Mass Polarization and Democratic Decline: Global Evidence from a Half-Century of Public Opinion

Supplementary Information

Isaac D. Mehlhaff*

March 4, 2022

Contents

*Graduate Student, The University of North Carolina at Chapel Hill; `mehlhaff@live.unc.edu`.

S1 Measurement Model

The main text briefly discusses the use of a Bayesian approach to building dynamic panels of mass polarization. Here, I fully explicate this model and describe its estimation and validation. My primary objective in fitting this measurement model is approximating a population-level distribution of ideology and party affect for each country-year. A Bayesian approach is especially useful here, as I am not actually interested in the point estimates of latent ideology or party affect, but rather their *distribution.* A measure of polarization can then be recovered through post-processing of these country-year latent distributions. A secondary objective is to collate the multitude of aggregate, country-level survey data available to scholars and correct for non-stochastic variation arising as a consequence of these data being splintered across time, space, and survey program. The following section describes how I employ dynamic latent variable models and an infinite Gaussian mixture model in a Bayesian framework to accomplish these objectives.

S1.1 Model

I first construct a dynamic latent variable model similar to that of Claassen (2019). I seek to preserve the discrete structure of survey data by modeling the number of respondents y_{itj} offering response option k to item j, rather than the proportion of respondents or some other derived measure (Caughey, O'Grady, and Warshaw 2019; Caughey and Warshaw 2015; Linzer 2013). These survey data, however, are imperfect; they provide noisy signals of aggregate public opinion and approximate, but do not capture directly, ideology or party affect in a given population. I instead use these survey data to recover latent ideology or party affect θ_{itk}, where i denotes country and t denotes time period (year, in this case). Thus,

$$y_{itj} \sim \text{Multinom}(n_{itj}, \pi_{itjk}), \tag{S1}$$

where n_{itj} is the number of observations collected for a survey item in a given country-year and π_{itjk} is a vector of probabilities $\pi_{itj1}, ..., \pi_{itjK}$ for that item's response categories. From here, there are two specification options. π_{itjk} can be modeled directly as a function of latent traits, or it can take a Dirichlet prior to allow for additional dispersion in survey responses beyond sampling error (e.g. Adida et al. 2016; Biemer 2010; Weisberg 2005). In the case of the Dirichlet-multinomial distribution,

$$\pi_{itjk} \sim \text{Dir}(\alpha_{itjk}), \tag{S2}$$

where α_{itjk} is a vector of concentration parameters $\alpha_{itj1}, ..., \alpha_{itjK}$. α_{itjk} can be reparameterized with an expectation parameter η_{itjk} and scale parameter ϕ,

$$
\begin{aligned}
\eta_{itjk} &= \frac{\alpha_{itjk}}{\sum_{k=1}^{K} \alpha_{itjk}}, \\
\phi &= \sum_{k=1}^{K} \alpha_{itjk}, \\
\rightarrow \alpha_{itjk} &= \eta_{itjk}\phi,
\end{aligned}
\tag{S3}
$$

where $\phi \sim \Gamma(4, 0.1)$ and η_{itjk} (or π_{itjk} in the models without the Dirichlet prior) is modeled as a function of latent traits.

In particular, it takes item bias effects λ_{jk}, item-country latent effects δ_{ijk}, item slopes γ_{jk}, and my primary quantity of interest, country-year latent effects θ_{itk}. I incorporate these parameters one by one to yield six models each for ideology and party affect (three multinomial models and three Dirichlet-multinomial models) but present only the four-parameter model here:

$$
\begin{aligned}
\eta_{itjk} &= \text{softmax}^{-1}(\lambda_{jk} + \delta_{ijk} + \gamma_{jk}\theta_{itk}), \\
&= \ln(\lambda_{jk} + \delta_{ijk} + \gamma_{jk}\theta_{itk}) + c.
\end{aligned}
\tag{S4}
$$

where $c = \ln(\sum_{k=1}^{K} e^{\eta_{itjk}})$. λ_{jk}, γ_{jk}, and δ_{ijk} are modeled hierarchically as a function of data with response options nested in survey items (for λ_{jk} and γ_{jk}) and response options nested in survey items and countries (for δ_{ijk}), making this a fully hierarchical linear model. Because item bias effects and item slopes may be correlated, they additionally take a bivariate normal distribution with correlation ρ:[1]

$$
\begin{aligned}
\begin{bmatrix} \lambda_{jk} \\ \gamma_{jk} \end{bmatrix} &\sim \text{N}\left(\begin{bmatrix} \mu_\lambda \\ \mu_\gamma \end{bmatrix}, \begin{bmatrix} \sigma_\lambda^2 & \rho\sigma_\lambda\sigma_\gamma \\ \rho\sigma_\lambda\sigma_\gamma & \sigma_\gamma^2 \end{bmatrix} \right), \\
\delta_{ijk} &\sim \text{N}(0, \sigma_\delta^2),
\end{aligned}
\tag{S5}
$$

[1]In the two- and three-parameter models, which do not take item slopes γ_{jk}, item bias effects λ_{jk} take a univariate normal prior $\lambda_{jk} \sim \text{N}(\mu_\lambda, \sigma_\lambda^2)$ and σ_λ^2 further takes a weakly informative $\text{N}^+(0, 2)$ prior.

where $\rho \sim \mathrm{N}^+(0,2)$ and the variance-covariance matrix is estimated using a Cholesky decomposition with an LKJ(2) prior (Lewandowski, Kurowicka, and Joe 2009). σ_δ^2 further takes a weakly informative $\mathrm{N}^+(0,2)$ prior.[2]

Finally, the country-year latent effects θ_{it} must be smoothed over time. I do this by simply specifying a random walk (Claassen 2019; Jackman 2005):

$$\theta_{itk} \sim \mathrm{N}(\theta_{i,t-1,k}, \sigma_\theta^2), \tag{S6}$$

where σ_θ^2, like other variance terms above, is held constant across countries, years, and response options; is estimated from the data; and takes a weakly informative $\mathrm{N}^+(0,2)$ prior.

