Quarterly Journal of Political Science, 2007, 2: 345–367

Rich State, Poor State, Red State, Blue State: What's the Matter with Connecticut?*

Andrew Gelman[1], Boris Shor[2], Joseph Bafumi[3] and David Park[4]

[1]*Department of Statistics and Department of Political Science, Columbia University, New York, USA; gelman@stat.columbia.edu; www.stat.columbia.edu/∼gelman*
[2]*Harris School of Public Policy Studies, University of Chicago, USA*
[3]*Department of Government, Dartmouth College, USA*
[4]*Department of Political Science, George Washington University, USA*

ABSTRACT

For decades, the Democrats have been viewed as the party of the poor, with the Republicans representing the rich. Recent presidential elections, however, have shown a reverse pattern, with Democrats performing well in the richer blue states in the northeast and coasts, and Republicans dominating in the red states in the middle of the country and the south. Through multilevel modeling of individual-level survey data and county- and state-level demographic and electoral data, we reconcile these patterns.

Furthermore, we find that income matters more in red America than in blue America. In poor states, rich people are much more likely than poor people to vote for the Republican presidential candidate, but in rich states (such as Connecticut), income has a very low correlation with vote preference.

* We thank the members of the Quantitative Political Science Discussion Group at Columbia University, participants at the Midwest Political Science Association meeting, Greg Wawro, Robert Erikson, Robert Shapiro, Elke Weber, Eric Johnson, Craig Newmark, Shigeo Hirano, David Epstein, Stephen Ansolabehere, Stuart Jordan, Dave Krantz, Nolan McCarty, participants at seminars at the George Mason University economics department and the Columbia University psychology department, and anonymous reviewers for helpful comments; Maryann Fiebach, Boliang Zhu, Brian Fogarty, Amanda Czerniawski, and Andres Centeno for research assistance; David Weaver and Eunseong Kim for data from the American Journalist Survey; David Leip for election data; Robert Shapiro and the Inter-university Consortium for Political and Social Research for poll data; and the National Science Foundation and the Columbia University Applied Statistics Center for financial support.

MS submitted 20 June 2006; *final version received 20 September 2007*
ISSN 1554-0626; DOI 10.1561/100.00006026
© 2007 A. Gelman, B. Shor, J. Bafumi and D. Park

Key methods used in this research are: (1) plots of repeated cross-sectional analyses, (2) varying-intercept, varying-slope multilevel models, and (3) a graph that simultaneously shows within-group and between-group patterns in a multilevel model. These statistical tools help us understand patterns of variation within and between states in a way that would not be possible from classical regressions or by looking at tables of coefficient estimates.

Keywords: Availability heuristic; ecological fallacy; hierarchical model; income and voting; multilevel model; presidential elections; public opinion; secret weapon; varying-slope model.

> I never said all Democrats are saloon-keepers. What I said is that all saloon-keepers are Democrats.
>
> —Horace Greeley, 1860

> Pat doesn't have a mink coat. But she does have a respectable Republican cloth coat.
>
> —Richard Nixon, 1952

> Like upscale areas everywhere, from Silicon Valley to Chicago's North Shore to suburban Connecticut, Montgomery County supported the Democratic ticket in last year's presidential election, by a margin of 63 percent to 34 percent.
>
> —David Brooks, 2001

> There is, for example, this large class of affluent professionals who are solidly Democratic. DataQuick Information Systems recently put out a list of 100 ZIP code areas where the median home price was above $500,000. By my count, at least 90 of these places — from the Upper West Side to Santa Monica — elect liberal Democrats.
>
> —David Brooks, 2004

> A lot of Bush's red zones can be traced to wealthy enclaves or sun-belt suburbs where tax cuts are king.
>
> —Matt Bai, 2001

> But in the Ipsos-Reid surveys, 38% of voters in "strong Bush" counties said that they had household incomes below $30,000, while 7% said that their families earned at least $100,000. In "strong Gore" counties, by contrast, only 29% of voters pegged their household income below $30,000, while 14% said that it was above $100,000.
>
> —James Barnes, 2002

DEMOCRATS AND REPUBLICANS, RICH, AND POOR: TWO PERSPECTIVES

Throughout the 20th century and even before, the Democratic Party in the United States has been viewed as representing the party of the lower classes and thus, by extension, the "average American." More recently, however, a different perspective has taken hold, in which the Democrats represent the elites rather than the masses. The view of Democrats as elitists began perhaps with the party's control over government planning during the New Deal era and was developed as a key theme by populist Republicans in the postwar period, including Nixon, Goldwater, and Reagan.[1] The Democratic party too has had many prominent populists, from Huey Long to Harry Truman to Jimmy Carter to Al Gore ("the people versus the powerful"), but this is less remarkable given its majority status during this period. Throughout, populism has been associated with the attitudes and interests of lower and middle-income Americans.

