JOSHUA D. CLINTON
Vanderbilt University
SIMON JACKMAN
Stanford University

To Simulate or NOMINATE?

Carroll et al. (2009) summarize the similarities and differences between the NOMINATE and IDEAL methods of fitting spatial voting models to binary roll-call data. As those authors note, for the class of problems with which either NOMINATE and the Bayesian quadratic-normal model can be used, the ideal point estimates almost always coincide, and when they do not, the discrepancy is due to the somewhat arbitrary identification and computational constraints imposed by each method. There are, however, many problems for which the Bayesian quadratic-normal model can be easily generalized, so as to address a broad array of questions and take advantage of additional data. Given the nature and source of the differences between NOMINATE and the Bayesian approach—as well as the fact that both approaches are approximations of the decision-making processes being modeled—we believe that it is preferable to choose the more flexible Bayesian approach.

We completely agree with Carroll, Lewis, Lo, Poole, and Rosenthal's (2009, 555) conclusion in their contribution to this issue of *Legislative Studies Quarterly* that "neither model [NOMINATE nor IDEAL] has a clear advantage over the other in the recovery of legislator locations or roll-call midpoints in either larger or small legislatures." We would suggest, however, that there are other important differences between the two methods of analyzing roll-call data. The Bayesian approach is more flexible than the family of NOMINATE estimators, as it is more easily adapted to account for additional data structures and alternative modeling assumptions. To answer the question that Carroll et al. pose—"Which (if any) approach is most appropriate in any given research situation?"—we believe that the Bayesian approach underlying IDEAL is more readily applicable to many research situations.

Before getting into details, we stress that this exchange is friendly. We regularly correspond with the other authors on matters regarding the analysis of roll calls: for example, the subtleties of identification, how to speed up computing for the roll-call problem, and using (and breaking) our respective computer programs.[1]

In the article that follows, we review the Bayesian underpinnings of our approach in Section 2. In Sections 3 and 5, we contrast the assumptions underlying both NOMINATE and our Bayesian approach, and in Section 4 we draw attention to some interesting features of NOMINATE. To introduce some notation and key concepts, we begin by briefly restating the operationalization of the Euclidean spatial voting model (Enelow and Hinich 1984) that we developed with Rivers (Clinton, Jackman, and Rivers 2004b) and which underlies IDEAL.

1. The Quadratic-Normal Voting Model

Roll call $j \in \{1, \ldots, m\}$ presents legislator $i \in \{1, \ldots, n\}$ with a choice between a "yea" position (ζ_j) and a "nay" position (ψ_j) with locations in $\mathbb{R}^d$, where d denotes the dimension of the policy space. Let $y_{ij} = 1$ if legislator i votes "yea" on the j-th roll call and $y_{ij} = 0$ if the legislator votes "nay." Abstentions generate missing data, which we consider to be missing at random in our analysis.[2] We assume legislators have quadratic utility functions, $U_i(\zeta_j) = -\left\|\zeta_j - \xi_i\right\|^2 + \eta_{ij}$ and $U_i(\psi_j) = -\left\|\psi_j - \xi_i\right\|^2 + v_{ij}$, where $\xi_i \in \mathbb{R}^d$ is the ideal point of legislator i, and η_{ij} and v_{ij} are stochastic elements of utility (errors), and $\|\bullet\|$ is the Euclidean norm. Utility maximization implies $y_{ij} = 1$ if $U_i(\zeta_j) > U_i(\psi_j)$ and $y_{ij} = 0$ otherwise. We complete the voting model by assuming that the errors η_{ij} and v_{ij} have a joint normal distribution with $E(\eta_{ij}) = E(v_{ij})$, $\text{var}(\eta_{ij} - v_{ij}) = \sigma_j^2$, and the errors are independent across both legislators and roll calls. Straightforward algebra yields $\pi_{ij} \equiv \Pr(y_{ij} = 1) = \Phi(\beta_j' \xi_i - \alpha_j)$, where $\beta_j = 2(\zeta_j - \psi_j)/\sigma_j$, $\alpha_j = (\zeta_j' \zeta_j - \psi_j' \psi_j)/\sigma_j$, and $\Phi(\bullet)$ denotes the normal distribution function. With these assumptions we have a probit model with an unobserved regressor, ξ_i, corresponding to the legislator's ideal point (a logit model results if the errors have extreme value distributions). In one dimension, the ratio $\tau_j = \alpha_j/\beta_j = (\zeta_j^2 - \psi_j^2)/2(\zeta_j - \psi_j) = (\zeta_j + \psi_j)/2$ is the cut-point, the point at which a legislator is indifferent between the proposals (absent stochastic sources of utility).

Assuming conditional independence of the votes across legislators and votes, given ideal points and vote-specific parameters, the likelihood is $p(Y \mid \mathrm{B}, \alpha, \Xi) = \prod_{i=1}^{n} \prod_{j=1}^{m} \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{1 - y_{ij}}$, where B is an m-by-d matrix of bill parameters (formed by stacking the β_j), $\alpha = (\alpha_1, \ldots, \alpha_m)'$ is a vector

of unknown intercept parameters, Ξ is an n-by-d matrix of ideal points (formed by stacking the ξ_i), and Y is the $n \times m$ matrix of observed votes with (i, j)th element y_{ij}.

Identification. When fitting a unidimensional model, we typically constrain the ideal points, ξ_i, to have mean 0 and standard deviation 1 across legislators (the "normalize" option in the implementation of IDEAL in R). This technique is sufficient for local identification in the sense that Rothenberg (1971) uses the term, ruling out observationally equivalent translations and rescalings of the parameters, save for a reflection of the parameters about the origin. We typically find that the lack of global identification is not a practical concern when we fit unidimensional models to datasets from recent U.S. Congresses. Because the posterior densities are usually estimated with reasonable levels of precision (the roll-call matrices are quite large, and the unidimensional model provides an excellent fit to data from recent U.S. Congresses), the Gibbs sampler never visits the "mirror image" mode of the posterior density. Identification is more delicate when we seek to fit higher-dimensional models (Rivers 2003).

Bayesian Inference. We adopt a Bayesian approach to inference for the unknown model parameters. That is, we seek to characterize the posterior density of the unknown parameters, which, via Bayes' Rule, is $p(B, \alpha, \Xi \mid Y) \propto p(Y \mid B, \alpha, \Xi)\, p(B, \alpha, \Xi)$, where the first density on the right-hand side is the likelihood and the second is the prior density. Absent information to the contrary, we generally assume complete *a priori* independence among the model parameters. For convenience, we usually specify a mean of 0 and normal densities for these priors. With the probit form of the voting model, we can use an easily implemented, although computationally intensive, data-augmented Gibbs sampling algorithm to sample from the posterior density (see Clinton, Jackman, and Rivers 2004b for details). Carroll et al. (2009, Section 3.1) have highlighted the importance of the choice of variances for these prior densities, a matter we explore in Section 5.

2. Bayesian Inference, or Something Else?

An important difference between NOMINATE and IDEAL is that the latter is Bayesian and the former is not. This difference is consequential because non-Bayesian, classical statistical inference is problematic in the roll-call setting. Jackman (2009) has made the case for Bayesian inference in the social sciences elsewhere, but we will provide a capsule of the argument here.

Classical statistical inference does not make sense in many social science applications, because the underlying idea of repeated random sampling from a population is untenable. Roll-call data provide a case in point. A typical roll-call analysis consists of gathering some or all recorded votes in a given legislature in a given time period. In what sense do these data constitute a random sample? What is the population from which this dataset can be considered one of many possible random samples? We think that the answers to these questions are not readily apparent.

In this way, roll-call analysis is vastly different from, say, the analysis of survey data, with which we can entertain the thought experiment of repeatedly sampling respondents from a well-defined population, administering the survey to them, and getting a different set of data and statistics each time. The possibility of sampling variability—the cornerstone of classical statistical inference—is reasonable in this survey setting, but it is more dubious in the roll-call setting.

Bayesian procedures ask, "Given *these* data, y, what should I now believe about an unknown quantity, θ?" Bayes' Rule tells us how to update prior beliefs regarding q in light of y to yield a posterior probability density of $p(\theta \mid y)$. Bayesian inference does not rest on the notion of repeated sampling, an idea that is, at best, sketchy in the roll-call setting. We prefer a Bayesian approach to analyzing roll calls because the underlying statistical foundations *make sense.*

We believe one ought to use a Bayesian approach, but we have yet to make the case for *the* specific Bayesian approach implemented in IDEAL. The probit, two-parameter item-response-theory (IRT) model in IDEAL follows from a set of assumptions about the form of legislators' utility functions. The specific choices we made were driven by our desire to generate an easily implemented estimator (for instance, the data-augmented Gibbs sampler) while staying close to the presentation of spatial voting models in the literature (for example, quadratic-loss utilities). Other assumptions are certainly possible, and, in fact, we provide a Bayesian version of NOMINATE in Section 6.

The other feature to recommend a Bayesian approach is its relative simplicity. Bayes' Rule states that this posterior density is proportional to the likelihood for the data, y, multiplied by the prior density for θ, $p(\theta)$. There is no more to Bayesian inference, and no less. There is no place for *ex post* or *ad hoc* "fixes" in a Bayesian analysis.

