

# Hierarchical models for estimating state and demographic trends in U.S. death penalty public opinion

Kenneth E. Shirley<sup>†</sup>

*AT&T Labs Research, Florham Park, USA*

Andrew Gelman

*Columbia University, New York, USA*

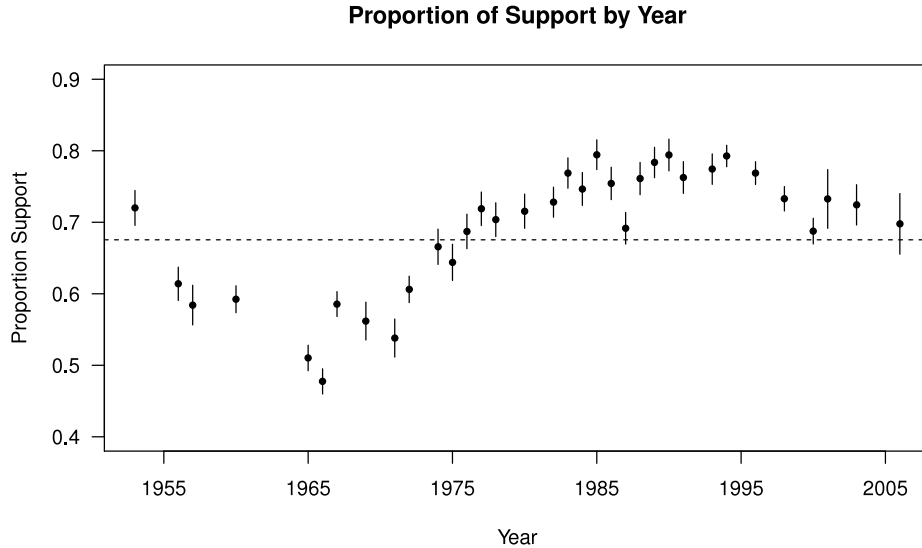
**Summary.** One of the longest running questions that has been regularly included in U.S. national public opinion polls is “Are you in favor of the death penalty for persons convicted of murder?” Because the death penalty is governed by state laws rather than federal laws, it is of special interest to know how public opinion varies by state, and how it has changed over time within each state. In this paper we combine dozens of national polls taken over a fifty-year span and fit a Bayesian multilevel logistic regression model to estimate support for the death penalty as a function of the year, the state, state-level variables, and various individual-level demographic variables. Among our findings were that support levels in northern and southern states have moved in opposite directions over the past 50 years, support among blacks has decreased relative to non-blacks, but at slightly different rates for men and women, and support among some education groups varies widely by region. Throughout the paper, we highlight the use of a variety of analytical and graphical tools for model understanding, including average predictive comparisons, finite population contrasts for overparameterized models, and graphical summaries of posterior distributions of group-level variance parameters.

*Keywords:* Gallup Survey, GSS, Hierarchical Bayes, MCMC, Multilevel Model, Political Science

## 1. Introduction

Capital punishment is perennially popular in the United States but is only legal in about two-thirds of the states (and is implemented rarely in many of these). To better understand the relationship between public opinion and policy, it would be desirable to know the support for the death penalty in each of the fifty states, and how this support has changed over time. To estimate state-level effects over time, it is necessary to control for the effects of national swings in public opinion as well as demographic variables, both of which are known to be large. As Figure 1 illustrates using poll data from the General Social Survey (GSS) and Gallup, national support for the death penalty has fluctuated substantially during the past fifty years, beginning with low support in the 1960s, increasing support throughout the 1970s (when capital punishment was ruled unconstitutional by the Supreme Court, followed by new rules under which the death penalty reappeared, one state at a time) high support in the 1980s (during which time a national concern about crime made the death penalty a prominent political issue), and finally, decreased support since the mid-1990s (when five states either explicitly illegalized or indirectly suspended the death penalty in part due to the exoneration of numerous death row inmates due to DNA evidence) (Baumgartner

<sup>†</sup>Most of this work was completed while the first author was a postdoc at Columbia University.



**Fig. 1.** The proportion of respondents who support the death penalty by year, with bars representing 95% confidence intervals, based on a combination of General Social Survey and Gallup polls. The dashed horizontal line represents the overall proportion of death penalty support across all years, which is 67.5%.

et al., 2008).<sup>‡</sup> Additionally, studies of individual surveys have repeatedly shown strong relationships between death penalty support and demographic variables such as sex, race, age, education, income, and religion, to name just a few (Fox et al., 1991; Ellsworth and Gross, 1994; Hanley, 2008).

Despite the extensive literature on the time series of aggregate death penalty support in the U.S. over the past fifty years, and also on individual-level predictors of death penalty support at specific time points (reviewed in detail in Section 2), there has been very little rigorous, simultaneous analysis of both. One of the challenges in modeling demographic, time series, and state-specific effects simultaneously is that even after combining multiple surveys, the data are still too sparse to estimate high-level interactions without some form of regularization, or shrinkage toward the mean. Furthermore, the national yearly fluctuations in death penalty support don't follow a straightforward pattern that can be embedded in a standard linear model. If such a model was fit to the differences between state-by-state levels of support and their average across states, this would be equivalent to treating the national average time series as if it were a known quantity, and the result would be the underestimation of the uncertainty of modeled quantities. Lax and Phillips (2009b) and Franklin (2001) discuss these problems in the context of public opinion polls and voting behavior.

Our approach to the problem is to fit a Bayesian multilevel logistic regression model to the data, where the model is overparameterized to allow for the simultaneous estimation of the national fluctuations in death penalty support as well as state-by-state deviations from

<sup>‡</sup>The issue even reached national politics, as in the 1988 presidential debate when death-penalty opponent Michael Dukakis was asked if his views would change if his wife were raped and murdered.

it. The hierarchical Bayesian approach handles the regularization required for sparse data via prior distributions for groups of parameters, and is a common approach for small area estimation (Rao, 2003). The overparameterization of the model is a consequence of modeling yearly effects using an AR(1) model, and also modeling state trends as linear deviations from the national average. We follow Gelfand and Sahu (1999) and use weakly informative priors for the unidentified parameters in order to speed convergence of the contrasts of exchangeable parameters within groups, rather than impose constraints to make the model identifiable.

The model can be fit using standard MCMC techniques, but assessing convergence, checking the fit of the model, and interpreting the parameter estimates requires a variety of non-standard tools for understanding complex hierarchical models. We highlight our use of these tools throughout our discussion of the fit of the model and the resulting conclusions about death penalty public opinion. First, when we analyze posterior samples, we compute finite population contrasts to compare different units within groups, and we focus our convergence assessments and posterior summaries on these quantities rather than the “raw” parameters from the model. Second, we include graphical summaries of the fitted model that visualize the variation among respondents within different groups, in order to discover which categorical predictors explain the most variation in the outcome. The large number of categorical predictors and resulting interaction terms in our model make the visual summary of variability absolutely necessary for model understanding. Last, we compute average predictive comparisons to provide an additional high-level summary of which predictors explain the most variation in death penalty support. This is how we compare, on an equal scale, how much death penalty support has changed as a function of time, state residency, and demographic variables – a novel comparison that requires the combination of pooling many surveys over time, fitting a complex model, and summarizing the model fit in a succinct way.

We find that public support for the death penalty is highly associated with certain demographic variables, such as sex, race, and education, which is consistent with previous research. Our model, however, provides novel estimates of how these effects have changed over time, and how they vary across states, especially with regard to opinion as a function of race and sex. We also find that support for the death penalty has changed significantly within certain states over time compared to the national average, holding constant the effects of demographic variables. In particular, we find that before the 1970s, capital punishment was more popular in the North than the South, a surprise given the current pattern in which the vast majority of executions are carried out in southern states. Additionally, we find that some of the variation among state trends can be explained by state-level variables including the legality of the death penalty and shifting partisan support over time, and the rest of the variation among states is explained by state-specific effects that we estimate with our model.

The models being developed and evaluated here are relevant not just for death sentencing and criminal justice but also more generally for studying the interactions between state-level opinion and policies, as discussed in literature including Erikson (1976), Datta et al. (1999), and Lax and Phillips (2009a). What is important is that the model allows for interactions between demographic, geographic, and time patterns, so that post-stratification can be done to estimate actual state-by-state support levels as accurately as possible (which is especially relevant for a public policy that faces a referendum vote).

The rest of the paper is organized as follows: Section 2 reviews previous work on death penalty opinions, Section 3 contains a description of the data to which we fit the model,

Section 4 contains a detailed description of the model, Section 5 contains a description of how we computed adjusted parameter estimates and post-processed our MCMC output, Section 6 contains a description of the model parameter estimates and other analysis of the fitted model, Section 7 discusses the goodness of the fit of the model via posterior predictive checks and residual plots, Section 8 discusses the results of out-of-sample predictions made by the model and similar competing models, and last, Section 9 contains a discussion of the results.

## 2. Public opinion on the death penalty

Opinion on capital punishment has received a large amount of attention in the political science literature for a variety of reasons. First, the death penalty has consistently been an issue of national interest since the earliest national opinion polls were conducted by Gallup in the mid 1930's; thus, there exists a large amount of historical data concerning death penalty public opinion. Second, there have been multiple Supreme Court decisions concerning death penalty laws that cite changing public opinion as a factor in determining whether aspects of the death penalty constitute "cruel and unusual punishment" (which is forbidden according to the eighth amendment in the U.S. Bill of Rights). In *Weems v. United States* in 1910 and later in *Trop v. Dulles* in 1958, Supreme Court decisions specifically pointed out that the definition of "cruel and unusual" can change over time according to societal standards (Vidmar and Ellsworth, 1974). Much later, in 2002, Supreme Court Justice John Paul Stevens specifically mentioned that public opinion polls provided insight concerning the public's feelings toward the death penalty for mentally retarded prisoners (Hanley, 2008). Although individuals rarely exert direct control over death penalty laws through state-level referendums, these Supreme Court precedents show that indirectly, death penalty public opinion affects public policy. A number of articles have explored this opinion-policy relationship (Erikson, 1976; Norrander, 2000).

The basic relationships between demographic variables and death penalty support have been well understood for decades, but higher-level interaction effects have been less studied. During the period in the 1970s when the Supreme Court was shaping modern death penalty policy, Vidmar and Ellsworth (1974) reviewed then-current public opinion of the death penalty based on a 1972 Gallup poll and previous work by Erskine (1970) tabulating poll results from the late 1960s. They found that higher support for death penalty was associated with respondents who were male, white, old, and less-educated. Analyses in the 1990s found similar results (Fox et al., 1991; Ellsworth and Gross, 1994). The question of whether these effects have changed over time has not been answered rigorously. Baumgartner et al. (2008) suggest that associations between demographic variables and death penalty support are mostly fixed over time, but they do not fit a quantitative model to back up this claim. Hanley (2008) considers changes in support as a function of sex and race over time, and claims that in the 1990s, support among all race  $\times$  sex subgroups was higher than it was in the 1970s, but does not consider the relative magnitudes of these differences (in Section 6 we show that while absolute levels of support may have been higher for all groups in the 1990s, the extent to which support among blacks was below the national average was increasing, and at slightly different rates for men and women). Hanley (2008) also points out that the relationship between age and support varied during the period from 1970-2000. Last, they found that the negative correlation between education level and support is not monotonic; the two least supportive educational groups are those without a high school degree, and

those with a graduate degree of some kind (these two groups occupy opposite ends of the education scale). Our results in Section 6 are consistent with this finding.