What is happening now? Do richer voters still support Republicans? If so, how can we understand the pattern that the Democratic do best in the richer blue states of the Northeast and West, while the Republicans dominate in the poorer red states in the south and between the coasts? And does living in a poor or rich state change individual vote preferences in some fashion? In other words, does context matter for individual voting behavior and, if so, how? We explore these questions by studying the relation between income and presidential vote preference, at the individual, county, and state levels. It turns out that the connections between income and voting in the United States are not simple; we find that rich and poor states differ in the relation between individual income and partisan preferences.

Perspectives from Social Science and the News Media

Census and opinion poll data since 1952 reveal that higher-income voters continue to support the Republicans in presidential elections.[2] However, higher-income states have in recent years favored the Democrats. The Republicans have the support of the richer voters within any given state but have more overall support in the poorer states. Thus, the identification of rich states with rich voters, or more generally, the "personification" of so-called red and blue states, is misleading. For example, in the context of the Brooks quotes above, within an "upscale" area that supports the Democrats, the more "upscale" voters are still likely to vote Republican.

The connection between income and support for conservative parties has long been noted and has attracted interest from political scientists and sociologists studying ideological polarization. McCarty *et al.* (2006) argue that partisanship and presidential vote

---

[1] See, for example, Buckley and Bozell (1954), Rovere (1959), McGirr (2001), Perlstein (2002), and Greenberg (2003).

[2] For example, from 2004 exit polls, Bush received 36% of the support of voters with incomes under $15,000, 41% with incomes between $15,000 and $30,000, monotonically increasing to 62% of those with incomes over $200,000.

choice have become more stratified by income over the past 50 years. Comparing survey respondents in the highest and lowest quintiles of income, they find that in 1956 and 1960, the proportion of Republican identifiers was only slightly higher in the highest than in the lowest quintile, but in 1992–2000, respondents in the highest quintile were more than twice as likely to identify as a Republican than were those in the lowest. Stonecash (2000, 2005) finds a growth of support for the Democrats since the 1970s among poor persons and in high-poverty areas. In contrast, Fiorina *et al.* (2005) find polarization of the political class but not of the general voting population, with only small differences in issue preferences when comparing voters in red and blue states. Fiorina *et al.*, however, do not discuss voting in relation to income, so our analysis supplements theirs by considering this variable. In an extensive analysis of opinion poll data, Ansolabehere *et al.* (2006) find voters to be most strongly motivated by economic issues, but they note that the connection between income and economic views can be weak. Brooks and Brady (1999) and Bartels (2006) find that income continues to be predictive of partisanship,[3] and Filer *et al.* (1993) studied the connection between income and voter turnout.

In contrast, media attention has focused on comparisons of states (and, to a lesser extent, counties), as illustrated by many of the quotations that lead off this article. We seek to simultaneously understand Republican strength among richer votes and in poorer states, and to study these trends over time. The journalists who see patterns on the red-and-blue map and the political scientists who analyze polls are talking past each other because they are looking at different levels of aggregation. Public perceptions of the two parties are important, and after setting the record straight on what is actually happening with income and voting, we consider some explanations from cognitive psychology for why misunderstandings about the correlations between income and vote preference could persist among otherwise well-informed observers.

Studying Patterns at the State Level

Comparing to previous studies of income and voting, our key contribution is to study patterns both within and between states, with both individual income and state-level income as predictors, using survey data on individuals and election and Census data for states and counties. The pattern that richer states support the Democrats is *not* a simple aggregation of rich voters supporting the Democrats. This can be viewed either as a debunking of the journalistic image of rich latte Democrats and poor Nascar Republicans — or as support for the journalistic images of political and cultural differences between red and blue states — differences which are not explained by differences in individuals' incomes.

