Engaging Experts:
Dealing with Divergent Elicited Priors in Political Science

Sarah B. Bouchat

15 October 2016

Not For Circulation or Citation

1 Introduction & Motivation

Priors over parameters constitute one of the most visible manifestations of the theoretical distinctions between Bayesian and frequentist statistics. The information contained within the specified prior, however, will depend upon the knowledge available to a given researcher. An elicited prior seeks to leverage the substantive knowledge of area experts, whether through interviews or published research, in order to improve the accuracy of posterior estimates. Elicited priors as a concept have been detailed since the early work of Savage (1971), Shuford and Brown (1975), Kadane and Wolfson (1998), and Garthwaite, Kadane, and O'Hagan (2005). Formally, Gill and Walker (2005, 841) discusses the elicited prior "as a means of drawing information from subject-area experts with the goal of constructing a probability structure that reflects their specific qualitative knowledge, and perhaps experiential intuition, about the studied effects." These aims and applications for elicited priors seem both reasonable and desirable, yet the 10 years since the publication of Gill and Walker's article have seen few instances of researchers, especially in political science, adopting elicited priors as a way to incorporate qualitative knowledge into quantitative research.

This paper argues that the limited application of elicited priors is due to both an unclear prescription for implementing the method and a limited vision for its application. The following sections seek to explore the range of studies from other disciplines that use elicited priors in order to understand both how they are used in practice and how they might be better used to realize the vision of qualitative and quantitative synergy that Gill and Walker present. An empirical section also outlines a specific process for undertaking work using elicited priors. This is illustrated with reference to applications that expand Gill and Walker's scope to encompass data-poor and authoritarian contexts in particular, where the limitations on quantitative data especially benefit from an elicited priors approach.

The concept of elicited priors shows particular promise in combining qualitative and quantitative analytical approaches. A typical modeling approach would rely only on the researcher's assessment of the appropriate array of variables, model type and structure, and method. Eliciting priors from those with expertise in an area offers an opportunity to gather more input for each of these choices from those who have substantive knowledge but are not directly related in the process of analysis. This approach has the added benefit, therefore, of maintaining transparency about assumptions in the modeling process. An elicited prior requires documentation of at least the qualifications of the source, whether in-person or published, and can therefore ease the process of replication and progression in similar veins of research.

Despite these advantages, the current work on elicited priors lacks guidance on how to elicit said priors, how to reconcile diverging beliefs, and under what circumstances or for what purposes they would be most useful. These omissions limit the use of elicited priors in the social sciences, and especially in political science. In particular, while the prospect of eliciting priors from experts with both experiential and scholarly insights adds to the appeal of the method, it is exactly the process of identifying "experts," eliciting their views in ways that accurately represent their assessments while being useful in the modeling process, and reconciling the diverging opinions of experts that makes elicited priors difficult to apply. This is especially true in authoritarian and poor data contexts, where the prior is simultaneously of greater importance to the modeling process and more contentious to construct as a result of the data-poor environment. For example, when soliciting input from government-sponsored and opposition-sponsored experts in authoritarian settings, there are no guidelines inherent to using elicited priors as presented by Gill et al. that would allow a researcher to adjudicate between strongly diverging opinions or assess their credibility. Likewise, no guidance is given regarding precisely *how* to elicit a prior—a process that could greatly influence outcomes.

This paper seeks to address these omissions and suggest ways in which elicited priors could be usefully applied, particularly in studies in authoritarian or data-poor contexts. The following section culls techniques from the literature using elicited priors to create suggestions for their broader application in social science. An empirical section follows, discussing strategies for implementing elicited priors and their implications in a situation with simulated data. Finally, the conclusions section will offer some next steps for refining these approaches and for broader application of elicited priors to a wider array of social scientific problems.

1.1 Priors in Bayesian Analysis

Often, researchers seeking to remain agnostic in the statistical analysis employ relatively diffuse priors. These "uninformative" priors are, however, literally informative at least two senses. First, they directly influence the conclusions to be drawn from the modeling process. This is especially true in cases of low data quality or quantity: diffuse priors limit the ability of the researcher to leverage the model structure in order to draw inferences about empirical phenomena. As Andrew Gelman writes, "[with] well-identified parameters and large sample sizes, reasonable choices of prior distributions will have minor effects on posterior inferences. ... If the sample size is small, or available data provide only indirect information about the parameters of interest, the prior distribution becomes more important" (Gelman 2002, 1634). While other approaches (e.g., clustering and estimating hierarchical models) often seek to address these issues, these statements still justify overarching concern with specifying reasonable priors where possible, with special consideration for instances with poor data. Specifying noninformative priors, furthermore, threatens the trajectory of scientific inquiry as studies build on each other, as Gill and Witko note: "It is also important to observe that the overwhelming proportion of prior distributions specified in published Bayesian social science work still avoids using reasonably informed priors, which unfortunately hurts the steady accumulation and progression of scientific knowledge" (Gill and Witko 2013, 462).

Second, these priors represent a strong positive claim that no useable information about a given θ exists with which to specify a more appropriate or precise prior. More generally, this approach reflects a normative position that favors the omission rather than the careful specification of assumptions throughout the modeling process.

An alternative approach would instead seek to take advantage of the significant knowledge accumulated across fields of study in order to select appropriate, informed priors. From the perspective of the researcher, this approach could represent an alternative agnosticism that does not require reliance on own knowledge or opinions, but rather documents the acquisition of information from expert sources. This notion effectively underlies the method of “eliciting” priors: whether through interviews with area experts or reference to published works, a researcher can develop appropriate priors for unknown θ that improve the prospects for modeling in a Bayesian framework.

2 Literature Review: Elicited Priors across Disciplines

The notion of elicited priors is not new, yet scholarship examining or employing the technique remains limited. Searching for “elicited prior” in the Web of Science database yields fewer than 50 relevant articles, most of which have substantive foci in biology or psychology rather than political science. In defining the origins of elicited priors, Gill and Walker (2005) reference previous terms such as “community of priors”—which incorporate opinions of both affirming and skeptical experts—and delineate elicited priors into a variety of types, including clinical priors, skeptical and enthusiastic priors, reference priors, etc. (843). Gill and Walker (2005, 844) summarize the three phases of research using elicited priors explained by Spetzler and Staël von Holstein (1975) as the deterministic, probabilistic, and informational phases. The deterministic phase, encompassing variable and expert selection, has costs, but is generally perceived as less challenging than the probabilistic phase, during which priors are actually elicited. Aside from determining data sources, the primary question addressed in the first stage is instead from how many experts one should elicit priors (Gill and Walker 2005, 844).

The informational phase follows, in which elicited responses are tested, evaluated, and scaled for consistency. Gill and Walker (2005) concur with Spetzler and Staël von Holstein (1975) that the methods of eliciting priors (“p-methods,” “v-methods,” “pv-methods,” etc.) present the greatest challenge, but concerns about how best to elicit priors from experts do not explain the very limited uptake of elicited prior methods overall in the social sciences. Rather, I suggest that more careful attention to the selection of “experts” and the calibration of responses opens possibilities for applying this method to new areas of research, specifically in authoritarian and data-poor settings. For example, rather than focusing on the purely statistical problem of how *many* experts should be selected, emphasizing the question of *who* is considered an expert and in what way opens new possibilities for identifying expertise and knowledge from which to generate priors.

2.1 Theoretical Contributions to Elicited Priors

While the use of elicited priors in scientific fields far surpasses its use in social science domains, its application remains somewhat limited by remaining skepticism of the "subjectivity" of the Bayesian approach more generally (Lele and Das 2000; Wang and Zhou 2009). Elicited priors, rather than overly diffuse or benchmark "objective" priors, have the potential to provide leverage in low-data circumstances and challenging modeling contexts, but also to substantively impact the quality of estimation in a positive way. In fact, Datta and Ghosh (1991) suggest that the elicited prior represents the "true" prior—presumably that, if sufficient expertise were available and applied, the resulting prior would be accurate. Still, problems remain with processes for engaging even trained individuals in statistical assessments. As Lin, Lin, and Raghubir (2004) note, people are susceptible to biases resulting from self-positivity, controllability of negative events, and especially order of elicitation—where previous questions are used to form assessments for later ones. Likewise, practically speaking, prior elicitation confronts the same issues as typical Bayesian analysis, where, for example, specifying priors for continuous parameters proves difficult since "[doing] so would require infinitely many prior probability judgments" (Dey and Birmiwal 1994).