Of course, the devil is in the details. What likelihood and priors will we assume? How will we carry out the computation? How will we summarize the posterior? The choice of which likelihood to fit to the data is one faced by all analysts, not only Bayesians, and the motivation for

our conventional choices appears in Section 1. The details of Bayesian computation for the roll-call problem are now relatively settled and uncontroversial (in contrast, we will raise some quibbles with the estimation and inference procedures in NOMINATE).

3. Differences in Functional Form

The analysis of roll-call data seeks to characterize revealed preferences. Researchers assume a model of voting in which preferences are embedded as parameters to be estimated from observed votes. Different assumptions about the underlying voting model and different estimation methods will lead investigators to different estimates of preferences. Even so, the excellent exposition of Carroll et al. demonstrates that similarities in the assumptions of the statistical models of NOMINATE and IDEAL tend to yield similar answers, particularly for large datasets that the models fit very well (as low-dimensional spatial voting models fit the data from recent U.S. Congresses).

The Bayesian/non-Bayesian distinction aside, the most fundamental difference between NOMINATE and IDEAL is that NOMINATE assumes that legislators' utility functions are scaled Gaussian functions, whereas IDEAL assumes that legislators' utilities are quadratic in Euclidean distance. As Carroll et al. point out (see their Figure 1), in a neighborhood of the legislator's ideal point, the Gaussian is quadratic, and so, in this local sense, NOMINATE and IDEAL share the same deterministic component of the utility function. This similarity explains why the estimates of ideal points are frequently near-identical but also why, when they differ, they typically differ at the extremes.

It is reasonable to ask if one utility function—or measurement model, per se—fits the data better than another. But this question is hard to answer because the Gaussian/logistic model used in NOMINATE and the quadratic-probit model in IDEAL do not nest via a set of parameter restrictions. Neither model is capable of out-of-sample predictions,[3] so comparing the in-sample performance of both models (see, for example, Carroll et al. 2009) is about as well as we can do. Yet any comparison of in-sample performance is muddied by the different identification strategies used by the two models.

Thus it is difficult to determine which set of functional form assumptions best fit observed roll calls. We doubt that solving this issue with roll-call data alone is feasible. Making fine distinctions about functional form with respect to unobservables is always difficult with observational data.[4] Determining the functional form of legislators' utility functions is a question of considerable interest, but we suspect

a quest for the "true" utility function would be costly, time consuming, and unable to yield the desired holy grail.

Even so, Carroll et al. suggest that NOMINATE has some resistance to outliers. To this end, they examined differences in the NOMINATE and IDEAL estimates of the ideal point of Senator Russ Feingold in the 109th Senate. To better understand this feature of NOMINATE, recall that the NOMINATE voting model is:

$$\pi_{ij} \equiv \Pr(y_{ij} = \text{"}Yea\text{"}) = \Lambda\left(\beta \exp\left[\frac{-w}{2}\left(\xi_i - Y_j\right)^2 \right] - \beta \exp\left[\frac{-w}{2}\left(\xi_i - N_j\right)^2 \right] \right), \quad (1)$$

where $\Lambda(\cdot) = \exp(\cdot)/(1 + \exp(\cdot))$ is the logistic distribution function, Y_j and N_j are the locations of the yea and nay positions for roll call j, and w and β are scaling parameters (Carroll et al. 2009, equation 2). These scaling parameters are estimated from the data by the NOMINATE algorithms, and, in this sense, NOMINATE fits a richer model to the data than the IRT model. In conjunction with the nonlinearity inherent in the Gaussian utility functions, these scaling parameters help NOMINATE dampen the influence of seemingly aberrant votes—although we are not confident that even a large dataset would yield much information about both w and β; we return to this question in Section 6.[5]

4. Additional Contrasts with NOMINATE

There are some additional differences between our Bayesian approach and NOMINATE that are not widely understood. Although none of these issues seem particularly consequential in the "easy" case of a large roll-call matrix well fit by a low-dimensional spatial voting model (such as provided by recent U.S. Congresses), we stress that: (a) the analysis of roll-call data involves numerous assumptions, regardless of the model used; (b) some peculiar assumptions underlie NOMINATE, shaping its output in ad hoc, informal ways; and (c) the incorporation of assumptions into our Bayesian approach is relatively transparent, parsimonious, and even spartan in comparison to the process in NOMINATE.

NOMINATE's point estimates are only approximately maximum-likelihood estimators, because NOMINATE obtains its estimates through a series of alternating, conditional maximum-likelihood steps.[6] In fact, it is possible that NOMINATE's point estimates correspond to saddlepoint solutions of its objective function. We lack good intuition regarding the practical consequences of NOMINATE being an approximate maximum-likelihood estimator.

In fairness, the point estimates produced by our Bayesian approach—the means of the respective marginal posterior densities—are only approximately Bayes estimates, because they are accompanied by Monte Carlo error (which modern computing power can make arbitrarily small). Nonetheless, we are more confident about the statistical basis of the output of IDEAL (Bayes estimates, point summaries of posterior densities) than we think we can be about the output of NOMINATE, particularly in light of the discussion to follow.

NOMINATE uses a correlation-based convergence criterion for determining when to cease iterations (Poole and Rosenthal 1997, 237). Between this feature and NOMINATE's alternating, conditional maximum-likelihood algorithm, NOMINATE may not find its way to the global maximum of its objective function, and difficulties in the optimization of that objective function—or even the nonuniqueness of a global maximum—can go undetected.

To be fair, our Bayesian approach explores the posterior density using Markov chain Monte Carlo (MCMC) methods, and there is always a risk that the analyst might stop a MCMC algorithm too soon. The determination of lower bounds on running times for MCMC algorithms is problem specific, and we can only suggest that users carefully check that any MCMC algorithm is generating a satisfactory exploration of the posterior density—for example, visiting the tails of the posterior density sufficiently often so as to generate good estimates of credible intervals.

NOMINATE makes *ad hoc* corrections to the distribution of estimated ideal points. According to Carroll et al., ". . . NOMINATE constrains the distance between those legislators located at –1 and 1 (the leftmost and rightmost positions) and their nearest neighbors not located at –1 and 1 to be no more than 0.1 unit (or 5% of the –1 to 1 scale). . . . In very small legislatures (fewer than 20 members), the constraint is not applied at all" (2009).

As far as we can tell, this constraint exists to make the resulting distributions of estimated ideal points look more plausible than they would otherwise; see Poole's discussion (2005, 155–57). This seemingly *ad hoc* constraint contrasts with the stark simplicity of Bayesian inference, in which "the posterior is proportional to the prior times the likelihood." Priors over ideal points or bill parameters may operate so as to constrain some ideal points to be closer or farther away from others, but priors are usually formulated to bring additional legislator-specific[7] or vote-specific[8] information to bear on the analysis, not to adjust the appearance of the *a posteriori* distribution of recovered ideal points.

NOMINATE uses a number of mechanisms to try to prevent the calculations from wandering into numerically unstable regions of the

parameter space. These measures include the defaults that any roll call retained for analysis have no fewer than 2.5% of the recorded votes on the losing side and that any legislator retained for analysis vote at least 20 times. Also, NOMINATE effectively constrains proposal and status-quo points and ideal points to lie in the [–1,1] interval when it fits a unidimensional model. These defaults—present as computational safeguards, more than anything—generate some of the same functionality provided by priors in our Bayesian analysis.

Even if NOMINATE were a maximum-likelihood estimator, the information matrix of all the parameters would be extremely large and difficult to compute and invert (Poole and Rosenthal 1997, 246). For many years, NOMINATE reported approximate standard errors, which were probably too small, but by a factor that was not particularly consequential in most large datasets. With the advent of faster computers, NOMINATE's standard errors are now generated by a simulation-based, parametric bootstrap procedure (Lewis and Poole 2004). Nonetheless, these standard errors have unusual properties in some cases we have examined, which we detail in Section 6. These issues arise even if we take the standard errors on their own terms, as estimates of variability in the estimates due to random sampling, our earlier critique of classical inference in the roll-call setting notwithstanding.

Some of these points are "legacy" issues, holdovers from the days when NOMINATE was developed for the analysis of large roll-call matrices on much slower computing hardware than we have at our disposal today.[9] Even so, they are differences worth considering when we evaluate the output of the two models. We are heartened by the Carroll et al. implementation of a MCMC version of NOMINATE, and we present a Bayesian version of NOMINATE that overcomes some of these issues in Section 6.

5. Priors in the Bayesian Analysis of Roll-Call Data

Carroll et al. draw particular attention to the sensitivity of the output of the Bayesian approach to the priors. We concede that the choice of the prior can be consequential. Our colleagues ably demonstrate circumstances in which posterior densities are sensitive to priors in roll-call-data analysis. Nonetheless, we would emphasize several points with respect to the choice of priors. First, the choice of prior is but one of several "subjective" decisions that must be made when analyzing roll-call data. Second, sensitivity to prior assumptions is inescapable, particularly if the roll-call data is "short" on one or both dimensions (the number of legislators and/or the number of votes), if some legislators

have extreme voting histories, if some votes are extremely lopsided, or if some combination of these scenarios exists. Third, when the roll-call matrix contains only a small amount of information about a legislator's ideal point or a particular roll call, prior densities can bring additional information to bear in a transparent, yet formal way.

That said, the analysis of roll-call data poses challenges for both classical and Bayesian analysis: the "bare bones" version of the model is not identified, and for extremely lopsided roll calls and legislators with extreme voting histories, the maximum-likelihood estimators of the corresponding parameters are not finite. Given these conditions, we should not be surprised that the results may be sensitive to the priors. Moreover, priors that seem "uninformative" might not be after we impose identifying restrictions. When conducting any Bayesian data analysis—but perhaps especially the analysis of roll-call data—we must think carefully about the implication of the priors being used.