---

[3] Manza and Brooks (1999, Chapter 3) show that the consistent correlation of high income with Republican vote masks changes in particular social and occupational groups (for example, professionals have moved toward the Democrats and self-employed persons toward the Republicans); here we focus on income, partly because of its relevance for government policy but especially because of its salience in current political discourse, an issue we return to at the end of this article.

We find that income matters more in red America than in blue America. In poor states, rich people are much more likely than poor people to vote for the Republican presidential candidate, but in rich states (such as Connecticut), income has almost no correlation with vote preference. The United States has red and blue voters, and red and blue states, but income cuts across them in different ways (a point noted by Alford (1963), in his study of social class and voting by region of the United States). As we demonstrate, the statistical technique of multilevel modeling allows us to understand the relation between income and vote among individuals, counties, and states. The final section of this article considers reasons for these patterns, along with psychological reasons why certain misunderstandings have persisted, and a discussion of the relevance of income/voting patterns to political perceptions.

The patterns of income and voting by state are politically important, and the common misperceptions of these patterns are also important. This article attempts to make sense of the data and also the misperceptions.

STUDYING THE RELATION BETWEEN INCOME AND VOTE PREFERENCES

Survey data show a small but persistent correlation between income and support for the Republican party, but at the aggregate level, it is the Democrats who do better in the richer states. Our strategy to understand these patterns is to study the relation between income and voting in four ways:

- Aggregate, by state: to what extent do richer states favor the Democrats?
- Nationally, at the level of the individual voter: to what extent do richer voters support the Republicans?
- Individual voters within states: to what extent do richer voters support the Republicans, within any given state? In other words, how much does context matter?
- Counties within states: to what extent do richer counties favor the Democrats, within any given state?

Patterns at these four levels have much different political interpretations from those supposed by confused political commentators. Most notably, the support for the Democrats in the richer states had led observers to view the typical Democrat as an upper-middle-class resident of a coastal metropolitan area, and the typical Republican as lower-middle-class and rural (see Brooks (2001)). That these claims have been overstated (see Frank (2004), and Issenberg (2004)) does not seem to lessen their appeal.

A multilevel strategy in understanding voting behavior is useful because we care about election outcomes as well as individual decisions. Elections are not simple cumulations of voter decisions (because of institutional features such as electoral rules and geographic boundaries, and the political decisions of parties and candidates), and so aggregate analysis should not be discarded simply by citing the ecological fallacy problem (Wright 1989). Trends of economic voting at multiple levels of analysis may or may not be similar, and their causes may or may not be similar.

As in Wright (1989), we consider the variation at each level of analysis. Income varies far more *within* states than average income does *between* states. Consequently, it is the within-state rather than the between-state effect of income that dominates the national patterns. In particular, a positive correlation of income and Republican voting within states, plus a negative correlation between states, combine to form a positive correlation among all voters.

We have both individual and aggregate data on income and votes. Thus, the statistical analysis is relatively straightforward, without the well-known problems that can arise when only aggregate data are available (Robinson 1950, Kramer 1983). For aggregate patterns, we use presidential election returns and Bureau of Economic Affairs data on average income by state and county. We estimate the relations between income and vote preferences for each presidential election from 1968 to 2004, with a particular goal of studying trends including any changes over time in the support for particular political parties.

For all the analyses, both simple and complex, we gain insight by replicating over several election years. Although obvious, this sort of replication is not always done, and when it is done, the resulting pile of analyses can seem too overwhelming to display. Time-series plots of data summaries and parameter estimates (as in many of the figures here) and repeated graphs (also called "small multiples"; see Bertin (1967) and Tufte (1990)) allow us to see patterns in a way that would be difficult using tabular representations (see Gelman *et al.* (2002)).[4]

Analysis of State and County Averages

We begin by fitting a state-level linear regression for each election year, predicting state support for the Republican candidate in the election from the average income in the state. Positive coefficients imply that richer states are supporting the Republicans more. To allow comparability over time, we adjust incomes in each year to 1996 dollars. In addition, we examine the coefficient of average income after controlling for percent African–American in the state. We also study average income and votes at the county level: within states, do the richer counties lean toward the Republicans or the Democrats?