In a sense, the lack of identifiable expertise in low-data settings, rather than giving rise to the implementation of diffuse priors, should give rise to a concerted effort to model our ignorance—a task Zaffalon (2005) identifies as a persistent challenge in Bayesian analysis. Accounting for prior ignorance and incomplete data, as Zaffalon explains, means that the true model of ignorance requires accounting for "all possible states of knowledge" (1005). One of the key related areas of inquiry in elicited priors is how to handle uncertainty or error within the prior itself. In general, studies posit an elicited prior π_0 in which there may be some "ϵ contamination" or error (Sivaganesan 1993).

Among these scientific papers, however, opinions about how best to elicit priors vary. Lele and Das (2000), for example, argue in favor of directly soliciting guess values:

> It is important to recognize that the concept of a prior probability distribution on the parameters of a statistical model is a statistical construct that is hard for most scientists to visualize. It is more natural for an expert to think in terms of the process

> under study and not in terms of statistical distributions over a parameter space. The sensible approach, then is to ask the expert to provide guess values for observable data, not a prior probability distribution. (Lele and Das 2000, 466)

In Lele and Das (2000)'s case, this suggests a hierarchical modeling strategy to combine experts' insights with the observed data under study.

The problem confronted by Lele and Das (2000)—sparse data in a spatial context but the potential for "soft" data in the form of expert opinion—mirrors the difficulty of conducting analyses in authoritarian or low-data situations. Lele and Das (2000) provide an analytical framework for addressing precisely this problem of "misleading" expert priors in their paper examining spatial hierarchical models with elicited priors. In particular, the authors want to account for the dependence between observed data values and elicited prior values when experts are asked for value estimates rather than providing complete prior distributions (468). "This dependence reflects the credibility of the expert," they argue, while noting that "[in] this way, even a misleading expert opinion can be informative" because "if we find that data elicited from an expert are negatively correlated with real data, this is useful information that can be used to suitably adjust our inferences" (468). To do this, Lele and Das imagine that an expert gives an opinion $\bar{E}$ about $\bar{Y}$ drawn from the distribution $f(\bar{y};\theta)$ where the θ parameters are unknown. This opinion $\bar{E}$ is itself distributed $g(\bar{e} \mid y;\eta)$, with η indicating the dependence between $\bar{Y}$ and $\bar{E}$. This leads Lele and Das to call η an "honesty parameter," measuring the "credibility" of an expert. In principle, however, such a parameter captures both the ability of the expert to think and express opinions statistically as well as the potential strategic misrepresentation of information and expertise.

2.2 Empirical Applications of Elicited Priors

A variety of disciplines, although primarily within the natural and biological sciences, have utilized elicited priors techniques for research. These studies offer a multitude of insights about how to implement an elicited priors approach, but also highlight the ways in which current practice in other disciplines is not currently well-suited to the study of authoritarian or low-data contexts.

Morris et al. (2013), for example, investigate the effects of vesicoamniotic shunts on lower urinary tract issues for fetuses where they have only 31 female subjects with singleton pregnancies and leverage elicited priors from 52 pediatric nephrologists, pediatric urologists, and fetal medicine specialists (Morris et al. 2013, 1500). The experts polled in the study largely agreed on the potential effects of treatment, although the authors note that it is "problematic" that these experts' opinions did not align with what is current common clinical practice. Their responses were pooled and averaged to create a prior distribution (1501). This study provides one of several examples in which expert opinions are equally weighted in the implementation of the prior. For social science purposes, however, two issues remain:

(1) How should expert opinions be weighted or aggregated if disagreement occurs?

(2) How should expertise coming from practical experience be reconciled, either with differing experience or with educated or credentialed expertise?

While studies like Morris et al. (2013) illustrate the disjuncture between expert opinion and practice, they do not resolve the concerns facing researchers who wish to use these methods in other social scientific settings.

The procedure for eliciting priors from experts also dialogues with these concerns for how expert priors might be adequately reconciled and combined. As Wheeler et al. (2014) describes, there are many possible ways of "eliciting" prior information. This can be done with previously published work or historical data as well as with interviews of living experts; the primary binding constraint is that the expert has not directly observed the data under current study (678). If the expert has directly observed the data in question, their prior is likely to reproduce these data rather than providing additional information that can aid in inference and analysis. Despite this constraint, there are still many options for how to implement prior elicitation, including eliciting priors about regression coefficients; the distribution of the dependent variable conditioned on fixed values of covariates; quantiles of the predicted dependent variable distribution, etc. (678).

Some of these methods present greater challenges than others in implementation. For example, eliciting beliefs regarding regression coefficients and their variance can prove difficult even for educated experts (Gill and Witko 2013, 462), as providing accurate estimates of statistical uncertainty of one's beliefs is a particular challenge (Albert et al. 2012, 504). A variety of methods also exist for quantile based direct elicitation of priors (Dey and Liu 2007). Regardless, as Wheeler et al. (2014) notes, "the incorporation of such information into the Bayesian modelling framework aligns with the philosophy of the scientific method, where knowledge that is available before collective data (prior) is used along with the observed current data (likelihood) to inform what we know now (posterior)" (Wheeler et al. 2014, 678).

The example offered by Daponte, Kadane, and Wolfson (1997) underscores this point. The authors use elicited priors to project the Iraqi Kurdish population from 1977–1990, and describe three critical advantages to conducting their demographic study in a Bayesian framework. First and foremost, they note that a Bayesian approach can promote communication among demographic researchers: "Making one's beliefs explicit using probability distributions allows other demographers to observe exactly how one views the sources of uncertainty in the phenomenon. Others can then know on what they agree or disagree. The reasons given for particular probability distributions can be an important source of insight" (1256). The authors also note that making these projects explicit in the form of probability distributions enhances their usability in a variety of applications. Finally, they distinguish the Bayesian approach and its advantages from more traditional methods: "Classical models either include or exclude a parameter about which no prior is expressed, which is often equivalent to expressing certainty about its value. Using probability distributions permits one to express states of knowledge in between these two alternatives" (1256). Much like other social scientific applications, as the authors describe, this study of the Iraqi Kurds "lacks high-quality data" and reflects incomplete information—conditions that align it in particular with the study of authoritarian regimes or other low-information environments in political science (1257). Each of these dimensions highlights the critical applicability of an elicited priors approach in an authoritarian or low-data context, and the significance of elicited priors as a source of information to adapt and improve statistical inference. How to integrate the priors offered by experts as part of a study, however, remains unaddressed by this work.

