  $\bar{x}$ .

**Proof.** By a result shown in the appendix, the  $t$  messages are equivalent to a single message  $\bar{x}$  from distribution  $N(\mu, \sigma_x^2/t)$ . Suppose  $\epsilon > 0$  and  $T = \frac{\sigma_x^2(|\mu_0 - \bar{x}| - \epsilon)}{\sigma_0^2 \epsilon}$ . Then  $t > T$  implies

$$\begin{aligned}
|\mu|x - \bar{x}| &= \left| \frac{\sigma_0^2 \bar{x} + (\sigma_x^2/t)\mu_0 - \bar{x}(\sigma_0^2 + \sigma_x^2/t)}{\sigma_0^2 + \sigma_x^2/t} \right| \\
&= \left| \frac{(\sigma_x^2/t)(\mu_0 - \bar{x})}{\sigma_0^2 + \sigma_x^2/t} \right| \\
&= \frac{\sigma_x^2 |\mu_0 - \bar{x}|}{t\sigma_0^2 + \sigma_x^2} \\
&< \frac{\sigma_x^2 |\mu_0 - \bar{x}|}{T\sigma_0^2 + \sigma_x^2} \\
&= \frac{\sigma_x^2 |\mu_0 - \bar{x}|}{\frac{\sigma_0^2 [\sigma_x^2 (\mu_0 - \bar{x}) - \epsilon]}{\sigma_0^2 \epsilon} + \sigma_x^2} \\
&= \frac{\sigma_x^2 |\mu_0 - \bar{x}|}{\frac{\sigma_x^2 (\mu_0 - \bar{x}) - \epsilon}{\epsilon} + \sigma_x^2} \\
&= \frac{\epsilon \sigma_x^2 |\mu_0 - \bar{x}|}{\sigma_x^2 (\mu_0 - \bar{x}) - \epsilon + \epsilon \sigma_x^2} \\
&= \epsilon.
\end{aligned}$$

□

**Discussion.** The proof reveals conditions under which any Bayesian updater's belief will converge to a subjectively defined  $\bar{x}$ . To some (e.g., Grynaviski 2006; Bartels 2002), it seems a short step to infer that Bayesians who initially disagree will come to agree with each other. In fact, the conditions set forth in the proof are quite stringent, and the conditions required for convergence to agreement are more stringent still.

First among the requirements for convergence to agreement is simply that people are Bayesian or that they adopt non-Bayesian updating rules that nevertheless permit convergence. This is a point too often elided by those who take nonconvergence as proof of partisan bias (e.g., Bartels 2002). There is ample laboratory evidence

that people are not Bayesian updaters, and while most documented non-Bayesian tendencies do nevertheless permit convergence (Phillips and Edwards 1966; Tversky and Kahneman 1971), some may not (Chapman and Chapman 1959; Hamm 1993).

The second requirement is that people with different priors perceive the new information in the same way: technically, they must agree on the value of  $\bar{x}$ . This rules out selective perception (unless people with different prior selectively perceive evidence in the same way—an unlikely circumstance, given that priors influence the strength and direction of selective perception). It also rules out cases in which people are updating evaluations that simply reflect different values. For example, if a Republican president's economic policies are consistent with Republican values and inconsistent with Democratic ones, we should not expect convergence even if members of both parties are Bayesians who have the same understanding of new economic information.

A third requirement likely to be violated is that people update exclusively on the basis of the same set of messages. If they do not—if, for example,  $i$  and  $j$  both update on the basis of  $x$ , but  $i$  also updates on the basis of messages that seem have a different sample mean—there is no reason to expect updating to cause agreement. This rules out *selective exposure*, whereby people with different views may systematically expose themselves chiefly to congenial news sources (Taber and Lodge 2006; Gentzkow and Shapiro 2006). It was once argued that selective exposure is uncommon—especially in politics—because people do not consciously seek to reinforce their views through their choice of media (Sears and Freedman 1967; Frey 1986). But selective exposure does not require a reinforcement motive (Katz 1968), and the heightened sorting of the electorate (Levendusky 2006) and splintering of the market for news into specialized niches (Prior 2007, Chapter 4) may have made it increasingly common.

A fourth requirement is that the parameter about which people are updating is constant over time. As I show below, weakening this assumption allows nonconvergence and polarization when people update their beliefs, even if they are updating in response to the same information and are interpreting that information in the same way.

Finally, complete convergence to agreement—the assumption that Bayesians will “inexorably come to see the world in the same way” (Grynaviski 2006, 331)—can only occur if updaters are responding to a set of messages so powerful that it causes them to completely ignore their prior beliefs. (Formally, the precision of the set of new messages must be infinitely greater than the precision of the prior beliefs.) Even relative to the other conditions, this is unrealistic. Bartels (1993) argues forcefully that people’s beliefs about candidates at the start of Presidential campaigns are far stronger than we usually imagine. And it is widely known that most Americans are exposed to only meager amounts of political news (Campbell et al. 1960; Zaller 1992; Delli Carpini and Keeter 1996; Prior 2007). The assumption may hold in the extremely long run, but by then, as Keynes noted, we’ll be dead.

Almost no interesting political predicaments satisfy all of these conditions. This suggests that the recent focus on convergence has been misplaced: the failure of Republicans and Democrats to evaluate the President in the same way or to otherwise share the same beliefs may be evidence of selective perception, but it may also be due to any of several other, quite likely factors. On the other hand, convergence becomes more likely as more of these conditions are met, and it would be quite surprising if Bayesian updating did *not* imply convergence of public opinion when all of them are met. That is why the next two proofs are interesting: they show that convergence may not occur under Bayesian updating if these conditions are only slightly relaxed.

## Nonconvergence and Polarization Under Bayesian Updating with Selective Perception

**Proposition 2: Nonconvergence and Polarization Under Bayesian Updating with Selective Perception.** Let voters  $i$  and  $j$  have prior belief  $\mu \sim N(\mu_0, \sigma_0^2)$ . Both update in response to  $\mathbf{x}$ , a set of  $T$  messages, under the presumption that  $x_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma_x^2) \forall i \in (1, \dots, T)$ . As in Equation 3.6,  $\mathbf{x}$  can be partitioned into  $\mathbf{x}_g$ , a set of  $t$  messages that favor the voters' prior, and  $\mathbf{x}_b$ , a set of  $(T - t)$  messages that contradict the voters' prior. Assume  $t \geq 1$  and  $(T - t) \geq 1$ . The objective mean of  $\mathbf{x}_g$  is  $\bar{x}_{g^*}$  and the objective mean of  $\mathbf{x}_b$  is  $\bar{x}_{b^*}$ ; again, these "objective" means are stipulated by the researcher. Assume  $\bar{x}_{g^*} \neq \bar{x}_{b^*}$ . Voter  $i$  updates by Equation 3.6 with selection weights  $\alpha_{gi}$  and  $\alpha_{bi}$ . Voter  $j$  updates by Equation 3.6 with selection weights  $\alpha_{gj}$  and  $\alpha_{bj}$ . If  $\alpha_{gi} \neq \alpha_{gj}$  or  $\alpha_{bi} \neq \alpha_{bj}$ , convergence to agreement cannot occur unless

$$(\alpha_{gi} - \alpha_{gj}) \left( \frac{t\bar{x}_{g^*}}{T} \right) = (\alpha_{bi} - \alpha_{bj}) \left( \frac{(T-t)\bar{x}_{b^*}}{T} \right),$$

and that will only occur by chance.

**Proof.** By Appendix A, updating in response to the messages in  $\mathbf{x}_g$  and  $\mathbf{x}_b$  is equivalent to updating in response to a single draw

$$\begin{aligned} \bar{x}_* &= \frac{\alpha_g \bar{x}_{g^*} \sigma_x^2 / (T-t) + \alpha_b \bar{x}_{b^*} \sigma_x^2 / t}{\sigma_x^2 / t + \sigma_x^2 / (T-t)} \\ &= \frac{\alpha_g \bar{x}_{g^*} \sigma_x^2 / (T-t) + \alpha_b \bar{x}_{b^*} \sigma_x^2 / t}{\sigma_x^2 T / n(T-t)} \\ &= \left( \alpha_g \bar{x}_{g^*} \sigma_x^2 / (T-t) + \alpha_b \bar{x}_{b^*} \sigma_x^2 / t \right) \frac{n(T-t)}{\sigma_x^2 T} \\ &= \frac{t\alpha_g \bar{x}_{g^*} + (T-t)\alpha_b \bar{x}_{b^*}}{T} \end{aligned}$$

from a normal distribution with variance  $\sigma_x^2/T$ . By Equation 3.3a, the posterior mean is

$$\mu|\mathbf{x} = \mu_0 \left( \frac{1/\sigma_0^2}{1/\sigma_0^2 + T/\sigma_x^2} \right) + \bar{x}_* \left( \frac{T/\sigma_x^2}{1/\sigma_0^2 + T/\sigma_x^2} \right).$$