National Analysis of Individual Voters

Our first individual-level analysis is a simple logistic regression modeling vote preference from the National Election Study (NES) polls taken during the month before each election (coding 1 = Republican, 0 = Democrat, excluding respondents who were undecided

[4] The method of repeated modeling, followed by time-series plots of estimates, is called the "secret weapon" (Gelman and Hill 2007) because it is so easy and powerful but yet rarely used as a data-analytic tool. We suspect that one reason for its rarity of use is that, once one acknowledges the time-series structure of a dataset, it is natural to want to take the next step and model that directly. In practice, however, there is a broad range of problems for which a cross-sectional analysis is informative, and for which a time-series display is appropriate to give a sense of trends. In our example, the secret weapon allows us to see how cross-sectional estimates for individual states and the entire United States vary over time. Expanding our multilevel models to include time series would be a major research undertaking that would require evaluation of additional time-series assumptions that are peripheral to our substantive research goals here.

or supported third-party candidates) on income. We summarize family income with a five-point quantile-based scale[5] used by the NES, which allows the results to be comparable over time. (However, individual income inequality has grown in recent decades, so coefficients for percentiles do not have a constant interpretation in terms of numerical relative incomes.)

We fit a separate logistic regression for each election year; if the coefficients are positive, this implies that Republicans were differentially supported by richer voters. We also see what happens when state indicators are included in the model, to see the predictive power of individual income within states. In addition, we examine the coefficient of income when additional predictors are added, including ethnicity (African–American or other), sex, age (18–29, 30–44, 45–64, or 65+), education, party identification, and ideology. However, our primary analyses use only income as a predictor, because our goal is to study differences between richer and poorer voters. Even if income effects were "explained" by other predictors, the correlations would still be real.

Analysis of Individuals within States

To study the relation of income to individual vote preferences, controlling for state, we fit from each election year's poll data[6] a multilevel logistic regression of vote preference on income (using the five-point scale noted in Footnote 5) and state. This varying-intercept model gives us 50 state-level coefficients allowing geographic variation in support for the Republican candidate in each election.[7] The coefficient for income then represents the extent to which Republicans are differentially supported by richer voters, within any given state.

Because we are interested in comparing states in different regions and of different income levels, we include region indicators and the average income within each state as group-level predictors. Including these predictor also increases the precision of the multilevel model fit, by reducing the residual error at the state level. As in the national poll analyses, we also examine the coefficients for income when ethnicity, sex, and age are included in the model. In addition, we consider models including the state-level Gini index to account for income inequality within states. We fit the multilevel models using

---

[5] The National Election Study uses 1 = 0–16 percentile, 2 = 17–33 percentile, 3 = 34–67 percentile, 4 = 68–95 percentile, and 5 = 96–100 percentile. We label these as $-2, -1, 0, 1, 2$, centering at zero so that we can more easily interpret the intercept terms of regressions that include income as a predictor.

[6] For 2000 and 2004, we fit using the National Annenberg Election Survey, which, with over 100,000 respondents, allows accurate estimation of the patterns in individual states. We also use news media exit polls: ABC News in 1984 and 1988, Voter Research and Surveys in 1992, Voter News Service in 1996 and 2000, and National Election Pool in 2004. These polls have disadvantage of a messier sampling scheme and use different income categories than the Annenberg and NES surveys. However, the exit polls have large sample sizes (even in small states) and provide an independent source of data with which to check our results.

[7] See Datta *et al.* (1999) for a similar analysis and Kreft and De Leeuw (1998), Snijders and Bosker (1999), and Raudenbush and Bryk (2002) for further discussion of multilevel models; and Gelman and Little (1997), Park *et al.* (2004), and Gelman and Hill (2007) for multilevel modeling of vote-preference data.

the `lmer()` function in the open-source statistical software package R (R Project 2000) and the Bayesian software package Bugs (Spiegelhalter *et al.* 1994, 2004) as linked from R (Sturtz *et al.* 2005).

We shall also fit varying-intercept, varying-slope models for individual income, but we defer these to the next section, following a thorough exploration of the models described so far.