Albert et al. (2012), on the other hand, tackle the issue of how best to aggregate elicited priors directly, proposing a hierarchical modeling approach to account for multiple sources of variation and the potential lack of independence between experts' assessments (503). For illustrative purposes, the authors define a sampling model containing observation $X \sim P_\theta$ where θ is an unknown parameter with prior π that is an "elaboration on a parametric family" such that $\pi \in \{\pi_\gamma, \gamma \in \Gamma\}$. γ is then estimated from the elicited priors (504). It is plausibly the case, as the authors note, that each expert polled provides a different γ, necessitating a procedure to combine these divergent priors. In the authors' review, pooling and averaging predominate as methods for combining these differing priors, where averaging "emphasizes the consensus on elicited quantities" and is advantageous in its "simplicity," but at the same time can "understate variation by ignoring uncertainty" and/or "mis-represent [sic] multiple modes" (Albert et al. 2012, 504). Pooling methods—whether linear or logarithmic—attempt to overcome some of these deficiencies by encompassing all values in a way that can be construed as an additive or multiplicative mixture (504). These divergent priors can then be combined using weights w_ℓ, such as by $\sum_{\ell=1}^{L} w_\ell \pi_{\gamma\ell}$ where ℓ indexes the individual expert whose prior π_γ is being combined with others (504). Devising these weights, however, as the authors report, has primarily been based on "p-values for evaluating how well expert assessments on seed variables align with empirical results" (504). This has the disadvantage, as the authors note, of embracing the "diversity" among the elicited priors without offering a notion of consensus or how individual experts might diverge from that consensus. More directly related to the study of authoritarian regimes and experts, however, this method of weighting expert opinions is intended to resolve discrepancies in experts' ability to offer statistical assessments or measures of uncertainty. In a situation of incomplete information both within the data and among the experts, assessing an individual expert's prior in alignment with the data may less appropriately capture their insights into the underlying data-generating process or motivations of actors captured within the data. Weighting expert opinions in this way, rather than understanding as well their potential incentives to offer misleading information or ability to only offer insights into one part of the "truth," means that existing methods for combining expert opinions are not well suited to use in low-data contexts or the study of authoritarianism in particular.

The method that Albert et al. (2012) propose essentially establishes a hyperprior on the unknown parameter θ. This process treats elicited information as "data" in the construction of a prior for the eventual analysis of interest. The prior probability distribution on θ, given by D_{elicit} gives rise to the following:

$$\pi(\gamma \mid D_{\text{elicit}}) \propto f(D_{\text{elicit}} \mid \gamma)\pi_0(\gamma)$$

The construction of the distribution of $\pi(\gamma \mid D_{\text{elicit}})$ from pooling requires a joint likelihood of expert opinions in order to account for dependence. The resulting model of the expert priors treats prior opinions as being sorted into classes J where members of the same class have the same distribution (Albert et al. 2012, 508). In this the authors acknowledge that smaller groups may represent the divergent opinion of a set of experts who are less represented in the population, for example, or who are less reliable, and the solution they propose is to assume a higher variance parameter for the distribution of a group j with those characteristics. This approach, while including important information or knowledge about the experts into the modeling exercise, relies on assumptions about the validity of the statements made by members of smaller groups of experts, which may not reflect a priori knowledge of experts in authoritarian settings. Likewise, this conception of groups of expert opinions is predicated on an implicit notion that some of these elicitations are accurate whereas others are flawed, rather than a conception that each contains some aspect of the true information.

O'Leary et al. (2009) reinforce that the elicitation method itself should reflect knowledge about the experts who might provide information; that is, it should take into account difficulties in engaging with experts as well as difficulties with experts providing accurate information. "Selection of an elicitation method," they write, "is determined by several issues. These include the expert's knowledge of statistics, their mapping skills, time available, access to experts and funding. The chosen elicitation method should balance the expert's knowledge of statistics and mapping with the output required" (396). These considerations serve as the basis for three different elicitation methods the article discusses, each with respect to a dataset concerning rock wallaby population estimates. The choice of elicitation tool, the type of elicitation (e.g., indirect P-method, direct, etc.), and the distribution form of the β coefficient primarily distinguish between these methods (388). This distinction is useful in highlighting that many methodological choices already exist in terms of adapting an elicted priors approach to a particular research question and context, but none specifically address the problem of how to aggregate priors across experts beyond the method of weighting with p-values for alignment between the elicited prior and the data that were discussed previously.

These problems notwithstanding, a precedent does already exist within political science for engaging "experts" beyond the credentialed class often sought after by scientists. Kendall, Nannicini, and Trebbi (2015) elicit multivariate distributions from voters reflecting their beliefs about incumbents' valence and ideology. This engagement of "experts" from a more general populous can allow the ultimate prior distribution used in the analysis of interest to reflect the diversity of information about the subject. Likewise, Small (2008) acknowledges that "it may be appropriate in many cases to elicit probabilities from different experts for different parts of the model" (1302). Likewise, Bakker (2009) utilizes elicited priors to enable experts to refine political party placements on a left–right spectrum. The broader application of an elicited priors approach to important problems in political science, however, rides on the ability of the approach to be adapted to other contexts and questions. The next section discusses how in particular this approach should be beneficial for the study of authoritarian political processes, and where its current practice falls short in offering clear signs of applicability.

3 Empirical Applications to Authoritarian and Low-Data Contexts

The aforementioned studies often engage expert opinion for priors in order to leverage the greater precision this lends to the analysis. An analogous application in political science relates to questions concerning authoritarian regimes or developing country contexts where low data and data manipulation/falsification are present. Bayesian analysis is often most appropriate for small-n studies, and more informative priors from elicitation can aid in resolving pathological problems arising from small data or complex modeling structures. This area therefore seems like a straightforward one in which to apply an elicited priors approach, but the current literature has several shortcomings that prevent the direct importation of elicited priors into a social science domain—especially where the study of authoritarian regimes or low-data contexts is concerned.

First, while some of the above studies offer more detailed accounts of how to elicit priors for scientific applications, these accounts have not been translated to a social science setting. This would be particularly important when expanding the notion of "expertise." Elicited priors are commonly thought to come from "experts" in a given field. While this may be a sufficient distinction in fields like medicine, where credentials come through schooling and certification, it is less clear for social science applications. When trying to understand regime behavior under authoritarianism, for example, one could conceivably ask professors who study the subject, bureaucrats who work under those conditions, activists who oppose such a system, or average citizens who live under authoritarian rule. Each of these groups has some degree of "expertise" to offer, but each would likely also have varying levels of statistical literacy to align with the protocols developed for eliciting priors in scientific contexts.

A second, and related, problem is the issue of diverging priors. The common procedure in the scientific literature, although not often discussed in detail, is to average elicited priors or to use a "consensus prior" after having elicited information from a series of experts. These aggregation rules intentionally eliminate or deemphasize information that diverges from the most common or popular prior. This may make sense for a panel of credentialed experts in a scientific field: for example, a researcher may be interested in the "state of the field" where the effect of a medication on prognosis is concerned. This rule makes less sense, however, in a circumstance like the one described previously, where the notion of expertise has been expanded to encompass several valid sources of information from a social scientific perspective. Particularly in authoritarian regimes, these different individuals may have widely varying perspectives on the issue at hand. These differing perspectives could be a result of different experiences, different relationships to the ruling regime or party (co-opted or oppositional), or different access to information in what is likely to be an environment rife with incomplete information.

A comprehensive analysis seeking to leverage this expertise nonetheless needs more guidance on how to incorporate and assess these divergent priors. Averaging the priors would discard potentially significant information, and arriving at a "consensus" could prove theoretically problematic. For example, imagine there are two individuals whose priors roughly overlap or correspond, but a third whose prior is significantly distinct. Should the researcher treat the two that correspond as the "consensus" or should the researcher be skeptical that those individuals have reproduced some sort of "party line" that intentionally obscures the underlying statistical relationship, whereas the third individual is telling the "truth"? A more ecumenical approach would seek to incorporate each of these, with the understanding that each potentially contains some part of an overall "truth."

Particularly when discussing social and political phenomena, "experts" may have strongly divergent opinions about the relationship between covariates and outcomes that, unlike in some scientific fields, cannot be resolved immediately by appealing to the data. Some of these experts may be more reliable than others in the sense of being able to provide statistical assessments, but in authoritarian contexts in particular, situations may arise in which experts have differing but not apparently false assessments of statistical likelihood as a function of either their desire to selectively provide (or misrepresent) information or their differing perceptions of truth caused by skewed information environments in authoritarian regimes.

In the following sections, I present an approach for addressing these concerns by providing an aggregation tool that can synthesize divergent expert opinions for Bayesian analysis. This approach does not seek to improve on methods for providing internal, statistical consistency of elicited priors, but rather seeks to deal with the divergence in potential elicited priors and the possible propensity for experts in authoritarian contexts to misrepresent the relationship between covariates and outcomes.