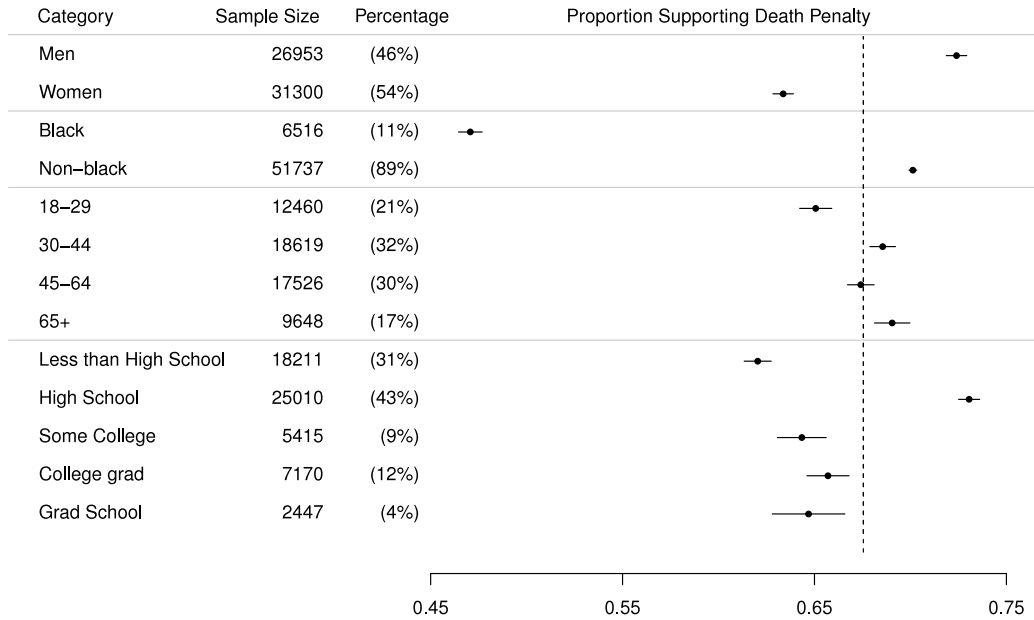
The most recent and thorough time series model of death penalty public opinion is found in Baumgartner et al. (2008). They focused on modeling long-term trends in death penalty public opinion as a function of changing media coverage, using an index of support for death penalty based on a weighted average of yearly changes in support of death penalty from 292 statewide and national surveys between 1953 and 2006 that asked about the death penalty using different question wordings. They found a relationship between changes in public opinion and the tone of media coverage and levels of crime. They don't, however, model individual responses simultaneously as a function of time and demographic variables.

### 3. The data

We put together data on public opinion of the death penalty in two stages. We started with the 21 General Social Survey (GSS) polls given between 1974 and 2000, all of which asked the question: "Do you favor or oppose the death penalty for persons convicted of murder?" Second, to increase the time span of our observed data, we included data from Gallup polls taken before 1974 and after 2000 that asked a similar question: "Are you in favor of the death penalty for persons convicted of murder?" We searched for these Gallup polls in the archive of the Roper Center for Public Opinion Research, and to our knowledge, we included every Gallup poll in the Roper archive that (1) asked the question of interest, and (2) was given before 1974 or after 2000, with the exception of the Gallup polls given in 1936 and 1937, which we excluded because they didn't include detailed information about the education level of each respondent. The source of the poll is not a significant factor in the level of support expressed by respondents: Baumgartner et al (2008) found that the correlation between support levels from the two polling sources, GSS and Gallup, is 0.90 for the years in which the both asked the question of interest, and Schuman and Presser (1981) show that having formal balance in a survey question (explicitly suggesting a positive *and* negative answer in the question) rarely affects the outcome. In all, we modeled data from 34 polls, all of which were taken in distinct years between 1953 and 2006, where the maximum number of years between consecutive polls was 5 years (between the 1960 and 1965 Gallup polls). The number of respondents per poll ranged from 445 to 3085, and the total number of responses was  $N = 58,253$ . We didn't weight the responses using the given survey weights because we include as predictors all the variables that were used to create the weights, and we compute post-stratified estimates ourselves (see Section 6).

Between 1953 and 2006, the proportion of poll respondents supporting the death penalty in a given year fluctuated between 47% and 79%, excluding those who had no opinion (the proportion of respondents with no opinion rarely exceeded 10%). Figure 1 contains a plot of death penalty support by year during this time. In each of the 34 polls we recorded the state of residence of each respondent, with the District of Columbia considered as the 51st state. The number of respondents per state per year was highly imbalanced, ranging from an average of 0.5 respondents per year in Hawaii (17 responses among the 34 surveys) to 170 respondents per year in California. We also classified states into 4 regions—Northeast, South, Midwest, and West—according to the U.S. Census state classification (with the District of Columbia included in the South). The other information we used for our models was demographic information about each respondent, consisting of the following four variables:

## EDA of Demographic Data



**Fig. 2.** Exploratory data analysis of the demographic variables. The points are the proportions of respondents in each group who supported the death penalty (pooled across all years and states), and the horizontal lines are 95% confidence intervals for these proportions. Of the four demographic variables recorded, the largest difference in death penalty support exists between blacks and non-blacks. Age and education, which are ordinal variables, have non-monotonic relationships with death penalty support. The dashed vertical line is the overall proportion of death penalty support across all subgroups, which is about 67.5%.

- Sex (male or female)
- Race (black or non-black)
- Age (a categorical variable with 4 levels, where 1 = 18–29, 2 = 30–44, 3 = 45–64, and 4 = 65 or older).
- Education (measured as the highest degree achieved by the respondent, a categorical variable with 5 levels, where 1 = Less than high school, 2 = High school, 3 = Some college or trade school, 4 = College graduate, and 5 = Graduate degree). From this point forward, we call this variable “Degree.”

Figure 2 displays some summary statistics of the distributions of each of these four variables, and includes the sample percentage of respondents in each main demographic category who supported the death penalty.

Last, we considered three state-level variables as potential explanatory variables in our model. The first two state-level variables are related to the state’s partisan political support, as measured by their support for Republican vs. Democratic presidential candidates throughout the years. Specifically, for each presidential election year from 1952–2004, we recorded the Republican share of the vote in each state, discarding third party votes, and

then we fit a linear regression model separately for each state to these percentages using time as the sole predictor. We recorded the estimated intercept and slope of this regression model for each state to create two state-level variables per state, which we call “Republican Share Intercept” and “Republican Share Slope.” We used this simplified predictor instead of using each individual state-time variable because presidential election years didn’t match the years of the surveys.

Third, for each year from 1953 to 2006, we recorded the percentage of years in this 54-year time span that the death penalty was legal in each state. There is only a moderate amount of variation in this variable across states. The death penalty was legal in 40 states from 1953-1972 (and in Oregon from 1953-1963), and then a supreme court decision, *Furman v. Georgia*, effectively rendered every state’s death penalty statute illegal in June, 1972. About 30 states rewrote their death penalty statute during the next five years, and seven other states rewrote their statutes at some point during the following 20 years.§ In 1976, another supreme court decision, *Gregg v. Georgia*, overturned the 1972 ruling and made the death penalty legal again.¶ For our definition of this variable, which we call “Legality,” if a state rewrote their death penalty statute between 1972 and 1976, we include these years as additional years in which the death penalty was legal for that state, even though technically every state was waiting for the court system to rule on the new version of the law during this time, and no executions were attempted. We code the variable this way to measure the degree to which each state supported the death penalty legislatively over this period; it is a slightly less precise variable than the strict percentage of years in which the death penalty was legal, but it increases the variation of this state-level variable, and it may be associated with public opinion. Across all states, the variable Legality has a maximum of 1 (Florida rewrote their death penalty law in 1972, the same year as *Furman v. Georgia*), and its minimum is 0 (10 states have never had a death penalty statute during this time span). The mean and standard deviation of Legality are 0.70 and 0.39, respectively.

In summary, there are six categorical variables: year, state, sex, race, age, and education. Age and education are actually ordinal, but we include them as unordered categories in our model. Of the  $54 \times 51 \times 2 \times 2 \times 4 \times 5 = 220,320$  distinct categories, we have at least one observation from only 24,103 of them. The high-level interaction effects between sets of these categorical variables requires a high ratio of parameters to data points, motivating the use of a regression model that allows for shrinkage, or regularization, such as a multilevel Bayesian model.

#### 4. The model

We fit a series of multilevel models to the the data with the main goal of understanding changes in public opinion of the death penalty for different states and demographic groups across time. We began by fitting some exploratory models using OLS regression to see which main effects and interaction effects were associated with the response, and we used this process to guide which effects to include in the smaller number of multilevel Bayesian models that we fit. We ultimately fit five multilevel Bayesian models of varying complexity to a portion of the data designated for training, and we measured the predictive accuracy of each

§Iowa, West Virginia, and the District of Columbia have never rewritten their death penalty statutes since 1972.

¶In both the 1972 decision and the 1976 decision, the Supreme Court ruled on a specific death penalty case, and their ruling set a precedent that was applicable to death penalty laws nation-wide.

model (in terms of deviance) on a test set. These models, which we will describe in detail in Section 8, are differentiated by the complexity of the interaction effects they included. In this section we present the simplest model that fit the data well (which we denote the “main model”), as measured by a combination of criteria including out-of-sample prediction error, posterior predictive checks, prior subject area knowledge, and interpretability. In Section 8 we describe the gains we achieved by using this model over others.

The first level of the main model states that

$$\begin{aligned}
p(Y_i = 1) = & \text{logit}^{-1} \left( \alpha_{(s,t)[i]}^{\text{state-year}} + \alpha_{(d,s)[i]}^{\text{degree-state}} + \alpha_{(a,s)[i]}^{\text{age-state}} + \delta_{(a,s)[i]}^{\text{age-state}} X_i^{\text{year}} \right. \\
& + \beta_{s[i]}^{\text{black-state}} X_i^{\text{black}} + \delta_{s[i]}^{\text{black-state}} X_i^{\text{black}} X_i^{\text{year}} \\
& + \beta_{s[i]}^{\text{female-state}} X_i^{\text{female}} + \delta_{s[i]}^{\text{female-state}} X_i^{\text{female}} X_i^{\text{year}} \\
& \left. + \beta_{s[i]}^{\text{black-female-state}} X_i^{\text{black}} X_i^{\text{female}} + \delta_{s[i]}^{\text{black-female-state}} X_i^{\text{black}} X_i^{\text{female}} X_i^{\text{year}} \right), (1)
\end{aligned}$$

for individual responses  $i = 1, \dots, 58253$ , states  $s = 1, \dots, 51$ , years  $t = 1, \dots, 54$ , degrees  $d = 1, \dots, 5$ , and ages  $a = 1, \dots, 4$ .  $X_i^{\text{year}}$  is the year of response  $i$  measured as a continuous variable, scaled to have a mean of zero and a standard deviation of 1 ( $X_i^{\text{year}} = 0$  corresponds to a response given in the mean survey year, 1980).  $X_i^{\text{black}}$  and  $X_i^{\text{female}}$  are likewise scaled to each have mean 0 and standard deviation 1 (which means that a black woman has the value  $(X^{\text{black}}, X^{\text{female}}) = (2.82, 0.93)$ , and a white man has the value  $(X^{\text{black}}, X^{\text{female}}) = (-0.35, -1.08)$ ). This coding scheme for race and sex is non-standard, but we use it because it effectively sets the baseline group,  $(X^{\text{black}}, X^{\text{female}}) = (0, 0)$ , to the population average rather than one particular race/sex combination. The subscript notation  $s[i]$  denotes the state of residence,  $s = 1, \dots, 51$ , for individual  $i$ .