RESULTS

Richer States Now Support the Democrats

We first present the comparison of red and blue states — more formally, regressions of Republican share of the two-party presidential vote on state average per-capita income. Figure 1(a) shows that, since the 1976 election, there has been a steady downward trend in the income coefficient over time. As time has gone on, richer states have increasingly favored the Democrats. So far, this fits with the "David Brooks" story of increasing elite support for the left, rather than the "Horace Greeley" story of elite support for the right. Rich, blue states such as California and New York are voting for Democratic presidential candidates, while poorer, red states like Alabama and Mississippi are voting Republican. For the past 20 years, the same patterns appear when fitting southern and non-southern states separately (Figures 1(b) and (c)).

There has been a trend of richer states supporting the Democrats. It makes sense that the red/blue issue has been more widely discussed in recent years, as this pattern has become increasingly clear.

We hypothesized that some of this variation could be explained by inequality. However, after refitting the models including the state-level Gini index of income inequality, we

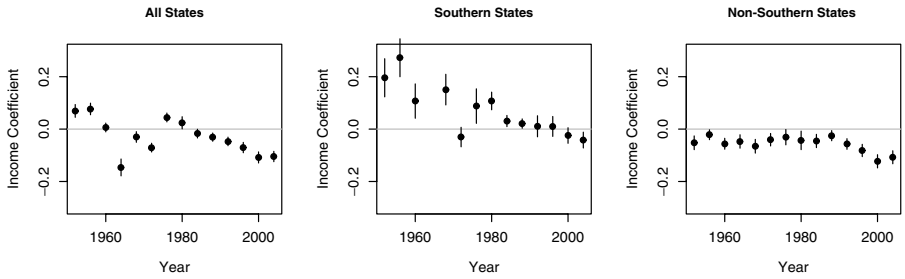

Figure 1. (a) Coefficients for average state income (in tens of thousands of 1996 dollars) in regressions predicting Republican vote share by state. The model w fit separately for each election year. Estimates and standard errors are shown. (b, c) Same model but fit separately to southern and non-southern states each year. In recent years, Republicans have done better in poor states than in rich states.

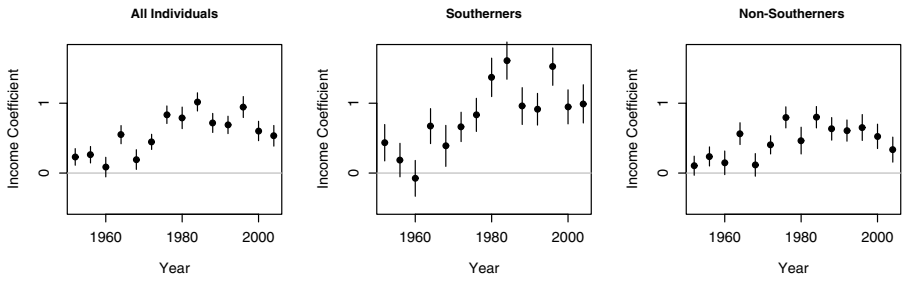

Figure 2. Coefficients and standard errors for income in logistic regressions of Republican vote, fit to NES data from each year. The positive coefficients indicate that higher-income voters have consistently supported the Republicans, a pattern that holds both within and outside the South.

found the coefficients for the Gini index to be essentially zero, and there was little change in the coefficients for state income.

Richer Voters Continue to Support the Republicans Overall and within States

We fit a logistic regression to the reported Republican presidential vote preference on personal income, fit separately to each presidential election since 1952. Figure 2 shows that higher-income people have been consistently more likely to vote Republican, especially since 1970.

We also fit the model controlling for states by fitting a multilevel logistic regression to each election year's NES data, with the 5-point income scale as an individual-level predictor and states as the groups.[8] The estimated coefficients for individual income over time looks much like Figure 2, implying that, on average, richer voters within states support the Republicans. When ethnicity, sex, education, and age are included in the model, the estimated coefficient for income decreases but still clearly remains positive.

Richer Counties Support the Republicans in Some States and the Democrats in Others

Richer counties used to support the Republicans, but this pattern has steadily declined to zero during the past 40 years. Patterns vary by region, however. In most southern states, rich counties voted for Republicans in the past and continue to do so. The southern states that support the Republicans most strongly show the highest coefficients — that is, the

[8] The NES uses cluster sampling and so, strictly speaking, the states in this analysis actually represent collections of sampled clusters. By ignoring the cluster sampling in the analysis, we may be understating standard errors. We are not so worried about this issue here because the results show a consistent pattern over time.

strongest relations between county income and Republican vote share. In contrast, in the western states, richer counties once tended to vote Republican, but now increasingly vote for Democrats (that is, the coefficients in the plots for these states are now negative). Trends in the midwest and northeast are more mixed.