The result of this dynamic latent variable model is a collection of vectors θ_{itk} modeled hierarchically and smoothed over time. The next challenge is to take these estimates of latent ideology and party affect and recover the distribution of those latent variables. Doing so is relatively simple: Pushing θ_{itk} back through the softmax function maps those latent estimates onto a simplex, which can then be passed to a multinomial distribution along with the total number of survey responses observed in each country n_{it} to generate $\tilde{y}_{it}$—the number of responses in each country-year-category corrected for item-category and country-item-category effects:

$$\tilde{y}_{it} \sim \mathrm{Multinom}(n_{it}, \mathrm{softmax}^{-1}(\theta_{itk})). \tag{S7}$$

These corrected data can now be used to approximate a distribution from which to estimate polarization. In particular, I treat $\tilde{y}_{it}$ as the outcome of a Gaussian mixture:

$$\tilde{y}_{it} = \sum_{c=1}^{C} \omega_{itc} \mathrm{N}(\mu_{itc}, \sigma_{itc}^2), \tag{S8}$$

where ω_{itc} gives the mixture weights and c is the component index. The parameters in (S8) take the following priors:

$$\begin{aligned} \mu_{itc} &\sim \mathrm{N}(\bar{y}_{it}, \sigma_{\mu_{itc}}^2), \\ \sigma_{itc}^2 &\sim \mathrm{N}^+(0, 0.5), \end{aligned} \tag{S9}$$

[2]Although inverse-gamma or half-Cauchy distributions are often preferred for this type of prior, recent work has moved toward using half-normals due to their computational tractability and numerical stability.

where $\bar{y}_{it}$ denotes the mean of $\tilde{y}_{it}$ for each country-year and $\sigma^2_{\mu_{itc}}$ further takes the weakly informative $\mathrm{N}^+(0,2)$ prior.

Because the hierarchical and dynamic nature of the data was accounted for in the dynamic latent variable model, no such construction is necessary for this mixture distribution; I can simply estimate the mixture weights ω_{itc}, means μ_{itc}, and standard deviations σ_{itc} for each country-year by fitting a separate mixture to that country-year's vector of data $\tilde{y}_{it}$.

A couple key problems remain. First, fitting a Gaussian mixture would require the *a priori* specification of the number of components C. It is not entirely clear what that number should be and, more perniciously, the appropriate number of latent components likely varies across countries and years. To account for this variation, I proceed nonparameterically and consider (S8) an infinite Gaussian mixture such that $C \to \infty$, implying that the model follows a Dirichlet process.[3] Note that this does not mean that the distribution of $\hat{y}_{it}$ will have infinitely many components. Rather, it will take a countably infinite set of component parameters, with most component weights approaching zero in the limit, leaving only those components within which most of the probability mass is contained. $\omega_{itc} \; \forall \; c \in 1, ..., C$ is therefore a sparse vector of length C, where the degree of sparsity is controlled by β in (S10).

But this poses a second problem: The component indicator variable that would typically be required in a Dirichlet process is computationally intractable, as discrete parameters are unsupported by most sampling algorithms. Marginalizing over that parameter, however, produces the notation in (S8), where the mixture weights ω_{itc} can be expressed as the outcome of a stick-breaking process:

$$\begin{aligned}
\omega_{itC} &= 1 - \sum_{c=1}^{C-1} \omega_{itc}, \\
\omega_{itc} &= \nu_{itc} \prod_{\ell=2}^{c-1} 1 - \nu_{it\ell}, \\
\omega_{it1} &= \nu_{it1}, \\
\nu_{it\ell} &\sim \mathrm{Beta}(1, \beta),
\end{aligned} \tag{S10}$$

where I specify $\beta = 4$. This is a relatively high value for β relative to typical implementations of stick-breaking processes, but the benefit of specifying a prior with so much probability density close to zero is

[3]Of course, a computer cannot hold any object with dimensions of infinite size. In practice, then, an upper bound must be placed on C. I reason that it is unlikely to uncover more than five distinct, important opinion clusters (indeed, prior knowledge suggests most cases will produce only two or three identifiable clusters), so I specify $C = 5$ to help conserve computing resources. Results suggest that this maximum number of components is sufficient, as ω_{it4} and ω_{it5} approach zero for most country-years.

that I can be more confident that whatever components are uncovered by the mixture model do, in fact, represent meaningful clusters of data.

The result of this infinite Gaussian mixture model is therefore a set of component means μ_{itc}, standard deviations σ^2_{itc}, and weights ω_{itc} for each component in each country-year. These parameters represent the location, dispersion, and size, respectively, of each opinion cluster. The emphasis on estimating both location and dispersion parameters is deliberate: Polarization is a function of both distance between groups and concentration within groups (Baldassarri and Bearman 2007; Esteban and Ray 1994; Ura and Ellis 2012), and these two dynamics can only be captured by fully parameterizing the latent distribution. I can therefore obtain a measure of polarization for each country-year by estimating the degree of polarization in the distribution parameterized by μ_{itc}, σ^2_{itc}, and ω_{itc}. I do this by applying the cluster-polarization coefficient (for details, see Mehlhaff 2021), which corrects for different numbers of opinion clusters across country-years.

