Appendix to "Partisan Bias and the Bayesian Ideal in the Study of Public Opinion"

John G. Bullock*

February 27, 2009

*Department of Political Science, Yale University. Please send comments to john.bullock@aya.yale.edu.

Proof of Proposition 1. It suffices to show that any updater's belief will converge to $\bar{x}$, the mean of the received messages. Let $\mathbf{x}$ be the sequence of messages $x_1, \ldots, x_n$ and $\bar{x}$ be the mean of these messages. Suppose $\epsilon > 0$ and $N = \frac{\sigma_x^2(|\hat{\mu}_0 - \bar{x}| - \epsilon)}{\sigma_0^2 \epsilon}$. Then $n > N$ implies

$$\begin{aligned}
|\hat{\mu}_n|\mathbf{x} - \bar{x}| &= \left| \frac{\sigma_0^2 \bar{x} + (\sigma_x^2/n)\hat{\mu}_0 - \bar{x}(\sigma_0^2 + \sigma_x^2/n)}{\sigma_0^2 + \sigma_x^2/n} \right| \\
&= \left| \frac{(\sigma_x^2/n)(\hat{\mu}_0 - \bar{x})}{\sigma_0^2 + \sigma_x^2/n} \right| \\
&< \frac{\sigma_x^2 \, |\hat{\mu}_0 - \bar{x}|}{N\sigma_0^2 + \sigma_x^2} \\
&= \frac{\epsilon \sigma_x^2 \, |\hat{\mu}_0 - \bar{x}|}{\sigma_x^2 \, (|\hat{\mu}_0 - \bar{x}| - \epsilon) + \epsilon \sigma_x^2} \\
&= \epsilon.
\end{aligned}$$

$\hat{\mu}_n|\mathbf{x}$ therefore converges surely to $\bar{x}$. □

Proposition 1 assumes normal priors and normal likelihoods because of the near-ubiquity of those assumptions in Bayesian public opinion studies. But the proposition is subsumed by several results that do not make strong distributional assumptions about the prior belief or the likelihood. Blackwell and Dubins (1962) show that convergence to agreement will occur when different priors are absolutely continuous with respect to each other, i.e., when they assign positive probability to exactly the same set of events. (Normal priors, like those considered in Proposition 1, are absolutely continuous with respect to each other.) Moreover, it is well-known that different Bayesian priors are consistent (and thus converge to agreement) under a wide range of conditions, chief of which are that the prior beliefs do not exclude the true parameter value as impossible, that the dimensionality of the parameter space is finite, and that the signals received are informative about the true parameter values (Doob 1949; Savage 1954, 46-50; Walker 1969; for discussions, see Lindley 1972, 61-64 and Diaconis and Freedman 1986).

Proof of Proposition 2. By a common result (e.g., Lee 2004, 38-39), one has the same posterior belief whether he updates in response to the n messages $x_1, \ldots, x_t, \ldots, x_n$ or in response to a single message $\bar{x} = \sum_{t=1}^{n} x_t/n$ with precision $\tau_{x*} = \sum_{t=1}^{n} \tau_{xt}$. Thus, by Equation 1a,

$$
\begin{aligned}
|\hat{\mu}_{Dn} - \hat{\mu}_{Rn}| &= \left|\hat{\mu}_{D0}\left(\frac{\tau_{D0}}{\tau_{D0} + \tau_{x*}}\right) - \hat{\mu}_{R0}\left(\frac{\tau_{R0}}{\tau_{R0} + \tau_{x*}}\right) + \bar{x}\left(\frac{\tau_{x*}}{\tau_{D0} + \tau_{x*}} - \frac{\tau_{x*}}{\tau_{R0} + \tau_{x*}}\right)\right| \\
&= \left|\frac{(\tau_{R0} + \tau_{x*})(\hat{\mu}_{D0}\tau_{D0})}{(\tau_{R0} + \tau_{x*})(\tau_{D0} + \tau_{x*})} - \frac{(\tau_{D0} + \tau_{x*})\hat{\mu}_{R0}\tau_{R0}}{(\tau_{D0} + \tau_{x*})(\tau_{R0} + \tau_{x*})} + \bar{x}\left(\frac{\tau_{x*}}{\tau_{D0} + \tau_{x*}} - \frac{\tau_{x*}}{\tau_{R0} + \tau_{x*}}\right)\right| \\
&= \frac{1}{(\tau_{D0} + \tau_{x*})(\tau_{R0} + \tau_{x*})}\left|(\hat{\mu}_{D0} - \hat{\mu}_{R0})\tau_{D0}\tau_{R0} + (\bar{x} - \hat{\mu}_{R0})\tau_{R0}\tau_{x*} + (\hat{\mu}_{D0} - \bar{x})\tau_{D0}\tau_{x*}\right|.
\end{aligned}
$$