Recall that in our Bayesian approach we typically obtain local identification for the unidimensional version of the model via a "mean = 0, variance = 1" restriction on cross-legislator distribution of the ideal points. That is, we begin with the unidentified model $y_{ij} \sim Bernoulli(\pi_{ij}), \Phi^{-1}(\pi_{ij}) = \beta_j \xi_i - \alpha_j, \xi_i \sim N(0, \sigma_\xi^2), (\beta_j, \alpha_j)' \sim N(\mathbf{0}, \sigma_\beta^2 \mathbf{I}_2)$. We then impose the normalization $\tilde{\xi}_i = (\xi_i - c)/m$, where $c = \bar{\xi}$ and $m = sd(\xi)$. Further, we seek $\tilde{\beta}_j$ and $\tilde{\alpha}_j$ such that $\xi_i \beta_j - \alpha_j = \tilde{\xi}_i \tilde{\beta}_j - \tilde{\alpha}_j$, from which we deduce that $\tilde{\beta}_j = \beta_j m$ and $\tilde{\alpha}_j = \alpha_j - \tilde{\beta}_j c/m = \alpha_j - \beta_j c$. *A priori*, we have $c = 0$ and (at least for a reasonably large n) $m = sd(\xi) \approx \sigma_\xi$, and so for the identified parameters we have $\tilde{\xi}_i \sim N(0,1)$, $\tilde{\beta}_j \sim N(0, \sigma_\xi^2 \sigma_\beta^2)$, and $\tilde{\alpha} \sim N(0, \sigma_\beta^2)$. Note that the prior variance for $\tilde{\beta}$ is the product $\sigma_\xi^2 \sigma_\beta^2$. If both σ_ξ^2 and σ_β^2 are set to large quantities, then the resulting prior variance for $\tilde{\beta}$ can be potentially massive, or at least much larger than what the user may have intended when specifying the variance of the prior of the unidentified parameter, β_j. An interesting case is one in which the user specifies $\sigma_\xi^2 = 1$, so $V(\tilde{\beta}_j) = V(\beta_j)$. The normalization from ξ_i to $\tilde{\xi}_i$ means that σ_ξ^2 is largely a redundant parameter: in the absence of any prior information over the ξ_i, we may as well set $\sigma_\xi^2 = 1$, such that $V(\tilde{\beta}_j) = \sigma_\beta^2$. The user-supplied value for this quantity corresponds to the variance over the identified parameter $\tilde{\beta}_j$.

Midpoint Parameters. Carroll et al. compared the performance of the two procedures in terms of recovery of roll-call midpoints. In the parameterization we employ in IDEAL, the midpoint between ζ_j and ψ_j is recovered as $\tau_j = \tilde{\alpha}_j / \tilde{\beta}_j$. This is a ratio of two random variables, each

of which are given normal priors centered on 0, which means that the implied prior on the bill midpoint, τ_j, has a heavy-tailed Cauchy form. Indeed, if $\tilde{\alpha}_j$ and $\tilde{\beta}_j$ have the same prior variance, then the prior 50% credible interval for τ_j is [–1, 1], but the prior 95% credible interval is [–12.7, 12.7] (recall that ideal points are normalized to have a mean of 0 and a standard deviation of 1). Thus, reasonable priors for the bill parameters generate ridiculously vague priors for the midpoints,[10] and if we really wish to perform inference with respect to the midpoints—and these are interesting parameters—then perhaps we need a different approach.[11]

Re-parameterize the Model in Terms of a Midpoint Parameter. The quadratic-normal model can be rewritten so that the midpoints appear as parameters in their own right: we rewrite the model as $\Phi^{-1}(\pi_{ij}) = \tilde{\beta}_j(\tilde{\xi}_i - \tau_j)$, where $\tau_j = \tilde{\alpha}_j / \tilde{\beta}_j$. We can then specify a prior for the midpoints directly: $\tau_j \sim N(0, 1.5^2)$ would be quite permissive relative to the a priori distribution of the ideal points, or we might use a prior that simply stipulates that the cut-points lie to the interior of the distribution of ideal points, $\tau_j \sim Unif\,(\min(\xi), \max(\xi))$. This alternative parameterization is trivial to deploy, especially when one uses a general-purpose program for Bayesian statistical modeling, such as BUGS (Spiegelhalter et al. 2003) or JAGS (Plummer 2009), for roll-call matrices of small to moderate size.

Not Bayesian Enough? Another possibility is a hierarchical model for τ_j or for the other model parameters. At various points, Carroll et al. have compared NOMINATE and IDEAL using the mean squared error (MSE) of estimates of parameters such as bill midpoints. If our goal is to reduce the MSE, then we know that the "shrinkage" induced by fitting a hierarchical model produces Bayesian posterior densities that have frequentist properties superior to classical estimators. The principal result, due to Stein (1956), is that a set of hierarchical Bayes estimates of normal means—each given by the mean of the respective marginal posterior density (with variance equal to the variance of the marginal posterior density)—has less total MSE than the classical estimators of the means. Accordingly, if instead of treating the bill midpoints as a series of fixed effects (to use some non-Bayesian terminology), we specified a hierarchical model for them,[12] then we should not be surprised to find superior performance in a total MSE sense. A similar argument holds for other parameters in the roll-call model. In short, we think it is quite possible that there exists a version of our Bayesian approach, one with hierarchical structure over midpoints and ideal points, that outperforms rival models on a total MSE criterion.

In fact, one might well argue that hierarchical modeling is (or ought to be) the default in Bayesian modeling[13] and that the "vanilla" analysis of roll-call data—in which the analyst typically possesses no prior knowledge to distinguish bill *j* from bill *k*—is a prime candidate for hierarchical modeling.[14] That is, perhaps IDEAL is not Bayesian enough! We present a hierarchical model for roll-call analysis in Section 5.2.

5.1. Example: Obama as "the Most Liberal Senator" in 2007

We explore the question of prior sensitivity by examining an interesting selection of roll calls from the U.S. Senate. In the midst of the 2008 presidential campaign, *National Journal* announced that Senator Barack Obama (D-IL) was the most liberal senator in 2007. *National Journal* had also pronounced Senator John Kerry (D-MA) the most liberal senator in 2004, a claim we examined in considerable detail at that time (Clinton, Jackman, and Rivers 2004a).

In both cases, a potentially serious threat to these conclusions is that the *National Journal* ratings ignore the fact that candidates for president miss a substantial fraction of votes. The votes analyzed by the *National Journal* in 2005, 2006, and 2007 are a small subset of the total number of recorded votes that year: 19%, 29%, and 22%, respectively. And the pattern of missingness in the votes selected by the *National Journal* is highly correlated with the presence of a presidential election. Senator John McCain (R-AZ) missed 4.2% and 4.9% of the votes scored by the *National Journal* in 2005 and 2006, respectively, but McCain missed 55.6% in 2007. Likewise, Senator Obama missed only 1.4% of all roll calls and none of the *National Journal* key votes in 2005 and 2006, respectively, but he missed 33.3% while campaigning during 2007. For Obama in 2007, we only have 66 recorded key votes; for McCain, we have only 44 key votes. The relatively high degree of missingness and the substantive importance of the analysis make it an excellent opportunity to look for the effect of priors on our Bayesian analysis. We also use these data to assess the plausibility of extending the basic model to incorporate additional information in Section 5.2. (An additional complication that we ignore at this point is that the bills scheduled for a vote when the presidential candidates were present may differ from those scheduled for when the candidates were absent, because the scheduling of issues in the Senate is endogenously determined—abstentions are not missing at random, as both NOMINATE and IDEAL assume.)

To demonstrate the consequences of adjusting the weight given to the prior relative to the likelihood function, we analyze the 99 *National*

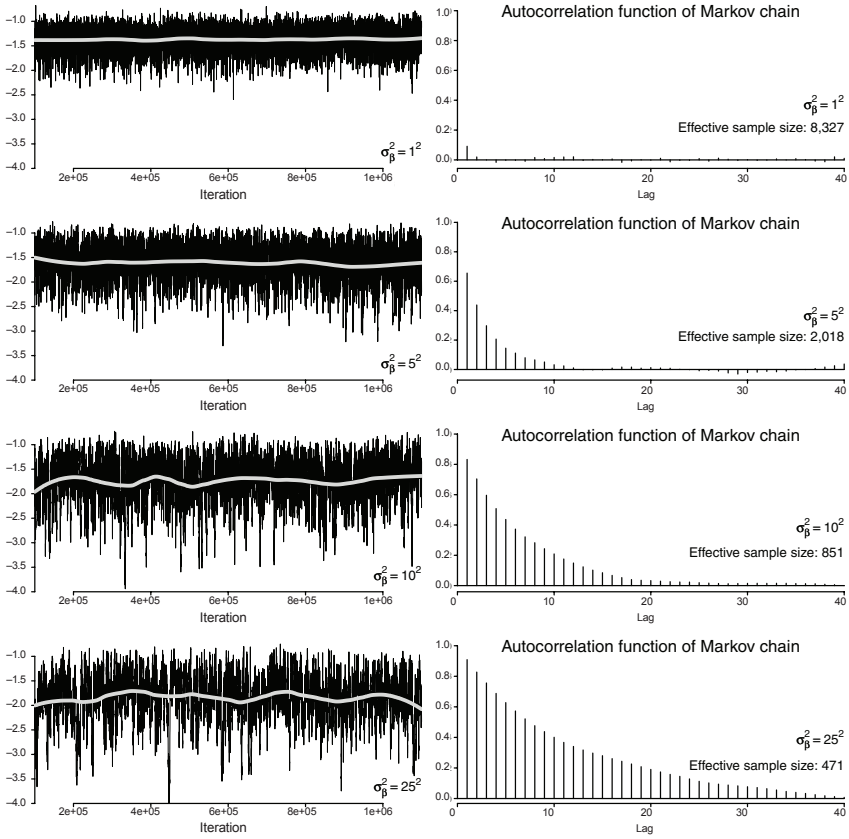
Journal key roll calls in 2007 using a one-dimensional model under four different priors: $\sigma_\beta^2 \in \{1, 5^2, 10^2, 25^2\}$. For each case, we employ the "mean = 0, variance = 1" normalization for the ideal points, ξ_i, discussed in Section 1. We set $\sigma_\xi^2 = 1$ such that the prior variance of the identified parameters $\tilde{\alpha}$ and $\tilde{\beta}$ is $\sigma_{\tilde{\beta}}^2 = \sigma_\beta^2 \sigma_\xi^2 = \sigma_\beta^2$. The prior is thus only a function of the σ_β^2 hyperparameter.