3.1 A Dirichlet-Based Method for Elicited Priors

In this section, I propose an approach for resolving the dilemma of prior elicitation and aggregation in social scientific contexts, and especially in studies of authoritarianism. Following this discussion, I provide examples of other uses for Dirichlet approaches as well as an example of an empirical circumstance where the approach could be applied with respect to elicited priors.

The method proposed in this paper adapts a Dirichlet process in order to resolve the issue of aggregating divergent elicited priors before implementing a Bayesian analysis. This Dirichlet-based method allows for multiple possible aggregation rules for elicited priors, enabling each of the priors to be "mixed" in a single distribution, and varying the concentration parameter of the Dirichlet distribution to seek a final prior that emphasizes greater consensus, equally weights all opinions, averages over opinions, etc. As Neal (2000) describes, "[mixtures] with a countably infinite number of components can reasonably be handled in a Bayesian framework by employing a prior distribution for mixing proportions, such as a Dirichlet process, that leads to a few of these components dominating" (249). This flexible framework for establishing prior distributions that carry substantial significance in a low data context is aimed to improve the integration of qualitative knowledge into quantitative assessments of particularly challenging contexts and questions. This approach facilitates a greater number of "new" kinds of experts serving as subjects for prior elicitation. Incorporating these differing kinds of expertise both enables better use of the information and expertise that already exists in authoritarian and other low-data contexts, as well as offers an opportunity for better engagement with qualitative researchers, who have long utilized this type of expertise in their own work.

The Dirichlet distribution is defined as

$$\frac{1}{B(\alpha)\prod_{i=1}^{K} x_i^{\alpha_i - 1}}$$

where α specifies a concentration parameter for the K "clusters" of the distribution. The Dirichlet as a mixture model of mixtures can be thought to play a "partitioning" role for clusters of observations. In the Chinese restaurant process example illustrating a Dirichlet, a customer entering the restaurant can be seated at a table already occupied by some diners, or at a new table. The probability of being seated at a given table corresponds to the distribution of occupants at one table relative to the others.

Applying this process to the elicited priors case, each elicited prior is separated into its constituent components: the proposed value for the dependent variable y as well as the values of independent variables x that corresponded to the y value. These proto "distributions" are treated as data for the purposes of the Dirichlet process. The Dirichlet will dynamically create clusters of observations, essentially finding moments of consensus since the population of these clusters is weighted by the existence of other data to occupy the cluster. Through an updating process to create these clusters, the Dirichlet process will eventually provide posteriors corresponding to each expert from whom a prior was elicited. These posteriors can then be used to create a prior distribution for the eventual analysis of the data of interest.

This framework is flexible, however, in terms of the specification and in terms of how input is weighted. An explicitly symmetric Dirichlet, which can be expressed as:

$$\frac{\Gamma(\alpha K)}{\Gamma(\alpha)^K}\prod_{i=1}^{K} x_i^{\alpha - 1}$$

would denote that all components of α are of equal value (equal weighting). Likewise, how the posterior information from the Dirichlet process is transformed for use as a prior in the data analysis leaves room for defining different aggregation rules: averaging, deemphasizing extreme values, deemphasizing consensus, etc. This process of influencing the aggregation itself reflects an implicit hyperprior by the researcher placed on the priors, and the aggregated prior, from the experts used for elicitation, specifying how much weight to place on different experts' opinions or on different components of the aggregation.

This proposed method of aggregation adds significant value relative to the existing methods of pooling or averaging. In addition to being better equipped to deal with the potentially divergent priors elicited in political contexts relative to medical or biological fields of study, this method provides a formal and transparent way to incorporate the researchers' beliefs about the credibility of a source, and to weigh the credibility of sources against each other and incorporate them into the final prior. This method also has the distinct potential to engage scholars across the qualitative-quantitative divide, as it does not resolve issues of expert selection but instead requires relying on the experience and knowledge of qualitative scholars who have previous familiarity with the case at hand to identify "expert" individuals and to evaluate their credibility.

3.2 Dirichlet Process Example

A common application of Dirichlet processes is in Latent Dirichlet Allocation (LDA). This process provides a model of a corpus, or body of textual works, where each document comprises a random mixture of latent "topics" and topics are distributions of words (Blei, Ng, and Jordan 2003, 996). For the basic LDA model with documents w in corpus D, as described by Blei, Ng, and Jordan, you select $N \sim \mathtt{Poisson}(\xi)$, $\theta \sim \mathtt{Dirichlet}(\alpha)$, and for each N words in w_n, you choose a topic $z_n \sim \mathtt{Multinomial}(\theta)$ and a word w_n from $p(w_n \mid z_n, \beta)$ (996). In this case the mix over latent topics θ is given with a Dirichlet having a concentration parameter of α. The ultimate allocation process seeks to understand how documents, which vary according to their words, can be sorted into a variety of topics. This is analogous to the mixture and allocation process that would occur in the elicited priors application of the Dirichlet process, where latent states correspond to experts' hidden dimensions of consensus. That is, the elicited priors are analogous to documents in the LDA example, which are comprised of components used to elicit them (e.g., across coefficients, according to quantiles, etc.). Experts offering these priors may have been subject to the same kinds of constraints (e.g., career or family concerns) or had access to or limitations from the same kinds of information (e.g., censorship, governing versus governed classes) that would shape the prior they gave. Allocating these into latent classes using the

Dirichlet allows the researcher to examine the groupings present in the elicitations in order to better weight and incorporate the contributions of each individual expert.

4 Dirichlet Allocation Example

One of the key contributions of this method is to allow researchers seeking to utilize expertise from a diverse and perhaps divided set of individuals to be able to aggregate and incorporate that divergent expertise into their prior distributions. The Dirichlet-based approach allows for substantively similar elicited priors to be clustered for more efficient estimation and a more representative final prior. Adequately capturing this diversity within the Dirichlet, however, depends on several parameters that ultimately come from the initial research design, including how many experts to engage and how many questions to ask and/or datapoints to collect for each covariate used in the final analysis.

In order to be able to calibrate the process of selecting a number of experts and how to elicit priors while also optimizing the construction of the aggregated prior, the Dirichlet allocation was moved to C++. The figures that follow illustrate how well the Dirichlet approach identifies clusters among priors as a function of the number of covariates and the number of questions asked. As these figures illustrate, the clustering process performs best when receiving enough, but not too much, information about experts' priors. For example, it performs equally well in returning the true clustering with 20 experts versus 10, as long as there were 10 questions for 10 covariates. Slightly greater uncertainty is introduced, however, when 15 questions are asked for 4 covariates, and even more uncertainty appears when fewer (10) questions are asked for 4 covariates. This uncertainty appears in the figures as gradations of light gray or light turquoise, whereas when two experts share a cluster their cell is turquoise, and when they do not, their cell is dark gray.

For example, in Figure 1, 20 experts are allocated into 5 clusters, where for example cell (1,1) is shaded turquoise because expert 1 shares a cluster with herself. Figures 2 and 3 demonstrate more uncertainty than Figure 1. In Figure 2, much more information is gleaned from the experts than corresponds to the lower number of covariates, leading to too fine-grained information for sorting. Figure 3's configuration performs better with fewer questions per expert.