S1.2 Identifying Restrictions

The models I present above contain several degeneracies that require restrictions to identify the model. First, the latent variable model defined in (S4) suffers from both additive and multiplicative aliasing (Bafumi et al. 2005). In the first case, a constant can be added to all terms without changing the model output. In the second case, either the item slopes γ_{jk} or the latent traits θ_{itk} can be multiplied by a constant and, if the other is divided by the same constant, the model output will not change. I identify the two-parameter models by fixing the first item bias effect to $\lambda_{1,k} = 0$; all subsequent λ_{jk} parameters can then be interpreted with respect to the fixed parameter as a baseline. I break the additive aliasing in the three-parameter models by fixing $\lambda_{1,k}$ and additionally specifying the mean of δ_{ijk} to be zero, ensuring that the latent traits θ_{itk} are the only parameters allowed to float freely. The four-parameter models are identified by fixing the means of μ_λ and μ_γ in the multivariate normal prior in (S5) to zero and one, respectively, and by constraining item slopes γ_{jk} to be positive. δ_{ijk} further retains a mean of zero. Finally, dynamic models like this one must impose some structure on the dynamic parameters θ_{itk} so the time series as a whole is anchored to some baseline value. I do this by placing a prior on the first vector of latent traits in each country, such that $\theta_{i1,k} \sim \mathrm{N}(0, 1)$.

Second, the mixture model defined in (S8) suffers from labeling degeneracy, as identical component distributions impose ambiguity about which parameters are associated with which component. One common way to break this degeneracy is to impose non-exchangeable, repulsive priors. However, it is not clear in this case what such priors should be, as the number of identified components in an infinite mixture model will vary. Fortunately, the inferences drawn from the model are also exchangeable. Components do not have meaningful labels; the only information that needs to be preserved is their location relative to each other. As a consequence of this symmetry, I can retain the exchangeable priors in (S9) but specify an ordering constraint such that $\mu_{it1} < \mu_{it2} < ... < \mu_{itC}$.

S1.3 Estimation

Typically, the default method for fitting a model like the one described above is to employ Markov chain Monte Carlo (MCMC) techniques to fully explore the parameter space. Unfortunately, even the most advanced MCMC methods can be extremely slow to converge when the posterior is complex or the number of data points is large, as are both the case here. Indeed, Caughey and Warshaw (2015) testify that their dynamic latent variable models needed up to several weeks to converge. Even leveraging recent advances in within-chain parallelization, I likewise estimate that the simplest two-parameter model would have taken approximately one week to converge, while the more complex four-parameter models would have needed well over one month.

Instead, I turn to variational inference to fit the models. Variational inference is a method of deterministic posterior approximation that is guaranteed to converge, easily assessed by convergence criteria, and finds the analytical posterior solution most closely resembling the true (analytically intractable) posterior by minimizing the Kullback-Leibler divergence between the two distributions (Grimmer 2011). Variational inference has become an indispensable tool in computer science and statistics (e.g. Airoldi et al. 2008; Blei, Ng, and Jordan 2003), but its application to political science has been more limited (for examples, see Grimmer 2013; Imai, Lo, and Olmsted 2016). To further reduce strain on computational resources, I split the model in two, fitting the latent variable models on the full data set and then fitting each country's mixture model separately, as countries are independent of each other once the latent variable models have been fit. In all cases, the evidence lower bound (ELBO)—the criterion used to monitor the variational algorithm—indicates convergence to the approximate posterior.

S1.4 Validation and Model Selection

The previous sections laid out six latent variable models each for ideology and party affect. This section validates the estimates obtained by those models and details how a particular model was selected for the calculation of polarization estimates. I conduct two internal validation exercises with mean absolute error and leave-one-out information criteria (LOOIC) (Vehtari, Gelman, and Gabry 2017) and two external validation exercises with mean absolute error and 80% credible interval (both equal-tailed (ETIC) and high-density

(HDIC)) coverage based on a randomly selected 75% training set and 25% test set.[4] Table S1 displays the results of these exercises.

Table S1: Internal and External Validation of Latent Variable Models

		Internal Validation		*External Validation*		
Variable	Model	MAE	LOOIC	MAE	ETIC	HDIC
Party Affect	2PL Multinomial	0.509	747300	13.091	0.899	0.893
	3PL Multinomial	14.847	102840338	16.069	0.889	0.889
	4PL Multinomial	0.629	1632477	14.287	0.897	0.891
	2PL Dirichlet-Multinomial	9.749	1415294	6.645	0.893	0.896
	3PL Dirichlet-Multinomial	6.115	1371934	6.779	0.892	0.896
	4PL Dirichlet-Multinomial	5.213	1340400	16.014	0.893	0.889
Ideology	2PL Multinomial	1.293	384842	32.708	0.757	0.696
	3PL Multinomial	1.309	403066	33.158	0.759	0.693
	4PL Multinomial	1.376	641674	33.952	0.759	0.685
	2PL Dirichlet-Multinomial	35.618	6786645	27.328	0.659	0.725
	3PL Dirichlet-Multinomial	10.591	6781720	27.425	0.659	0.724
	4PL Dirichlet-Multinomial	20.722	6769666	47.921	0.853	0.674

Beginning with internal validation, it is clear that the Dirichlet-multinomial specifications introduce greater error into the in-sample predictions. With the exception of the three-parameter affect model, the specifications without the Dirichlet prior display MAE values that are orders of magnitude lower than those produced by the Dirichlet-multinomial specifications. This result is echoed by the LOOIC, which have no intrinsic meaning themselves but can be used to compare models with respect to each other, with lower values indicating a better fit.

However, there is generally a tradeoff between internal and external fit—models can be well-fit to the data on which they are trained but may display poor out-of-sample predictive power. That is the case here. Looking at the external validation results, the Dirichlet-multinomial specifications typically display substantially lower MAE than the multinomial specifications, with the exception of the four-parameter ideology model. The models are more evenly matched in evaluating ETIC and HDIC. The affect models' coverage differs by, at most, one percentage point. The ideology models' coverage displays more variation, but the models that perform well on ETIC perform worse on HDIC, and vice versa.