Figure 1 displays trace plots and autocorrelation functions for the output of the MCMC algorithm sampling from the marginal posterior density of Obama's ideal point, under the four different priors.[15] As the prior information about $\tilde{\beta}$ and $\tilde{\alpha}$ becomes vaguer (σ_β^2 gets larger), the performance of the MCMC algorithm degrades, with larger and more slowly decaying autocorrelations. This degradation occurs because as the precision of the prior information about the bill parameters eases, the posterior densities for these parameters become more diffuse; in turn, since the $\tilde{\beta}$ and $\tilde{\alpha}$ parameters contribute information about the ideal points, the posterior density for Obama's ideal point skews to the left, even with the identifying restriction that the ideal points have a mean of 0 and a variance of 1.

Figure 2 summarizes the marginal posterior density for Obama's ideal point in the panels on the left. The histograms summarize the 10,000 draws from the posterior density of Obama's ideal point retained from a 1.1 million-iteration run of IDEAL. The phenomenon we described earlier is clearly apparent. Over the four prior specifications for the bill parameters, the mean of the posterior density for Obama's ideal point moves about one-half of a standard deviation of the ideal-point distribution to the left, from –1.37 to –1.84. The increase in the left skew of the posterior density is also apparent, and the standard deviation of the posterior density almost doubles as we increase the standard deviation of the prior density by a factor of 25.

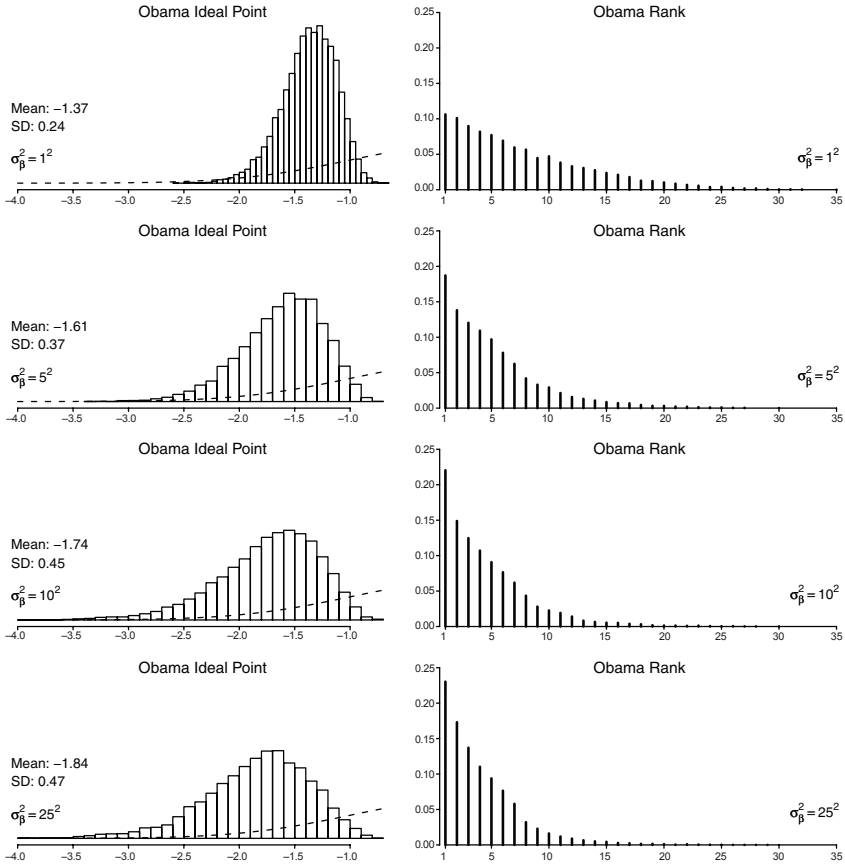
The impact of the change in the prior on the inferences we might draw about Obama's rank can be assessed if we examine the right column of panels in Figure 2. As we assume less about the bill parameters (increasing their prior variance σ_β^2), we become less certain about where to locate Obama's ideal point (left panels of Figure 2), but we become slightly more confident that Obama's ideal point occupies rank 1. The posterior mass function over the order statistic of Obama's ideal point is quite diffuse with the relatively informative prior given by setting $\sigma_\beta^2 = 1$ (top row of Figure 2); we assign only a .11 probability to Obama being "the most liberal senator" with this prior. As we assume less about the bills, inducing the leftward drift and skew in the posterior density over Obama's ideal point, we see the probability that Obama occupies rank 1 increasing to .19 ($\sigma_\beta^2 = 5^2$), .22 ($\sigma_\beta^2 = 10^2$), and .23 ($\sigma_\beta^2 = 25^2$).

FIGURE 1
Trace Plots and Autocorrelation Functions
of Markov Chain Monte Carlo Output for Obama's Ideal Point
under Different Prior Specifications



Note: Ideal points, ξ_i, were constrained to have a mean of 0 and a variance of 1 across senators, such that σ_β^2 is the only "free" hyperparameter in the specification of the prior. For the trace plots, the gray line shows a moving average. "Effective sample size" is a measure of the inefficiency of the MCMC algorithm; for instance, with $\sigma_\beta^2 = 10^2$, the 10,000 iterations (thinning one million iterations by a factor of 100) yield an estimate of the mean of the posterior density that has as much Monte Carlo error as we would get from approximately 851 independent draws from the posterior density.

FIGURE 2
Marginal Posterior Densities for Obama's Ideal Point
and Posterior Mass Functions for the Rank of Obama's Ideal Point,
under Different Prior Specifications



Note: Ideal points, ξ, have a mean of 0 and a variance of 1 across senators, such that σ_β^2 is the only "free" hyperparameter in the specification of the prior. The histograms on the left summarize 10,000 draws from the marginal posterior density for Obama's ideal point; the dotted line indicates the $N(0,1)$ prior for the ideal point.

Obama never has the highest probability of being the most liberal senator; Bernie Sanders (I-VT) is always rated as being more liberal than any other senator. Moreover, even with the least amount of prior information about the bills, no one senator is unambiguously "the most liberal." For instance, Sanders's posterior probability of being the most liberal senator never exceeds .30 in these specifications, and as many as seven senators can be assigned a non-negligible posterior probability of occupying rank 1.

In sum, it is hardly the case that the results "fall apart" or are "totally driven by the prior," but it is apparent that the less we are prepared to assume about the properties of the votes *a priori*, the less we wind up knowing, relatively, about Obama's ideal point *a posteriori*. This limitation is a subtle feature of the Bayesian IRT model, and one that Carroll et al. rightfully bring to light. Nonetheless, for priors of the sort $\sigma_\beta^2 > 5^2$, the results are relatively robust. Moving from the relatively tight priors with $\sigma_\beta^2 = 1$ to $\sigma_\beta^2 = 5^2$ produces the biggest changes in the results, and even these changes are not massive in any absolute or substantive sense.

In addition, only the estimated rank order of extremists experiences any real sensitivity to changes in the prior variance. Recall that Obama not only missed a substantial number of votes, but also recorded a relatively extreme voting history. In contrast, McCain, despite missing nearly 20% more votes than Obama, is assigned nearly the same rank regardless of the prior variance assumed.

To the extent that priors matter in the estimation of the basic IDEAL model, they matter most for the extreme legislators. For almost all of the others, however, the data dominate the prior. Because the estimates of interest are typically relatively centrally located—medians of the floor, committees, parties, and such—so long as the roll calls have a reasonable number of cut-points in this central region of the policy space, the choice of prior variance over the bill parameters will not be substantively consequential. Moreover, it is trivial to assess the sensitivity of the results to the priors.

5.2. An Informative Prior via Hierarchical Modeling

When constructing its rankings in 2007, the *National Journal* assumed that only votes cast in 2007 were relevant for estimations of senators' ideology in 2007. These data are not particularly informative about the ideal points of some legislators, either, because presidential candidates missed a much larger percentage of votes than did other senators. An example shows how the Bayesian approach can use additional information to supplement the analysis of roll-call data.