The main feature of the priors (or higher-level models) for the parameters in Equation 1 is that every group of state-level parameters is normally distributed around a regional mean. Additionally, the state-year interaction effects have a structured model of their own. There is a lot of repetition in the setup of the priors, so here we write down some of them in full, and later we explain how the rest of the priors are analogous to these.

$$\begin{aligned}
\alpha_{(s,t)}^{\text{state-year}} & \sim N(\alpha_t^{\text{year}} + \alpha_s^{\text{state}} + \delta_s^{\text{state}} X_t^{\text{year}}, \sigma_{\text{state-year}}^2), \\
\alpha_t^{\text{year}} & \sim N(\mu + \mu_\delta X_t^{\text{year}} + \phi(\alpha_{t-1}^{\text{year}} - \mu - \mu_\delta X_{t-1}^{\text{year}}), \sigma_{\text{year}}^2), \\
\alpha_1^{\text{year}} & \sim N(\mu + \mu_\delta X_1^{\text{year}}, \sigma_{\text{year}}^2 / (1 - \phi^2)), \\
\alpha_s^{\text{state}} & \sim N(\alpha_{r[s]}^{\text{region}} + \beta \mathbf{X}_s^{\text{state}}, \sigma_{\text{state}, r[s]}^2), \\
\delta_s^{\text{state}} & \sim N(\delta_{r[s]}^{\text{region}} + \gamma \mathbf{Z}_s^{\text{state}}, \tau_{\text{state}, r[s]}^2), \\
\alpha_{(a,s)}^{\text{age-state}} & \sim N(\alpha_{(a,r[s])}^{\text{age-region}}, \sigma_{\text{age-state}, (a,r[s])}^2), \\
\alpha_{(r,s)}^{\text{age-region}} & \sim N(\alpha_a^{\text{age}}, \sigma_{\text{age-region}, a}^2), \\
\beta_s^{\text{black-state}} & \sim N(\beta_{r[s]}^{\text{black-region}}, \sigma_{\text{black-state}, r[s]}^2), \\
\beta_r^{\text{black-region}} & \sim N(\beta^{\text{black}}, \sigma_{\text{black-region}}^2).
\end{aligned} \tag{2}$$

$$\tag{3}$$

We use  $N(0, 5^2)$  priors for  $\mu$  and  $\mu_\delta$  (the national average mean and trend), for  $\beta_j$  and  $\gamma_j$  for  $j = 1, 2$  (the effects of state-level variables on state slopes and intercepts), and for  $\beta^{\text{black}}$  (the effect of race). The regional intercepts and slopes ( $\alpha_r^{\text{region}}$  and  $\delta_r^{\text{region}}$ ) and the set of age effects ( $\alpha_a^{\text{age}}$ ) are given normal priors with a mean of zero and (unknown)



variances,  $\sigma_{\text{region}}^2$ ,  $\tau_{\text{region}}^2$ , and  $\sigma_{\text{age}}^2$ , respectively. The AR(1) parameter for the yearly effects,  $\phi$ , is given a  $U(-1,1)$  prior. We also specify that the prior distribution for every standard deviation parameter is a half- $t$  distribution with scale 5, and 3 degrees of freedom. There are 110 such parameters in the model.  $\mathbf{X}^{\text{state}}$  and  $\mathbf{Z}^{\text{state}}$  denote the  $(51 \times 2)$  matrices of state-level covariates that affect the state intercepts and slopes, respectively, where  $\mathbf{X}_s^{\text{state}} = (\text{Republican Share Intercept, Legality})_s$ , and  $\mathbf{Z}_s^{\text{state}} = (\text{Republican Share Slope, Legality})_s$  for states  $s = 1, \dots, 51$ .

The rest of the prior distributions we use are identical in structure to some of those listed above. First, the prior distributions for the degree-state intercepts,  $\alpha_{(d,s)}^{\text{degree-state}}$ , are exactly the same as the priors for  $\alpha_{(a,s)}^{\text{age-state}}$ , where “degree” replaces “age” in every specification, and there are five levels of degree effects (rather than four for the age effects). Next, the prior distributions for the age-state slopes,  $\delta_{(a,s)}^{\text{age-state}}$ , are exactly the same as the priors for the age-state intercepts, except that slope parameters,  $\delta$ , replace the intercept parameters,  $\alpha$ , and the standard deviation parameters are denoted by  $\tau$  rather than  $\sigma$ . Last, there are five additional sets of race-sex effects whose priors are not listed above. Each set has a prior distribution identical in structure to the prior distribution for  $\beta_s^{\text{black}}$ , where intercepts,  $\beta$ , and slopes,  $\delta$ , for individual states are normally distributed around regional means, which are, in turn, normally distributed around the grand mean, which is given a weakly informative  $N(0, 5^2)$  prior. The supplemental files contain a graphical illustration of the full model in the form of a directed acyclic graph (DAG).

We opt to model the national average yearly effects,  $\alpha_t^{\text{year}}$ , as an AR(1) process with a linear trend, where the differences between individual states and this national average yearly pattern are modeled as linear (on the logistic scale). We assume the AR(1) process is stationary so that we can estimate the overall mean across years (rather than conditioning on the first year, or anchoring the mean to a given year as we would have to do if the process weren’t stationary). We chose not to expand the AR(1) model to individual states because we felt the assumption that each state was individually stationary might not be realistic. We discuss the goodness-of-fit of the linearity assumption for states in Section 7.

We fit the model using JAGS (Plummer, 2003), which implements a mix of Gibbs sampling, slice sampling, and Metropolis jumping, and we performed all pre- and post-processing in R. Before we ran the MCMC, we computed the binomial count of those who supported the death penalty for each observed state-year-demographic 6-way combination (there were 24,103 unique state-year-demographic combinations with at least one observation in the data), so as to save time in the model fitting by modeling the sufficient statistics rather than each individual data point.

## 5. MCMC Convergence and Post-processing

We ran the MCMC algorithm on three separate chains for 10,000 iterations each, and we saved every fifth iteration among the last 5,000 to form a posterior sample of size 1,000 for each of the three chains. (We thinned the output only for convenience of manipulating a smaller amount of posterior output in R).

One of the main challenges to understanding the raw output from the MCMC algorithm is that the model we fit is overparameterized (see Gelfand and Sahu (1999) and Chapter 19.4 of Gelman and Hill (2007) for background on intentionally overparameterized models). This overparameterization comes in two different varieties, and in each case we disentangle non-identifiable parameters using parameter adjustments in the post-processing stage of

model-fitting. The first case is simple: We adjust our MCMC output to reflect finite population contrasts, because for many of our groups of parameters (such as groups of states, age levels, and education levels), the members of the group constitute the entire population, rather than a sample from a larger population. Doing this allows us to make more precise inferences about differences between observed units in a given group (Gelman and Hill, 2007). So, for example, when a group of parameters is distributed around an unknown mean, such as the set of race intercepts,  $\beta_s^{\text{black-state}}$  and their mean,  $\beta^{\text{black}}$ , we compute an adjusted version of each parameter:

$$\beta_s^{\text{black-state}'} = \beta_s^{\text{black-state}} - \beta^{\text{black-state}}, \quad (4)$$

$$\beta^{\text{black}'} = \beta^{\text{black-state}}. \quad (5)$$

where the dot subscript denotes the (unweighted) mean of a vector (weights are relevant for post-stratification, as described in Section 6.4, but not for computing finite population contrasts). The final set of these adjustments is made slightly more complicated by the nesting of state effects within regions, but the basic principle remains the same.

The second form of overparameterization is slightly more complicated to disentangle, and is related to the model for the time trend of death penalty support. One of the key components of our model that allows us to simultaneously estimate national fluctuations in opinion as well as state-by-state trends is the AR(1) model we use for the yearly effects, which treats the years as different levels of a categorical variable. We also modeled state effects as linear time trends with a mean linear trend,  $\mu_\delta$ . Such an overparameterized model is used because it preserves the uncertainty in the overall year-to-year fluctuations in death penalty support while simultaneously modeling state deviations from this pattern in a parsimonious way. The result is a lack of identifiability among the yearly effects and the mean slope. In order to capture the true mean slope, we compute the adjusted version of  $\mu_\delta$  by summing over every component of the model that allows for a linear trend in death penalty support:

$$\mu'_\delta = \frac{\sum_{t=1}^T X_t^{\text{year}} (\alpha_{.t}^{\text{state-year}} - \alpha_{..}^{\text{state-year}})}{\sum_{t=1}^T (X_t^{\text{year}})^2} + \delta_{..}^{\text{age-state}}. \quad (6)$$

The first term on the right side of Equation 6 is the estimated slope of death penalty support embedded within the state-year effects, and the second term is the mean slope across all age-state combinations. Similarly, to capture more precise estimates of the yearly effects, subtracting out the mean slope, we compute the adjusted yearly effects:

$$\alpha_t^{\text{year}'} = \alpha_{.t}^{\text{state-year}} - \alpha_{..}^{\text{state-year}} - (\mu'_\delta - \delta_{..}^{\text{age-state}}) X_t^{\text{year}}. \quad (7)$$

These adjustments result in more precise estimates of the quantities of interest, which are sets of centered parameters whose mean is zero, and the accompanying means themselves, which are adjusted to be identifiable. The full set of adjustments that we make is detailed in the supplementary materials.

We computed the potential scale reduction factor (Gelman and Rubin, 1992) and the effective sample size for each adjusted parameter (there were 4099 of them). The range of the potential scale reduction factors was (1.00, 1.03), indicating that the chains mixed well on all measured dimensions. The effective sample sizes of these adjusted parameters ranged from about 200 to 3000 (where an effective sample size of 3000 means that the autocorrelation of the three chains for a given parameter was virtually zero).

## 6. Results

We discuss our results in four sections: First we discuss our inference about the national average of support for the death penalty over time, which is modeled as a linear trend with first-order autoregressive yearly effects. Second, we investigate trends in support related to states and regions, including the effects of state-level variables on these trends. Third, we look at the association between demographic variables and death penalty support. Last, we compute estimates of actual support in each state and year by using post-stratification weights calculated from census data, to account for the varying demographic compositions of each state in each year. In all the discussions of the fit of the model, we include a variety of graphical summaries of the parameter estimates to visualize our inferences.