As previously alluded to, it is important to keep in mind that selecting a model requires the balancing of tradeoffs; one model may not have a clear advantage over the others on all criteria. I ultimately selected the three-parameter Dirichlet-multinomial specification for both party affect and ideology. These models

[4] Credible interval coverage (CIC) measures the accuracy of the model's uncertainty estimates by calculating the percentage of observations contained in the out-of-sample credible interval. CIC indicates a good fit when the value is as close as possible to the width of the interval—80% in this case. Values well over 0.8 indicate standard errors that are too wide, and values well under 0.8 indicate standard errors that are too narrow.

display low-to-middling levels of internal MAE, external MAE, and LOOIC, suggesting that they offer a good middle ground between a model that is overfit on in-sample data (like the mulitnomial models) and a model that contains high levels of uncertainty (like the two- and four-parameter Dirichlet-multinomial models). Additionally, these chosen models' ETIC and HDIC do not differ significantly from the next closest specifications. I use the output from these models to proceed through the rest of the model defined above.

S2 Data Sources and Manipulation

I take data from all available national public opinion survey projects that ask about ideology on the left-right scale and feelings or attitudes toward parties. I exclude items asking about feelings toward political leaders, as these may tap into a different attitude (e.g. someone can feel very close to their party but dislike their chosen candidate). Whenever included by the survey project, I apply weights to make the data nationally representative.

The following survey projects have at least one country-year included in the data set: World Values Survey, Latinobarómetro, Comparative Study of Electoral Systems, Eurobarometer, AmericasBarometer, Comparative National Election Project, European Values Study, Central and Eastern Eurobarometer, Consolidation of Democracy in Central and Eastern Europe 1990-2001, Candidate Countries Eurobarometer, Pew Global Attitudes, Australian Election Study, Canadian Election Study, Danish Election Study, Icelandic National Election Study, Israel National Election Study, Italian National Election Survey, Dutch Parliamentary Election Study, New Zealand Election Study, Statistics Norway Election Survey, Statistics Sweden Election Study, Swiss Election Study, British Election Study, Parliamentary Election Belarus 1995, Democratic Attitudes in Belarus 2002, Croatian National Election Study, Party Preferences Czech Republic, Hungarian Election Study, Polish National Election Study, Election Study Serbia, Political and Social Attitudes in Serbia 2002, Current Problems of Slovakia 1999, Slovenian Public Opinion Survey 1997, American National Election Studies, General Social Survey, and Politbarometer.

As stated above, I endeavor to preserve the integrity of these data as much as possible, so I engage in minimal data manipulation. First, I scale the values of each response option to be on the same scale, allowing the latent variable model to estimate country- and item-level variation in specific response options across items with different numbers of response options. This requires me to make an assumption of cardinality, but because the latent variable model operates on categories and not values, the assumption applies only to the mixture model. Further, I believe this assumption is not an especially strong one and it is routinely employed by scholars calculating statistics such as mean or standard deviation from these data. After the data values have been scaled, I calculate the number of respondents offering each response option, weighted to be nationally representative when possible. These weighted counts are fed directly to the latent variable model.