Divergence requires

$$
|\hat{\mu}_{D0} - \hat{\mu}_{R0}| < \frac{1}{(\tau_{D0} + \tau_{x*})(\tau_{R0} + \tau_{x*})}\left|(\hat{\mu}_{D0} - \hat{\mu}_{R0})\tau_{D0}\tau_{R0} + (\bar{x} - \hat{\mu}_{R0})\tau_{R0}\tau_{x*} + (\hat{\mu}_{D0} - \bar{x})\tau_{D0}\tau_{x*}\right|,
$$

which implies $|\hat{\mu}_{D0} - \hat{\mu}_{R0}|\,(\tau_{D0}+\tau_{x*})(\tau_{R0}+\tau_{x*}) < |(\hat{\mu}_{D0} - \hat{\mu}_{R0})\tau_{D0}\tau_{R0} + (\bar{x} - \hat{\mu}_{R0})\tau_{R0}\tau_{x*} + (\hat{\mu}_{D0} - \bar{x})\tau_{D0}\tau_{x*}|$. If $\hat{\mu}_{D0} \geq \hat{\mu}_{R0}$, this inequality holds only when $(\bar{x} - \hat{\mu}_{R0})\tau_{R0}\tau_{x*} + (\hat{\mu}_{D0} - \bar{x})\tau_{D0}\tau_{x*}$ is greater than a or less than b. And if $\hat{\mu}_{D0} \leq \hat{\mu}_{R0}$, this inequality holds only when $(\bar{x} - \hat{\mu}_{R0})\tau_{R0}\tau_{x*} + (\hat{\mu}_{D0} - \bar{x})\tau_{D0}\tau_{x*}$ is less than a or greater than b. □

Proof of Corollary to Proposition 2. Proof of (a) by contradiction. Without loss of generality, assume $\hat{\mu}_{D0} > \hat{\mu}_{R0}$. Assume divergence even though $\bar{x} \in [\hat{\mu}_{R0}, \hat{\mu}_{D0}]$. By Proposition 2, one of two conditions must be satisfied if divergence is to occur. First, it will occur if $(\hat{\mu}_{D0} - \bar{x})\tau_{D0}\tau_{x*} + (\bar{x} - \hat{\mu}_{R0})\tau_{R0}\tau_{x*} > (\hat{\mu}_{D0} - \hat{\mu}_{R0})\,[(\tau_{D0} + \tau_{x*})\,(\tau_{R0} + \tau_{x*}) - \tau_{D0}\tau_{R0}] \Rightarrow 0 > (\mu_{D0} - \bar{x})(\tau_{R0}\tau_{x*} + \tau_{x*}^2) + (\bar{x} - \hat{\mu}_{R0})(\tau_{D0}\tau_{x*} + \tau_{x*}^2)$, which is impossible because $(\mu_{D0} - \bar{x})$, $(\bar{x} - \hat{\mu}_{R0})$, and the precisions are all nonnegative. Second, divergence will occur if $(\hat{\mu}_{D0} - \bar{x})\tau_{D0}\tau_{x*} + (\bar{x} - \hat{\mu}_{R0})\tau_{R0}\tau_{x*} < (\hat{\mu}_{R0} - \hat{\mu}_{D0})\,[(\tau_{D0} + \tau_{x*})\,(\tau_{R0} + \tau_{x*}) + \tau_{D0}\tau_{R0}]$. But this inequality cannot hold, either, because $(\hat{\mu}_{D0} - \bar{x})\tau_{D0}\tau_{x*} + (\bar{x} - \hat{\mu}_{R0})\tau_{R0}\tau_{x*} \geq 0$ and $(\hat{\mu}_{R0} - \hat{\mu}_{D0})\,[(\tau_{D0} + \tau_{x*})\,(\tau_{R0} + \tau_{x*}) + \tau_{D0}\tau_{R0}] \leq 0$. This establishes that divergence cannot occur when $\bar{x} \in [\min\{\hat{\mu}_{D0}, \hat{\mu}_{R0}\}, \max\{\hat{\mu}_{D0}, \hat{\mu}_{R0}\}]$.