Another way to understand these patterns is to compare counties within richer blue Democratic-leaning states and poorer red Republican-leaning states. In deep-red southern states such as Oklahoma, Texas, and Mississippi, the richer counties support the Republicans and poorer counties support the Democrats. In contrast, consider the states nearest major national media: New York, Maryland, Virginia, and California. In these particular states, the richer counties showed a slight tendency to support the Democrats.

Thus, amusingly, national journalists have noticed a pattern (richer counties supporting the Democrats) that is concentrated in the states where these journalists live. For example, Brooks (2001) compared a rich county in Maryland to a poor county in Pennsylvania. Had he compared counties within states such as Oklahoma, he would likely have noticed an opposite pattern.

MODELING STATE-LEVEL DIFFERENCES IN INDIVIDUAL-LEVEL PATTERNS

Reconciling the Individual and Aggregate Results

Many observers have been misled by the seemingly contradictory pattern of richer states supporting Democrats but richer voters supporting Republicans. As we have seen with our hierarchical model, richer voters support Republicans within states as well as overall; thus direct comparisons of voters for the two parties do not fit the red–blue stereotype. However, the income and voting differences between red and blue states are real.

To better visualize this puzzling pattern, we construct a graph that simultaneously displays variation within and between states. Figure 3 shows three lines, representing the probability of support for Bush in 2000 and 2004 for each of the five income categories in each of three states — Connecticut (the richest state, which supported Gore and then Kerry), Ohio (an intermediate state, which was closely contested), and Mississippi (the poorest state, which supported Bush). The three lines show the estimated probability from the multilevel logistic regression (the lines are, in fact, portions of logistic curves, shifted by different amounts corresponding to the varying intercept in the model).

Figure 3 shows a statistical resolution of the red–blue paradox. Within each state, income is positively correlated with Republican vote choice, but average income varies by state. For each of the three states in the plot, the open circles show the relative proportion of households in each income category (as compared to national averages), and the solid circle shows the average income level and estimated average support for Bush in the state. The Bush-supporting states have more lower-income people, and as a result there is a negative correlation between average state income and state support for Bush, even amid the positive slope for each state. The poor people in red (Republican-leaning) states tend to be Democrats; the rich people in blue (Democrat-leaning) states tend to

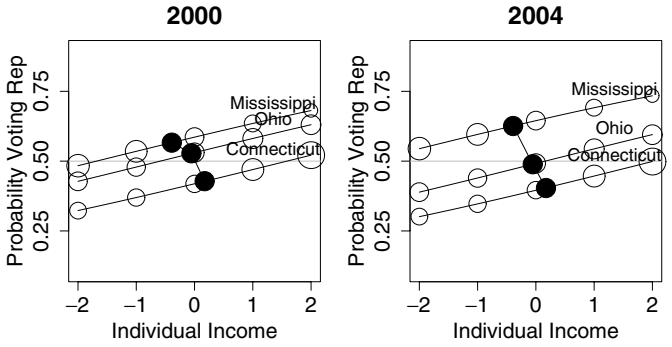

Figure 3. The paradox is no paradox. From the multilevel logistic regression model fit to Annenberg poll data from 2000 to 2004: probability of supporting Bush as a function of income category, for a rich state (Connecticut), a middle-income state (Ohio), and a poor state (Mississippi). The open circles show the relative proportion (as compared to national averages) of households in each income category in each of the three states, and the solid circles show the average income level and estimated average support for Bush for each state. Within each state, richer people are more likely to vote Republican, but the states with higher income give more support to the Democrats.

be Republicans. Income matters, but so does geography. Individual income is a positive predictor, and state average income is a negative predictor, of Republican presidential vote support. The graph (which is related to plots developed for examining variation in medical statistics; see Baker and Kramer (2001), and Wainer (2002)) simultaneously displays variation within and between states that would be difficult to see simply by studying regression coefficients.