Assume  $\epsilon > 0$ . By Proposition 1,  $|\mu|\mathbf{x} - \bar{x}_*| < \epsilon$  for  $T > \frac{\sigma_x^2(\mu_0 - \bar{x}_* - \epsilon)}{\sigma_0^2 \epsilon}$ , i.e.,  $\mu|\mathbf{x}$  will converge to  $\bar{x}_*$ . It only remains to show that  $\bar{x}_*$  takes on different values for  $i$  and  $j$  when they have different selection weights:

$$\begin{aligned} \bar{x}_{*i} - \bar{x}_{*j} &= \frac{t\alpha_{gi}\bar{x}_{g*} + (T-t)\alpha_{bi}\bar{x}_{b*} - t\alpha_{gj}\bar{x}_{g*} + (T-t)\alpha_{bj}\bar{x}_{b*}}{T} \\ &= (\alpha_{gi} - \alpha_{gj}) \left( \frac{t\bar{x}_{g*}}{T} \right) - (\alpha_{bi} - \alpha_{bj}) \left( \frac{(T-t)\bar{x}_{b*}}{T} \right). \end{aligned}$$

where  $\bar{x}_{*i}$  is the value of  $\bar{x}_*$  using  $i$ 's selection weights and  $\bar{x}_{*j}$  is the value of  $\bar{x}_*$  using  $j$ 's selection weights. By assumption,  $t$  and  $T - t$  are positive, and  $\bar{x}_{g*} \neq \bar{x}_{b*}$ , so the posterior beliefs of  $i$  and  $j$  will never converge to agreement unless

$$(\alpha_{gi} - \alpha_{gj}) \left( \frac{t\bar{x}_{g*}}{T} \right) = (\alpha_{bi} - \alpha_{bj}) \left( \frac{(T-t)\bar{x}_{b*}}{T} \right). \quad \square$$

**Discussion.** The proof shows that Bayesians with the same prior beliefs who update under selective perception will generally have different prior beliefs; it is therefore not just a nonconvergence result but a polarization result, too. Of course, nonconvergence is no less likely when  $i$  and  $j$  have different prior beliefs: it depends entirely on the difference between  $\bar{x}_{*i}$  and  $\bar{x}_{*j}$ , and not at all on the difference between prior

beliefs. And even when  $i$  and  $j$  have different priors, selective perception will lead to polarization when

$$|\mu_{0i} - \mu_{0j}| < \left| \left( \alpha_{gi} - \alpha_{gj} \right) \left( \frac{t\bar{x}_{g^*}}{T} \right) - \left( \alpha_{bi} - \alpha_{bj} \right) \left( \frac{(T-t)\bar{x}_{b^*}}{T} \right) \right|.$$

### Convergence and Polarization under Kalman-Filter Updating

The models considered to this point assume that people are trying to learn about  $\mu$ , a quality of the political environment that does not change. Achen (1992), for example, assumes that the Democratic and Republican Parties offer benefits to each voter that oscillate over time around a mean benefit level that never changes, and he uses the assumption to justify his use of the normal-normal model to study changes in party identification. Such fixed-mean assumptions are apt when we are trying to learn about history and perhaps when we are trying to update our beliefs over the short term. But they are inappropriate when the parameter of interest changes over time. *Pace* Achen, the net benefit that I derive from a party changes as my views or economic status change. 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 increasingly under the sway of constituents whose views I oppose. In all of these cases, the constant-parameter assumption is an approximation at best.

Suppose that a Bayesian is trying to learn about a parameter that changes according to the rule

$$\alpha_t = \gamma\alpha_{t-1} + \epsilon_\alpha, \quad \epsilon_\alpha \sim N(0, \sigma_\alpha^2) \tag{3.8}$$

## 112 Bayesian Updating of Political Beliefs

where  $t$  is the current period,  $\gamma$  is a known autoregressive parameter, and  $\epsilon_\alpha$  is a disturbance term with known, finite, nonzero variance  $\sigma_\alpha^2$ . Let  $x_1, \dots, x_{t-1}, x_t$  denote the observed values of the parameter of interest at times  $1, \dots, t-1, t$ . The relationship between  $x_t$  and  $\alpha_t$  is

$$x_t = \alpha_t + \epsilon_x, \quad \epsilon_x \sim N(0, \sigma_x^2) \quad (3.9)$$

where  $\sigma_x^2$  is a known, finite, nonzero variance term. (Following Gerber and Green 1998 and Green, Gerber, and de Boef 1999,  $\gamma$ ,  $\sigma_\alpha^2$ , and  $\sigma_x^2$  are held constant for simplicity of exposition. But the model described in this section can easily accommodate the case in which they change over time. See Meinhold and Singpurwalla 1983 for an example.)

His initial belief is

$$(\alpha_0 | \gamma, \sigma_\alpha^2, \sigma_x^2) \sim N(\hat{\alpha}_0, P_0).$$

Looking forward to period 1, his prior belief about the parameter of interest is governed by Equation 3.8:

$$(\alpha_1 | \gamma, \sigma_\alpha^2, \sigma_x^2) \sim N(\gamma \hat{\alpha}_0, \gamma^2 P_0 + \sigma_\alpha^2).$$



But after receiving message  $x_1$ , he updates. Because both his prior belief and the perceived likelihood of  $x_1$  are normal, he updates according to Equations 3.3a and 3.3b:

$$(\alpha_1 | \gamma, \sigma_\alpha^2, \sigma_x^2, x_1) \sim N(\hat{\alpha}_1, P_1), \text{ where}$$

$$\hat{\alpha}_1 = \gamma \hat{\alpha}_0 \left[ \frac{1/(\gamma^2 P_0 + \sigma_\alpha^2)}{1/(\gamma^2 P_0 + \sigma_\alpha^2) + 1/\sigma_x^2} \right] + x_1 \left[ \frac{1/\sigma_x^2}{1/(\gamma^2 P_0 + \sigma_\alpha^2) + 1/\sigma_x^2} \right]$$

$$P_1 = \frac{1}{1/(\gamma^2 P_0 + \sigma_\alpha^2) + 1/\sigma_x^2}.$$

And generally,

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

$$\hat{\alpha}_t = \gamma \hat{\alpha}_{t-1} \left[ \frac{1/(\gamma^2 P_{t-1} + \sigma_\alpha^2)}{1/(\gamma^2 P_{t-1} + \sigma_\alpha^2) + 1/\sigma_x^2} \right] + x_t \left[ \frac{1/\sigma_x^2}{1/(\gamma^2 P_{t-1} + \sigma_\alpha^2) + 1/\sigma_x^2} \right] \quad (3.10)$$

$$P_t = \frac{1}{1/(\gamma^2 P_{t-1} + \sigma_\alpha^2) + 1/\sigma_x^2}, \quad (3.11)$$

and where  $\mathbf{x}_{t-1}$  is the vector of messages  $x_1, \dots, x_{t-1}$ . Equations 3.10 and 3.11 are known as the *Kalman filter algorithm* after Kalman (1960) and Kalman and Bucy (1961), who show that Equation 3.10 yields the expected value of  $\alpha_t$  under the assumption of normal errors. If the normality assumption is relaxed, the Kalman filter estimator of  $\alpha_t$  remains best (i.e., least-squares-minimizing) among all linear estimators. Harvey (1989) and Beck (1990) contain extensive descriptions of the Kalman filter and its properties. Goussev (2004) argues from a neurological perspective that the human brain unconsciously uses it to update probabilities. Meinhold and Singpurwalla (1983) provide a lucid introduction to it from a Bayesian point of view. Gerber and Green (1998) note that the normal-normal model is a special

case of the Kalman filter model in which  $\gamma = 1$  and  $\sigma_\alpha^2 = 0$ . And given that the Kalman filter model is a Bayesian updating model, one of its surprising implications is that people's beliefs may diverge even if they are updating in response to the same messages and interpreting those messages in the same way.

**Proposition 3A: Polarization Can Occur under Kalman Filtering Even in the Absence of Selective Perception and Selective Exposure.** Assume that  $\alpha_t$ ,  $\sigma_\alpha^2$ ,  $\gamma$ ,  $x_t$ , and  $\sigma_x^2$  are as described above. Let  $K_t = \frac{1/\sigma_x^2}{1/(\gamma^2 P_{t-1} + \sigma_\alpha^2) + 1/\sigma_x^2}$ . Voter  $i$ 's belief about  $\alpha_0$  is  $N(\hat{\alpha}_{i0}, P_0)$ . Voter  $j$ 's belief about  $\alpha_0$  is  $N(\hat{\alpha}_{j0}, P_0)$ . They are exposed to one message at each stage  $t$ ; the entire set of messages is  $\mathbf{x}_t$ . They update their beliefs according to Equations 3.10 and 3.11. If  $\hat{\alpha}_{i0} \neq \hat{\alpha}_{j0}$ , their beliefs diverge from time  $t$  to time  $t + 1$  if and only if  $(1 - K_{t+1})|\gamma| > 1$ .

**Proof.** We begin with a lemma: the Kalman filter estimator  $\hat{\alpha}_t$  can be written as a linear function of  $\hat{\alpha}_0$ ,

$$\hat{\alpha}_t = c_t \hat{\alpha}_0 + \mathbf{f}_t' \mathbf{x}_t, \quad (3.12)$$

where  $c_t = \prod_{i=1}^t (1 - K_i)\gamma$ ,  $\mathbf{f}_t'$  is a row vector, and  $\mathbf{x}_t$  is the column vector of messages  $x_1, \dots, x_t$ . (See the appendix for a proof.)