S3 Liberal Democracy Plots

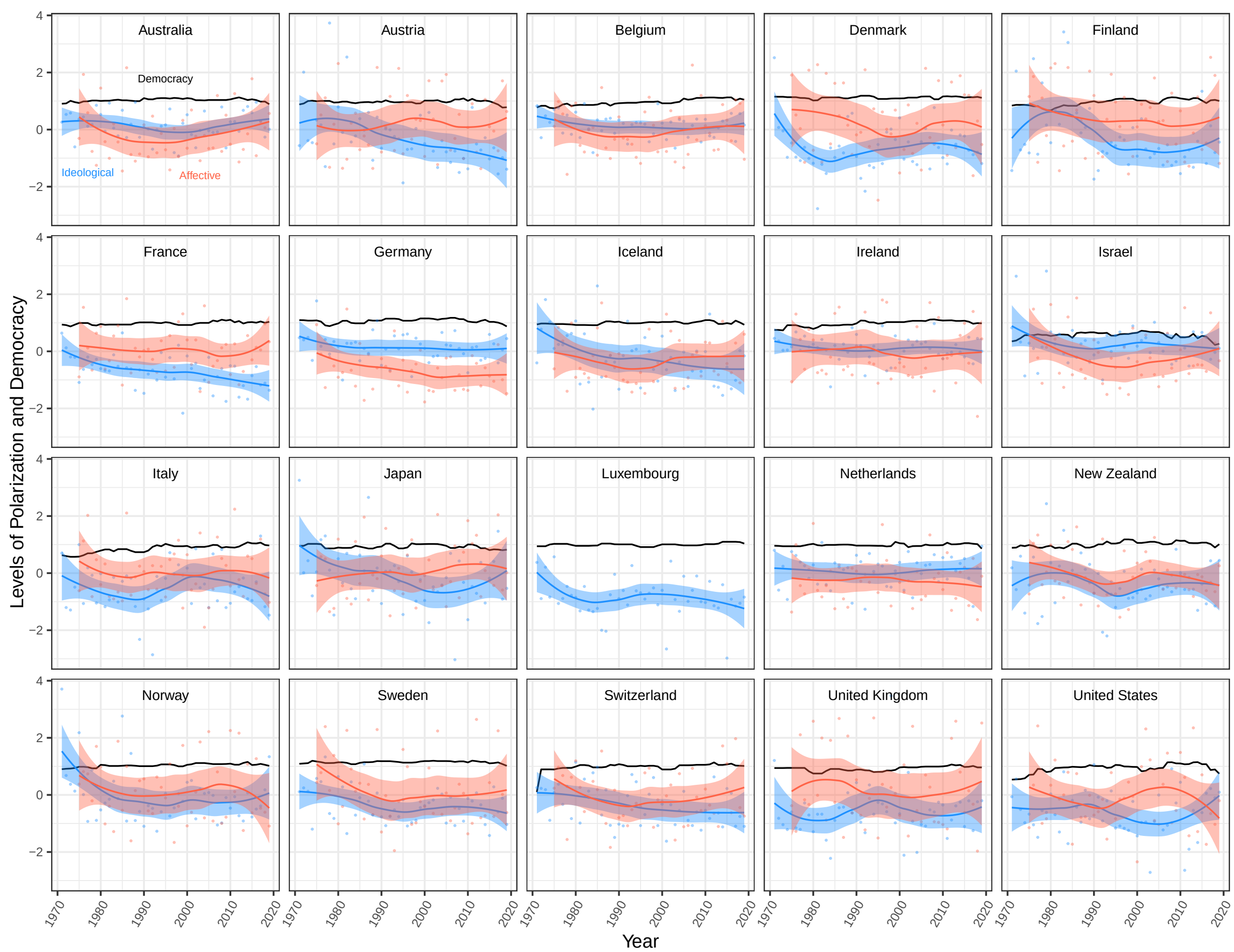

Figure S1: Estimates of Liberal Democracy, Ideological Polarization, and Affective Polarization in States Classified as Liberal Democracies in 1971. Each variable is scaled to be $\sim N(0,1)$ for ease of comparison. Ideological and affective polarization lines are fit to the data using LOESS; bands represent 95% confidence intervals. These trend lines are for visualization purposes only and are not used in any analysis.

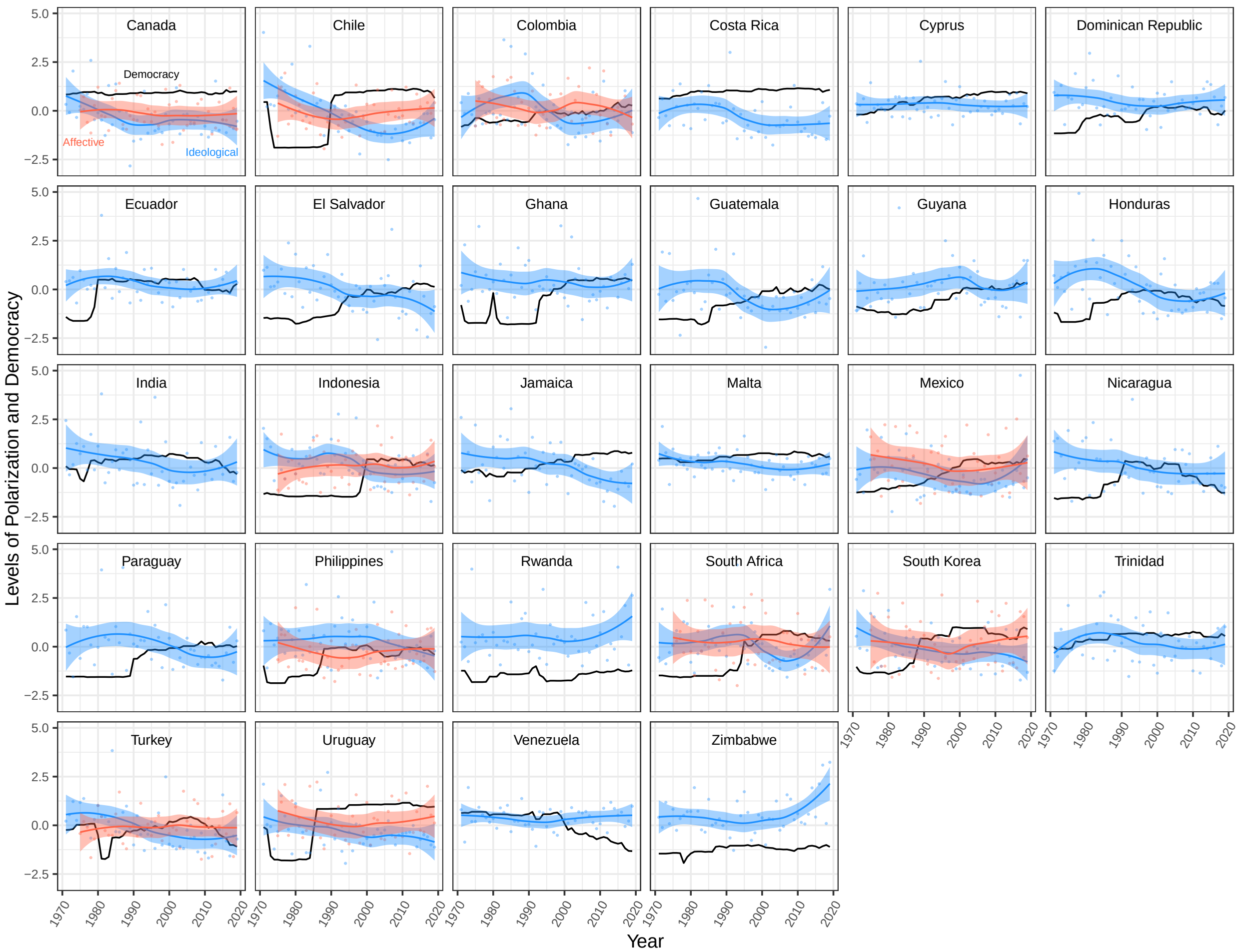

Figure S2: Estimates of Liberal Democracy, Ideological Polarization, and Affective Polarization in States Classified as Hybrid Regimes in 1971. Each variable is scaled to be $\sim N(0,1)$ for ease of comparison. Ideological and affective polarization lines are fit to the data using LOESS; bands represent 95% confidence intervals. These trend lines are for visualization purposes only and are not used in any analysis.