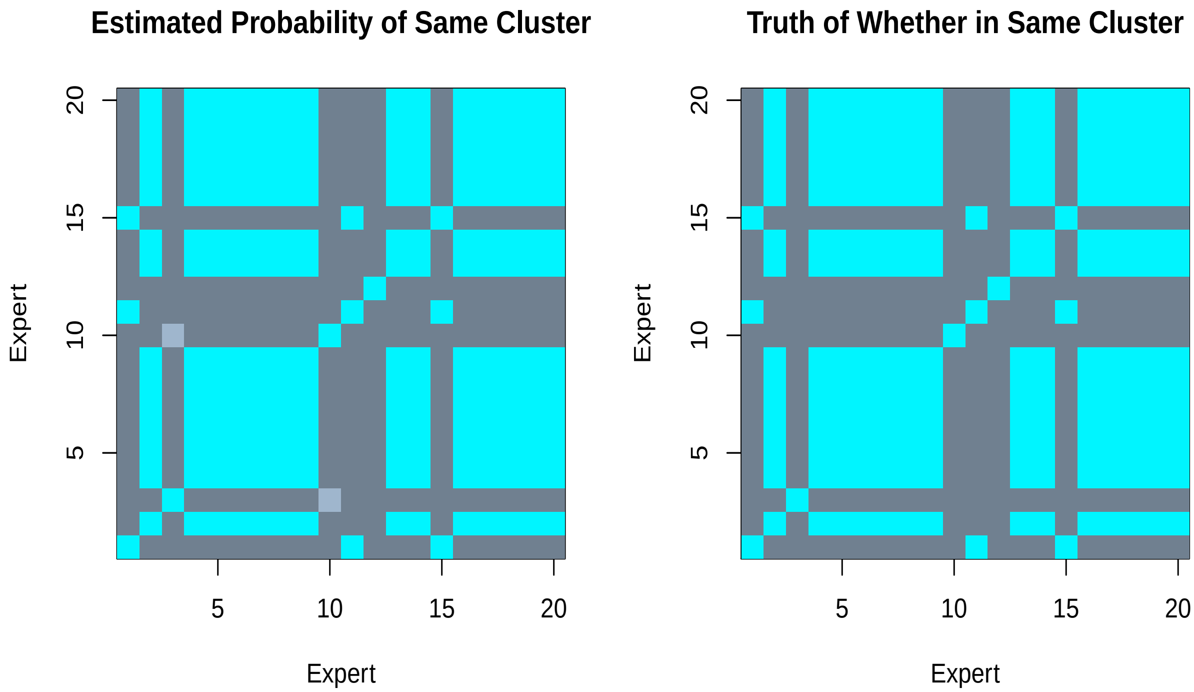

Figure 1: Cluster Allocation with 20 experts, 10 questions, 10 covariates

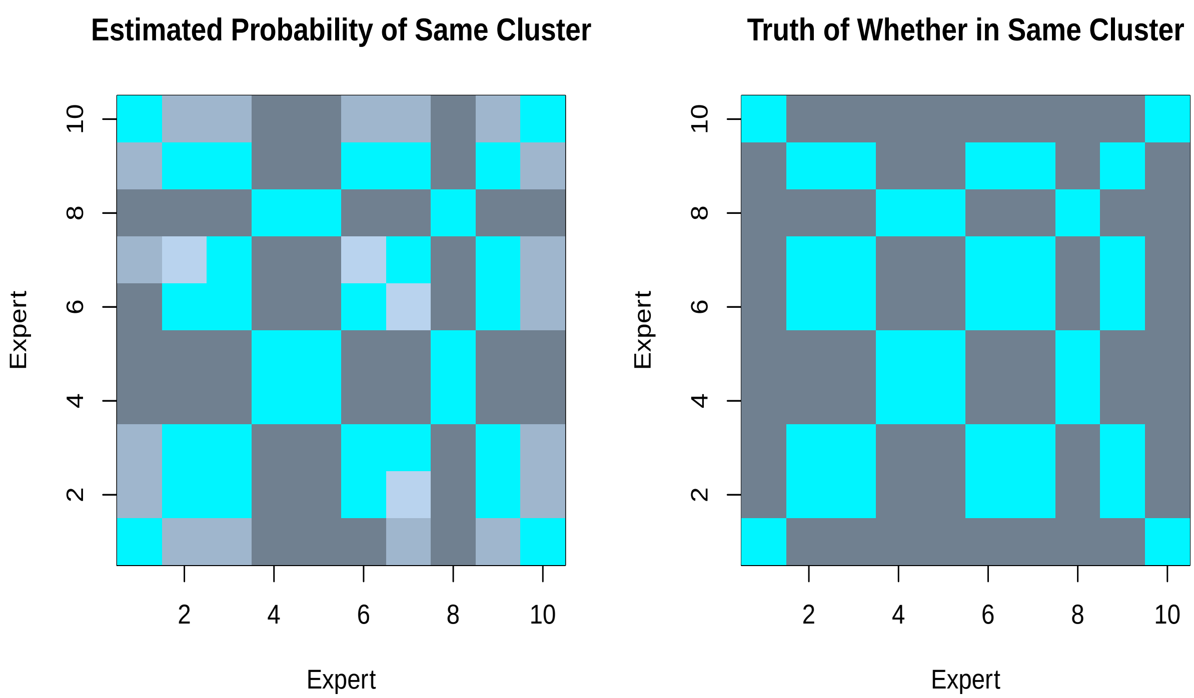

Figure 2: Cluster Allocation with 10 experts, 15 questions, 4 covariates

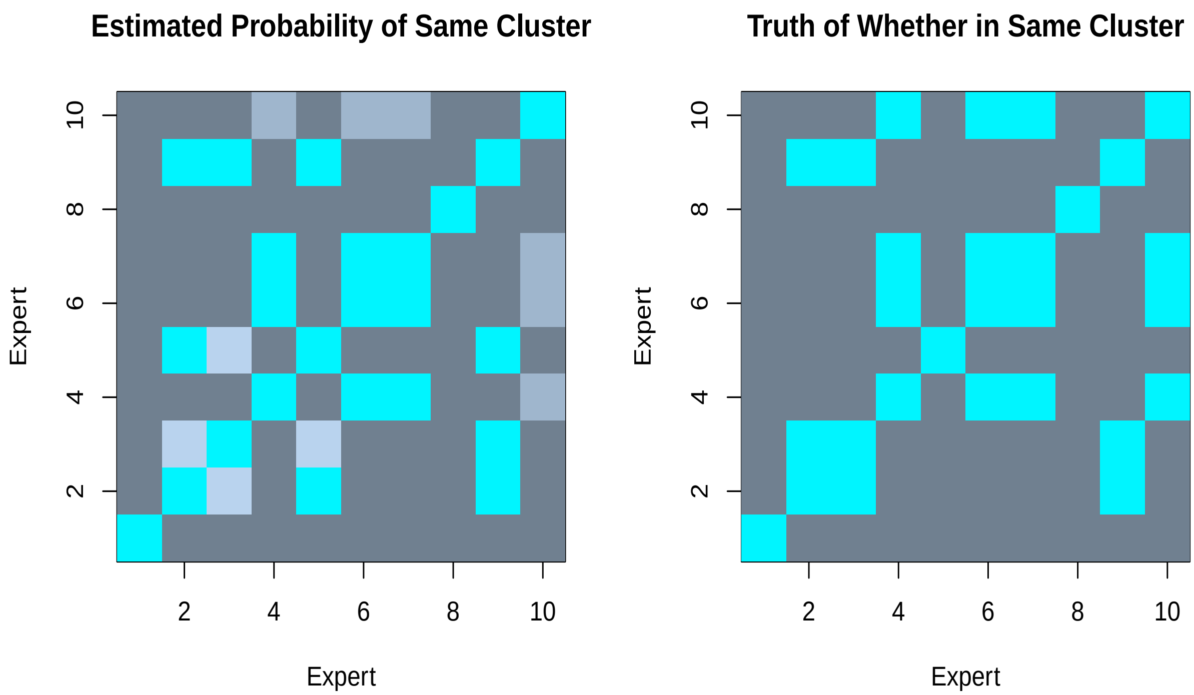

Figure 3: Cluster Allocation with 10 experts, 10 questions, 4 covariates

5 Dirichlet Process Aggregation Applied: Western and Jackman (1994)

In order to illustrate the flexibility of this approach in incorporating multiple divergent priors, I take up the analysis conducted by Bruce Western and Simon Jackman in their 1994 paper, "Bayesian Inference for Comparative Research." In the paper, the authors highlight the many benefits of Bayesian analysis for comparative politics research. In particular, they emphasize that the use of priors to encapsulate differing ideas about the state of the world and combine them with data to draw inferences formalizes a process already undertaken in comparative politics research, albeit less transparently and with less ability to directly adjudicate between competing views. The authors illustrate this argument by drawing on a then-recent debate between Michael Wallerstein and John Stephens concerning the most important factors giving rise to unionization in advanced industrialized democracies.