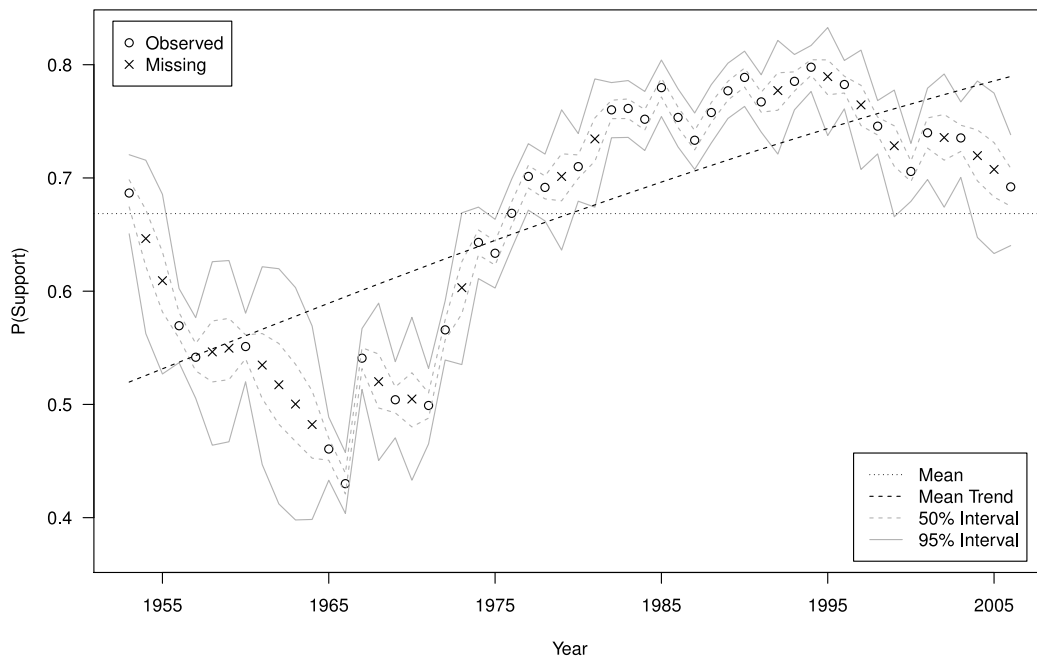
### 6.1. National average trends and yearly effects

The quantity  $\text{logit}^{-1}(\mu' + \mu'_\delta X_t^{\text{year}} + \alpha_t^{\text{year}'})$  is the mean support for the death penalty in year  $t$  for a respondent from a state with average support, and with average demographic variables. By “average” demographic variables, we mean that the respondent’s age and education level are average across levels, and their values of  $X^{\text{black}}$  and  $X^{\text{female}}$  are (the impossible values of) zero. This means that the national average yearly trend will not correspond directly to any particular baseline group, but is the average across all subgroups defined by the other variables in the model (where the average is taken on the logistic scale, and then converted to the probability scale).

The mean proportion of support for the death penalty for the average respondent in this time period is estimated to be about 66.8% (calculated as the posterior mean of  $\text{logit}^{-1}(\mu')$ ), and the mean linear change in support per year is estimated to be about 0.50%. This means that the linear component of support for the death penalty for the average respondent increased by about 1% every 2 years during the years 1953-2006, or by a total of about 27% during this time span. This trend is modeled as being linear on the logistic scale ( $\mu'_\delta = 0.37 \pm 0.03$ ), which means that the actual *proportion* of people supporting the death penalty is not technically modeled as linear. The difference, however, is slight: on the probability scale, the estimated curve has a slope of about 0.58% at the beginning of the time span (1953-54), and a slope of about 0.39% at the end of the time span (2005-06).

The yearly effects, which are modeled as deviations from the linear time trend, follow an AR(1) model (on the logistic scale) where  $\hat{\phi} = 0.92$  and the standard error of  $\phi$  is about 0.06. The posterior distribution of  $\phi$  is skewed to the left since it is bounded on the right by 1. The year-to-year standard deviation,  $\sigma_{\text{year}}$ , is about 0.17 ( $\pm 0.03$ ) on the logistic scale. On the probability scale, this means that the estimated standard deviation of the change in one year’s proportion of support, given the previous year’s proportion, is about 3-4%. The marginal standard deviation of the estimated yearly support for an average respondent is about 11.1% (with a mean, as we said earlier, of about 66.8%).

Figure 3 shows posterior means and intervals for the proportion of death penalty support in each year for an average respondent, including the posterior means of the intercept and slope of the linear trend. Each yearly estimate is essentially a weighted average of the observed proportion of support in that year, the linear trend across all years, and the proportions in neighboring years.



**Fig. 3.** National average trend and yearly effects. The points are posterior means of the estimated proportion of support for an average respondent in each year from 1953-2006. The circular points denote the years in which we have survey data, and the x-shaped points denote years in which there was no survey data. As expected, the intervals are wider for years of missing data, and are especially wide when there is a multi-year gap between consecutive surveys.

## 6.2. State and regional trends

There was substantial variation in the trends among states. To estimate the difference between a given state’s trend and the national average trend, we computed, for each state  $s$  and year  $t$ , the difference between the estimated state trend, on the probability scale, including the effects of state-level variables, and the estimated national average trend on the probability scale:

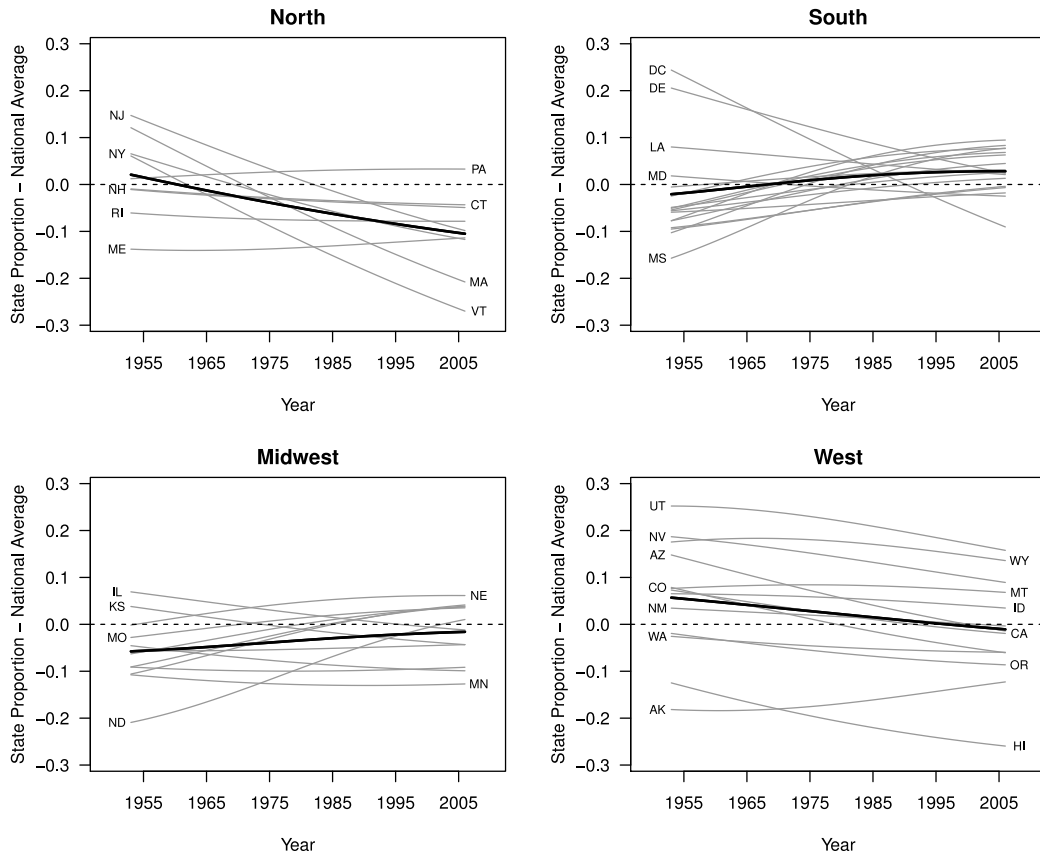
$$\begin{aligned}
 & \text{logit}^{-1}(\mu' + X_t^{\text{year}} \mu'_\delta) && \text{(National average trend)} \\
 & -\text{logit}^{-1}(\alpha_{r[s]}^{\text{region}'} + \delta_{r[s]}^{\text{region}'} X_t^{\text{year}} && \text{(Regional trend)} \\
 & \quad + \alpha_s^{\text{state}'} + \delta_s^{\text{state}'} X_t^{\text{year}} && \text{(State trend)} \\
 & \quad + \boldsymbol{\beta} \mathbf{X}_s^{\text{state}} + \boldsymbol{\gamma} \mathbf{Z}_s^{\text{state}} X_t^{\text{year}}), && \text{(State-level variables trend)}
 \end{aligned}$$

where  $r[s]$  is the region of state  $s$ , and the full definitions of the adjusted parameters are available in the supplementary materials. In this comparison we ignore the state-year effects for each state, and the national yearly effects, focusing solely on differences in estimated trends. Figure 4 plots these estimated differences, grouping states by region. The different patterns of variation between the four regions are clear—the western states are the most variable in their levels of support throughout the entire time span (with Utah respondents showing relatively high support and Hawaii respondents showing relatively low support), and the northern states are the most variable in their slopes (with Massachusetts and Maine exhibiting relatively low and high slopes, respectively). The average support among western and northern states decreased over time relative to the national average. The midwestern states are somewhat less variable in their slopes, and, along with southern states, gradually increased their support over the time span, relative to the average. Interestingly, only three out of 51 states – Massachusetts, Vermont, and DC – have an estimated trend that is decreasing *absolutely* (after including the positive national average trend). The states with the fastest increasing estimated trends were all in the south: Mississippi, Alabama, and Georgia.

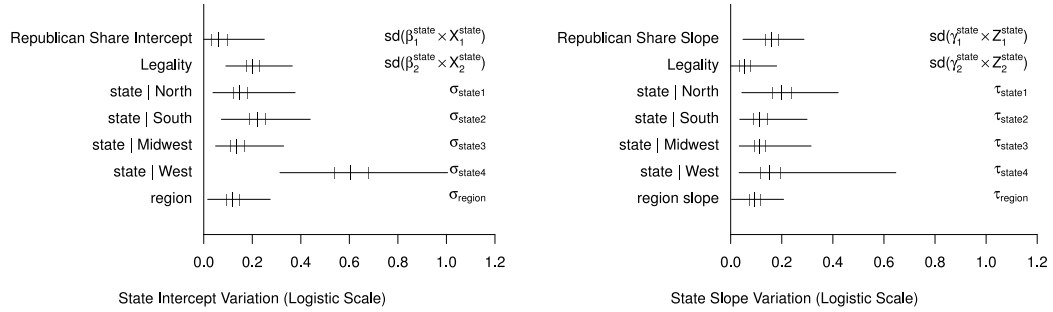
Our model explains the variation among state trends pictured in Figure 4 using three types of variables: State-level variables, regional effects, and state-specific variation. Figure 5 summarizes the amount of variation explained by each of these three sets of variables using point estimates and intervals of the group-level standard deviation estimates. We describe these effects in more detail in the following two subsections.

### 6.2.1. Effects of state-level variables

The estimated effects of the state-level variables “Republican Share Intercept” and “Legality” on the intercepts of the state trends ( $\boldsymbol{\beta}$  from Equation 3) are  $-0.06$  and  $0.20$ , with standard errors of  $0.06$  and  $0.04$ , respectively. The first estimate shows that higher average levels of support for Republican presidential candidates during this time span are associated with lower levels of death penalty support, but the interval estimate of this effect contains zero, so it is not statistically significant. On the other hand, the estimated effect of legality of the death penalty on state intercepts is large and statistically significant, indicating that states where the death penalty was legal during a longer proportion of the time span 1953-2006 also show higher average levels of support for the death penalty.



**Fig. 4.** Posterior means of the differences in trends of support between states and the national average for a respondent of “average” demographics (see Section 6.1), grouped by region, where the dark lines are regional means and the trends include the effects of state-level variables. The curves are not linear because of the transformation from the logistic scale to the probability scale. All four plots are on the same scale to highlight the differences in variation among states between the four regions. Support among western states is the most variable throughout the time span observed here, while the state slopes vary the most among northern states. Not every state could be labeled here; please see the Table 2 of the supplementary materials for slopes and intercepts describing all 51 estimated trend differences.