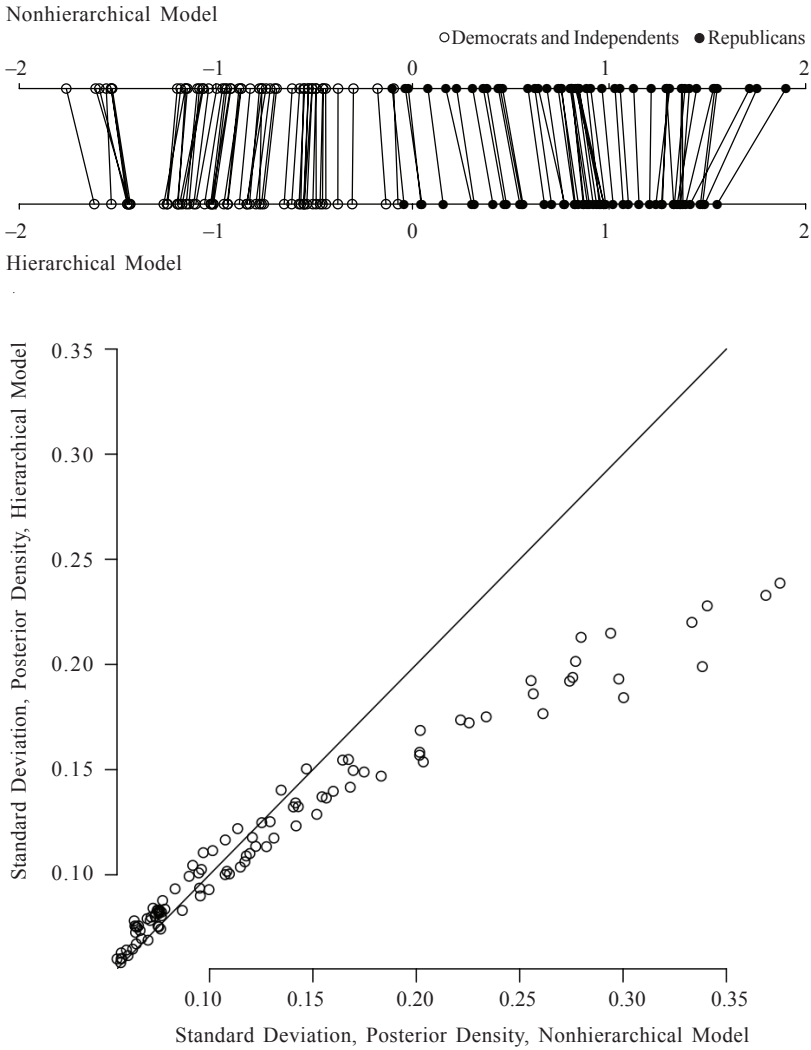
The simplest way to incorporate additional information is to pool roll calls across time (in this case, years). We can then make the strong assumption that all ideal points are constant across time, effectively increasing the amount of information from which we make inferences about the ideal points. This method is similar to the approach we adopted in our earlier work examining the 2003 *National Journal* ratings (Clinton, Jackman, and Rivers 2004a).

Here we use a different strategy, exploiting information about legislators via the following hierarchical model. Our model is very simple, exploiting two attributes of legislators: their party affiliations, and the political complexions of their respective states, as measured by the share of the vote won by John Kerry in the 2004 presidential election. We incorporate this information with the following hierarchical model for the ideal points: $\xi_i \sim N(\mu_\xi, \omega^2)$, $\mu_\xi = \gamma_0 + R_i\gamma_1 + K_{s(i)}\gamma_2$, $\boldsymbol{\gamma} \sim N(\mathbf{0}, 10^2 \cdot \mathbf{I}_3)$, $\omega \sim Unif(0,1)$, where $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2)'$ and ω^2 are hyperparameters, R_i is a binary indicator (1 if senator i is Republican and 0 otherwise), and $K_{s(i)}$ is Kerry's share of the vote in state s. The last two stochastic components of the hierarchical model specify vague priors over the hyperparameters. This model allows the observed characteristics R_i and $K_{s(i)}$ to contribute information to the analysis, but it uses the vague priors on the hyperparameters to reflect *a priori* uncertainty on our part as to how the information in R_i and $K_{s(i)}$ shapes inferences about ξ_i. This model is perhaps too simplistic to be substantively demonstrative, but it illustrates the larger point that our Bayesian model can be easily extended to incorporate additional information.

As usual, we imposed the identifying normalization that the ideal points have a mean of 0 and variance of 1 across legislators. This restriction sets an upper bound on ω and results in transformations of the $\boldsymbol{\gamma}$ and ω parameters.[16] For the bill parameters, we used the prior $(\beta_j, \alpha_j)' \sim \mathrm{N}(\mathbf{0}, 5^2 \cdot \mathbf{I}_2)$. We implemented this model in JAGS, which can generate several thousand samples from the posterior density for this model and for this small dataset quite quickly. The sampled values of the identified parameters reveal that the sampler in JAGS generates an efficient exploration of the posterior density for this problem.

We compared the posterior densities over ideal points produced by this hierarchical model with the results of the nonhierarchical model fit with $\sigma_\beta^2 = 5^2$. Figure 3 presents the two sets of estimates, with lines connecting them so as to highlight the shrinkage we usually see as a consequence of fitting a hierarchical model. The hierarchical model tends to pull together the estimated ideal points of senators with similar covariate values, especially when the information in the data about ideal points is not particularly rich (for example, for extremists, or senators

FIGURE 3
Comparison of Posterior Means of Ideal Points
and Comparison of Standard Deviations of Posterior Densities



Note: The top panel shows posterior means of ideal points for hierarchical and nonhierarchical models fit to 99 *National Journal* key votes of 2007. The lower panel shows standard deviations of marginal posterior densities of ideal points for hierarchical and nonhierarchical models fit to 99 *National Journal* key votes of 2007. The solid line is at 45 degrees.

with relatively short voting histories, or some combination thereof). The identifying restriction that the ideal points have a mean of 0 and variance of 1 means that there cannot be big differences between the hierarchical and nonhierarchical models with respect to estimates of ideal points.

In the lower panel of Figure 3, we present the standard deviations of the two sets of marginal posterior densities for the ideal points: one obtained from the hierarchical model (vertical axis), the other from the nonhierarchical model (horizontal axis). The comparison highlights the gains of utilizing the information in the covariates. The hierarchical model generally produces more precise estimates of the ideal points, especially for ideal points estimated relatively imprecisely by the standard, nonhierarchical model. Note what we have accomplished here: politically relevant sources of information about the ideal points have entered the analysis but not in a particularly heavy-handed way (we deployed vague priors for the hyperparameters), considerably boosting the precision of the inferences drawn about ideal points and quantities that are functions of the ideal points, such as rank orderings.

Also of interest are the $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ parameters, the hyperparameters attaching to the Republican indicator variables (R_i) and the Kerry vote shares ($K_{s(i)}$). The posterior mean (and 95% highest probability density interval) for $\tilde{\gamma}_1$ is 1.51 (1.35, 1.69) and –.032 (–.042, –.021) for $\tilde{\gamma}_2$. Consistent with the strong separation of ideal points by party, these parameter estimates suggest that, on average, senators from the same state but different parties will have ideal points that differ by 1.5 standard deviations of the ideal-point distribution. Moreover, these two variables account for a great deal of the variation across legislators in ideal points: the posterior density for the residual standard deviation, $\tilde{\omega}$, has a mean of .35 (.30, .41).

6. Bayesian NOMINATE

We fit these same data—the 99 *National Journal* key Senate roll calls for 2007—using NOMINATE as implemented in the R package W-NOMINATE (Poole et al. 2007). The default data-retention options in W-NOMINATE use all 101 senators and all 99 roll calls for analysis. We fit a one-dimensional model, with a polarity option that results in Democrats on the left and Republicans on the right. We used 100 parametric bootstrap samples to obtain reasonable estimates of confidence intervals and standard errors.

Table 1 reports point estimates and bootstrapped standard errors for the ten most liberal and ten most conservative senators (namely, the ten lowest and ten highest $\hat{\xi}_i$). Fifteen senators appear at the

TABLE 1
Ten Most Liberal and Ten Most Conservative Senators,
W-NOMINATE Point Estimates of Ideal Points
and Bootstrapped Standard Errors,
for 99 *National Journal* Key Votes of 2007

Most Liberal	$\hat{\xi}_i$	$se(\hat{\xi}_i)$	Most Conservative	$\hat{\xi}_i$	$se(\hat{\xi}_i)$
Biden (D-DE)	–1.00	1.03	Demint (R-SC)	1.00	1.00
Lautenberg (D-NJ)	–1.00	1.03	Coburn (R-OK)	1.00	1.13
Boxer (D-CA)	–1.00	1.22	Burr (R-NC)	1.00	1.01
Sanders (I-VT)	–1.00	1.11	Allard (R-CO)	1.00	1.06
Menendez (D-NJ)	–1.00	1.03	Kyl (R-AZ)	1.00	1.15
Obama (D-IL)	–1.00	1.04	Inhofe (R-OK)	1.00	1.08
Whitehouse (D-RI)	–1.00	1.03	Bunning (R-KY)	1.00	1.15
Feingold (D-WI)	–1.00	1.13	Vitter (R-LA)	0.99	1.05
Leahy (D-VT)	–0.92	1.07	Ensign (R-NV)	0.98	1.05
Stabenow (D-MI)	–0.91	1.01	Enzi (R-WY)	0.96	0.98

±1 boundary of NOMINATE's parameter space, including Obama, who is placed at the lower limit of –1 along with Senators Biden, Lautenberg, Boxer, Sanders, Menendez, Whitehouse, and Feingold. It would seem that the only way to distinguish among these eight senators with estimated ideal points of –1—say, if we wanted to induce a rank ordering over senators—is to look at the standard errors. Presumably, if two (or more) legislators have the same estimated ideal points, then the legislator whose ideal point has the largest standard error should be ranked lower than the other. This order would make sense if the sampling distribution of the estimated ideal points was symmetric, with the sampling distribution being normal, at least asymptotically.