Western and Jackman's aim is very similar to my project in this paper: namely, to emphasize the importance of priors as instantiations of knowledge, and to highlight the challenge of dealing with differing perspectives even within a prior probability framework. In the context of the Wallerstein/Stephens debate, this concern is applied to the study of union density—union members as a percent of the labor force. While Wallerstein argued for the size of the labor force as the most critical determinant, Stephens favored industrial concentration as an explanation. As Western and Jackman note, these two variables are nearly perfectly negatively correlated (Jackman and Western 1994, 416). Furthermore, the sample of interest—20 industrialized nations in a single year—both suggests against a frequentist approach and increases the challenge of the collinearity between the favored explanatory variables. Drawing on this debate, and with reference to outside sources, Western and Jackman construct plausible prior means and variances for Stephens and Wallerstein for each of three explanatory variables relative to unionization: left government, log labor force size, and economic concentration. Each of these prior means and the corresponding precisions are represented in the figure below.

The most notable difference, as suggested by their written works, is that Wallerstein believes logged labor force should have a strong negative effect on unionization (Stephens believes it has no effect) and that in turn Stephens believes economic concentration has a strong positive effect on unionization, whereas Wallerstein believes it has none.

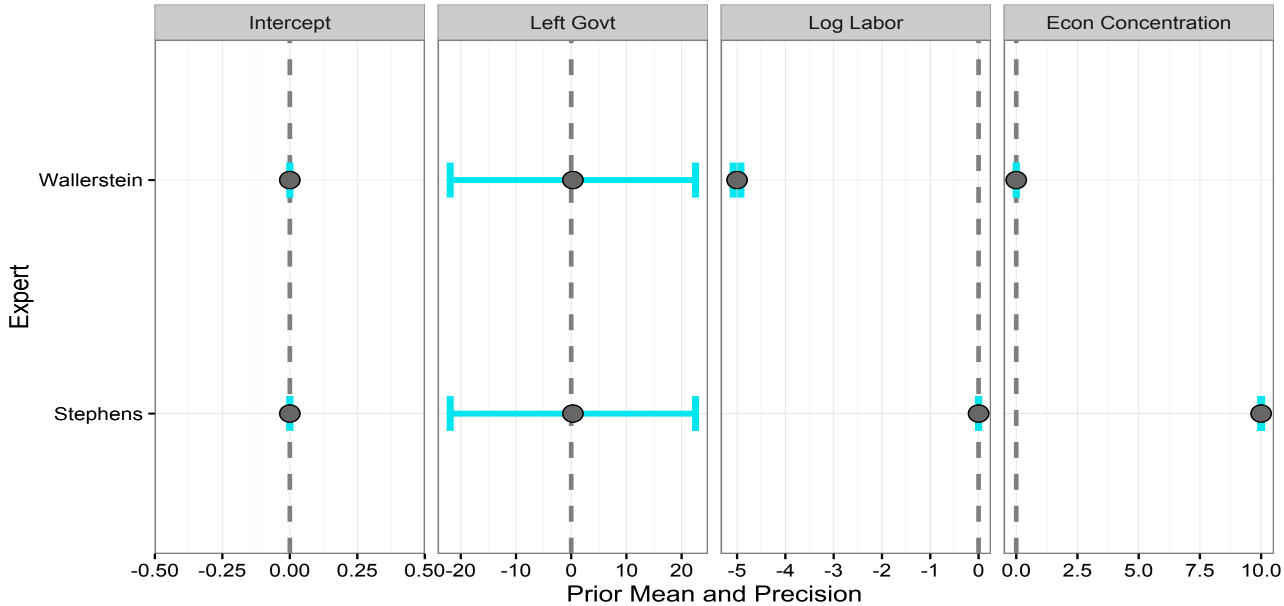

Figure 4: Wallerstein and Stephens' Priors and Precisions as Shown in Jackman and Western (1994)

In the subsequent analysis, Jackman and Western use each of these priors respectively to conduct a simple Bayesian linear regression

$$\mathrm{Unionization} \sim \alpha + \mathrm{Leftgovt} + \mathrm{LogLabor} + \mathrm{EconConcent} + \epsilon$$

using 20 observations of country-level data on unionization and each of the explanatory variables. They first estimate the regression with uninformative priors to find a baseline result, and then apply Wallerstein's prior and Stephens' prior separately in turn to compare the results. Their findings illustrate both the power of priors to shape the conclusions we draw from our data analyses, and the need for better techniques to incorporate what might be divergent priors into analyses. While their paper emphasizes each of these priors separately to illustrate the divergent results, experts such as Wallerstein and Stephens may have equally valid insights into the research problem, or may have pieces of the same picture. How should these differing views then be reconciled?

5.1 Extending Western and Jackman

To illustrate the efficacy of the Dirichlet-process-based method I propose for aggregating priors, I reassess the data used by Western and Jackman using a slightly larger body of hypothetical priors. Rather than simply using the two paradigmatic examples of Stephens and Wallerstein, I construct hypothetical priors for "experts" who may favor any of the explanatory variables identified by Western and Jackman, as reasonable scholars of the literature may have a diverse set of beliefs about key explanatory factors. For example, while Western and Jackman treat "left government" as simply a control variable, my analysis assumes that some experts might consider left governments to play a significant role in determining unionization. To reflect this diversity of possible perspectives, I construct priors for 10 hypothetical experts, including Wallerstein and Stephens. In addition to these two iconic scholars, I suppose that there is someone skeptical of economic explanations—one who believes that government policies solely determine unionization; two different camps of communist views, one governmentalist that believes a liberal state can support unionization and another more classical, believing that only a large labor force could solidify the aims of the proletariat and lead to organization; a neoliberal who believes that the competition arising from economic concentration (industrialization) should decrease unionization; two labor-supporting experts who believe that the labor force is most strongly determining but who disagree about the magnitude; and two "uncertain" experts, who can agree on the same explanatory variables as the other experts but who believe the magnitude of effects is small.

The prior means and precisions chosen for each of these hypothetical experts is reflected in the figure below. The variances selected in the original Jackman and Western (1994) are often quite large, and others are quite small, and I follow this convention in choosing variances corresponding to the priors of the new hypothetical experts. Precisions ($\frac{1}{\sigma^2}$) are shown for visual clarity. As is evidenced in the figure, these prior means and precisions all seem like reasonable positions that experts on the issue of unionization might hold, but at the same time they reflect considerable diversity.

One approach to estimating a model with such diverse opinions of experts would generate separate estimates, with each expert's prior used in succession, and then compare results. This framework is reasonable if we believe that while each prior is legitimate, only one provides