Proof by contradiction: assume some  $t$  such that beliefs at  $t + 1$  are not more polarized than beliefs at  $t$  even though  $(1 - K_{t+1})|\gamma| > 1$ . Note that  $0 < 1 - K_t < 1 \forall t$ . Then

$$\begin{aligned} & |\hat{\alpha}_{it} - \hat{\alpha}_{jt}| - |\hat{\alpha}_{i,t+1} - \hat{\alpha}_{j,t+1}| \geq 0 \\ \Rightarrow & |c_t \hat{\alpha}_{i0} + \mathbf{f}'_t \mathbf{x}_t - (c_t \hat{\alpha}_{j0} + \mathbf{f}'_t \mathbf{x}_t)| - |(c_{t+1} \hat{\alpha}_{i0} + \mathbf{f}'_{t+1} \mathbf{x}_{t+1}) - (c_{t+1} \hat{\alpha}_{j0} + \mathbf{f}'_{t+1} \mathbf{x}_{t+1})| \geq 0 \\ \Rightarrow & |c_t| - |c_{t+1}| > 0 \\ \Rightarrow & \prod_{i=1}^t (1 - K_i) |\gamma| - (1 - K_{t+1}) |\gamma| \prod_{i=1}^t (1 - K_i) |\gamma| \geq 0 \\ \Rightarrow & (1 - K_{t+1}) |\gamma| \leq 1, \end{aligned}$$

which is a contradiction. This establishes that divergence occurs between times  $t$  and  $t + 1$  if  $(1 - K_{t+1})|\gamma| > 1$ . Now assume some  $t$  such that beliefs at  $t + 1$  are less alike than beliefs at  $t$  even though  $(1 - K_{t+1})|\gamma| \leq 1$ . Then

$$\begin{aligned} & |\hat{\alpha}_{it} - \hat{\alpha}_{jt}| - |\hat{\alpha}_{i,t+1} \hat{\alpha}_{j,t+1}| < 0 \\ \Rightarrow & \prod_{i=1}^t (1 - K_i) |\gamma| - (1 - K_{t+1}) |\gamma| \prod_{i=1}^t (1 - K_i) |\gamma| < 0 \\ \Rightarrow & (1 - K_{t+1}) |\gamma| > 1, \end{aligned}$$

which is a contradiction. This establishes that divergence occurs only if

$$(1 - K_{t+1}) |\gamma| > 1. \quad \square$$

**Proposition 3B: Convergence to Agreement Occurs Eventually under the Kalman Filter.** Assume the conditions of Proposition 3A. At some period  $t$ ,  $K_t$  will reach a steady state. After that point, polarization will not be possible.

**Proof.** We begin with a lemma:  $K_t$  will gradually converge to

$$K = \frac{\gamma^2 - c - 1 + \sqrt{(-\gamma^2 + c + 1)^2 + 4c\gamma^2}}{2\gamma^2},$$

where  $c = \sigma_a^2/\sigma_x^2$  (Gerber and Green 1998; see the appendix for a proof). By Proposition 3A, divergence occurs if and only if  $(1 - K_{t+1})|\gamma| > 1$ , but this is not possible when  $K_{t+1} = K$ .

**Proof by contradiction.**  $(1 - K_{t+1})|\gamma| > 1$  implies  $|\gamma| > 1$ , because  $0 < K_t < 1 \forall t$ .

If  $\gamma > 1$ , divergence in the steady state implies

$$\begin{aligned} & \left(1 - \frac{\gamma^2 - c - 1 + \sqrt{(-\gamma^2 + c + 1)^2 + 4c\gamma^2}}{2\gamma^2}\right)\gamma > 1 \\ \Rightarrow \gamma - \frac{\gamma^2 - c - 1 + \sqrt{(-\gamma^2 + c + 1)^2 + 4c\gamma^2}}{2\gamma} & > 1 \\ \Rightarrow -(\gamma^2 - c - 1 + \sqrt{(-\gamma^2 + c + 1)^2 + 4c\gamma^2}) & > (1 - \gamma)2\gamma \\ \Rightarrow c + 1 - \sqrt{(-\gamma^2 + c + 1)^2 + 4c\gamma^2} & > 2\gamma - \gamma^2 \\ \Rightarrow c + 1 - \sqrt{\gamma^4 + 2c\gamma^2 - 2\gamma^2 + c^2 + 2c + 1} & > 2\gamma - \gamma^2 \\ \Rightarrow \gamma^2 - 2\gamma + c + 1 & > \sqrt{\gamma^4 + 2c\gamma^2 - 2\gamma^2 + c^2 + 2c + 1} \\ \Rightarrow \gamma^4 - 4\gamma^3 + 2c\gamma^2 + 6\gamma^2 - 4c\gamma - 4\gamma + c^2 + 2c + 1 & > \gamma^4 + 2c\gamma^2 - 2\gamma^2 + c^2 + 2c + 1 \\ \Rightarrow -4\gamma^3 + 6\gamma^2 - 4c\gamma - 4\gamma & > -2\gamma^2 \\ \Rightarrow -4\gamma^3 + 8\gamma^2 - 4\gamma c - 4\gamma & > 0. \end{aligned}$$

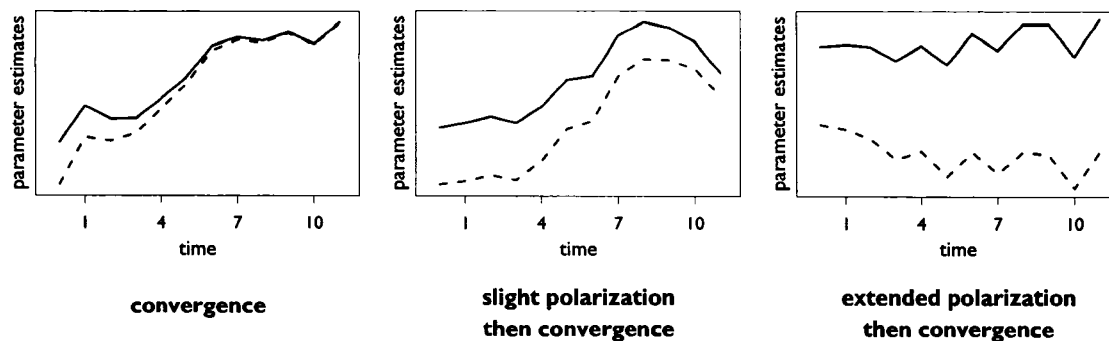
This implies  $c < -\gamma^2 + 2\gamma - 1$ . But  $-\gamma^2 + 2\gamma - 1 \leq 0 \forall \gamma$ , and  $c$  must be positive because it is a ratio of positive variances. Contradiction.

If  $\gamma < -1$ ,

$$\begin{aligned}
 & \left( 1 - \frac{\gamma^2 - c - 1 + \sqrt{(-\gamma^2 + c + 1)^2 + 4c\gamma^2}}{2\gamma^2} \right) |\gamma| > 1 \\
 \Rightarrow & |\gamma| - \frac{\gamma^2 - c - 1 + \sqrt{(-\gamma^2 + c + 1)^2 + 4c\gamma^2}}{2|\gamma|} > 1 \\
 \Rightarrow & - \left( \gamma^2 - c - 1 + \sqrt{(-\gamma^2 + c + 1)^2 + 4c\gamma^2} \right) > (1 - |\gamma|)2|\gamma| \\
 \Rightarrow & - \left( \gamma^2 - c - 1 + \sqrt{(-\gamma^2 + c + 1)^2 + 4c\gamma^2} \right) > 2|\gamma| - 2\gamma^2 \\
 \Rightarrow & c + 1 - \sqrt{(-\gamma^2 + c + 1)^2 + 4c\gamma^2} > 2|\gamma| - \gamma^2 \\
 \Rightarrow & \gamma^2 - 2|\gamma| + c + 1 > \sqrt{\gamma^4 + 2c\gamma^2 - 2\gamma^2 + c^2 + 2c + 1} \\
 \Rightarrow & \gamma^4 - 4|\gamma|^3 + 6\gamma^2 + 2\gamma^2 c - 4|\gamma| - 4|\gamma|c + c^2 + 2c + 1 > \gamma^4 + 2c\gamma^2 - 2\gamma^2 + c^2 + 2c + 1 \\
 \Rightarrow & -4|\gamma|^3 + 8\gamma^2 - 4|\gamma|c - 4|\gamma| > 0.
 \end{aligned}$$

This implies  $c < -\gamma^2 - 2\gamma - 1$ . But  $-\gamma^2 - 2\gamma - 1 \leq 0 \forall \gamma$ , and  $c$  must be positive because it is a ratio of variances. Contradiction.  $\square$

**Discussion.** The proof of Proposition 3A shows that polarization can happen even in the absence of selective exposure or perception: it occurs when updating causes people who disagree to weight their prior beliefs more heavily than they did before. And this occurs when  $(1 - K_{t+1})|\gamma| > 1$ . Because  $K_t$  lies between 0 and 1 for all values of  $t$ ,  $(1 - K_{t+1})$  must also lie between 0 and 1, and  $|\gamma| > 1$  is therefore a necessary condition for polarization. It may seem, moreover, that bigger values of  $|\gamma|$  produce more polarization. But this is not generally so, because  $\gamma$  enters into the definition of  $K_t$ , too, and the algorithm that determines  $K_t$  quickly “catches up” to offset the size of  $|\gamma|$  in  $(1 - K_t)|\gamma|$ . Proposition 3B shows that the algorithm catches up completely when  $K_t$  reaches its steady state  $K$ : it is impossible for  $(1 - K)|\gamma|$  to be greater than 1.