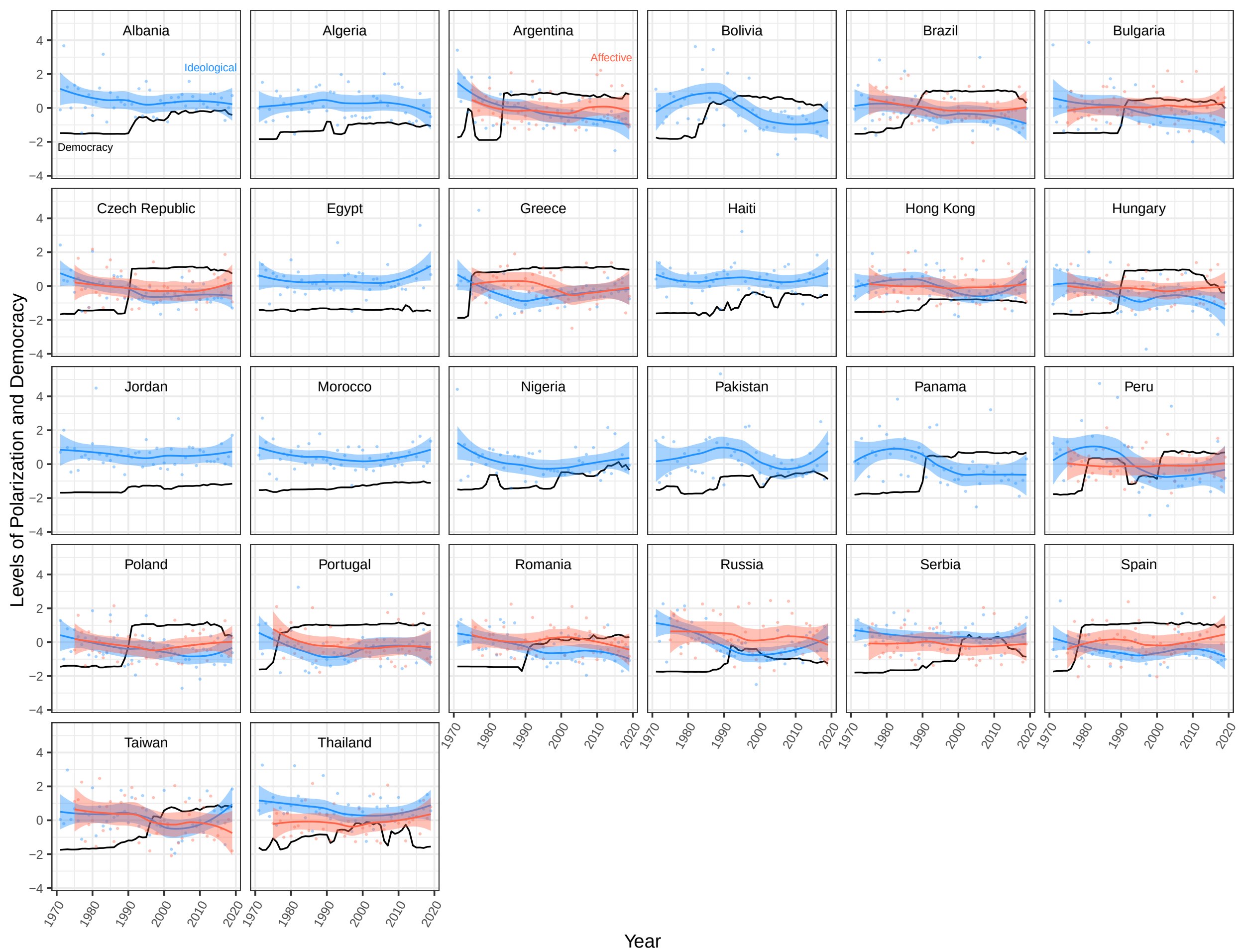

Figure S3: Estimates of Liberal Democracy, Ideological Polarization, and Affective Polarization in States Classified as Autocracies in 1971. Each variable is scaled to be $\sim N(0, 1)$ for ease of comparison. Ideological and affective polarization lines are fit to the data using LOESS; bands represent 95% confidence intervals. These trend lines are for visualization purposes only and are not used in any analysis.

S4 Full Model Results

Table S2: Naive OLS Models of Polarization and Democracy

	Dependent variable:			
	Electoral	Liberal	Electoral	Liberal
	(1)	(2)	(3)	(4)
Ideological	-0.093^{*}	-0.081^{*}		
	(0.010)	(0.010)		
Affective			-0.037^{*}	-0.035^{*}
			(0.013)	(0.012)
Presidential	-0.212^{*}	-0.287^{*}	0.025	-0.051
	(0.024)	(0.023)	(0.034)	(0.032)
GDP	0.546^{*}	0.611^{*}	0.766^{*}	0.845^{*}
	(0.013)	(0.012)	(0.021)	(0.021)
Resources	-0.032^{*}	-0.034^{*}	-0.402^{*}	-0.119
	(0.010)	(0.010)	(0.166)	(0.161)
Muslim	-0.206^{*}	-0.165^{*}	-0.084^{*}	-0.122^{*}
	(0.011)	(0.010)	(0.020)	(0.020)
Intercept	0.216^{*}	0.233^{*}	-0.033	0.031
	(0.015)	(0.014)	(0.042)	(0.041)
Observations	3,405	3,405	1,920	1,920
R^2	0.618	0.678	0.528	0.611
Adjusted R^2	0.618	0.677	0.527	0.610
Residual Std. Error	0.584 (df = 3399)	0.552 (df = 3399)	0.556 (df = 1914)	0.538 (df = 1914)

Note: *p<0.05. Values in parentheses give standard errors. All real-valued variables scaled to $\sim N(0, 1)$.