**Fig. 5.** The black, central vertical hash marks are point estimates of group-level finite-population standard deviation parameters, and the gray hash marks and horizontal lines represent 50% and 95% intervals for these parameters. In the left plot, it is clear that the intercepts of the western state-specific effects (not including state-level variables) have the largest variability (where the intercepts refer to support levels in the mean year, 1980), and “Legality” explains nearly as large a proportion of total variation as the state-specific intercepts among the southern states, which comprise the second-most variable set of state-specific intercepts among the four regions. The right plot illustrates that the state-specific slopes for northern states account for the most variation among all sets of predictors related to state slopes.

A different story emerges when we look at the effects of state-level variables on the *rate of change* of death penalty support over time (the parameters  $\gamma$  from Equation 4). In this case, the estimated effect of the slope of Republican presidential support over this time span on the slope of state-level death penalty support is 0.16 with an sd of 0.03. That is, states that increased their relative support for Republican presidential candidates over this time span also tended to increase their death penalty support during this time, relative to the national trend, and the association is large and statistically significant. The converse is also (necessarily) true: states that decreased their relative support for Republican presidential candidates during this time period also tended to decrease their level of support for the death penalty relative to the national average. The legality of the death penalty has a statistically insignificant interaction with the slope of state-level support for the death penalty (the estimated effect is 0.05 with an SD of 0.03).

From Figure 5, you can see that these state-level variables explain a substantial amount of the variation in state trends—comparable to the amount of variation at the state and region levels that is accounted for by the sets of state-specific variables.

6.2.2. Regional effects and state-specific effects

The overall state trends also depend on regional effects, and state-specific trends centered at zero (in their slopes and intercepts) within each region. The posterior means of both sets of these effects are illustrated in Figure 1 of the supplementary file “Supplementary-figures-and-tables.pdf”. The state-specific varying intercepts and slopes estimate the predictive effects of residing in a given state that are not already explained by whatever state-level variables are included in the model—in our case, each state’s Republican voting trend and proportion of years of legality.

Recall that the amount of variability among the state-specific intercepts and slopes is

allowed to be different for each of the four regions; in fact, the data justify these separate estimates of variability (previous models assuming equal variance across regions didn't fit as well). The posterior means of the standard deviations of the state-specific varying intercepts and slopes are, in order of the regions (North, South, Midwest, West),  $\hat{\sigma}_{\text{state}} = (0.15, 0.22, 0.14, 0.61)$ , and  $\hat{\tau}_{\text{state}} = (0.20, 0.12, 0.12, 0.16)$ . Recall that Figure 4 displayed state trends that included the effects of state-level variables. Here we are describing the residual trends in each state attributed to unmeasured state-level variables. The regional means of the state-specific slopes and intercepts are  $\hat{\alpha}^{\text{region}} = (-0.10, -0.03, -0.05, 0.16)$ , and  $\hat{\delta}^{\text{region}} = (-0.02, -0.01, 0.11, -0.11)$ . In other words, these are the estimated intercepts and slopes of death penalty support relative to the national average for a respondent of average demographics from a random state in one of these regions. The average of the state-specific varying slopes for northern states is almost zero, whereas most northern states have strong decreasing trends of support compared to the national average as pictured in Figure 4. This is because most of the decline in support for the death penalty among northern states is accounted for by their decreased support for Republican presidential candidates during this time span. This isn't a *causal* effect, but it is merely an association between Republican vote share and death penalty support.

We'll make a few more comments here regarding state-specific trends:

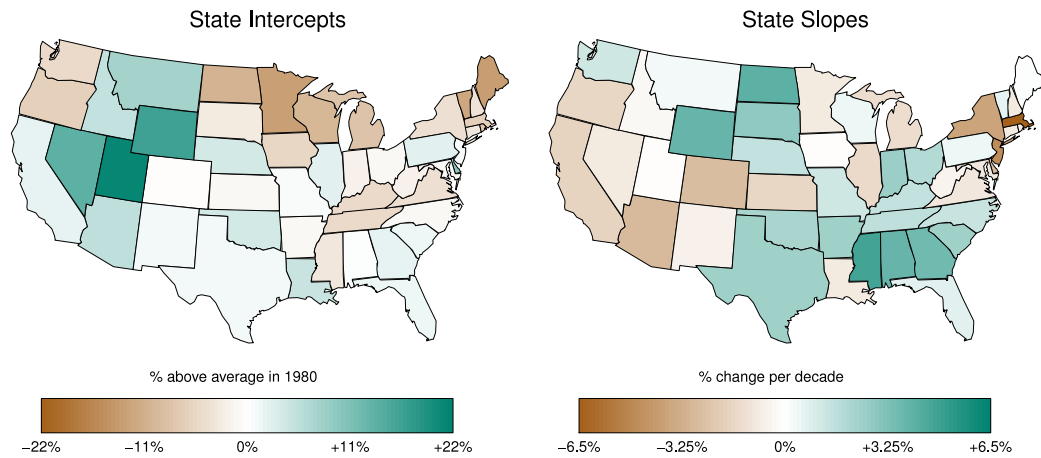
- Among the 9 northern states, Maine and Rhode Island have positive state-specific slopes, and relatively flat slopes when the effects of state-level variables are included. The rest of the northern states have relatively flat or negative state-specific slopes, and negative slopes when the effects of state-level variables are included. The three northern states with the fastest-decreasing support for the death penalty over this time period are Massachusetts, New Jersey, and Vermont. Most of Vermont's decrease in death penalty support can be accounted for by Vermont's declining support for Republican presidential candidates over this time period.
- The state-specific slopes of the western states show little variation. Their overall levels of support, though, are highly variable, with Hawaii and Alaska showing low support for the death penalty, and Utah and Wyoming showing high support. Most of the lower levels of support in Hawaii and Alaska can be accounted for in the model by the fact that the death penalty has never been legal in those states, whereas for the rest of the western states, it has been legal during most of the time span in question.
- Although the District of Columbia and Delaware are classified by the U.S. Census Bureau as southern, their slopes are much more similar to the northern states' slopes. Their slopes are negative, indicating declining support for the death penalty relative to the national average. This is not entirely surprising, since they are geographically contiguous to the northern states. We further discuss regional effects in Section 9.

Last, we visualize the intercepts and slopes for each state relative to the national average using colored maps in Figure 6. Some regional correlations are visible in the maps, but overall the maps make it clear that the variation between the states is greater than that between the regions.

### 6.2.3. *State-year interaction effects*

Our model also includes state-year interaction effects to account for additional variation on the state-year level. For example, when a highly-publicized crime occurs, or when a murder





**Fig. 6.** The slopes and intercepts of the state trends in death penalty support relative to the national average (67%) are visualized here on maps of the United States using a color scale. Time was scaled to have a mean of zero, which means that the intercepts are estimated levels of support for the mean year of the sample, which is 1980 (meaning that the left map would be colored differently for a comparison in a different year). Alaska and Hawaii are omitted.

trial receives a large amount of media attention, it is plausible that death penalty support in that state at that time could experience a relatively sudden increase or decrease that would not be captured by the state trends that are included in our model. The posterior mean of the standard deviation of the state-year interaction effects,  $\sigma_{\text{state-year}}$ , was 0.27 with a standard error of about 0.02; in other words, the state-year interaction effects explain a substantial amount of variation in the response, and the estimate of their variability is precise. The precision of the estimate of  $\sigma_{\text{state-year}}$  is partially a result of the large number of state-year interactions that are contained in the model (there are  $51 \times 54$  of them)—it is fairly easy to estimate the variability of such a large set of parameters compared to, say, the variability of the slopes of states within a region, where there are only about 10 parameters in the group.

A few examples of individual state-year effects are visualized in Figure 12 (in Section 7). They are re-centered (see the supplement on parameter adjustments) so that the mean of the state-year effects is zero for each year and state. Their general characteristic is that they are a weighted average of the level of support in a given state and year and the mean level of support for that state according to its state-level trend. The amount of shrinkage in the estimate (from the observed level of support in a given year and state toward the state-level trend) depends on the sample size for that given year and state, where larger samples shrink less.

### 6.3. Demographic effects and trends

Individual demographic variables also explain a substantial amount of the variation in death penalty support during the time span 1953-2006. Recall from Section 4 that we model the effects of race, sex and their two-way interaction as a linear trend on the logistic scale, and we allow these trends to vary by state (for a total of  $2 \times 2 \times 51$  intercepts, and the same

number of slopes). We model the effects of age on death penalty support as a linear trend, where the linear trends vary by state, resulting in  $4 \times 51$  age-state lines that are estimated. Last, we also model the effects of education (measured by the highest degree earned by the respondent) on the intercept of the logit of the probability of support, and we allow these effects to vary for each state (for a total of  $5 \times 51$  degree-state effects).

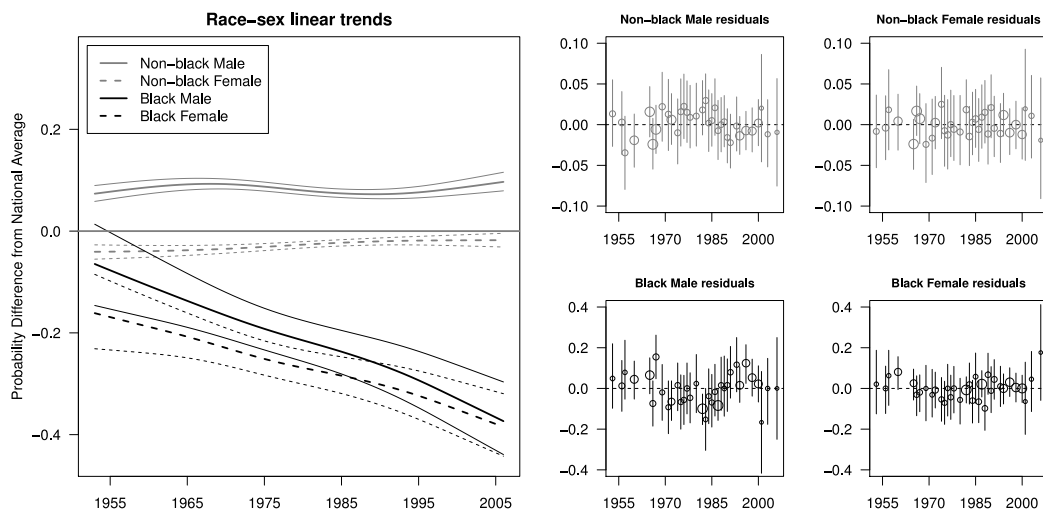
### 6.3.1. Trends related to race and sex

The left panel of Figure 7 shows the estimated difference between the level of support among individuals of each of the four race-sex combinations that we consider (black/non-black  $\times$  female/male) compared to their weighted average (i.e. the national average).