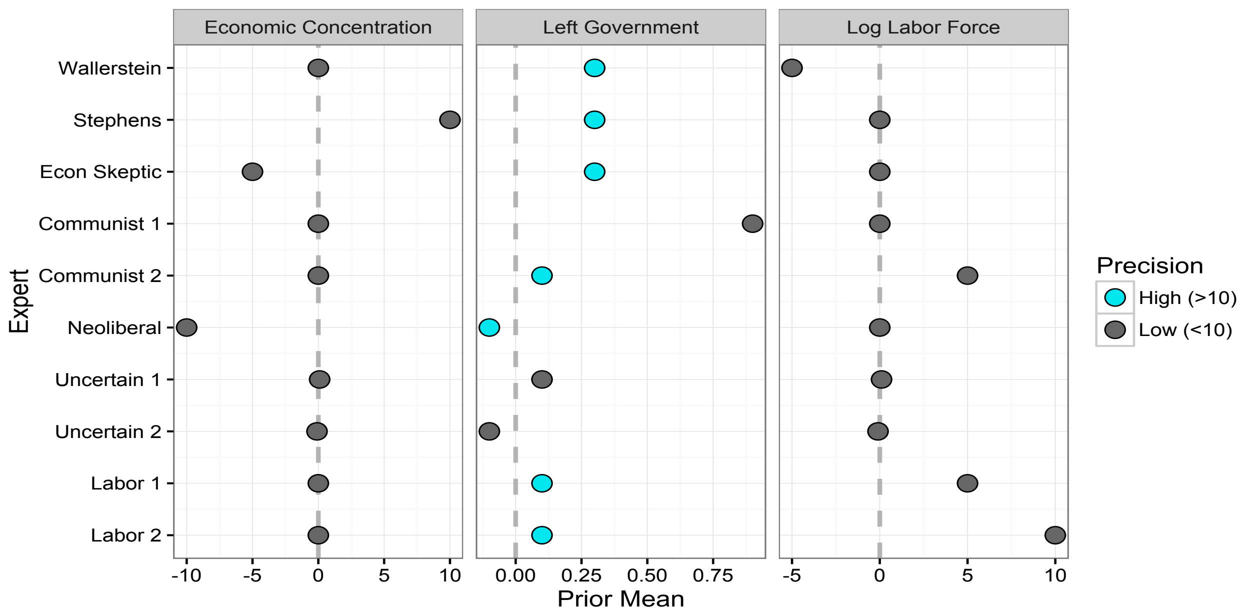

Figure 5: Hypothetical Expert Priors and Precisions

the "true" answer. As is especially true in studies of authoritarianism, where the information environment is incredibly fragmented, however, one can also imagine a circumstance in which each of these expert's priors contains some element of the truth, but having an aggregate picture would be preferable.

Current research in elicited priors recognizes this tradeoff to some extent. Because expert priors are often elicited in focus group settings, rather than in separate sessions, however, this problem is addressed by attempting to achieve a consensus prior, or trying to pool or average the priors of the attending experts. This approach, I argue, loses considerable information, and a Dirichlet-based method of aggregation should perform better, particularly when expert opinions are divergent. To illustrate this with the Jackman and Western (1994) data, I estimate the model from the original paper, and demonstrate the resulting posterior means and credible intervals when the 10 hypothetical priors are averaged versus when they are clustered in a Dirichlet process. I omit a comparison to pooling because the result would depend upon researcher-chosen weights for expert opinions.

The results in Figure 6 demonstrate the ability of the Dirichlet process aggregation to better incorporate a series of divergent priors into a common analysis of unionization, relative to averaging priors. The intercepts across both models do not vary because the intercepts for all priors

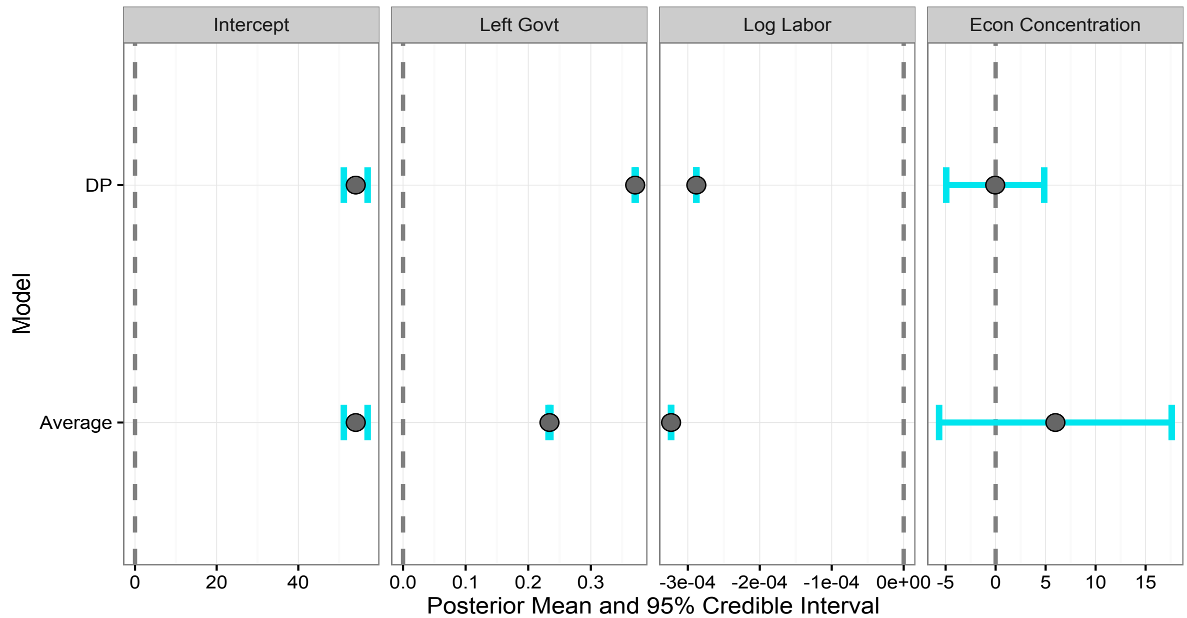

Figure 6: Averaging vs. Dirichlet Process Prior Results

were set to be equal. The results do differ, however, in their estimation of the three explanatory variables. In estimating an effect for left governments, for example, the Dirichlet process model recovers a coefficient similar to the high-precision prior means supplied by Wallerstein and Stephens, who share a cluster, whereas the model that averages priors gives slightly more weight to the views of communist and neoliberal experts. In particular, the averaging model may be assigning more weight to the prior of "Communist 2," who in the Dirichlet setup is in a cluster by herself. A similar dynamic appears in the log labor force coefficients, which in both models reflect the prior means that were near zero in most cases, but which differ in that greater weight is assigned to the "Labor 1" and "Labor 2" experts in the Dirichlet process, who partially share a cluster. Finally, while the coefficients for economic concentration appear the most extreme, their confidence intervals do overlap. Even so, the Dirichlet process posterior mean centers on zero, as is reflected in most of the prior means.

These results indicate that the Dirichlet process effectively handles the diversity of perspectives that experts may bring to an empirical analysis and reflect in their prior distributions. At the same time, its results appear less volatile than those of the averaging model because the Dirichlet process clustering is able to incorporate extreme views while identifying latent credibility in the data structure that serves to diffuse less informative extreme perspectives. Even so, the Dirichlet framework is very flexible: the standard setup provides equal weighting to clusters, but a researcher themselves may have priors about the credibility or reliability of experts. The researcher's priors in turn can be instantiated as a hyperior, providing different weights for the clusters created through the Dirichlet process. In this way, even with as few as 20 observations, the information held by experts in the field, as well as researchers investigating current questions, can be effectively leveraged to conduct data analyses.

6 Conclusions and Next Steps

An elicited priors approach, as the literature discussed above demonstrates, provides incredibly useful additional insight in Bayesian analysis across a wide array of disciplines. In applying the method within political science, however, examples are much more limited. As I have argued, this is likely due to the fact that the current body of literature does not provide adequate guidance for adapting elicited priors to social science settings. In particular, because "expertise" can be a much broader and more nebulous concept in social scientific contexts, researchers are likely to encounter greater diversity in the priors they elicit. Current techniques do not adequately justify methods for aggregating these divergent opinions, nor do they offer concrete methods for adjudicating the value of some priors versus others beyond the statistical accuracy of the statements.

For researchers in the social sciences, and particularly political science, more work is needed to be able to adapt an elicited priors approach to relevant research questions. This is particularly the case for work that addresses authoritarian regimes or takes place in low-data environments. These settings, often synonymous with small-n work and poor data or information quality, are precisely the types of settings where a Bayesian approach should predominate because of its ability to more easily handle modestly sized data. In these settings, however, "experts" are likely to diverge more significantly in their opinions due to differing access to information and biases resulting from political status. Accounting for these differences in an elicited priors framework requires moving beyond the most common aggregation methods currently available (pooling and averaging).