Proof of (b) by direct calculation. If $\sigma^2_{D0} = \sigma^2_{R0}$, the condition for divergence given in Proposition 2 reduces to $(\hat{\mu}_{D0} - \hat{\mu}_{R0})\tau_{D0}\tau_{x*} \notin [\min\{a', b'\}, \max\{a', b'\}]$, where $a' = (\hat{\mu}_{D0} - \hat{\mu}_{R0})[(\tau_{D0} + \tau_{x*})^2 - \tau^2_{D0}]$ and $b' = (\hat{\mu}_{R0} - \hat{\mu}_{D0})[(\tau_{D0} + \tau_{x*})^2 + \tau^2_{D0}]$. This cannot be: if $\hat{\mu}_{D0} > \hat{\mu}_{R0}$, $b' < (\hat{\mu}_{D0} - \hat{\mu}_{R0})\tau_{D0}\tau_{x*} < a'$; if $\hat{\mu}_{D0} < \hat{\mu}_{R0}$, $a' < (\hat{\mu}_{D0} - \hat{\mu}_{R0})\tau_{D0}\tau_{x*} < b'$; if $\hat{\mu}_{D0} = \hat{\mu}_{R0}$, $(\hat{\mu}_{D0} - \hat{\mu}_{R0})\tau_{D0}\tau_{x*} = a' = b' = 0$. □

Proof of Proposition 3. Proof by cases. First, assume $\bar{x} > \max\{\hat{\mu}_{D0}, \hat{\mu}_{R0}\}$. Then by Equation 1a, $\hat{\mu}_{Dn} > \hat{\mu}_{D0}$ and $\hat{\mu}_{Rn} > \hat{\mu}_{R0}$, so $(\hat{\mu}_{Dn} - \hat{\mu}_{D0}) / (\hat{\mu}_{Rn} - \hat{\mu}_{R0}) > 0$. Next, assume $\bar{x} < \min\{\hat{\mu}_{D0}, \hat{\mu}_{R0}\}$. Then by Equation 1a, $\hat{\mu}_{Dn} < \hat{\mu}_{D0}$ and $\hat{\mu}_{Rn} < \hat{\mu}_{R0}$, so $(\hat{\mu}_{Dn} - \hat{\mu}_{D0}) / (\hat{\mu}_{Rn} - \hat{\mu}_{R0}) > 0$. By the Corollary to Proposition 2, we need not consider the case in which $\bar{x} \in [\min\{\hat{\mu}_{D0}, \hat{\mu}_{R0}\}, \max\{\hat{\mu}_{D0}, \hat{\mu}_{R0}\}]$.□

Proof of Proposition 4. We begin with three lemmas:

Lemma 1. The Kalman filter estimator $\hat{\mu}_t$ can be written as a linear function of the mean of the prior belief, $\hat{\mu}_0$, and the new messages that have been received, $\mathbf{x_t}$: $\hat{\mu}_t = c_t\hat{\mu}_0 + \mathbf{f'_t}\mathbf{x_t}$, where $c_t = \prod_{i=1}^{t}(1 - K_i)\gamma$, $K_i = \sigma^2_i/\sigma^2_x$ is the "Kalman gain," $\mathbf{f'_t}$ is a $1 \times t$ vector of coefficients, and $\mathbf{x_t}$ is the $t \times 1$ vector of messages $x_1, \ldots, x_t$. (See Gerber and Green 1998, 805 for a proof specific to the Kalman filter; for a general statement, see Theorem 2 in Diaconis and Ylvisaker 1979. Note a clerical error in Gerber and Green 1998: the article has $c_t = \prod_{i=1}^{t}(1 - K_i)\gamma^t$ instead of the correct $c_t = \prod_{i=1}^{t}(1 - K_i)\gamma$.)