**Figure 3.4: Convergence and Polarization under Kalman Filter Updating.** All panels depict simulated belief updating by two hypothetical voters when  $\gamma = 1.1$ ,  $\alpha_0 = 1$ , the voter represented by the solid line has prior belief  $N(1, 1)$ , and the voter represented by the dashed line has prior belief  $N(-1, 1)$ . The scale of the y axes differs across panels to make differences between the voters apparent.

In the leftmost panel,  $\sigma_\alpha^2 = .7$  and  $\sigma_x^2 = 4$ . We see the intuitive pattern of belief updating under Bayes' Theorem: subjects who receive the same messages and interpret them in the same way draw closer to agreement every time they update. The convergence is represented by the vertical distance between the solid and dashed line, which diminishes in each period.

The middle panel depicts updating when  $\sigma_\alpha^2 = .5$  and  $\sigma_x^2 = 25$ . Even though voters are receiving the same messages and interpreting them in the same way, their beliefs diverge (slightly) from period 0 to 1 and again from period 1 to period 2. Not until period 4 is the distance between their beliefs smaller than the distance between their priors. This pattern is exacerbated in the rightmost panel, where  $\sigma_\alpha^2 = .19$  and  $\sigma_x^2 = 100$ . Here, voters' views diverge continuously in each of the first eleven periods; not until the 22nd period (unshown) is the distance between their beliefs smaller than the distance between their priors.

In practice, polarization is sustained longest when  $|\gamma|$  is barely greater than 1 and when  $\sigma_x^2$  is far smaller than  $\sigma_\alpha^2$ . (See Figure 3.4.) The former condition occurs in politics whenever a parameter of interest is trending slowly away from zero.<sup>7</sup> Almost all political parameters of interest trend away from zero sometimes, and some (e.g., population, racial tolerance, per capita income in developed countries) seem to trend slowly away from zero for very long periods of time. The second condition,  $\sigma_x^2 \ll \sigma_\alpha^2$ , exists whenever the true variation in a parameter is slight but the quality of the

<sup>7</sup> For parameters that have no natural scale, zero is arbitrarily defined. For example, in the case of the President's honesty, zero might correspond to a survey answer of "neither honest nor dishonest."

messages that we receive about it is poor. This is likely when secretive or totalitarian states exert control over the press. Schumpeter (1942) argues that it is also a condition endemic to democratic politics: although there are knowable political truths, the feedback that we receive about our political decisions is so poor as to verge on useless. Presume, for example, that we want to know whether the President is a good steward of the economy. How are we to know? If we think that we have observed improvement in the economy, it might be attributed to the President's policies, or the lagged effects of his predecessor's policies, or the actions of incumbents at other levels of government, or the lagged effects of their predecessors, or the Federal Reserve, or wholly apolitical factors. It is rarely easy to tell who is responsible. And this elides the difficulty of simply knowing when the economy has improved. Even a free and robust media will not be of much help in these cases: no matter how precise the signals they send about the state of the economy, signals about who deserves credit and blame are unavoidably vague.<sup>8</sup>

Bartels (2002, 123) tells us that "it is *failure* to converge that requires explanation within the Bayesian framework," and he fingers selective perception by Democrats and Republicans as the culprit in their failure to agree on a host of actual matters. He may be right, but the case is far from cinched, and his characterization of Bayesian updating is too strong. Bayesian updating models imply only partial convergence of only some beliefs, and only under fantastically rare conditions do they imply that people who disagree will "inexorably come to see the world in the same way." The logic of Bayesian updating alone provides no reasonable expectation of

---

<sup>8</sup> Indeed,  $\sigma_x^2 \ll \sigma_a^2$  in the case of attributions of praise and blame is the root of Schumpeter's (1942, 262) contention that "the typical citizen drops down to a lower level of mental performance as soon as he enters the political field." The problem is not that political man is stupid but that causality in politics is so complex and feedback about political decisions is so poor. Man's "lower level of mental performance" is due to his failure to sufficiently adjust for the poor quality of information at his disposal.

convergence—not during a campaign, not even over a lifetime. Of course, agreement may occur. But enduring disagreement is no proof that people are not Bayesian. Still less is it proof of partisan bias.

## **Discussion: Bayes' Theorem as a Normative Standard of Belief Updating**

As a way of describing how citizens actually update their beliefs, Bayesian updating models have come in for a wealth of criticism. Much laboratory research shows that people do not update as Bayesians (Phillips and Edwards 1966; Tversky and Kahneman 1971; though see Koehler 1996). And some political scientists believe that the Theorem cannot accommodate ordinary features of public opinion about politics (Taber and Lodge 2006; Fischle 2000). This essay shows, to the contrary, that important features of public opinion about politics are readily accommodated by Bayesian updating. Political events are often surprising, and Bayesian updating can reflect that surprise by causing people to hold their views less confidently. Partisan bias affects people's views through selective perception or misjudgments of the credibility of media outlets, and Bayesian updating can easily accommodate this. Most importantly, people who disagree about politics are rarely moved to agreement by even a flood of evidence. Exposure to the evidence may even cause their views to draw further apart. Contrary to what some have written, Bayes' Theorem can easily accommodate enduring disagreement and polarization, too. None of this means that Bayesian updating models perfectly capture every facet of political decision-making—no models do—but it does mean that they are better than many suppose.



But it is precisely this flexibility that makes Bayesian updating inadequate as a standard of rational thinking about politics. One need only consider all that it permits. There is no limit to the amount of evidence that a Bayesian may ignore or misconstrue, for Bayes' Theorem applies only to updating, not to the collection or interpretation of evidence. And even in the absence of perceptual biases—even if people are interpreting and using new information in exactly the same way—Bayes' Theorem permits their views to polarize. No standard of rational belief revision should be so permissive.

The problem is not that the Theorem is irrational; as we have seen, it entails abiding by axioms of probability so fundamental that they should be a component of any standard of rational thinking. The problem is that the Theorem is not restrictive enough: it permits what we should reject as irrational. Political scientists interested in constructing standards of rational political thinking will do well to couple it with restrictions on the interpretation of political evidence. Bayes' Theorem should be part of any normative criteria by which we judge rational political thinking—but only a part.

## Appendix H: Proofs of Several Bayesian Updating Results

### Updating in Response to Many Normal Messages is Equivalent to Updating in Response to a Single, More Precise Normal Message

Updating sequentially in response to many messages is equivalent to updating once in response to a single, more precise message. I show this result for normal-normal updating—first for the case in which all the messages are drawn from the same distribution, then for the case in which different messages are drawn from distributions with different variances.

Formally, assume (as in the normal-normal model) that we are trying to learn the mean of a distribution whose variance is known. Let the prior belief about the mean be  $\mu \sim N(\mu_0, \sigma_0^2)$ . A random sample of messages  $\mathbf{x} = (x_1, \dots, x_n)$  is drawn independently from a distribution that is presumed to be  $N(\mu, \sigma_x^2)$ . Because Bayes' Theorem requires that the posterior be proportional to the prior times the likelihood of the new messages, updating sequentially requires that

$$\begin{aligned}
 p(\mu|\mathbf{x}) &\propto \text{prior} \times \mathcal{L}(x_1|\mu) \times \dots \times \mathcal{L}(x_n|\mu) \\
 &= \text{prior} \times \exp\left[-\frac{(x_1 - \mu)^2}{2\sigma_x^2}\right] \times \dots \times \exp\left[-\frac{(x_n - \mu)^2}{2\sigma_x^2}\right] \\
 &= \text{prior} \times \exp\left[-\frac{n}{2\sigma_x^2}\left(\mu^2 - 2\mu\bar{x} + \frac{x_1^2 + \dots + x_n^2}{n}\right)\right] \\
 &= \text{prior} \times \exp\left[-\frac{n}{2\sigma_x^2}\left(\mu^2 - 2\mu\bar{x} + \bar{x}^2\right)\right] \times \exp\left[-\frac{n}{2\sigma_x^2}\left(\frac{x_1^2 + \dots + x_n^2}{n} - \bar{x}^2\right)\right].
 \end{aligned}$$

The part that does not depend on  $\mu$  is constant. Absorbing it into the proportionality constant, we get

$$p(\mu|\mathbf{x}) \propto \text{prior} \times \exp\left[-\frac{1}{2}\left(\frac{(\bar{x} - \mu)^2}{\sigma_x^2/n}\right)\right].$$

The rightmost term is the likelihood of a single observation from a normal distribution with mean  $\mu$  and variance  $\sigma_x^2/n$ .