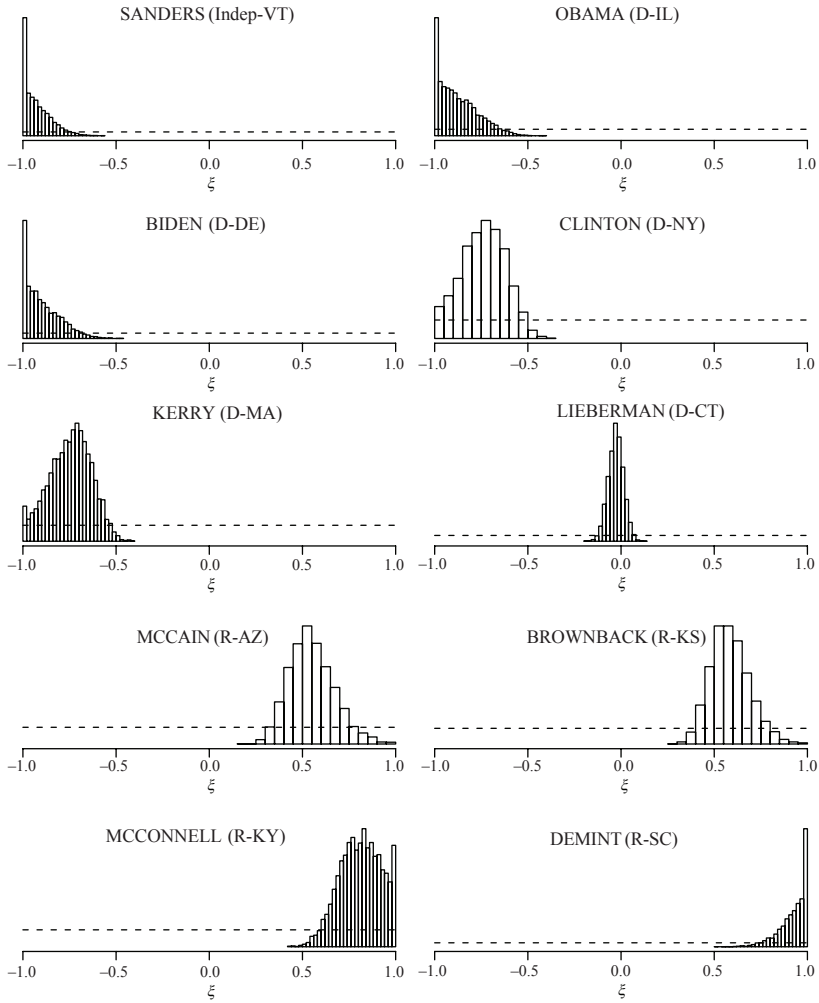
But what are we to make of the output of W-NOMINATE in this case, with 15 point estimates at the edge of the parameter space, each with a reasonably large standard error? In one dimension, NOMINATE constrains the ideal points to lie in the interval [–1,1], so the sampling distribution of $\hat{\xi}_i$ cannot possibly be normal, for then events that the model rules out by assumption would be assigned positive probability. For example, if $\hat{\xi}_i = -1.0$,, then any uncertainty must be to the right. Cases to the interior of the parameter space present difficulties, as well. What should we make of the fact that Senator Stabenow has an estimated ideal point of –.91 with a standard error of 1.01? It is difficult to know how to interpret these estimates and their standard errors.

A fully Bayesian analysis clears up these issues. Taking the NOMINATE model as presented by Carroll et al. and restated in our equation (1), we use the prior $\xi_i \sim Unif(-1, 1)$, but in addition we work with a (locally) identified likelihood by normalizing the ξ_i to range from –1 to 1 across legislators. As before, we denote the identified ideal points as $\tilde{\xi}_i$. We place uniform priors on the Y_j and N_j parameters, constraining them to lie between $\min(\xi_i)$ and $\max(\xi_i)$. We also transform Y_j and N_j to $\tilde{Y}_j$ and $\tilde{N}_j$, respectively, and the scaling parameter w to $\tilde{w}$. The β parameter is not affected by these identifying normalizations.[17] We also specify priors for the scaling parameters β and w—that is, $\beta \sim Unif\,(0, 1000)$ and $w \sim Unif\,(0, 1)$—and use the general-purpose Bayesian analysis computer program JAGS (Plummer 2009) to sample from the posterior density of the model parameters. For this small data-set, a general-purpose program suffices, as we trade off programming time for run time. We initialize the algorithm with estimates of β and w and ideal points from W-NOMINATE, and we generate start values for the Y_j and N_j parameters by running logistic regressions of each roll call on the W-NOMINATE ideal-point estimates and transforming the estimated slope and intercept parameters.[18] The resulting exploration of the posterior density of the model parameters is quite efficient with respect to the key parameters, the (identified) ideal points $\tilde{\xi}_i$, and bill parameters $\tilde{Y}_j$ and $\tilde{N}_j$. The usual convergence diagnostics suggest the MCMC algorithm deployed by JAGS may be stopped after about 10,000 iterations.

The resulting posterior densities for the ideal points of selected legislators are represented in Figure 4. For legislators with considerably high levels of missing data (say, Obama and Biden), the posterior densities are pushed up against the edge of the parameter space, as are the densities for legislators with relatively one-sided voting histories, such as Sanders or Demint (R-SC). These posterior densities may look odd, but this presentation is merely a consequence of the particular identifying restriction in NOMINATE; IDEAL or any other model would produce similar-looking posterior densities for the ideal points of extremist legislators if it employed the normalization of –1 to 1 used by NOMINATE.

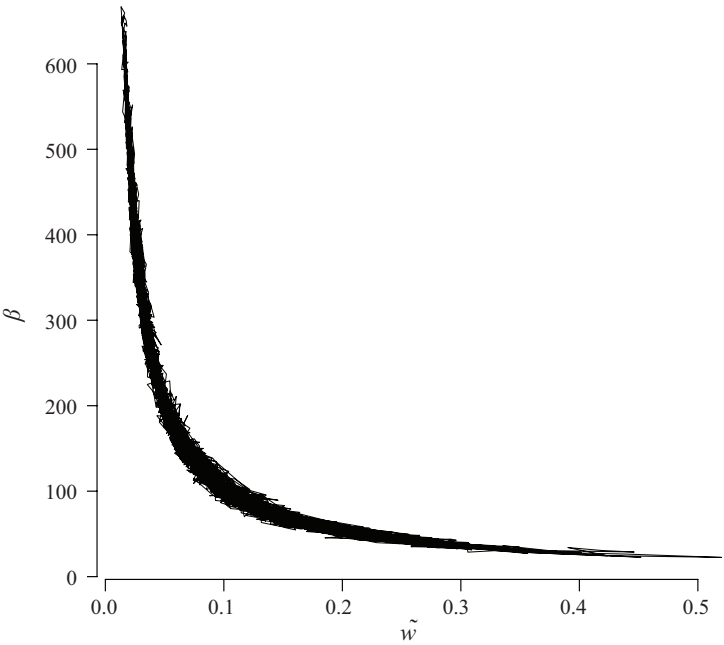
The analyst must decide how to summarize the posterior densities generated by this Bayesian implementation of NOMINATE. Many of these marginal posterior densities are not normal, and so the mean and standard deviation are not sufficient summary statistics. What kind of point estimate makes sense? Should the analyst report the posterior mean or the maximum *a posteriori* value, or simply provide the marginal

FIGURE 4
Marginal Posterior Densities for Legislative Ideal Points
Bayesian Implementation of W-NOMINATE



Note: The histograms summarize 10,000 draws from the marginal posterior density of the ideal point of the indicated Senator. The dotted line is the *Unif* [–1,1] prior for each ideal point.

FIGURE 5
Two-Dimensional Trace Plot of Samples
from the Joint Posterior Density for $\tilde{w}$ and β,
based on 150,000 MCMC Iterates, NOMINATE Model Fit
to 99 *National Journal* Key Votes of 2007



histograms, as we have done here? Providing a point estimate and a 95% highest probability density interval might be the best option, with the asymmetry in the HPD interval a signal to the alert reader that the corresponding posterior density is skewed, perhaps considerably so.

We also note that the MCMC algorithm performs *extremely* inefficiently with respect to the scaling parameters β and $\tilde{w}$ in NOMINATE. Even after 150,000 iterations of the sampler, we obtained the output depicted in Figure 5. The MCMC algorithm generates a very slow exploration of the joint posterior density for these parameters, with the joint posterior density of these two parameters having a distinct "banana" shape; a large slice of the parameter space has high posterior probability. There is not much information about these parameters in these data and one of the two scaling parameters could be redundant.