Let $f'_{t[i]}$ be the i_{th} element of $\mathbf{f'_t}$. Then $f'_{t[i]} = \gamma^{t-i}K_i \prod_{j=i+1}^{t}(1 - K_j)$ for $i \in 1, \ldots, t-1$, and $f'_{t[i]} = K_i$ for $i = t$. Proof by induction: for $i = 1$, $\hat{\mu}_1 = \gamma\hat{\mu}_0(1 - K_1) + x_1K_1$ (by Equation 3a), so $f'_{1[1]} = K_1$. Assume that the statement is true for $i = t$. Then for $i = t + 1$,

$$\begin{aligned}\hat{\mu}_{t+1} &= \gamma\hat{\mu}_t(1 - K_{t+1}) + x_{t+1}K_{t+1} \\ &= \gamma(1 - K_{t+1})\left(c_t\hat{\mu}_0 + \mathbf{f'_t}\mathbf{x_t}\right) + x_{t+1}K_{t+1} \\ &= \gamma(1 - K_{t+1})\left(c_t\hat{\mu}_0\right) + \mathbf{f'_{t+1}}\mathbf{x_{t+1}},\end{aligned}$$

where the last element of $\mathbf{f'_{t+1}}$ is K_{t+1} and the other elements are given by $f'_{t+1[i]} = f'_{t[i]}(1-K_{t+1})\gamma = \left[\gamma^{t-i} K_i \prod_{j=i+1}^{t}(1-K_j)\right](1-K_{t+1})\gamma = \gamma^{t+1-i} K_i \prod_{j=i+1}^{t+1}(1-K_j)$. □

Lemma 2.

$$\text{plim } K_t = K = \frac{\gamma^2 - h - 1 + \sqrt{(-\gamma^2+h+1)^2 + 4h\gamma^2}}{2\gamma^2}$$

where $h = \sigma_\mu^2/\sigma_x^2$. (See Gerber and Green 1998 for a proof.)

Lemma 3. $0 \leq (1-K)\gamma < 1$. By definition, $K < 1$ and $\gamma \geq 0$. It follows immediately that $0 \leq (1-K)\gamma$. For the claim $(1-K)\gamma < 1$, proof by contradiction: assume $(1-K)\gamma \geq 1$. Note that $0 < K < 1$ because $K_t = \sigma_t^2/\sigma_x^2 = \frac{1/\sigma_x^2}{1/(\gamma^2\sigma_{t-1}^2+\sigma_\mu^2)+1/\sigma_x^2}$. This implies $0 < (1-K) < 1$. Therefore, $(1-K)\gamma \geq 1$ implies $\gamma > 1$. Now,

$$(1-K)\,\gamma \geq 1$$

$$\Rightarrow \left(1 - \frac{\gamma^2 - h - 1 + \sqrt{(-\gamma^2+h+1)^2 + 4h\gamma^2}}{2\gamma^2}\right)\gamma \geq 1$$

$$\Rightarrow -4\gamma^3 + 8\gamma^2 - 4\gamma h - 4\gamma \geq 0.$$

This implies $h \leq -\gamma^2 + 2\gamma - 1$. But $-\gamma^2 + 2\gamma - 1 < 0 \;\forall\; \gamma \notin [-1, 1]$, and h must be nonnegative because it is a ratio of variances. Contradiction. □

With these lemmas in hand, the proof is straightforward. We need to show plim $(\hat{\mu}_{Dt} - \hat{\mu}_{Rt}) = 0$. By Lemma 1, the difference between the means of updaters' beliefs at any time t is $\hat{\mu}_{Dt} - \hat{\mu}_{Rt} = \hat{\mu}_{D0}\prod_{i=1}^{t}(1-K_{Di})\gamma - \hat{\mu}_{R0}\prod_{i=1}^{t}(1-K_{Ri})\gamma + \left(\mathbf{f'_{Dt}} - \mathbf{f'_{Rt}}\right)\mathbf{x_t}$. Because plim $y\prod_{i=1}^{\infty} z_i = 0$ for any constant y if plim $|z_i| < 1$, plim $\hat{\mu}_{D0}\prod_{i=1}^{t}(1-K_{Di})\gamma = 0$ if plim $|(1-K_{Di})\gamma| < 1$. And by Lemma 3, it is. This establishes plim $\hat{\mu}_{D0}\prod_{i=1}^{t}(1-K_{Di})\gamma = 0$, and by the same logic, plim $\hat{\mu}_{R0}\prod_{i=1}^{t}(1-K_{Ri})\gamma = 0$.