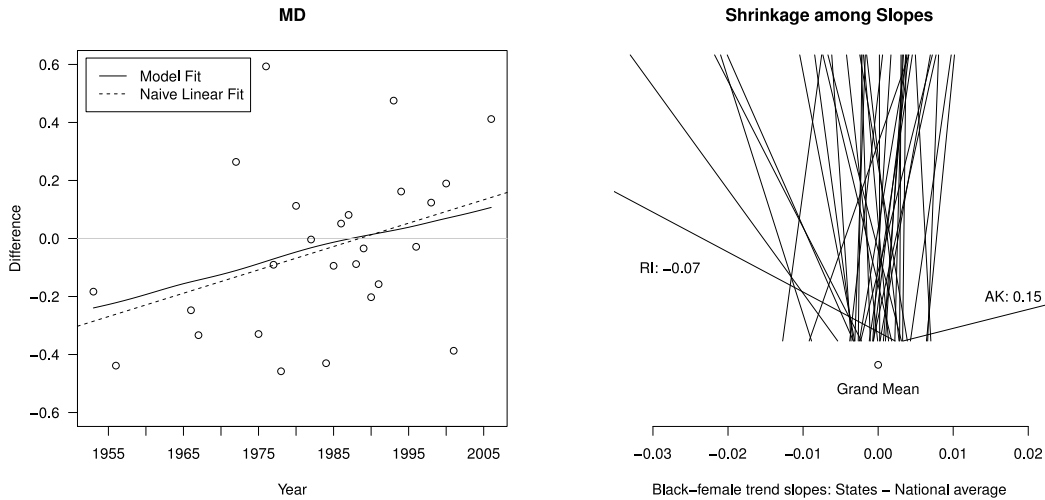
Black males have shown the sharpest relative decline in support over this time period, with an average decrease in support of about 1% every two years compared to the national average, starting with 6% lower than average estimated support in 1953, declining to estimated levels of support that were about 37% below average in 2006. Black females have also decreased their relative support over time, at almost the same rate, but not quite as steeply; their estimated support decreased from about 16% below average to about 38% below average. Averaged across these years, black females showed the least support over time for the death penalty of the four race-sex combinations that we consider here. Non-black males showed the highest average levels of support over this time period, increasing their estimated support from about 6% above average to about 8% above average. Last, non-black females began this time period with estimated support about 5% below average, and increased their relative support by a total of about 2% over the time period.

These trends among race-sex groups are allowed to vary by state, where the variability of the state-specific trends for each race-sex subgroup is estimated separately for each of the four U.S. regions. In general, there was more variability in race-sex trends between states within the same region than there was between the regions themselves. There was a particularly large amount of variation between the intercepts of racial trends among northern states ( $\sigma_{\text{black-state}_1} \approx 0.11 \pm 0.06$ ), and also between the slopes of racial trends among southern states ( $\tau_{\text{black-state}_2} \approx 0.14 \pm 0.02$ ). We won't summarize each group of state-specific trends here, but a full set of parameters estimates is available in the supplementary material.

To check the fit of the model with regard to individual states (see Figure 7 for residual plots across all states), we plotted the observed difference between the support for death penalty in a given state among a particular race-sex group and the national average support over time among that race-sex group. Figure 8 shows this comparison for Maryland, the state with the fastest increasing support among black females of all the southern states (the region where there was a lot of variation among state slopes). The raw data show an increasing trend over time, just as the model fit suggests. Figure 8 also shows the shrinkage of the estimated slopes by comparing them to "naive" slopes estimated from the raw differences in percentages between each state's support among a given race-sex group and the national average. The multilevel Bayesian model generally shrinks the estimated slopes toward zero, as it does for Maryland. Here, Maryland is used just as an example – this type of residual plot can and should be used to check fitted trends for all states. Note that some of the state slopes in the right plot of Figure 8 switch sides with respect to the grand mean – this surprising behavior is possible because some of the "naive" slopes were confounded by the age and degree of the respondents. Once the effects of these variables were accounted for, the slope among a race-sex subgroup was found to be on the other side



**Fig. 7.** The left plot shows the estimated differences between support levels over time for each of the four race-sex demographic groups compared to the national average (for an average state and average levels of age and education). The three lines associated with each group represent the mean and the boundaries of the 95% interval for the difference in support for that group and their average over time. The right set of plots show the residuals of support (observed - fitted) for each race-sex group over time, conditional on the observed states of residence, ages, and educational levels of each group. The vertical lines are 95% posterior intervals for the estimated differences, and the point sizes are proportional to the sample sizes for each race/sex/year combination. The y-axes of the residual plots are different for blacks and non-blacks, to better visualize the residuals.



**Fig. 8.** The left plot contains the observed difference between support among black women in Maryland over time and the support of black women across the nation over time. The dotted line is least-squares regression line through the differences in percentages (the “naive” fit we mention), and the dark line is the difference estimated by the model (for a person of average age and education). The slope and intercept of the model fit are both shrunk slightly toward zero. The right plot shows the shrinkage of estimated slopes like those pictured in the left plot. The upper ends of the line segments in the plot form the set of 51 estimated slopes from fitting a linear model to the difference in observed percentages of support among black women between a given state and national average (equivalent to the dotted line in the left plot); the lower ends of the line segments are the estimated slopes for each state from the multilevel Bayesian model, and the bottom point is the grand mean. The naive slopes for two states, Rhode Island and Alaska, did not easily fit into the plotting region, so the plot was truncated on both sides, and the lines for these states are labeled individually.

of the mean from the “naive” slope.

### 6.3.2. Age and Education Effects

We model a separate linear time trend (on the logistic scale) for each state and age category, resulting in a total of  $4 \times 51$  estimated trends. In general, the variation in trends across states was greater than that across age categories.

Averaging across states, we find that 18-29 year-olds supported the death penalty the least (about 3% less than average), and their average support did not change over time. 30-44 year-olds have shown the most support for the death penalty on average (about 1.8% above average), and their support remained steadily above average for the whole time period. There are minor trends visible in support among the other two age groups. 45-64 year-olds showed increasing relative support for the death penalty over time, increasing their support by about 6% on average over the 54-year time span. Respondents over 65 years old decreased their relative support by about 5% over the 54-year time span. Note that this does not imply that the opinion of any particular 45-64 year-old, for example, changed over time – these are only changes over time in the average opinions of people in a particular age group at that time. To track the evolving opinions of individual cohorts

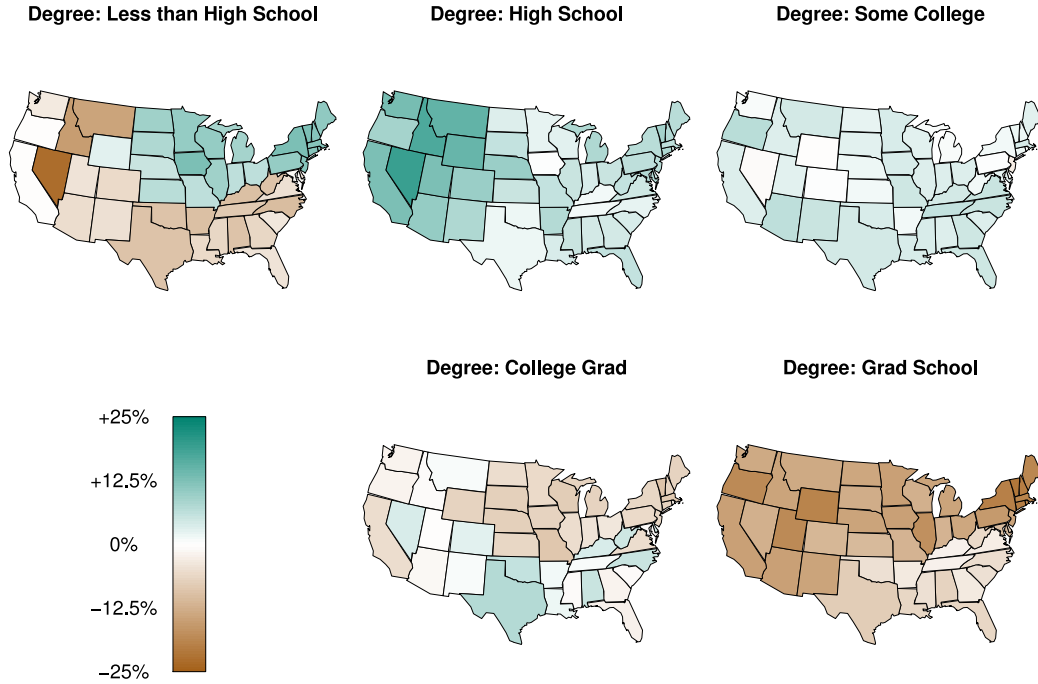
of people, one could look at estimates of relative support for 18-29 year-olds in the 1960s, 30-44 year-olds in the 1970s, 45-65 year-olds in the 1980s and 1990s, and people over 65 in the 2000s. The main drawback to estimating cohort trends using this data is that the age categories are coarse, so we only really observe about four estimates of opinion for the cohort described above, and even fewer for any other cohorts.

When we look at the estimated trends for each state within each age category, we find that the largest amount of variation in a group of intercepts is among western states in the 30-44 year-old age category, where  $\hat{\sigma}_{\text{age-state}_{(2,4)}} = 0.26$ , and the estimated differences between a state's mean (1980) support for this age group and that state's overall mean support ranges from -4% (Idaho) to +6% (Hawaii). There is also a substantial amount of variation among the slopes of 45-64 year-olds in the midwestern states, for example, where  $\hat{\tau}_{\text{age-state}_{(3,3)}} = 0.24$ , and among the slopes of 65+ year-olds in the western states, where  $\hat{\tau}_{\text{age-state}_{(4,4)}} = 0.30$ . We don't discuss specific hypotheses regarding trends among age-state cohorts in this paper, but if further investigation were to be done, we suggest plotting the raw data vs. the model fit in a single figure as a tool for further understanding, similar to how Figure 8 displayed the trend in support among one particular race-sex group (black females) in a given state (Maryland).

The education level of a respondent is measured by the highest degree they earned, and we model the effects of degree on death penalty support separately for each state. Degree level explains a substantial amount of variation in death penalty support, and its interaction with state of residence is also a strong predictor.

Figure 9 illustrates the differences among states for each degree category with colored maps. The five maps that correspond to the different degree levels illustrate the state-to-state variation in public opinion within each degree category:

- Death penalty support between states varies the most for respondents with less than a high school education, compared to the other educational categories. On average, respondents with this degree level support the death penalty at a level equal to the national average, but in some states (Vermont, New York, and Iowa, for example) respondents in this degree group support the death penalty about 12% more than average, and in other states (Idaho, North Carolina, and Montana), about 12% fewer respondents support the death penalty than the national average. Nevada is the most extreme state—respondents there with less than a high school degree support the death penalty 22% less than the national average. The standard deviation of the differences (by state) for this degree category is about 9%, compared to about 3% to 5% for the other four degree categories.
- Respondents whose highest degree is high school support the death penalty about 7% more than average—the highest level of support across the degree categories.
- Respondents whose highest degree is a graduate degree support the death penalty at much lower levels than average (about 12% lower than average); they are the degree group that differs from average the most.
- We don't find evidence for any time trend for degree categories; residual plots illustrating this are in Figure 2 of the supplementary file "Supplementary-tables-and-figures.pdf".



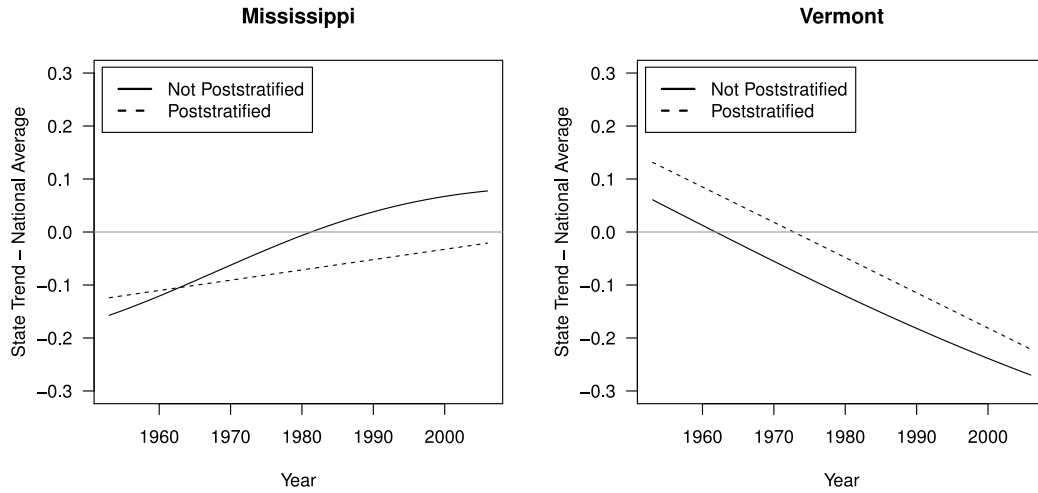
**Fig. 9.** Each state in each map is shaded from brown to blue, where brown indicates a level of support 25% below the national average, and blue indicates a level of support 25% above the national average. White shading indicates support equal to the national average. The most variation across states within a single degree level occurs for respondents with less than a high school degree.