The case of updating with messages from distributions with different variances is very similar. Suppose that messages  $\mathbf{x}_1 = (x_1, \dots, x_n)$  have mean  $\bar{x}_1$  and are presumed to be drawn independently from the  $N(\mu, \sigma_{x_1}^2)$  distribution, while messages  $\mathbf{x}_2 = (x_{n+1}, \dots, x_N)$  have mean  $\bar{x}_2$  and are presumed to be drawn independently from the  $N(\mu, \sigma_{x_2}^2)$  distribution. By the previous result, this is equivalent to updating on the basis of just two messages:

$$p(\mu|\mathbf{x}_1, \mathbf{x}_2) \propto \text{prior} \times \exp\left[-\frac{1}{2}\left(\frac{(\bar{x}_1 - \mu)^2}{\sigma_{x_1}^2/n}\right)\right] \times \exp\left[-\frac{1}{2}\left(\frac{(\bar{x}_2 - \mu)^2}{\sigma_{x_2}^2/(N-n)}\right)\right].$$

Let  $\sigma_{1*}^2 = \sigma_{x_1}^2/n$  and  $\sigma_{2*}^2 = \sigma_{x_2}^2/(N-n)$ . Then

$$\begin{aligned} p(\mu|\mathbf{x}_1, \mathbf{x}_2) &\propto \text{prior} \times \exp\left[-\frac{1}{2}\left(\frac{(\bar{x}_1 - \mu)^2}{\sigma_{1*}^2} + \frac{(\bar{x}_2 - \mu)^2}{\sigma_{2*}^2}\right)\right] \\ &= \text{prior} \times \exp\left[-\frac{1}{2}\left(\frac{(\bar{x}_1 - \mu)^2 \sigma_{2*}^2 + (\bar{x}_2 - \mu)^2 \sigma_{1*}^2}{\sigma_{1*}^2 \sigma_{2*}^2}\right)\right] \\ &= \text{prior} \times \exp\left[-\frac{1}{2}\left(\frac{(\bar{x}_1^2 - 2\mu\bar{x}_1 + \mu^2) \sigma_{2*}^2 + (\bar{x}_2^2 - 2\mu\bar{x}_2 + \mu^2) \sigma_{1*}^2}{\sigma_{1*}^2 \sigma_{2*}^2}\right)\right] \\ &= \text{prior} \times \exp\left[-\frac{1}{2}\left(\frac{\mu^2(\sigma_{1*}^2 + \sigma_{2*}^2) - 2\mu(\bar{x}_1\sigma_{2*}^2 + \bar{x}_2\sigma_{1*}^2) + \bar{x}_1^2\sigma_{2*}^2 + \bar{x}_2^2\sigma_{1*}^2}{\sigma_{1*}^2 \sigma_{2*}^2}\right)\right] \\ &= \text{prior} \times \exp\left[-\frac{1}{2}\left(\mu^2 - \frac{2\mu(\bar{x}_1\sigma_{2*}^2 + \bar{x}_2\sigma_{1*}^2)}{\sigma_{1*}^2 + \sigma_{2*}^2} + \frac{\bar{x}_1^2\sigma_{2*}^2 + \bar{x}_2^2\sigma_{1*}^2}{\sigma_{1*}^2 + \sigma_{2*}^2}\right)\left(\frac{\sigma_{1*}^2 + \sigma_{2*}^2}{\sigma_{1*}^2 \sigma_{2*}^2}\right)\right]. \end{aligned}$$

Absorbing the part that does not depend on  $\mu$  into the constant of proportionality, we get

$$p(\mu|\mathbf{x}_1, \mathbf{x}_2) \propto \text{prior} \times \exp \left[ -\frac{1}{2(\sigma_{1*}^2 \sigma_{2*}^2)/(\sigma_{1*}^2 + \sigma_{2*}^2)} \left( \mu - \frac{\bar{x}_1 \sigma_{2*}^2 + \bar{x}_2 \sigma_{1*}^2}{\sigma_{1*}^2 + \sigma_{2*}^2} \right)^2 \right].$$

The rightmost term is the kernel of a normal density with mean  $\frac{\bar{x}_1 \sigma_{2*}^2 + \bar{x}_2 \sigma_{1*}^2}{\sigma_{1*}^2 + \sigma_{2*}^2}$  and variance  $\frac{\sigma_{1*}^2 \sigma_{2*}^2}{\sigma_{1*}^2 + \sigma_{2*}^2}$ .

### **Under Kalman Filtering, the Posterior Estimate of the Mean is a Linear Function of the Prior Belief and Messages Received**

Equation 3.3a shows that when a normally distributed prior belief is updated in response to information that is perceived to be normal, the Bayes estimate of the posterior mean is a linear combination of the prior estimate of the mean and the new information. Diaconis and Ylvisaker (1979) show that this is not peculiar to updating with normal priors and likelihoods: whenever the prior belief has a distribution in the exponential family and is conjugate to the distribution of the new information, the Bayes estimate of the posterior mean is a convex combination of the prior estimate of the mean and the new information. In this appendix, I calculate the weights of the linear combination for the Kalman filter estimator  $\hat{\alpha}_t$  of random variable  $\alpha_t$ , which are defined on pages 111-113.

The Kalman filter estimator  $\hat{\alpha}_t$  can be written as a linear function of  $\hat{\alpha}_0$ ,  $\hat{\alpha}_t = c_t \hat{\alpha}_0 + \mathbf{f}'_t \mathbf{x}_t$ , where  $c_t = \prod_{i=1}^t (1 - K_i) \gamma$ ,  $K_t$  is the “Kalman gain” defined on page 114,  $\mathbf{x}_t$  is the vector of messages  $(x_1, \dots, x_t)$ , and  $\mathbf{f}'_t$  is a row vector of weights on the components of  $\mathbf{x}_t$ .

Proof by induction. The claim is true for  $t = 1$ :

$$\begin{aligned}\hat{\alpha}_1 &= \gamma\hat{\alpha}_0 + K_1(x_1 - \gamma\hat{\alpha}_0) \\ &= (\gamma - K_1\gamma)\hat{\alpha}_0 + K_1x_1 \\ &= c_1\hat{\alpha}_0 + \mathbf{f}'_1\mathbf{x}_1,\end{aligned}$$

with  $c_1 = (\gamma - K_1\gamma) = \prod_{i=1}^1(1 - K_i)\gamma$  and  $\mathbf{f}_1 = K_1$ . Assuming the claim is true for  $t$ , it is also true for  $t + 1$ :

$$\begin{aligned}\hat{\alpha}_{t+1} &= \gamma\hat{\alpha}_t + K_{t+1}(x_{t+1} - \gamma\hat{\alpha}_t) \\ &= (1 - K_{t+1})\gamma\hat{\alpha}_t + K_{t+1}x_{t+1} \\ &= (1 - K_{t+1})\gamma [c_t\hat{\alpha}_0 + \mathbf{f}'_t\mathbf{x}_t] + K_{t+1}x_{t+1} \\ &= [(1 - K_{t+1})\gamma c_t] \hat{\alpha}_0 + (1 - K_{t+1})\gamma\mathbf{f}'_t\mathbf{x}_t + K_{t+1}x_{t+1} \\ &= c_{t+1}\hat{\alpha}_0 + \mathbf{f}'_{t+1}\mathbf{x}_{t+1},\end{aligned}$$

with  $c_{t+1} = (1 - K_{t+1})\gamma c_t$  and  $\mathbf{f}_{t+1}$  a vector in which the first  $t$  elements are given by  $(1 - K_{t+1})\gamma\mathbf{f}_t$  and in which the  $(N + 1)$ th element is  $K_{t+1}$ .  $\square$

This proof follows the similar proof in Gerber and Green (1998). They are not identical because (perhaps due to a copyediting error) the earlier works reports  $c_t = \prod_{i=1}^t(1 - K_i)\gamma^t$ , which is not quite right.

## K Has a Steady State

According to Gerber and Green (1998), the Kalman weight  $K_t$  in Kalman-filter updating stabilizes at a unique value:

$$K = \frac{-[c + (1 - \gamma^2)] + \sqrt{[c + (1 - \gamma^2)]^2 + 4c\gamma^2}}{2\gamma^2}$$

where  $c = \sigma_\alpha^2 / \sigma_x^2$ .

**Proof.**  $K_t = P_t / \sigma_x^2$ , so behavior of  $P_t$  implies behavior of  $K_t$ . To find the steady state of  $P_t$  (and thus of  $K_t$ ), we need to find the value for which  $P_t = P_{t+1}$ :

$$\begin{aligned} P_t &= \sigma_x^2 K_{t+1} \\ &= \sigma_x^2 (\gamma^2 P_t + \sigma_\alpha^2) / (\gamma^2 P_t + \sigma_\alpha^2 + \sigma_x^2) \\ P_t (\gamma^2 P_t + \sigma_\alpha^2 + \sigma_x^2) &= P_t \gamma^2 \sigma_x^2 + \sigma_x^2 \sigma_\alpha^2 \\ P_t^2 \gamma^2 + P_t (\sigma_\alpha^2 + \sigma_x^2 - \gamma^2 \sigma_x^2) - \sigma_x^2 \sigma_\alpha^2 &= 0. \end{aligned}$$

This is just a quadratic equation, so

$$P = \frac{-(\sigma_\alpha^2 + \sigma_x^2 - \gamma^2 \sigma_x^2) \pm \sqrt{(\sigma_\alpha^2 + \sigma_x^2 - \gamma^2 \sigma_x^2)^2 - 4\gamma^2(-\sigma_x^2 \sigma_\alpha^2)}}{2\gamma^2}.$$

The denominator is positive. The numerator must also be positive, then, because  $P$  must be positive.