If we only looked at the output of the NOMINATE algorithm and the point estimates for w and β, then we would not notice that

these parameters are barely identified. Our Bayesian analysis leads us to explore the posterior density for these and other parameters in the model. With the flat priors we adopt here for all parameters, our use of MCMC methods generates a (random) exploration of the likelihood, revealing the dependency between these two parameters in the NOMINATE model.

7. Extending IDEAL: Alternative Models

A relative advantage of the item-response model underlying the IDEAL estimator over the class of NOMINATE estimators is that the IRT model can be more easily altered to account for alternative data—often in ways that are not easily possible using NOMINATE. This flexibility underlies many applications of the model in the literature, spanning many subfields of political science. Our examples illustrate that when comparing IDEAL and NOMINATE, another relevant point to consider is the relative flexibility of the model underlying IDEAL and the relative ease of estimating alternative models. Although we recognize that scholars are, thankfully, never restricted to a single estimator, we suggest that familiarity with the class of IRT models will likely yield greater payoffs than familiarity with NOMINATE, because of the flexibility of the class of models.

For example, Quinn (2004) has discussed the extension to nonbinary, including continuous, data, and variants of the model have been used to measure such diverse topics as the level of democracy in a country (Jackman and Treier 2008), the ideology of U.S. congressional districts (Levendusky, Pope, and Jackman 2008), voting behavior when abstention is an informative choice (Voeten 2000), the ideology of political moderates in the U.S. electorate (Hillygus and Treier 2009), and the ideology of federal agencies in the United States (Clinton and Lewis 2008).

Another extension involves the estimation of "large n, small m" data containing missing observations (that is, nonrectangular data sets). Recently, to assess congruence of preferences, several surveys queried citizens or federal executives about a select set of issues that have come before Congress. The resulting datasets have often contained a large number of respondents (who are treated as "legislators") and relatively few questions (which function as "roll calls"). In addition, many respondents have been unable or unwilling to answer all of the questions they were asked. Estimating data with these properties in NOMINATE can be problematic. As a result, scholars have used the item-response model underlying IDEAL to examine spatial voting in the

2004 U.S. presidential election (Jessee 2009), the congruence of citizen and legislator preferences (Bafumi and Herron 2009), and the relationship of federal agencies to Congress (Bertelli et al. 2009).

A related elaboration appears in Clinton's work with Lapinski (2006), which featured a multirater measurement model to assess legislative significance in a vein similar to Mayhew's research (1991). Because the dataset contained significant amounts of missing data (not every chronicler of congressional activity summarized every period of congressional activity), it was impossible for Clinton and Lapinksi to generate estimates using NOMINATE without substantially altering the NOMINATE code. Moreover, one could incorporate additional statute-level information that was both available and likely correlated with the significance of the enacted statutes if one used a hierarchical IRT, but such extension would present considerable difficulties for NOMINATE.

Finally, it is possible to impose informative priors on the item parameters to structurally estimate a game form. Clinton and Meirowitz (2004) have done so with their analysis of the Compromise of 1790. Treier and Pope (2009) did so to assess the enactment of the Great Compromise in the Constitutional Convention of 1787, as did Jeong (2008) for the contemporary U.S. Congress. Structurally estimating game forms and imposing the required constraints is relatively straightforward if one uses the item-response model, but it is much more difficult to use NOMINATE without customizing the program for the precise application of interest.

While we do not claim that NOMINATE cannot accommodate these data structures and incorporate additional information, we would argue that making the required adjustments to the underlying statistical measurement model is much easier in the item-response framework—as is evidenced by the growing quantity of scholarship making use of Bayesian IRT models to analyze such problems. In the absence of compelling evidence regarding the relative desirability of the two models on statistical grounds, we consider the flexibility of the item-response framework to be an advantage of IDEAL over NOMINATE.

8. Conclusion

As Carroll et. al and we have made clear, there are many similarities between NOMINATE and IDEAL. There is one difference, however, that we believe is rather consequential and worth highlighting: the Bayesian approach implemented in IDEAL is more flexible than NOMINATE and can be adjusted to take advantage of additional data and theory. The NOMINATE family of algorithms was designed to

solve a particular problem using hardware from "a time of less-powerful computing." Large (if not massive) roll-call matrices were supplied as input, and point estimates of legislative ideal points were output, determined by slightly idiosyncratic operationalization of the Euclidean spatial voting model in NOMINATE (that is, scaled Gaussian utility functions). In contrast, when developing the Bayesian approach, we began with the observation that if we assume quadratic utilities (an assumption widely used in other social-science applications), then the Euclidean voting model can be represented as a two-parameter item-response model. Our choice to use this model was quite deliberate; it opens up the possibility of exploiting some of the many developments in the Bayesian analysis of IRT models in psychometrics and statistics (see, for example, van der Linden and Hambleton 1997).

Viewed from this perspective, the NOMINATE algorithms are something of a "closed shop," with extensions or modifications of the scaled-Gaussian/logistic-voting model all but impossible for anyone other than the small set of individuals familiar with the NOMINATE source code. On the other hand, NOMINATE and IDEAL provide near-identical answers for the class of problems for which it is possible to run both NOMINATE and IDEAL, and this correlation is heartening. We believe the fact that the Bayesian approach can accommodate a whole class of problems for which a NOMINATE equivalent does not (yet?) exist is an important advantage to consider when choosing which class of measurement models to use.

Joshua D. Clinton <josh.clinton@vanderbilt.edu> *is Associate Professor of Political Science and Co-Director of the Center for the Study of Democratic Institutions at Vanderbilt University, 301 Calhoun Hall, Box 1817 Station B, Nashville, TN 37235. Simon Jackman* <jackman@stanford.edu> *is Professor of Political Science, Encina Hall, Stanford University, Stanford, CA 94305. Jackman's work on this article was completed while he was a Visiting Professor at the United States Studies Centre, University of Sydney.*

NOTES

1. Indeed, we thank Doug Rivers, Jeff Lewis, and Keith Poole—as well as Andrew Martin and Kevin Quinn—for what they have all taught us about the analysis of roll-call data.

2. Rosas and Shomer (2008) critique the assumption that abstentions are missing at random. Carrubba, Gabel, and Hug (2008) address the possibility of a strategic agenda.

3. As an aside, the inability to form out-of-sample predictions is an under-appreciated feature of roll-call analysis: that is, statistical operationalizations of spatial voting models do not yield out-of-sample predictions. Making predictions for voting on a new roll call requires knowing the location of the yea and nay positions for that roll call; making predictions for a new legislator requires knowing the legislator's ideal point.

4. Experiments may be a better way to proceed here, since some of these parameters can be fixed, or considered fixed. Exploring how decision makers evaluate competing options may provide us with some evidence of the shape of the underlying utility function (see, for example, Grosser and Morton 2009).

5. We are grateful to Carroll et al. for highlighting this facet of NOMINATE in such a detailed fashion. Their discussion has stimulated our thinking about how we might introduce some resistance to outliers into the Bayesian model. One possibility would be to replace the normal density with a t_v density, where v is an unknown degrees-of-freedom parameter. The t_v density is a scale-mixture of normals, with the scale (variance) parameters of the normals following an inverse-gamma density; the normal model is a limiting case, as $v \rightarrow \infty$. If we suspect scale heterogeneity in the stochastic component of the legislator utility functions—say, either across legislators or across roll calls, and these are surely plausible hypotheses—but we lack firm beliefs as to the nature or source of that heterogeneity, then the t model is a simple, yet parametric way for us to tackle this possibility. We have not yet implemented this model, nor are we sure that it will yield results that differ markedly from some extant extensions of the IRT model (e.g., Bafumi et al. 2005) or that the data will convey much information about the degrees-of-freedom parameter, v (cf. Albert and Chib 1993).

6. See Poole and Rosenthal (1997, 237), Poole (2005, 110), or the clear exposition in Carroll et al. (2009).

7. For example, one may wish to incorporate a prior belief that the ideal points of legislators a and b are likely to be similar because the legislators represent districts with similar demographic or political characteristics, or that past roll-call voting conveys information about the ideal point underlying contemporaneous roll-call voting.

8. One might incorporate, for example, a prior belief that the parameters of roll calls j and k are similar because the roll calls deal with similar subject matter, have similar sets of sponsors, or something along these lines.

9. Indeed, the Markov chain Monte Carlo algorithms underlying IDEAL are only feasible on desktop computers with hardware available from the early 2000s or more recently. Poole and Rosenthal were analyzing large roll matrices for some time prior to this!

10. The key factor is that the prior variance for the numerator, $\tilde{\alpha}_j$, equals the prior variance for the denominator, $\tilde{\beta}_j$. We could set both variances to small values consistent with precise beliefs about $\tilde{\alpha}_j$ and $\tilde{\beta}_j$ and still obtain the Cauchy-like prior over τ_j.