#### 6.4. Post-stratification Estimates of Death Penalty Support

To estimate the actual proportion of a given state's population that supports the death penalty in a given year, we compute post-stratification weights from census data to adjust for the demographic composition of each state in each year. First, we gathered data from each available census during the span of the surveys, which included decadal censuses from 1950 to 2000 and yearly censuses from 2001 to 2006. Then, for each census, we counted the number of people in each of the  $(51 \times 2 \times 2 \times 5 \times 4)$  cells of the population defined by state, race, sex, degree, and age, respectively. To estimate the population of each cell in years between censuses, we used a linear interpolation. We denote the resulting estimates of the population within each cell as  $N_{stbfda}$ , for states  $s = 1, \dots, 51$ , years  $t = 1953, \dots, 2006$ , races  $b = 1, 2$ , sexes  $f = 1, 2$ , education levels  $d = 1, \dots, 5$ , and age categories  $a = 1, \dots, 4$ .

Then, to estimate the actual proportion of people who support the death penalty in a given state and year, denoted  $\theta_{st}^{\text{post-stratified}}$ , we computed the mean of the posterior distribution of  $P(Y_{stbfda} = 1)$  from the model (Equation 1) for each post-stratification cell, denoted  $\hat{\theta}_{stbfda}$ , and we adjusted it according to its post-stratification weight and summed over the race, sex, age, and education categories:

$$\theta_{st}^{\text{post-stratified}} = \sum_{b,f,d,a} \hat{\theta}_{stbfda} \times \frac{N_{stbfda}}{\sum_{b,f,d,a} N_{stbfda}}. \quad (8)$$



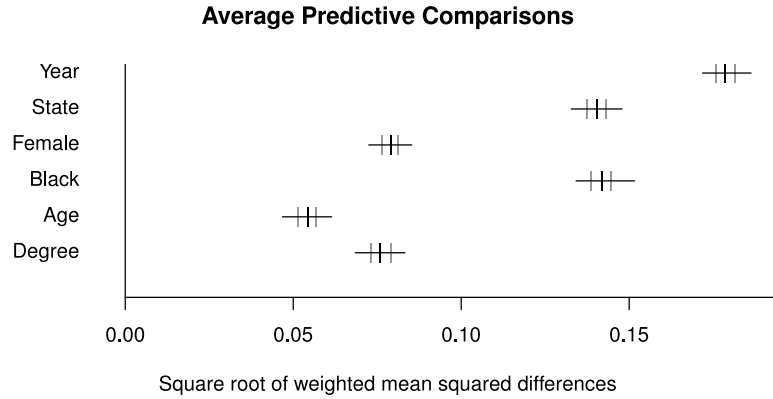
**Fig. 10.** The difference between raw and post-stratified estimates of state trends (where the national average trend is subtracted out) for two states where these differences are large. In both cases, the difference is due mostly to the proportion of black residents of these states – blacks show lower support for the death penalty, and comprise a larger-than-average proportion of Mississippi’s population, and a smaller-than-average proportion of Vermont’s population. The differences in slopes is further affected by the estimated trends among blacks in each state.

Figure 10 shows the difference between raw and post-stratified estimates of state trends for two states where this difference was large: Mississippi and Vermont. In both cases, the difference between the two types of estimates can largely be explained by the proportion of black residents in each state: much larger than average for Mississippi, and much smaller than average for Vermont. Across all states and years, the standard deviation of the difference between raw and post-stratified estimates of support is 13%, which is similar to the magnitude of the effects of some of the strongest predictors in the model (such as year, state, and race – see Figure 11 in Section 9). We include the full set of post-stratified estimates of support for all states and years in the supplementary materials.

### 6.5. Average Predictive Comparisons

One useful visual summary of the model we fit is given in Figure 11, which displays average predictive comparisons (Gelman and Pardoe, 2007), or APCs, for each of the six main input variables in our model. These are estimates of the magnitudes of the effects of changing from one value of an input variable to another value, while holding all the other input variables constant at their observed values. APCs are useful for nonlinear models with interactions, such as the model presented here, because in such models the raw regression coefficients are not easily interpretable on the outcome scale (as opposed to linear regression without interactions, where the coefficients themselves act as APCs).

To compute the APC interval for the variable “Year”, for example, we do the following for each posterior sample: For each respondent,  $i = 1, \dots, n$ , and for each year  $t = 1, \dots, T$ , compute the squared difference in respondent  $i$ ’s estimated death penalty support if he/she had responded to the survey in year  $t$ ,  $E(Y_i | X_i^{\text{year}} = t, V_i)$ , vs. the *actual* estimated



**Fig. 11.** Average predictive comparisons for the six input variables: year, state, sex, race, age, and education. These are estimates of the expected differences in death penalty support between two respondents with different values of one particular input variable, but similar values of all others. The vertical hash marks are the point estimates and 50% intervals for the differences, and the width of the horizontal lines indicate the 95% intervals.

support for respondent  $i$ ,  $E(Y_i | X_i^{\text{year}}, V_i)$ , where  $V_i$  denotes the vector of values of the other five input variables for respondent  $i$ . Next, weight these  $T$  squared differences by a measure of how many of the other five input variables would likely be shared in common by a respondent in year  $t$  and a respondent in year  $X_i^{\text{year}}$ . This is done to downweight the squared differences between pairs of years in which it would be unlikely to find two similar respondents in our data (see Section 4 of Gelman and Pardoe (2007) for more details). Last, take the square root of the mean of these weighted squared differences across all respondents. We did this for all 3,000 posterior samples, for all six main variables, and plotted the resulting 50% and 95% APC intervals in Figure 11.

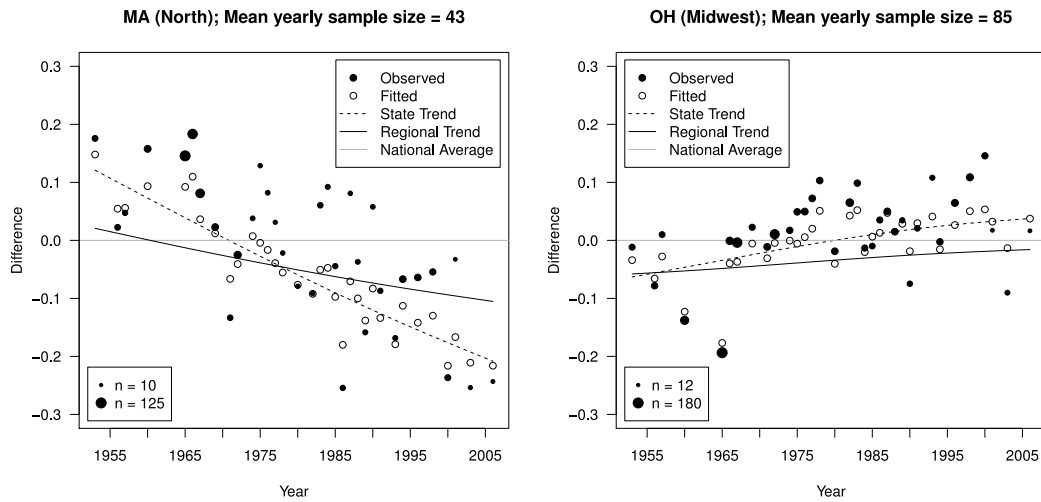
Our conclusion is that the survey year is associated with the most variation in death penalty support among all input variables, and that the estimated difference in probability of death penalty support between a randomly chosen respondent in our data and a similar respondent from a different year is about 18%. This estimate accounts for all the different ways in which the survey year affects death penalty support, from the national average time trend and yearly effects to the interactions between survey year and various demographic variables. The variables that account for the next two largest differences are state and race, and each accounts for about a 14% difference in support levels. Then, sex, education, and age account for about 8%, 8%, and 5% differences in death penalty support. In the case of all these variables, the 95% intervals have a width of only about 1.5%.

## 7. Goodness of fit checks

We have already included plots of model estimates vs. observed data in Figures 7 and 8, showing the fit of the model with respect to trends among racial and gender-based groups. In this section, we display graphical checks of the fit of the model with respect to state trends, state-year interaction effects, and trends among educational groups.

Figure 12 illustrates the fit of the model with respect to the difference between the





**Fig. 12.** Differences between individual states and the national average over time on the probability scale, with estimates from the fitted model plotted with observed data. The model-based estimates are of the estimated support for an individual of “average” demographics. Point sizes are proportional to the sample sizes for each state-year.

national average level of death penalty support and support in two specific states, Massachusetts and Ohio, over time. For both states, the assumption of a linear trend explaining the differences between state-level opinion and national opinion seems reasonable. There may be some dependence from year to year for Ohio (especially in the late 1970’s, when a group of about 4-5 consecutive years were all above the estimated state trend line), but on the whole, the independence assumption for the state-year interaction effects looks realistic. In years where more data are observed (where the points are drawn in proportion to the sample size for that state-year), the estimated probabilities do not shrink as far toward the state trend line (1965, for example). In years where there is little data, the estimated probabilities are pooled almost all the way to the state trend line (Massachusetts in 2001, for example). Last, when there is no data from a given year, the estimated percentages for a given state lie exactly on the state trend line.

We considered extending the model to allow for the effects of degree to vary across time. To investigate whether this would be likely to improve the fit of the model, we performed a posterior predictive check in which we simulated data for each respondent, and compared the predicted levels of support for each degree level over time, holding the other variables constant at their observed levels. The results are displayed in the five residuals plots in Figure 2 of the supplementary file “Supplementary-tables-and-figures.pdf”. There don’t appear to be any patterns across time among the residuals in any of the degree-level groups that would indicate that our model is missing an important time trend. There appear to be influential points at the extremes of the x-axis for the high school degree category and the college degree category, but on the whole, the 95% intervals have approximately the correct level of coverage.

## 8. Out-of-sample predictions

To guide our model selection, we estimated the predictive accuracy of the five multilevel Bayesian models that we fit in terms of the posterior mean of the deviance, where the deviance of model  $m$  for posterior sample  $g$ ,  $D_m^{(g)}$ , is given by

$$D_m^{(g)} = -2 \sum_{i=1}^n \left( Y_i \log P(Y_i | \theta_m^{(g)}) + (1 - Y_i) \log(1 - P(Y_i | \theta_m^{(g)})) \right),$$

and  $\theta_m^{(g)}$  is the vector of all model parameters for iteration  $g$  in model  $m$ . First, we randomly divided our data set into a training set and a test set, containing 80% and 20% of the individual survey responses, respectively. Since there is no widely agreed upon general rule for how to split data into training and test sets (Hastie et al., 2008), we choose a split (80/20) that lies within the range of previous analyses, e.g. Breiman (2001). We fit each model to the training data and made out-of-sample predictions on the test data.