$\sigma_\alpha^2$ ,  $\sigma_x^2$ , and  $\gamma^2$  must all be positive. The numerator will thus only be positive if it is  $-(\sigma_\alpha^2 + \sigma_x^2 - \gamma^2\sigma_x^2) + \sqrt{(\sigma_\alpha^2 + \sigma_x^2 - \gamma^2\sigma_x^2)^2 - 4\gamma^2(-\sigma_x^2\sigma_\alpha^2)}$ . In other words, we replace  $\pm$  with  $+$ :

$$\begin{aligned}
 P &= \frac{-(\sigma_\alpha^2 + \sigma_x^2 - \gamma^2\sigma_x^2) + \sqrt{(\sigma_\alpha^2 + \sigma_x^2 - \gamma^2\sigma_x^2)^2 - 4\gamma^2(-\sigma_x^2\sigma_\alpha^2)}}{2\gamma^2} \\
 K &= \frac{-(\sigma_\alpha^2 + \sigma_x^2(1 - \gamma^2)) + \sqrt{(\sigma_\alpha^2 + \sigma_x^2 - \gamma^2\sigma_x^2)^2 + 4\gamma^2\sigma_x^2\sigma_\alpha^2}}{\sigma_x^2 2\gamma^2} \\
 &= \frac{-(c + (1 - \gamma^2))}{2\gamma^2} + \frac{\sqrt{(\sigma_\alpha^2 + \sigma_x^2 - \gamma^2\sigma_x^2)^2 + 4\gamma^2\sigma_x^2\sigma_\alpha^2}}{\sigma_x^2 2\gamma^2} \\
 &= \frac{\gamma^2 - c - 1}{2\gamma^2} + \frac{\sqrt{\frac{(\sigma_\alpha^2 + \sigma_x^2 - \gamma^2\sigma_x^2)^2}{(\sigma_x^2)^2} + 4c\gamma^2}}{2\gamma^2} \\
 &= \frac{\gamma^2 - c - 1}{2\gamma^2} + \frac{\sqrt{\frac{(\sigma_\alpha^2)^2 + (\sigma_x^2)^2 + \gamma^4(\sigma_x^2)^2 + 2\sigma_\alpha^2\sigma_x^2 - 2\sigma_\alpha^2\gamma^2\sigma_x^2 - 2\gamma^2(\sigma_x^2)^2}{(\sigma_x^2)^2} + 4c\gamma^2}}{2\gamma^2} \\
 &= \frac{\gamma^2 - c - 1 + \sqrt{(c - \gamma^2 + 1)^2 + 4c\gamma^2}}{2\gamma^2}.
 \end{aligned}$$

That is, after some stage  $t$ , the newest observation always receives the same weight ( $K$ ) when updating. By extension, one's prior always receives the same weight ( $1 - K$ ) when updating.

This result is contrary to the normal-normal model, in which information received at previous stages is always reflected in the prior. As time passes, the weight placed on the prior under the normal-normal model always increases, because the prior always reflects more information than it did in the past. Consequently, the weight placed on new messages always decreases.

The Kalman filter is more realistic because it implies that very old information is either forgotten or discounted: when the parameter of interest is changing over time, a very old message is not as important as a new message, even if both come from

equally credible sources. That is why the weight placed on the prior when updating under the Kalman filter model does not always increase over time. See Gerber and Green (1998) for an extensive discussion of this difference between the normal-normal and Kalman-filter models.



# Bibliography

- Abelson, Robert P. 1986. "Beliefs Are Like Possessions." *Journal for the Theory of Social Behaviour* 16 (October): 224-50.
- 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. 2002. "Parental Socialization and Rational Party Identification." *Political Behavior* 24 (June): 151-70.
- Achen, Christopher H. 2003. "Toward a New Political Methodology: Microfoundations and ART." *Annual Review of Political Science* 5 (June): 423-50.
- 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.
- American Presidency Project. 2006 03 01. "Political Party Platforms."  
<http://www.presidency.ucsb.edu/platforms.php>.
- Anderson, Craig A., B. Lynn New, and James R. Speer. 1985. "Argument Availability as a Mediator of Social Theory Perseverance." *Social Cognition* 3 (3): 235-49.

- Anderson, Craig A., Mark R. Lepper, and Lee Ross. 1980. "The Perseverance of Social Theories: The Role of Explanation in the Persistence of Discredited Information." *Journal of Personality and Social Psychology* 39 (December): 1037-49.
- Ansolabehere, Stephen, James M. Snyder, Jr., and Charles Stewart, III. 2001. "Candidate Positioning in U.S. House Elections." *American Journal of Political Science* 45 (January): 136-59.
- Ansolabehere, Stephen, and Shanto Iyengar. 1995. *Going Negative: How Political Advertisements Shrink and Polarize the Electorate*. New York: Free Press.
- Ansolabehere, Stephen, Shigeo Hirano, James M. Snyder, Jr., and Michiko Ueda. 2006. "Party and Incumbency Cues in Voting: Are They Substitutes?" *Quarterly Journal of Political Science* 1 (March): 119-37.
- Barge, Matthew. 9 August 2005. "NARAL Falsely Accuses Supreme Court Nominee Roberts." <http://www.factcheck.org/article340.html>.
- Bargh, John A., Shelly Chaiken, Rajen Govender, and Felicia Pratto. 1992. "The Generality of the Automatic Attitude Activation Effect." *Journal of Personality and Social Psychology* 62 (June): 893-912.
- 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.
- Bernardo, José. M., and A. F. M. Smith. 1994. *Bayesian Theory*. New York: Wiley.

- Bochner, Stephen, and Chester A. Insko. 1966. "Communicator Discrepancy, Source Credibility, and Opinion Change." *Journal of Personality and Social Psychology* 4 (December): 614-21.
- Bound, John, David A. Jaeger, and Regina M. Baker. 1995. "Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable Is Weak." *Journal of the American Statistical Association* 90 (June): 443-50.
- Box, George E. P., and George C. Tiao. 1973. *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison-Wesley.
- Brader, Ted. 2006. *Campaigning for Hearts and Minds: How Emotional Appeals in Political Ads Work*. Chicago: University of Chicago Press.
- Bullock, John G. 2006. "Partisanship and the Enduring Effects of False Political Information." Presented at the Annual Conference of the Midwest Political Science Association, Chicago.
- Callander, Steven, and Simon Wilkie. 2007. "Lies, Damned Lies, and Political Campaigns." *Games and Economic Behavior*.
- Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. 1960. *The American Voter*. Chicago: University of Chicago Press.
- Chaiken, Shelly. 1979. "Communicator Physical Attractiveness and Persuasion." *Journal of Personality and Social Psychology* 37 (August): 1387-97.
- Chapman, Loren J., and Jean P. Chapman. 1959. "Atmosphere Effect Re-examined." *Journal of Experimental Psychology* 58 (September): 220-26.

- Chong, Dennis, and James N. Druckman. 2006. "Democratic Competition and Public Opinion." Presented at the Annual Meeting of the American Political Science Association, Philadelphia.
- Cohen, Geoffrey L. 2003. "Party Over Policy: The Dominating Impact of Group Influence on Political Beliefs." *Journal of Personality and Social Psychology* 85 (November): 808-22.
- Cohen, Joshua. 1989. "Deliberation and Democratic Legitimacy." In *The Good Polity*, ed. Alan Hamlin and Phillip Pettit. Oxford: Blackwell.
- Conover, Pamela Johnston, and Stanley Feldman. 1989. "Candidate Perception in an Ambiguous World: Campaigns, Cues, and Inference Processes." *American Journal of Political Science* 33 (November): 912-40.
- Davis, Otto A., Melvin J. Hinich, and Peter C. Ordeshook. 1970. "An Expository Development of a Mathematical Model of the Electoral Process." *American Political Science Review* 64 (June): 426-48.
- de Finetti, Bruno. 1974. *Theory of Probability: A Critical Introductory Treatment*. Trans. Antonio Machi and Adrian F. Smith. New York: Wiley.
- Delli Carpini, Michael X., and Scott Keeter. 1996. *What Americans Know about Politics and Why It Matters*. New Haven, CT: Yale University Press.
- Diaconis, Persi, and David Freedman. 1986. "On the Consistency of Bayes Estimates." *Annals of Statistics* 14 (March): 1-26.
- Diaconis, Persi, and Donald Ylvisaker. 1979. "Conjugate Priors for Exponential Families." *Annals of Statistics* 7 (March): 269-81.

- Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: HarperCollins.
- Druckman, James N. 2001. "The Implications of Framing Effects for Citizen Competence." *Political Behavior* 23 (September): 225-56.
- Druckman, James N. 2004. "Political Preference Formation: Competition, Deliberation, and the (Ir)relevance of Framing Effects." *American Political Science Review* 98 (November): 671-86.
- Druckman, James N., and Kjersten R. Nelson. 2003. "Framing and Deliberation: How Citizens' Conversations Limit Elite Influence." *American Journal of Political Science* 47 (October): 729-45.
- Duelfer, Charles. 2004. "Comprehensive Report of the Special Advisor to the Director of Central Intelligence on Iraq's Weapons of Mass Destruction." [https://www.cia.gov/cia/reports/iraq\\_wmd\\_2004/](https://www.cia.gov/cia/reports/iraq_wmd_2004/).
- Epley, Nicholas, and David Dunning. 2000. "Feeling 'Holier than Thou': Are Self-Serving Assessments Produced by Errors in Self- or Social Prediction?" *Journal of Personality and Social Psychology* 79 (December): 861-75.
- Fischle, Mark. 2000. "Mass Response to the Lewinsky Scandal: Motivated Reasoning or Bayesian Updating?" *Political Psychology* 21 (March): 135-59.
- Freedman, David A. 1991. "Statistical Models and Shoe Leather." *Sociological Methodology* 21: 291-313.
- Frey, Dieter. 1986. "Recent Research on Selective Exposure to Information." *Advances in Experimental Social Psychology* 19: 41-80.