Table S3: Time Series Models of Polarization and Democracy

	Dependent variable:							
	Electoral		Liberal		Electoral		Liberal	
	Pooled OLS				System GMM			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Electoral_{t-1}	1.169*	1.208*			1.141*	1.095*		
	(0.035)	(0.042)			(0.030)	(0.048)		
Electoral_{t-2}	−0.227*	−0.278*			−0.248*	−0.319*		
	(0.034)	(0.040)			(0.030)	(0.034)		
Liberal_{t-1}			1.212*	1.225*			1.193*	1.135*
			(0.034)	(0.038)			(0.029)	(0.041)
Liberal_{t-2}			−0.256*	−0.282*			−0.278*	−0.326*
			(0.033)	(0.038)			(0.032)	(0.033)
Ideological_{t-1}	−0.002		−0.001		−0.011		−0.007	
	(0.003)		(0.003)		(0.007)		(0.006)	
Affective_{t-1}		0.002		0.0003		−0.013*		−0.013*
		(0.004)		(0.003)		(0.006)		(0.006)
$\text{Presidential}_{t-1}$	−0.002	0.012	−0.007	−0.001	0.023*	0.016	0.013	−0.001
	(0.007)	(0.011)	(0.006)	(0.009)	(0.008)	(0.027)	(0.007)	(0.026)
GDP_{t-1}	0.026*	0.041*	0.021*	0.033*	0.063*	0.160*	0.056*	0.151*
	(0.006)	(0.012)	(0.005)	(0.011)	(0.020)	(0.038)	(0.021)	(0.038)
Resources_{t-1}	−0.0004	0.012	−0.001	0.037	0.001	−0.108	0.0001	−0.090
	(0.002)	(0.041)	(0.001)	(0.038)	(0.003)	(0.067)	(0.002)	(0.055)
Muslim	−0.017*	−0.012	−0.012*	−0.014*	−0.025*	−0.025	−0.016*	−0.030
	(0.003)	(0.008)	(0.002)	(0.007)	(0.008)	(0.017)	(0.006)	(0.018)
Intercept	0.020*	0.018	0.018*	0.026*				
	(0.005)	(0.013)	(0.004)	(0.011)				
N Observations	3404	1919	3404	1919	6721	3790	6721	3790
N Units	86	48	86	48	92	92	92	92
N Time Periods	16-44	24-44	16-44	24-44	49	49	49	49

Note: *p<0.05. Values in parentheses give panel-corrected (OLS) or heteroskedasticity-consistent (GMM) standard errors. All real-valued variables scaled to $\sim N(0, 1)$.

S5 Binomial Logit Model Results

One additional strategy for capturing the extent to which polarization and backsliding are related is to determine whether democracy decreased in each country year relative to the prior year and evaluate how polarization contributes to whether or not democracy declines, regardless of how big or small the decline is. I pursue this strategy here by assigning each country-year a value of 1 if its level of democracy decreased relative to the prior year and a value of 0 otherwise. I then fit a binomial logit to these data with the temporal structure defined in (2) in the main text. Results, presented in Table S4, are consistent with what I find in other model specifications: polarization has a minimal effect on whether democracy declines in any given country-year. The coefficients associated with ideological and affective polarization never achieve statistical significance at the $p < 0.05$ level.

Table S4: Binomial Logit Models of Polarization and Backsliding

	Dependent variable:			
	Electoral	Liberal	Electoral	Liberal
	(1)	(2)	(3)	(4)
Ideological$_{t-1}$	-0.087^{*}	-0.051		
	(0.038)	(0.037)		
Affective$_{t-1}$			-0.042	-0.003
			(0.048)	(0.047)
Presidential$_{t-1}$	0.064	0.047	0.290^{*}	0.276^{*}
	(0.086)	(0.085)	(0.126)	(0.125)
GDP$_{t-1}$	-0.029	0.003	0.189^{*}	0.231^{*}
	(0.046)	(0.046)	(0.081)	(0.080)
Resources$_{t-1}$	-0.025	-0.018	-0.847	-0.341
	(0.038)	(0.037)	(0.649)	(0.617)
Muslim	0.020	0.005	0.195^{*}	0.212^{*}
	(0.037)	(0.037)	(0.075)	(0.075)
Intercept	-0.582^{*}	-0.467^{*}	-0.897^{*}	-0.697^{*}
	(0.055)	(0.054)	(0.164)	(0.158)
Observations	3,405	3,405	1,920	1,920
Log Likelihood	−2,235.188	−2,278.402	−1,250.404	−1,275.970
Akaike Inf. Crit.	4,482.376	4,568.803	2,512.809	2,563.939

Note: *p<0.05. Values in parentheses give standard errors. All real-valued variables scaled to $\sim N(0, 1)$.

S6 Analysis by Regime Type

Figure S4 presents results of the negligible effects analysis in the main text, stratified by the type of regime each state was classified as at the beginning of the time series. This analysis reveals some interesting variation: Although affective polarization is very weakly related to democracy in the full sample, this result appears to be a consequence of the phenomenon having different effects in different contexts. In particular, affective polarization appears to have a much more sizable effect in hybrid regimes and autocracies—contexts lacking in strong (or any) democratic institutions. Indeed, the 90% confidence intervals eclipse the lower bound for meaningful effects in two of these cases, suggesting that affective polarization may, indeed, have a detrimental causal impact on democracy in certain contexts.