In this paper, I have proposed a Dirichlet-based framework for addressing these diverging priors, borrowing a technology commonly applied in text-as-data type analyses. This method has greater flexibility and transparency, and can allow the researcher to aggregate priors according to different latent categories of consensus without only relying on the priors' statistical validity in reference to the data. This method stands to be particularly useful in settings where small-n and/or a complex modeling structure require informative priors, and where in principle experts exist to offer prior information but they are an as-yet untapped resource. As the example with the Jackman and Western (1994) data demonstrates, the Dirichlet-based approach competently deals with the diversity of potential expert opinions, and facilitates estimating a model even with sparse data.

The approach suggested here and the emphasis on eliciting priors may have its own downsides. In particular, elicitation of priors and the documentation of the research process in low-data contexts is likely to be difficult and more time consuming than other approaches. The aim, however, is to generate an approach that would both facilitate the use of elicited priors more broadly, while also particularly championing elicited priors as a tool for use in challenging research contexts where the lack of data or absence of identifiable expertise are particularly problematic. This goal is especially well-suited to closing part of the quantitative/qualitative divide that has emerged in some areas of political science, since the elicited priors approach would allow quantitative scholars to still conduct analyses while partnering with qualitative scholars whose deep contextual knowledge could aid in the identification of experts and the elicitation process as well.

6.1 Additions and Extensions

To further the analysis included in this paper, I aim to conduct a series of simulations that will illustrate the applications of the proposed method, and its performance relative to current methods such as pooling and averaging. In particular, this will involve simulating both continuous and discrete models/data with a series of generated priors "elicited from experts." The simulation will show how differing distributions of experts' information (e.g., divergent versus convergent, large versus small numbers of experts, large versus small number of data points elicited, etc.) contribute to the estimation process with each of these methods for prior elicitation.

To guide this empirical assessment of the proposed model, however, I also aim to provide a formal model of information elicitation. A general model of information elicitation is useful as a baseline for developing expectations about the type of information we as researchers receive when using an elicited priors approach, especially in an authoritarian setting. In particular, the authoritarian context for elicitation means that the information environment is constrained: individuals have limited access to information and their understanding of situations and consequences may be biased as a result. Whether and what information an "expert" offers during elicitation is a product of this constrained information environment, and this condition in particular should shape how information from different "experts" is and can be combined to better understand social processes at work. Constructing a formal model of why, whether, and what information individuals are likely to offer during elicitation given authoritarian constraints can provide more rigorous expectations for what methodology to employ when eliciting and aggregating these priors.

References

Albert, Isabelle, Sophie Donnet, Chantal Guihenneuc-Jouyaux, Samantha Low-Choy, Kerrie Mengersen, and Judith Rousseau. 2012. "Combining Expert Opinions in Prior Elicitation." *Bayesian Analysis* 7 (3): 503–531.

Bakker, Ryan. 2009. "Re-measuring left-right: A comparison of SEM and Bayesian measurement models for extracting left-right party placements." *Electoral Studies* 28 (3): 413–421.

Blei, David, Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3:993–1022.

Daponte, Beth Osborne, Joseph B. Kadane, and Lara J. Wolfson. 1997. "Bayesian Demography: Projecting the Iraqi Kurdish Population, 1977–1990." *Journal of the American Statistical Association* 92 (440): 1256–1267.

Datta, G.S., and M. Ghosh. 1991. "Asymptotic Optimality of Hierarchical Bayes Estimators and Predictors." *Journal of Statistical Planning and Inference* 29 (3): 229–243.

Dey, Dipak K., and Junfeng Liu. 2007. "A Quantitative Study of Quantile Based Direct Prior Elicitation from Expert Opinion." *Bayesian Analysis* 2 (1): 137–166.

Dey, D.K., and L.R. Birmiwal. 1994. "Robust Bayesian Analysis Using Divergence Measures." *Statistics and Probability Letters* 20 (4): 287–294.

Garthwaite, Paul H., Joseph B. Kadane, and Anthony O'Hagan. 2005. "Statistical Methods for Eliciting Probability Distributions." *Journal of the American Statistical Association* 100 (470): 680–700.

Gelman, Andrew. 2002. "Prior Distribution." In *Encyclopedia of Environmetrics,* edited by Abdel H. El-Shaarawi and Walter W. Piegorsch, 3:1634–1637. Chichester: John Wiley & Sons.

Gill, Jeff, and Lee D. Walker. 2005. "Elicited priors for Bayesian model specifications in political science research." *Journal of Politics* 67 (3): 841–872.

Gill, Jeff, and Christopher Witko. 2013. "Bayesian Analytical Methods: A Methodological Prescription for Public Administration." *Journal of Public Administration Research and Theory* 23 (2): 457–494.

Jackman, Simon, and Bruce Western. 1994. "Bayesian Inference for Comparative Research." *American Political Science Review* 88 (2): 412–423.

Kadane, Joseph B., and Lara J. Wolfson. 1998. "Experiences in Elicitation." *Journal of the Royal Statistical Society Series D (The Statistician)* 47 (1): 3–19.

Kendall, Chad, Tommaso Nannicini, and Francesco Trebbi. 2015. "How Do Voters Respond to Information? Evidence from a Randomized Campaign." *American Economic Review* 105 (1): 322–353.

Lele, S.R., and A. Das. 2000. "Elicited data and incorporation of expert opinion for statistical inference in spatial studies." *Mathematical Geology* 32 (4): 465–487.

Lin, C.H., Y.C. Lin, and P. Raghubir. 2004. "The interaction between order of elicitation and event controllability on the self-positivity bias." In *Advances in Consumer Research, Volume XXXI,* edited by B.E. Kahn and M.F. Luce, 31:523–529.

Morris, Rachel K., Gemma L. Malin, Elisabeth Quinlan-Jones, Lee J. Middleton, Karla Hemming, Danielle Burke, Jane P. Daniels, Khalid S. Khan, Jon Deeks, and Mark D. Kilby. 2013. "Percutaneous vesicoamniotic shunting versus conservative management for fetal lower urinary tract obstruction PLUTO: a randomised trial." *Lancet* 382 (9903): 1496–1506.

Neal, Radford M. 2000. "Markov Chain Sampling MMethod for Dirichlet Process Mixture Models." *Journal of Computational and Graphical Statistics* 9 (2): 249–265.

O'Leary, Rebecca A., Samantha Low Choy, Justine V. Murray, Mary Kynn, Robert Denham, Tara G. Martin, and Kerrie Mengersen. 2009. "Comparison of three expert elicitation methods for logistic regression on predicting the presence of the threatened brush-tailed rock-wallaby Petrogale penicillata." *Environmentrics* 20 (4): 379–398.

Savage, Leonard J. 1971. "Elicitation of Personal Probabilities and Expectations." *Journal of the American Statistical Association* 66 (336): 783–801.

Shuford, Emir, and Thomas A. Brown. 1975. "Elicitation of Personal Probabilities and Their Assessment." *Instructional Science* 4:137–188.

Sivaganesan, S. 1993. "Range of the Posterior Probability of an Interval for Priors with Unimodality Preserving Contaminations." *Annals of the Institute of Statistical Mathematics* 45 (1): 187–199.

Small, Mitchell J. 2008. "Methods for assessing uncertainty in fundamental assumptions and associated models for cancer risk assessment." *Risk Analysis* 28 (5): 1289–1307.

Spetzler, Carl S., and Carl-Axel S. Staël von Holstein. 1975. "Probability Encoding in Decision Analysis." *Management Science* 22:340–358.

Wang, Zhongming, and Hui Zhou. 2009. "A General Method of Prior Elicitation in Bayesian Reliability Analysis." In *Proceedings of the 2009 8th International Conference on Reliability, Maintainability and Safety, Vols. I and II,* edited by R. Kang, H.Z. Huang, M.J. Zuo, and Q. Miao, 415–419.

Wheeler, David C., Antonio Paez, Jamie Spinney, and Lance A. Waller. 2014. "A Bayesian approach to hedonic price analysis." *Papers in Regional Science* 93 (3): 663–683.

Zaffalon, M. 2005. "Credible classification for environmental problems." *Environmental Modelling and Software* 20 (8): 1003–1012.