11. One approach is to express prior variances for $\tilde{\alpha}_j$ and $\tilde{\beta}_j$ such that the resulting implied prior for τ_j looks reasonable. For instance, if $\tilde{\alpha}_j \sim N(0, 1)$ and $\tilde{\beta}_j \sim N(0, 10^2)$, then the prior 95% credible interval for the bill midpoints $\tau_j = \tilde{\alpha}_j / \tilde{\beta}_j$ is about [–1.3, 1.3], which, if anything, is too restrictive, since ideal points are normalized to have a mean of 0 and a standard deviation of 1 across legislators. When we consider that $\pi_{ij} = \mathrm{I}(\xi_i = 0) = \Phi(\tilde{\alpha}_j)$ is the probability that a legislator located at the mean of the distribution of ideal points votes for proposal *j,* we see that the prior $\tilde{\alpha}_j \sim N(0, 1)$ induces a uniform density on $\pi_{ij} = \mathrm{I}(\xi_i = 0)$, which would not seem unreasonable but appears to generate a too restrictive prior on τ_j. Holding the prior variance of $\tilde{\alpha}_j$ constant at 1.0, we can generate a more permissive prior on by *decreasing* the prior variance of $\tilde{\beta}_j$. For example, with $\tilde{\alpha}_j \sim N(0, 1)$ and $\tilde{\beta}_j \sim N(0, 6^2)$, a prior 95% bound for the bill midpoints, τ_j, would be about [2.1, 2.1], which would span most legislators. Assuming a N(0,1) density across legislators (which holds, at least *a priori*), we would conclude that a 95% bound on τ_j ranges from the 1.8 to the 98.2 percentiles of the distribution of ideal points.

12. We might denote such a hierarchical model generically as $\tau \sim p_\tau(\theta_\tau)$, where p_τ is a density indexed by parameter(s) θ_τ. A simple, specific example might be $\tau_j \sim N(\mu_\tau, \omega_\tau^2)$ where μ_τ and ω_τ^2 are hyperparameters with their own prior densities.

13. Strong statements to this effect appear in Gelman et al. 2004, Gelman and Hill 2007, and Jackman 2009.

14. That is, the roll calls are *exchangeable* and therefore we can treat them as if they were generated by a common stochastic process indexed by hyperparameters, and so forth.

15. For each set of priors, we ran the data-augmented Gibbs sampler in IDEAL for 1.1 million iterations, discarding the first 100,000 iterations and saving every one-hundredth iteration of the remaining one million iterations. The sampler was initialized with all Democratic senators at –1 and all Republicans at 1. We generated start values for the bill parameters by running a probit of each roll call on these initial values for the ideal points.

16. If we transform from unidentified ideal points, ξ_i, to identified parameters, $\tilde{\xi}_i = (\xi_i = c)/m$, then we have $\tilde{\gamma}_1 - \gamma_1/m$ (and similarly for γ_2), $\tilde{\gamma}_0 = (\gamma_0 - c)/m$, and $\tilde{\omega} = \omega/m$.

17. The unidentified version of the NOMINATE model includes likelihood contributions of the sort $w(\xi_i - Y_j)^2$ and $w(\xi_i - N_j)^2$. We transformed ξ_i to the identified parameters $\tilde{\xi}_i = (\xi_i + c)m$, where c and m are solutions to the equations $\min(\xi_i) + c = -1/m$ and $\max(\xi_i) + c = 1/m$. The transformation from ξ_i to $\tilde{\xi}_i$ implies the transformations from Y_j to $\tilde{Y}_j = (Y_j + c)m$ (with a similar transformation for N_j) and $\tilde{w} = w/m^2$. That is, $w(\xi_i - Y_j)^2 = \tilde{w}(\tilde{\xi}_i - \tilde{Y}_j)^2$, with a similar transformation for the likelihood contribution involving N_j.

18. The logistic regression of $\boldsymbol{y}_j$ on $\hat{\boldsymbol{\xi}}$ yields an intercept $\hat{\alpha}_{j0}$ and a slope $\hat{\alpha}_{j1}$, $j = 1, \ldots, m$. We then generate start values $Y_j^{(0)} = -\hat{\alpha}_{j0}/\hat{\alpha}_{j1} + \hat{\alpha}_{j1}/4$ and $N_j^{(0)} = -\hat{\alpha}_{j0}/\hat{\alpha}_{j1} - \hat{\alpha}_{j1}/4$.

REFERENCES

Albert, James H., and Siddhartha Chib. 1993. "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association* 88: 669–79.

Bafumi, Joseph, and Michael Herron. 2009. "Preference Aggregation, Representation, and Elected American Political Institutions." Working Paper. Dartmouth College.

Bafumi, Joseph, Andrew Gelman, David K. Park, and Noah Kaplan. 2005. "Practical Issues in Implementing and Understanding Bayesian Ideal Point Estimation." *Political Analysis* 13: 171–87.

Bertelli, Anthony, Joshua D. Clinton, Christian Grose, David Lewis, and David Nixon. 2009. "Situating Agencies and Congress: The Ideology of Federal Executives and their Agencies." Working Paper. Vanderbilt University.

Carroll, Royce, James Lo, Jeffrey Lewis, Keith Poole, and Howard Rosenthal. 2009. "Comparing NOMINATE and IDEAL: Points of Difference and Monte Carlo Tests." *Legislative Studies Quarterly*. 34: 555–91.

Carrubba, Clifford, Matthew Gabel, and Simon Hug. 2008. "Legislative Voting Behavior, Seen and Unseen: A Theory of Roll-call Vote Selection." *Legislative Studies Quarterly* 33: 543–72.

Clinton, Joshua D., and John S. Lapinski. 2006. "Measuring Legislative Accomplishment, 1877–1994." *American Journal of Political Science* 50: 232–49.

Clinton, Joshua D., and David E. Lewis. 2008. "Expert Opinion, Agency Characteristics, and Agency Preferences." *Political Analysis* 16: 3–20.

Clinton, Joshua D., and Adam Meirowitz. 2004. "Testing Accounts of Legislative Strategic Voting: The Compromise of 1790." *American Journal of Political Science* 48: 675–89.

Clinton, Joshua D., Simon Jackman, and Douglas Rivers. 2004a. "'The Most Liberal Senator'?: Analyzing and Interpreting Congressional Roll Calls." *PS: Political Science and Politics* 37: 805–11.

Clinton, Joshua D., Simon Jackman, and Douglas Rivers. 2004b. "The Statistical Analysis of Roll Call Data." *American Political Science Review* 98: 355–70.

Enelow, James M., and Melvin J. Hinich. 1984. *The Spatial Theory of Voting: An Introduction*. New York: Cambridge University Press.

Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 2004. *Bayesian Data Analysis*. 2d ed. Boca Raton, FL: Chapman and Hall.

Grosser, Jens, and Becky Morton. 2009. "NOMINATE in the Laboratory." Working Paper. Florida State University.

Hillygus, Sunshine, and Shawn Treier. 2009. "Front and Center? The Policy Attitudes of Ideological Moderates." Working Paper. Harvard University.

Jackman, Simon. 2009. *Bayesian Analysis for the Social Sciences*. Hoboken, NJ: Wiley.

Jackman, Simon, and Shawn Treier. 2008. "Democracy as a Latent Variable." *American Journal of Political Science* 52: 201–17.

Jeong, Gyung-Ho. 2008. "Testing the Predictions of the Multidimensional Spatial Voting Model with Roll Call Data." *Political Analysis* 16: 179–96.

Jessee, Stephen. 2009. "Spatial Voting in the 2004 Presidential Election." *American Political Science Review* 103: 59–81.

Levendusky, Matthew, Jeremy Pope, and Simon Jackman. 2008. "Measuring District Preferences with Implications for the Study of U.S. Elections." *Journal of Politics* 70: 736–53.

Lewis, Jeffrey B., and Keith T. Poole. 2004. "Measuring Bias and Uncertainty in Ideal Point Estimates via the Parametric Bootstrap." *Political Analysis* 12: 105–27.

Mayhew, David R. 1991. *Divided We Govern: Policy Control, Lawmaking, and Investigations, 1946–1990*. New Haven, CT: Yale University Press.

Plummer, Martyn. 2009. *rjags: Bayesian graphical models using MCMC*. R package version 1.0.3-8. http://mcmc-jags.sourceforge.net (accessed April 15, 2009).

Poole, Keith T. 2005. *Spatial Models of Parliamentary Voting*. New York: Cambridge University Press.

Poole, Keith T., and Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll Call Voting*. New York: Oxford University Press.

Poole, Keith, Jeffrey Lewis, James Lo, and Royce Carroll. 2007. *W-NOMINATE: W-NOMINATE Roll Call Analysis Software.* R package version 0.93. http:/cran.r-project.org/web/packages/wnominate/index.html (accessed April 15, 2009).

Quinn, Kevin M. 2004. "Bayesian Factor Analysis for Mixed Ordinal and Continuous Responses." *Political Analysis* 12: 338–53.

Rivers, Douglas. 2003. "Identification of Multidimensional Item-Response Models." Working Paper. Stanford University.

Rosas, Guillermo, and Yael Shomer. 2008. "Models of Nonresponse in Legislative Politics." *Legislative Studies Quarterly* 33: 572–602.

Rothenberg, Thomas J. 1971. "Identification in Parametric Models." *Econometrica* 39: 577–91.

Spiegelhalter, David J., Andrew Thomas, Nicky Best, and Dave Lunn. 2003. *WinBUGS User Manual Version 1.4*. Cambridge, UK: MRC Biostatistics Unit.

Stein, C. 1956. "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution." In *Proceedings of the Third Berkeley Symposium of Mathematical Statistics and Probability*. Vol. 1 Berkeley, CA: University of California Press.

Treier, Shawn, and Jeremy Pope. 2009. "Reconsidering the Great Compromise at the Federal Convention of 1787: Deliberation and Agenda Effects on the Senate and Slavery." Working Paper. University of Minnesota.

van der Linden, Wim J., and Ronald K. Hambleton. 1997. *Handbook of Modern Item Response Theory*. New York: Springer.

Voeten, Eric. 2000. "Clashes in the Assembly." *International Organization* 54: 185–215.