## 134 Bibliography

- 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*.
- Gellman, Barton. 2004 January 7. "Iraq's Arsenal Was Only on Paper; Since Gulf War, Nonconventional Weapons Never Got Past the Planning Stage." *Washington Post* p. A1.
- Gellman, Barton, and Walter Pincus. 2003 August 10. "Depiction of Threat Outgrew Supporting Evidence." *Washington Post* p. A1.
- Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 2004. *Bayesian Data Analysis*. 2nd ed. Boca Raton, FL: Chapman & Hall.
- Gentzkow, Matthew, and Jesse M. Shapiro. 2006. "What Drives Media Slant? Evidence from U.S. Daily Newspapers." University of Chicago. Manuscript.
- 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.
- Gertz v. Schmidt*. 1974. 418 U.S. 323.
- Goussev, Valeri. 2004. "Does the Brain Implement the Kalman Filter?" *Behavioral and Brain Sciences* 27 (June): 404-05.

- Green, Donald P., Alan S. Gerber, and Suzanna L. de Boef. 1999. "Tracking Opinion over Time: A Method for Reducing Sampling Error." *Public Opinion Quarterly* 63 (Summer): 178-92.
- Groseclose, Tim. 2001. "A Model of Candidate Location When One Candidate Has a Valence Advantage." *American Journal of Political Science* 45 (October): 862-86.
- Grynaviski, Jeffrey D. 2006. "A Bayesian Learning Model with Applications to Party Identification." *Journal of Theoretical Politics* 18 (July): 323-46.
- Hamm, Robert M. 1993. "Explanations for Common Responses to the Blue/Green Cab Probabilistic Inference Word Problem." *Psychological Reports* 72 (1): 219-42.
- Harvey, Andrew C. 1989. *Forecasting, Structural Time Series and the Kalman Filter*. New York: Cambridge University Press.
- Higgins, E. Tory. 1996. "Knowledge Activation: Accessibility, Applicability, and Salience." In *Social Psychology: Handbook of Basic Principles*, ed. E. Tory Higgins and Arie W. Kruglanski. New York: Guilford Press.
- Hovland, Carl I., and Muzafer Sherif. 1957. "Assimilation and Contrast Effects in Reactions to Communication and Attitude Change." *Journal of Abnormal Social Psychology* 55 (September): 244-52.
- Husted, Thomas A., Lawrence W. Kenny, and Rebecca B. Morton. 1995. "Constituent Errors in Assessing Their Senators." *Public Choice* 83 (June): 251-71.
- Isikoff, Michael, and John Barry. 2005 May 9. "Gitmo: SouthCom Showdown." *Newsweek* p. 10. <http://www.msnbc.msn.com/id/7693014/site/newsweek/>.
- Iyengar, Shanto, and Nicholas A. Valentino. 2000. "Who Says What? Source Credibility as a Mediator of Campaign Advertising." In *Elements of Reason*,

ed. Arthur Lupia, Mathew D. McCubbins, and Samuel L. Popkin. New York: Cambridge University Press.

Jackman, Simon. 2005. "Pooling the Polls Over an Election Campaign." *Australian Journal of Political Science* 40 (December): 499-517.

Jacobson, Gary C. 2006. *A Divider, Not a Uniter: George W. Bush and the American People*. Upper Saddle River, NJ: Pearson.

Jamieson, Kathleen Hall, and Paul Waldman. 2003. *The Press Effect: Politicians, Journalists, and the Stories that Shape the Political World*. New York: Oxford University Press.

Jeffrey, Richard. 2004. *Subjective Probability: The Real Thing*. New York: Cambridge University Press.

Jessee, Stephen A., and Douglas Rivers. 2007. "Political Judgment Under Uncertainty: Heuristics and Biases." Stanford University. Manuscript.

Jones, Edward E., and Richard E. Nisbett. 1972. "The Actor and the Observer: Divergent Perceptions of the Causes of the Behavior." In *Attribution: Perceiving the Causes of Behavior*, ed. Edward E. Jones, David E. Kanouse, Harold H. Kelley, Richard E. Nisbett, Stuart Valins, and Bernard Weiner. Morristown, NJ: General Learning Press.

Kalman, Rudolph E. 1960. "A New Approach to Linear Filtering and Prediction Problems." *Journal of Basic Engineering, Series D* 82: 35-45.

Kalman, Rudolph E., and Richard S. Bucy. 1961. "New Results in Linear Filtering and Prediction Theory." *Journal of Basic Engineering, Series D* 83 (March): 95-108.



- Kam, Cindy D. 2005. "Who Toes the Party Line? Cues, Values, and Individual Differences." *Political Behavior* 27 (June): 163-82.
- Katz, Elihu. 1968. "On Reopening the Question of Selectivity in Exposure to Mass Communication." In *Theories of Cognitive Consistency: A Sourcebook*, ed. Robert P. Abelson, Elliot Aronson, William J. McGuire, Theodore M. Newcomb, Milton J. Rosenberg, and Percy H. Tannenbaum. Chicago: Rand McNally.
- Keenan, Nancy. 11 August 2005. "Rebuttal to FactCheck.org Analysis of NARAL Pro-Choice America's TV Ad."  
<http://www.factcheck.org/UploadedFiles/Rebuttal-letter-from-Nancy-Keenan-to-FactCheck.org-8-11-05.pdf>.
- Kelley, Harold H. 1973. "The Processes of Causal Attribution." *American Psychologist* 28 (February): 107-28.
- Kennan, George F. 1997. *George F. Kennan and the Origins of Containment, 1944-1946: The Kennan-Lukacs Correspondence*. Columbia, Missouri: University of Missouri Press.
- Kerkhof, Peter. 1999. "Applying the Unimodel to Political Persuasion." *Psychological Inquiry* 10 (2): 137-40.
- Kingdon, John W. 1984. *Agendas, Alternative, and Public Policies*. Boston: Little, Brown.
- Knight, Jack, and James Johnson. 1997. "What Sort of Political Equality Does Deliberative Democracy Require?" In *Deliberative Democracy*, ed. James Bohman and William Rehg. Cambridge: MIT Press.
- Koehler, Jonathan J. 1996. "The Base Rate Fallacy Reconsidered: Descriptive, Normative, and Methodological Challenges." *Behavioral and Brain Sciences* 19 (March): 1-53.

## 138 Bibliography

- Krosnick, Jon A., and Donald R. Kinder. 1990. "Altering the Foundations of Support for the President through Priming." *American Political Science Review* 84 (June): 497-512.
- Kuklinski, James H., Paul J. Quirk, Jennifer Jerit, David Schwieder, and Robert F. Rich. 2000. "Misinformation and the Currency of Democratic Citizenship." *Journal of Politics* 62 (August): 790-816.
- Kunda, Ziva. 1990. "The Case for Motivated Reasoning." *Psychological Bulletin* 108 (November): 480-98.
- Kurtz, Howard. 16 May 2005. "Newsweek Apologizes: Inaccurate Report on Koran Led to Riots." *Washington Post* p. A01.
- Lau, Richard P., and David P. Redlawsk. 2001. "Advantages and Disadvantages of Heuristics in Political Decision Making." *American Journal of Political Science* 45 (October): 951-71.
- Leamer, Edward E. 1978. *Specification Searches*. New York: Wiley.
- Leamer, Edward E. 1983. "Let's Take the Con out of Econometrics." *American Economic Review* 73 (1): 31-43.
- Lepper, Mark R., Lee Ross, and Richard R. Lau. 1986. "Persistence of Inaccurate Beliefs about the Self: Perseverance Effects in the Classroom." *Journal of Personality and Social Psychology* 50 (March): 482-91.
- Levendusky, Matthew S. 2006. "Sorting in the U.S. Mass Electorate." Ph.D. diss. Stanford University.
- Lieb, David A. 2005 April 7. "Missouri Medicaid Cuts Sent to Blunt." *Associated Press*.

- Lodge, Milton, and Charles Taber. 2000. "Three Steps toward a Theory of Motivated Political Reasoning." In *Elements of Reason*, ed. Arthur Lupia, Mathew D. McCubbins, and Samuel L. Popkin. New York: Cambridge University Press.
- Lohmann, Suzanne. 1994. "The Dynamics of Informational Cascades: The Monday Demonstrations in Leipzig, East Germany, 1989-91." *World Politics* 47 (October): 42-101.
- Londregan, John B. 2000. *Legislative Institutions and Ideology in Chile*. New York: Cambridge University Press.
- 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 (November): 2098-2109.
- Luce, R. Duncan. 1995. "Four Tensions Concerning Mathematical Modeling in Psychology." *Annual Review of Psychology* 46: 1-26.
- Lupia, Arthur. 2002. "Who Can Persuade Whom? Implications from the Nexus of Psychology and Rational Choice Theory." In *Thinking about Political Psychology*, ed. James H. Kuklinski. New York: Cambridge University Press.
- Lupia, Arthur. 2006. "When Can Politicians Scare Citizens Into Supporting Bad Policies? A Theory of Incentives with Fear-Based Content." University of Michigan at Ann Arbor. Typescript.
- Lupia, Arthur, and Mathew D. McCubbins. 1998. *The Democratic Dilemma: Can Citizens Learn What They Need to Know?* New York: Cambridge University Press.
- Lupia, Arthur, Mathew D. McCubbins, and Samuel L. Popkin, eds. 2000. *Elements of Reason*. New York: Cambridge University Press.