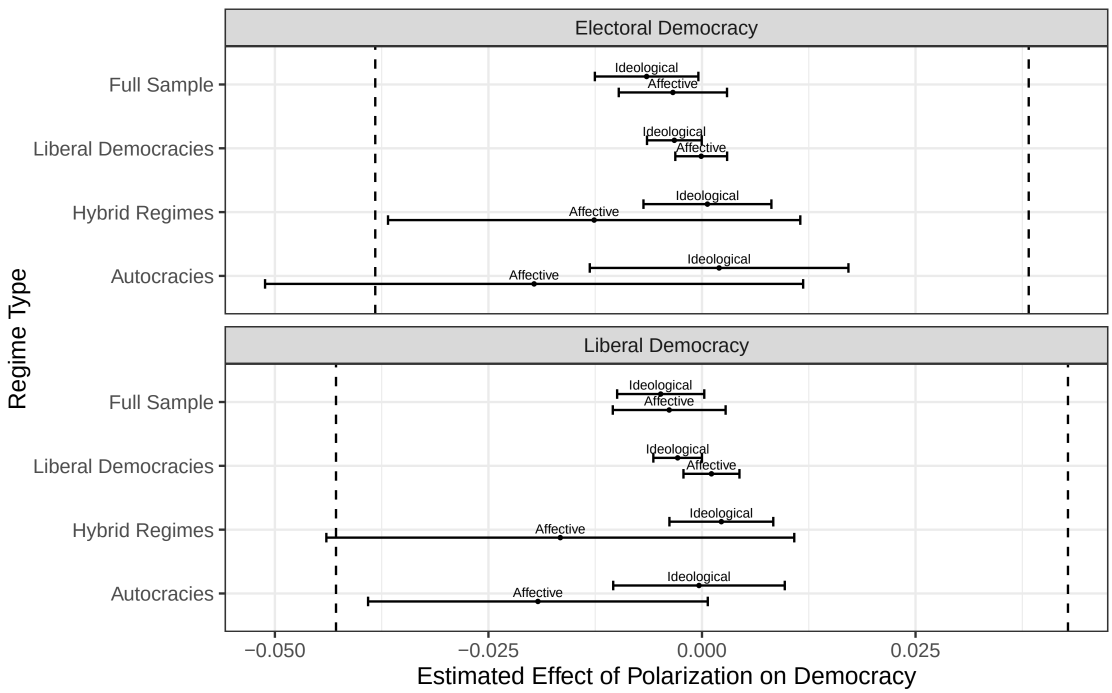

Figure S4: Testing for a Negligible Effect of Polarization on Democracy. Point estimates correspond to δ in (2) and (3) in the main text. Error bars give 90% confidence intervals. Dotted lines represent $-\tau$ and τ for each dimension of democracy.

References

Adida, Claire L. et al. (Oct. 2016). "Who's Asking? Interviewer Coethnicity Effects in African Survey Data". In: *Comparative Political Studies* 49.12, pp. 1630–1660. DOI: 10.1177/0010414016633487.

Airoldi, Edoardo M. et al. (2008). "Mixed Membership Stochastic Blockmodels". In: *Journal of Machine Learning Research* 9, pp. 1981–2014.

Bafumi, Joseph et al. (2005). "Practical Issues in Implementing and Understanding Bayesian Ideal Point Estimation". In: *Political Analysis* 13.2, pp. 171–187. DOI: 10.1093/pan/mpi010.

Baldassarri, Delia and Peter Bearman (Oct. 2007). "Dynamics of Political Polarization". In: *American Sociological Review* 72.5, pp. 784–811.

Biemer, Paul P. (Jan. 2010). "Total Survey Error: Design, Implementation, and Evaluation". In: *Public Opinion Quarterly* 74.5, pp. 817–848. DOI: 10.1093/poq/nfq058.

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

Caughey, Devin, Tom O'Grady, and Christopher Warshaw (Aug. 2019). "Policy Ideology in European Mass Publics, 1981–2016". In: *American Political Science Review* 113.3, pp. 674–693. DOI: 10.1017/S0003055419000157.

Caughey, Devin and Christopher Warshaw (2015). "Dynamic Estimation of Latent Opinion Using a Hierarchical Group-Level IRT Model". In: *Political Analysis* 23.2, pp. 197–211. DOI: 10.1093/pan/mpu021.

Claassen, Christopher (Jan. 2019). "Estimating Smooth Country–Year Panels of Public Opinion". In: *Political Analysis* 27.1, pp. 1–20. DOI: 10.1017/pan.2018.32.

Esteban, Joan-María and Debraj Ray (July 1994). "On the Measurement of Polarization". In: *Econometrica* 62.4, pp. 819–851.

Grimmer, Justin (2011). "An Introduction to Bayesian Inference via Variational Approximations". In: *Political Analysis* 19.1, pp. 32–47. DOI: 10.1093/pan/mpq027.

— (July 2013). "Appropriators Not Position Takers: The Distorting Effects of Electoral Incentives on Congressional Representation". In: *American Journal of Political Science* 57.3, pp. 624–642. DOI: 10.1111/ajps.12000.

Imai, Kosuke, James Lo, and Jonathan Olmsted (Nov. 2016). "Fast Estimation of Ideal Points with Massive Data". In: *American Political Science Review* 110.4, pp. 631–656. DOI: 10.1017/S000305541600037X.

Jackman, Simon (Dec. 2005). "Pooling the Polls Over an Election Campaign". In: *Australian Journal of Political Science* 40.4, pp. 499–517. DOI: 10.1080/10361140500302472.

Lewandowski, Daniel, Dorota Kurowicka, and Harry Joe (Oct. 2009). "Generating Random Correlation Matrices Based on Vines and Extended Onion Method". In: *Journal of Multivariate Analysis* 100.9, pp. 1989–2001.

Linzer, Drew A. (Mar. 2013). "Dynamic Bayesian Forecasting of Presidential Elections in the States". In: *Journal of the American Statistical Association* 108.501, pp. 124–134. DOI: 10.1080/01621459.2012.737735.

Mehlhaff, Isaac D. (June 2021). "A Group-Based Approach to Measuring Polarization". The University of North Carolina at Chapel Hill.

Ura, Joseph Daniel and Christopher R. Ellis (Jan. 2012). "Partisan Moods: Polarization and the Dynamics of Mass Party Preferences". In: *The Journal of Politics* 74.1, pp. 277–291.

Vehtari, Aki, Andrew Gelman, and Jonah Gabry (Sept. 2017). "Practical Bayesian Model Evaluation Using Leave-One-Out Cross-Validation and WAIC". In: *Statistics and Computing* 27.5, pp. 1413–1432.

Weisberg, Herbert F. (2005). *The Total Survey Error Approach: A Guide to the New Science of Survey Research*. Chicago: The University of Chicago Press.