We now have plim $(\hat{\mu}_{Dt} - \hat{\mu}_{Rt}) =$ plim $\left(\mathbf{f'_{Dt}} - \mathbf{f'_{Rt}}\right)\mathbf{x_t}$. This equals 0 if plim $\left(\mathbf{f'_{Dt}} - \mathbf{f'_{Rt}}\right) = \mathbf{0}$, i.e., if $\lim_{t\to\infty} f'_{Dt[i]} - f'_{Rt[i]} = 0$ for all i. It does. For $i \in 1,\ldots,t-1$, $f'_{Dt[i]} = K_{Di}\prod_{j=i+1}^{t}(1-K_{Dj})\gamma$ by Lemma 1, and because plim $|(1-K_{Di})\gamma| < 1$ (by Lemma 3), plim $f'_{Dt[i]} = 0$ for $i \in 1,\ldots,t-1$. By the same logic, plim $f'_{Rt[i]} = 0$ for $i \in 1,\ldots,t-1$, so $\lim_{t\to\infty} f'_{Dt[i]} - f'_{Rt[i]} = 0 - 0$ for $i \in 1,\ldots,t-1$.

For $i = t$, $\text{plim } f'_{Dt[i]} = f'_{Rt[i]} = K$ by Lemma 2. This establishes $\lim_{t\to\infty} f'_{Dt[i]} - f'_{Rt[i]} = 0$ for all i, completing the proof. □

Proof of Proposition 5. Proof by contradiction. Let $z = K_{Du}(x_u - \gamma\hat{\mu}_{Dt}) - K_{Ru}(x_u - \gamma\hat{\mu}_{Rt})$. Assume no divergence even though $z \notin [\min\{a, b\}, \max\{a, b\}]$. Then

$$
\begin{aligned}
|\hat{\mu}_{Dt} - \hat{\mu}_{Rt}| &\geq |\hat{\mu}_{Du} - \hat{\mu}_{Ru}| \\
&= |\gamma\hat{\mu}_{Dt}(1 - K_{Du}) + x_u K_{Du} - \gamma\hat{\mu}_{Rt}(1 - K_{Ru}) - x_u K_{Ru}| \\
&= |\gamma(\hat{\mu}_{Dt} - \hat{\mu}_{Rt}) + z|\,.
\end{aligned}
$$

If $\hat{\mu}_{Dt} \geq \hat{\mu}_{Rt}$, $b \leq a$, and the inequality holds only when $z \geq b$ and $z \leq a$. If $\hat{\mu}_{Dt} \leq \hat{\mu}_{Rt}$, $a \leq b$, and the inequality holds only when $z \geq a$ and $z \leq b$. Either case is a contradiction. This establishes that divergence occurs if $z \notin [\min\{a, b\}, \max\{a, b\}]$.

Now assume divergence even though $z \in [\min\{a, b\}, \max\{a, b\}]$. Then

$$
\begin{aligned}
|\hat{\mu}_{Dt} - \hat{\mu}_{Rt}| &< |\hat{\mu}_{Du} - \hat{\mu}_{Ru}| \\
&= |\gamma(\hat{\mu}_{Dt} - \hat{\mu}_{Rt}) + z|\,.
\end{aligned}
$$

If $\hat{\mu}_{Dt} \geq \hat{\mu}_{Rt}$, $b \leq a$, and the inequality holds only when $z > a$ or $z < b$. If $\hat{\mu}_{Dt} \leq \hat{\mu}_{Rt}$, $a \leq b$, and the inequality holds only when $z < a$ or $z > b$. Either case is a contradiction. This establishes that divergence occurs only if $z \notin [\min\{a, b\}, \max\{a, b\}]$. □

Proof of Corollary to Proposition 5. Suppose $\sigma^2_{Dt} = \sigma^2_{Rt}$. Then $K_{Du} = \sigma^2_{Du}/\sigma^2_x = K_{Ru} = \sigma^2_{Ru}/\sigma^2_x = K_u$. In this case, divergence requires

$$
\begin{aligned}
|\hat{\mu}_{Dt} - \hat{\mu}_{Rt}| &< |\hat{\mu}_{Du} - \hat{\mu}_{Ru}| \\
&= |\gamma\hat{\mu}_{Dt}(1 - K_u) - \gamma\hat{\mu}_{Rt}(1 - K_u)| \\
&= |\gamma(\hat{\mu}_{Dt} - \hat{\mu}_{Rt}) + K_u\gamma(\hat{\mu}_{Rt} - \hat{\mu}_{Dt})| \\
&= (1 - K_u)\gamma\,|\hat{\mu}_{Dt} - \hat{\mu}_{Rt}| \\
\Rightarrow 1 &< (1 - K_u)\gamma.
\end{aligned}
$$