Varying-Intercept, Varying-Slope Model

As we have just seen, the varying-intercept multilevel model allows us to understand the positive correlation of individual income with Republican support, in the context of countervailing patterns between states. Our next step is to allow the relation between income and voting to vary by state. We fit a multilevel varying-intercept, varying-slope logistic regression:

$$\Pr(y_i = 1) = \text{logit}^{-1}(\alpha_{s[i]} + \beta_{s[i]} x_i), \quad \text{for } i = 1, \ldots, n, \tag{1}$$

where $s[i]$ represents the state where respondent i lives and x_i is the respondent's income (on the -2 to $+2$ scale). The state-level intercepts and slopes that are themselves modeled given average state incomes and region indicators, with group-level errors (that is, unexplained state-level variation in intercepts and slopes) having mean 0 and covariance matrix estimated from data. By including region and average income as state-level predictors, we are not requiring the intercepts and slopes to vary linearly by income

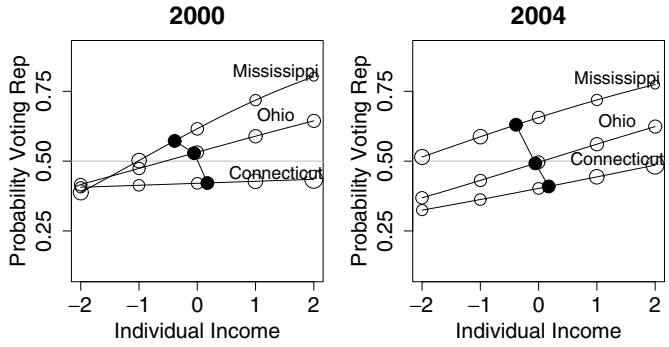

Figure 4. From the multilevel logistic regression model with varying intercepts and slopes fit to Annenberg poll data from 2000 to 2004: probability of supporting Bush as a function of income category, for a rich state (Connecticut), a middle-income state (Ohio), and a poor state (Mississippi). The open circles show the relative proportion (as compared to national averages) of households in each income category in each of the three states, and the solid circles show the average income level and estimated average support for Bush for each state. Income is a very strong predictor of vote preference in Mississippi, a weaker predictor in Ohio, and only weakly predicts vote choice at all in Connecticut. See Figure 5 for estimated slopes in all 50 states, and compare to Figure 3, in which the state slopes are constrained to be equal.

within region — the error terms ϵ_s allow for deviation from the model — but rather are allowing the model to find such linear relations to the extent they are supported by the data.

From this new model, we indeed find strong variation among states in the role of income in predicting vote preferences. Figure 4 recreates Figure 3 with the estimated varying intercepts and slopes. As before, we see generally positive slopes within states and a negative slope between states. What is new, though, is a systematic pattern of the within-state slopes, with the steepest slope in the poorest state — Mississippi — and the shallowest slope in the richest state — Connecticut.

In addition, the varying-intercept, varying-slope model improves the fit compared to the simpler model in which only intercepts vary. In addition to being clear from the consistent patterns in the graphs, a formal comparison shows that allowing the slopes to vary reduces the deviance information criterion (DIC) by 74 and 53 for the analyses from 2000 to 2004, respectively.[9]

Figure 5 shows the estimated slopes for all 50 states and reveals a clear pattern, with high coefficients — steep slopes — in poor states and low coefficients in rich

[9] DIC is a measure of fit that automatically adjusts for the number of parameters in a model; a decrease in DIC implies an estimated improvement in a model's out-of-sample predictions, not merely an improved fit to observed data (Spiegelhalter *et al.* 2004).

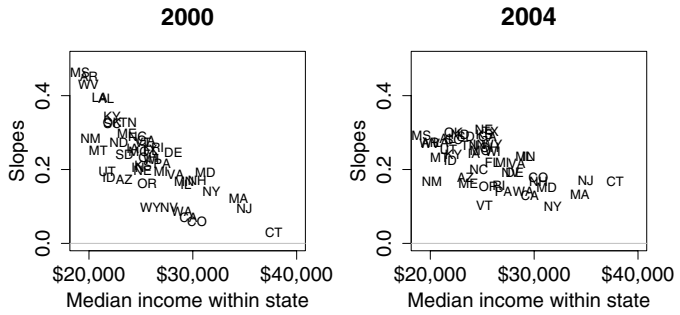

Figure 5. Estimated coefficients for income within state plotted vs. average state income, for the varying-intercept, varying-slope multilevel model fit to the Annenberg survey data from 2000 to 2004.

states. Income matters more in red America than in blue America. Or, to put it another way, being in a red or blue state matters more for rich voters than for poor voters.