Of the five multilevel Bayesian models we fit to the data, the fourth model, which we call the “main” model, is the one described in Equation 1 and pictured by the DAG in the supplemental materials. The first, most basic model contained main effects for the four demographic variables and state-year interaction terms, where the state-year interaction terms were centered at the sum of their respective year and state main effects, and each group of main effect parameters had a prior mean of zero and a half- $t$  prior distribution on its standard deviation (the same as described in Section 4). The effective number of parameters,  $p_D$ , for this model, as estimated by the difference between the posterior mean of the log-likelihood and the log-likelihood evaluated at the posterior mean of the parameters in the likelihood equation (Spiegelhalter et al., 2002), was about 500.

The second model contained state-year interaction terms, but this time with the exact same prior structure as the main model—that is, this model included state trends, state-level variables, and an AR(1) prior distribution (with a linear component) for the yearly effects. The rest of the demographic main effects were given the same structure and priors as in the first model—means of zero, and half- $t$  prior distributions on their group-level standard deviations. This model had fewer effective parameters than the first one (about 434 compared to 500) because of the partial pooling induced by the structured prior on the state-year interaction terms, and the inclusion of state-level variables (which help to induce even more pooling on the state-specific slopes and intercepts, decreasing the estimate of  $p_D$ ). This model fit the training data worse than the first model, but made better predictions on the test set.

The third and fifth models were similar to the main model, with only minor modifications. The third model omitted the age-state slopes, and was otherwise identical to the main model. This model performed slightly worse on the test set than the main model, and substantially worse on the training set. The fifth model included all of the parameters in the main model, and also included four additional sets of two-way interactions: (sex, age), (sex, degree), (race, age), and (race, degree). The inclusion of each of these two-way interactions improved the predictive performance of the models fit during exploratory data analysis using OLS regression, which is why we tried including them in a multilevel Bayesian model. The result of their addition to the main model, though, was a poorer fit than the main model on both the training and test data sets.

It is possible that including only a subset of these additional interaction effects (or other unexplored interactions not mentioned here) could have slightly improved the fit on the

**Table 1.** Post-stratified estimates of support levels among the four race-sex subgroups over time.

Race-Sex Subgroup	1955	1965	1975	1985	1995	2005
Non-black Men	0.70	0.58	0.74	0.85	0.85	0.79
Non-black Women	0.61	0.47	0.63	0.77	0.78	0.69
Black Men	0.50	0.33	0.46	0.59	0.56	0.42
Black Women	0.42	0.27	0.39	0.54	0.52	0.40

test set, but our goal was not solely to make the best predictions, but rather to understand changes in opinion over *time* (which is not a component of any of the additional interaction effects in the fifth model), and to do so by emphasizing thorough graphical checks of the fit of the model. Thus, we proceeded with our analysis of the main model. Table 1 in the supplementary file “Supplementary-tables-and-figures.pdf” contains details of the deviance, effective number of parameters, and DIC of the multilevel Bayesian models we fit.

## 9. Discussion

We fit a multilevel Bayesian model to 58,253 individual responses to the question “Are you in favor of the death penalty for persons convicted of murder?” using data from a 54-year time span and including demographic and state-level variables. The use of a structured prior distribution on the yearly effects allowed us to simultaneously estimate their variation, while also estimating various main effects, trends, and interaction effects that shed new light on certain relationships between demographic variables and death penalty support, and on trends in death penalty support among states.

We found that blacks have decreased their support over time dramatically compared to non-blacks, with the support among black men showing a faster relative decrease, on average, than among black women. This pattern was previously difficult to identify because of nationwide fluctuations in death penalty support over time (see Figure 3) and sampling variation. Hanley (2008) points out that support was higher in the 1990s than in the 1970s among many subgroups of the population, including blacks, despite high-profile court cases in the late 1980’s that found evidence for bias against blacks in death sentencing (which presumably would result in lower support among blacks). Although support among blacks was higher in the 1990s than in the 1970s in absolute terms, this increase in support was observed for both blacks and non-blacks (i.e. it can be explained by the increased national support for death penalty in the 1990s). According to our model, the relative support among blacks compared to non-blacks was in the middle of a steady decline – one that was steeper, on average, for black males than black females (See Figure 7). Further, Hanley (2008) claim that “since 1990 black men have been more supportive of the death penalty than black women, after a period of fluctuation in the late 1970s and 1980s.” Whether or not this fluctuation was real (or just the result of sampling variability), our analysis shows that the gap in support among black men compared to black women is shrinking, on average, from a difference of about 10% in the 1950s to a difference of just 1% in the mid-2000s. For a summary of support levels among race-sex subgroups over time, in addition to Figure 7, which plots estimated *relative* trends in support among subgroups, Table 9 contains post-stratified estimates of *absolute* support levels among race-sex subgroups.

We also found that death penalty support has grown faster in states where support for Republican presidential candidates has grown during the past 50 years, and average

support over time has been higher in states where the death penalty has been legal for a larger proportion of the past 50 years. In future work, each of these state-level variables could be refined or modified to discover additional, potentially stronger associations between state-level partisan support, state-level policy, and individual death penalty support. For example, partisan support at the state level for governors or state legislatures may be more strongly associated with death penalty support than partisan presidential support. Since partisan support is also a dynamic process, a joint model for partisan support and death penalty support could reveal new associations between these two phenomena over time. Also, rather than looking only at whether the death penalty is legal in a given state, it is possible that other state-level variables such as the crime rate, the number of death sentences, or the number executions carried out in a given state and year could be associated with individual-level support for the death penalty in that state.

We chose to model state-specific slopes and intercepts as being centered around their regional means, according to a U.S. Census-based four-region breakdown of the 51 states. We chose to use this four-region breakdown of the states because it is relatively simple, well-known, and often cited in studies of public opinion by region in the U.S. (Gelman, 2008; Lax and Phillips, 2009a). Alternative groupings of states into regions, or alternative models to allow for correlations between states, however, could be of interest for future work. The U.S. Census contains alternative regional groupings of the states into nine groups of states, which might allow for more variation between regions than what we observed with our model. One could also fit a model in which state effects are correlated with the effects of neighboring states, or one could fit a latent class model, to infer groups of similarly opinionated states from the data, rather than using pre-defined regions. Such an investigation into correlations between groups of states would be interesting, but was beyond the scope of this paper.

Another modification to the model that might be useful in future investigations would be to allow for dependence of the state-year effects across time within states. If each state's state-year effects followed an AR(1) model, for example, then perhaps the varying level of dependence from year to year for each state would provide new information about how public opinion varies across time within states. For example, the opinions of respondents in some states might very stable, if that state had highly autocorrelated state-year effects. On the other hand, their opinions might very prone to yearly swings, if that state had low or negative autocorrelation between state-year effects, perhaps in reaction to local events or policy changes that were related to death penalty.

Future models could also include additional survey data and, if necessary, model differences in data collected by different organizations. One could also attempt to gather more demographic variables for each respondent, such as political party affiliation, income, or religion, to better understand the factors that predict attitudes on this issue.

Last, regarding the type of model that we fit, it could also make sense to consider non-parametric regression methods such as Bayesian Additive Regression Trees (Chipman et al., 2010), perhaps in combination with existing regression models to pull out any structure in the data beyond what is captured by our logistic regression.

To return to the political questions that motivated this work: public opinion is, presumably, both a cause and a consequence of policies on capital punishment. In order to study these connections, researchers need measures of state-level opinion. In a study of state-level attitudes on gay rights, Lax and Phillips (2009a) showed a level of responsiveness of legislators' policy to opinion that was surprising given some of the earlier literature on state politics. The present research goes further by modeling trends at the state level using sparse data.

## 10. Supplementary Materials

- (a) DAG.pdf: A directed acyclic graph (DAG) of the full model
- (b) Adjusted-parameters.pdf: A description of the calculations to derive the adjusted parameters.
- (c) Supplementary-figures-and-tables.pdf: A plot of the state-specific slopes and intercepts by region, and tables containing model comparison statistics and state effects.
- (d) estimates.RData: Interval estimates of the level of support for the death penalty for each possible combination of (state, year, race, sex, degree, age).
- (e) varnames.RData: Variable names that correspond to the dimensions of the array in the file estimates.RData.
- (f) poststrat.RData: Post-stratified estimates of the level of support for the death penalty by state and year.

## References

- Baumgartner, F. R., S. L. De Boef, and A. E. Boydston (2008). *The Decline of the Death Penalty and the Discovery of Innocence*. Cambridge University Press.
- Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science* 16(3), 199–231.
- Chipman, H. A., E. I. George, and R. E. McCulloch (2010). Bart: Bayesian additive regression trees. *Annals of Applied Statistics* 4(1), 266–298.
- Datta, G. S., P. Lahiri, R. Maiti, and K. L. Lu (1999). Hierarchical bayes estimation of unemployment rates for the states of the u.s. *Journal of the American Statistical Association* 94(448), 1074–1082.
- Ellsworth, P. C. and S. R. Gross (1994). Hardening of the attitudes: Americans' views on the death penalty. *Journal of Social Issues* 50(2), 19–52.
- Erikson, R. (1976, February). The relationship between public opinion and state policy: A new look based on some forgotten data. *American Journal of Political Science* 20(1), 25–36.
- Erskine, H. (1970). The polls: Capital punishment. *Public Opinion Quarterly* 34, 290–307.
- Fox, J. A., M. L. Radelet, and J. L. Bonsteel (1991). Death penalty opinion in the post-furman years. *New York University Review of Law and Social Change* 28, 499–528.
- Franklin, C. (2001). Pre-election polls in nation and state: A dynamic bayesian hierarchical model. Technical report, University of Wisconsin.
- Gelfand, A. E. and S. K. Sahu (1999, March). Identifiability, improper priors, and gibbs sampling for generalized linear models. *Journal of the American Statistical Association* 94(445), 247–253.
- Gelman, A. (2008). *Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way They Do*. Princeton University Press.

- Gelman, A. and J. Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- Gelman, A. and I. Pardoe (2007). Average predictive comparisons for models with nonlinearity, interactions, and variance components. *Sociological Methodology* 37(1), 23–51.
- Gelman, A. and D. Rubin (1992, November). Inference from iterative simulation using multiple sequences. *Statistical Science* 7(4), 457–472.
- Hanley, J. (2008). *Public Opinion and Constitutional Controversy*, Chapter 5: The Death Penalty, pp. 108–138. Oxford University Press.
- Hastie, T., R. Tibshirani, and J. Friedman (2008). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Volume 2. Springer.
- Lax, J. R. and J. H. Phillips (2009a, August). Gay rights in the states: Public opinion and policy responsiveness. *American Political Science Review* 103(3), 367–386.
- Lax, J. R. and J. H. Phillips (2009b, January). How should we estimate public opinion in the states? *American Journal of Political Science* 53(1), 107–121.
- Norrander, B. (2000, December). The multi-layered impact of public opinion on capital punishment implementation in the american states. *Political Research Quarterly* 53(4), 771–793.
- Plummer, M. (2003, March 20–22, Vienna, Austria). Jags: A program for analysis of bayesian graphical models using gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*.
- Rao, J. N. K. (2003). *Small Area Estimation*. John Wiley and Sons.
- Schuman, H. and S. Presser (1981). *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. Sage Publications, Inc.
- Spiegelhalter, D. J., N. G. Best, C. B. P., and A. Van der Linde (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B* 64(4), 583–616.
- Vidmar, N. and P. Ellsworth (1974, June). Public opinion and the death penalty. *Stanford Law Review* 26(6), 1245–1270.