This condition does not depend on the value of x_u. Thus, divergence when learning about a changing condition requires neither $\sigma^2_{Dt} = \sigma^2_{Rt}$ nor $x_u \in [\min\{\hat{\mu}_{Dt}, \hat{\mu}_{Rt}\}, \max\{\hat{\mu}_{Dt}, \hat{\mu}_{Rt}\}]$. □

Proof of Proposition 6. Proof by cases. By definition, polarization occurs between t and u iff divergence occurs between t and u and $(\hat{\mu}_{Du} - \hat{\mu}_{Dt})/(\hat{\mu}_{Ru} - \hat{\mu}_{Rt}) < 0$. Assume $\hat{\mu}_{Du} > \hat{\mu}_{Dt}$. Polarization under this condition implies $\hat{\mu}_{Ru} < \hat{\mu}_{Rt}$. We have $\hat{\mu}_{Du} > \hat{\mu}_{Dt} \Rightarrow \gamma\hat{\mu}_{Dt} - \gamma\hat{\mu}_{Dt}K_{Du} + x_uK_{Du} > \hat{\mu}_{Dt} \Rightarrow K_{Du}(x_u - \gamma\hat{\mu}_{Dt}) > (1 - \gamma)\hat{\mu}_{Dt}$ and $\hat{\mu}_{Ru} < \hat{\mu}_{Rt} \Rightarrow \gamma\hat{\mu}_{Rt} - \gamma\hat{\mu}_{Rt}K_{Ru} + x_uK_{Ru} < \hat{\mu}_{Rt} \Rightarrow K_{Ru}(x_u - \gamma\hat{\mu}_{Rt}) < (1 - \gamma)\hat{\mu}_{Rt}$.

Now assume $\hat{\mu}_{Du} < \hat{\mu}_{Dt}$. Polarization under this condition implies $\hat{\mu}_{Ru} > \hat{\mu}_{Rt}$. We have $\hat{\mu}_{Du} < \hat{\mu}_{Dt} \Rightarrow \gamma\hat{\mu}_{Dt} - \gamma\hat{\mu}_{Dt}K_{Du} + x_uK_{Du} < \hat{\mu}_{Dt} \Rightarrow K_{Du}(x_u - \gamma\hat{\mu}_{Dt}) < (1 - \gamma)\hat{\mu}_{Dt}$ and $\hat{\mu}_{Ru} > \hat{\mu}_{Rt} \Rightarrow \gamma\hat{\mu}_{Rt} - \gamma\hat{\mu}_{Rt}K_{Ru} + x_uK_{Ru} > \hat{\mu}_{Rt} \Rightarrow K_{Ru}(x_u - \gamma\hat{\mu}_{Rt}) > (1 - \gamma)\hat{\mu}_{Rt}$.

Now assume $\hat{\mu}_{Du} = \hat{\mu}_{Dt}$. By definition, polarization cannot occur under this condition, because $(\hat{\mu}_{Du} - \hat{\mu}_{Dt})/(\hat{\mu}_{Ru} - \hat{\mu}_{Rt})$ cannot be negative. □

References

Blackwell, David, and Lester Dubins. 1962. "Merging of Opinions with Increasing Information." *Annals of Mathematical Statistics* 33 (September): 882-86.

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.

Doob, Joseph L. 1949. "Application of the Theory of Martingales." *Le Calcul des Probabilites et ses Applications*: 22-28.

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

Lee, Peter M. 2004. *Bayesian Statistics: An Introduction*. 3rd ed. London: Arnold.

Lindley, D. V. 1972. *Bayesian Statistics: A Review*. Philadelphia: Society for Industrial and Applied Mathematics.

Savage, Leonard J. 1954. *The Foundations of Statistics*. New York: Wiley.

Walker, A. M. 1969. "On the Asymptotic Behaviour of Posterior Distributions." *Journal of the Royal Statistical Society, Series B* 31 (1): 80-88.