- Luskin, Robert C., James S. Fishkin, and Roger Jowell. 2002. "Considered Opinions: Deliberative Polling in Britain." *British Journal of Political Science* 32 (July): 455-87.
- McGuire, William J. 1964. "Inducing Resistance to Persuasion: Some Contemporary Approaches." *Advances in Experimental Social Psychology* 1: 191-229.
- McGuire, William J. 1969. "The Nature of Attitudes and Attitude Change." In *Handbook of Social Psychology*, ed. Gardner Lindzey and Elliot Aronson. 2nd ed. Reading, MA: Addison-Wesley.
- Meinhold, Richard J., and Nozer D. Singpurwalla. 1983. "Understanding the Kalman Filter." *American Statistician* 37 (May): 123-27.
- Mondak, Jeffery J. 1993a. "Public Opinion and Heuristic Processing of Source Cues." *Political Behavior* 15 (June): 167-92.
- Mondak, Jeffery J. 1993b. "Source Cues and Policy Approval: The Cognitive Dynamics of Public Support for the Reagan Agenda." *American Journal of Political Science* 37 (February): 186-212.
- NARAL Pro-Choice America. 2005. "Speaking Out." Television advertisement. <http://www.factcheck.org/video/Speaking-Out.wmv>.
- Neill, Daniel B. 2005. "Cascade Effects in Heterogeneous Populations." *Rationality and Society* 17 (2): 191-241.
- Newell, Allen, J.C. Shaw, and Herbert A. Simon. 1958. "Chess-Playing Programs and the Problem of Complexity." *IBM Journal of Research and Development* 2 (October): 320-35.

- Nielsen Media Research. 18 August 2005. "Supreme Court Political Ads Virtually Absent From Television." Press release.
- Ottati, Victor C. 1990. "Determinants of Political Judgments: The Joint Influence of Normative and Heuristic Rules of Inference." *Political Behavior* 12 (June): 159-79.
- Page, Benjamin I., and Richard A. Brody. 1972. "Policy Voting and the Electoral Process: The Vietnam War Issue." *American Political Science Review* 66 (September): 979-95.
- Petrocik, John R. 1996. "Issue Ownership in Presidential Elections, with a 1980 Case Study." *American Journal of Political Science* 40 (August): 825-50.
- Petty, Richard E., and John T. Cacioppo. 1986. *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*. New York: Springer-Verlag.
- Phillips, Lawrence D., and Ward Edwards. 1966. "Conservatism in a Simple Probability Inference Task." *Journal of Experimental Psychology* 72 (September): 346-54.
- Poole, Keith T., and Howard L. Rosenthal. 2006. *Ideology and Congress*. Piscataway, NJ: Transaction Publishers.
- Popkin, Samuel L. 1994. *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns*. 2nd ed. Chicago: University of Chicago Press.
- Pratt, John W., Howard Raiffa, and Robert Schlaifer. 1964. "The Foundations of Decision under Uncertainty: An Elementary Exposition." *Journal of the American Statistical Association* 59 (June): 353-75.
- Press, S. James. 2003. *Subjective and Objective Bayesian Statistics*. 2nd ed. Hoboken, NJ: Wiley.

## 142 Bibliography

- Prior, Markus. 2007. *Post-Broadcast Democracy*. New York: Cambridge University Press.
- Program on International Policy Attitudes. 2003 February 21. "Americans on Iraq and the U.N. Inspections II." [http://www.pipa.org/OnlineReports/Iraq/IraqUNInsp2\\_Feb03/IraqUNInsp2%20Feb03%20rpt.pdf](http://www.pipa.org/OnlineReports/Iraq/IraqUNInsp2_Feb03/IraqUNInsp2%20Feb03%20rpt.pdf).
- Rahn, Wendy M. 1993. "The Role of Partisan Stereotypes in Information Processing about Political Candidates." *American Journal of Political Science* 37 (May): 472-96.
- Rawls, John. 1993. *Political Liberalism*. New York: Columbia University Press.
- Ross, Lee, and Andrew Ward. 1996. "Naive Realism in Everyday Life: Implications for Social Conflict and Misunderstanding." In *Values and Knowledge*, ed. T. Brown, E. Reed, and E. Turiel. New Jersey: Lawrence Erlbaum.
- Ross, Lee, Mark R. Lepper, and Michael Hubbard. 1975. "Perseverance in Self-Perception and Social Perception: Biased Attributional Processes in the Debriefing Paradigm." *Journal of Personality and Social Psychology* 32 (May): 880-92.
- Savage, Charlie. 2005 June 4. "Pentagon Details Koran Mistreatment." *Boston Globe* p. A1.
- Savage, Leonard J. 1954. *The Foundations of Statistics*. New York: Wiley.
- Savage, Leonard J. 1964. "The Foundations of Statistics Reconsidered." In *Studies in Subjective Probability*, ed. Henry E. Kyburg, Jr. and Howard E. Smokler. New York: Wiley.

- Schaffner, Brian F., and Matthew J. Streb. 2002. "The Partisan Heuristic in Low-Information Elections." *Public Opinion Quarterly* 66 (Winter): 559-81.
- Schuman, Howard, and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. San Diego: Academic Press.
- Schumpeter, Joseph A. 1942. *Capitalism, Socialism, and Democracy*. New York: Harper & Brothers.
- Schwieder, David W., and Paul J. Quirk. 2004. "Missed Cues: Judgment Heuristics and Citizen Competence." Presented at the Annual Conference of the Midwest Political Science Association, Chicago.
- Sears, David O., and Jonathan L. Freedman. 1967. "Selective Exposure to Information: A Critical Review." *Public Opinion Quarterly* 31 (Summer): 194-213.
- Simon, Herbert A. 1955. "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics* 69 (February): 99-118.
- Simon, Herbert A. 1959. "Theories of Decision-Making in Economics and Behavioral Science." *American Economic Review* 49 (June): 253-83.
- Simon, Herbert A., and Allen Newell. 1958. "Heuristic Problem Solving: The Next Advance in Operations Research." *Operations Research* 6 (January-February): 1-10.
- Smith, Edward E. 1990. "Categorization." In *Thinking: An Invitation to Cognitive Science*, ed. Daniel N. Osherson and Edward E. Smith. Cambridge, MA: MIT Press.
- Smith, Edward E., and Douglas L. Medin. 1981. *Categories and Concepts*. Cambridge, MA: Harvard University Press.

## 144 Bibliography

- Sniderman, Paul M., and Sean M. Theriault. 2004. "The Structure of Political Argument and the Logic of Issue Framing." In *Studies in Public Opinion: Attitudes, Nonattitudes, Measurement Error, and Change*, ed. Willem E. Saris and Paul M. Sniderman. Princeton, New Jersey: Princeton University Press.
- 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.
- Stokes, Donald E. 1963. "Spatial Models of Party Competition." *American Political Science Review* 57 (June): 368-77.
- Strasser, Helmut. 1981. "Consistency of Maximum Likelihood and Bayes Estimates." *Annals of Statistics* 9 (September): 1107-1113.
- Taber, Charles S., and Milton Lodge. 2006. "Motivated Skepticism in the Evaluation of Political Beliefs." *American Journal of Political Science* 50 (July): 755-69.
- Tessin, Jeff. 2006. "Cues Given, Cues Received: How Candidates Use Shortcuts When Voters Need Them Most." Presented at the Annual Meeting of the Midwest Political Science Association, Chicago.
- Tetlock, Philip E. 2005. *Expert Political Judgment*. Princeton: Princeton University Press.
- Tversky, Amos, and Daniel Kahneman. 1971. "Belief in the Law of Small Numbers." *Psychological Bulletin* 76 (August): 105-10.
- van Houweling, Robert P., and Paul M. Sniderman. 2004. "The Political Logic of a Downsian Space." Presented at the Annual Meeting of the Midwest Political Science Association, Chicago.



- Webster, James G. 2005. "Beneath the Veneer of Fragmentation: Television Audience Polarization in a Multichannel World." *Journal of Communication* 55 (June): 366-82.
- White House. 2002 February 14. "President Announces Clear Skies & Global Climate Change Initiatives." Speech. <http://www.whitehouse.gov/news/releases/2002/02/20020214-5.html>.
- Whitney v. California*. 1927. 274 U.S. 357.
- Williams, Daniel, and Kamran Khan. 2005 May 28. "Muslims Rally Over Koran Report." *Washington Post* p. A14.
- Zaller, John R. 1992. *The Nature and Origins of Mass Opinion*. New York: Cambridge University Press.
- Zanna, Mark P., and John K. Rempel. 1988. "Attitudes: A New Look at an Old Concept." In *The Social Psychology of Knowledge*, ed. Daniel Bar-Tal and Arie W. Kruglanski. New York: Cambridge University Press.
- 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.