The large sample size of the Annenberg survey makes it easy to estimate a varying-slope model. However, the survey was not done before 2000. To see how varying state income effects have changed over time, we turn to exit polls. Figure 6 replicates Figure 4 for the years 1984–2004. The generally positive slopes within states have persisted for decades, but only since 1992, and especially since 1996, have systematic differences between rich and poor states become so clear.[10]

Figures 7 and 8 show the estimated intercepts α_s and slopes β_s in Model 1 as a function of average state income for each presidential election year since1984. The estimates vary from year to year, but we again see the strongest patterns since 1996: poor states have become consistently more Republican, and the coefficients for income have been higher in these states.[11]

We performed some model checking with both the Annenberg and exit polls, comparing individual states to the fitted models. A natural concern is nonlinearity or even

---

[10] A problem with fitting state-by-state models here is that the exit polls use cluster sampling (see Footnote 6), and so technically our logistic regressions, which assume independence among data within a state, is inappropriate. Essentially, we must interpret the resulting estimates for each state as applying to the selected clusters rather than to the entire state. We trust the general patterns, however, because we are interested in the general patterns of income and vote preference comparing rich and poor states.

[11] The multilevel model shrinks the state estimates toward the estimated group-level regression lines. In a year such as 2000 where intercepts and slopes are shrunk very strongly toward the fitted lines, this does not mean that we are certain that all 50 states fall along these lines, but rather that the data are *consistent* with the fitted lines, and the multilevel model finds this pattern. That is, we believe there is a strong (negative) correlation between intercept and average state income, and between slope and average state income, even though any of the particular states might not fall exactly on these fitted lines.

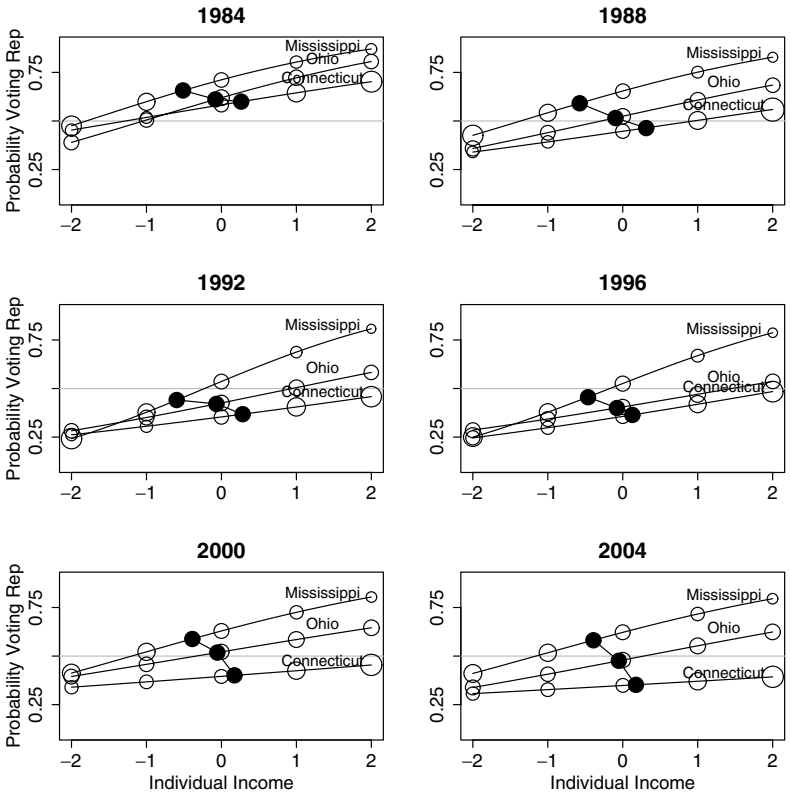

Figure 6. Using exit poll data from 1984 to 2004, results for the varying-intercept, varying-slope multilevel logistic regression. The curves show the probability of supporting Bush as a function of income category, within states that are poor, middle-income, and rich.

non-monotonicity in the relation between income and Republican voting, either in aggregate or within states. In most states there were no serious departures from approximate linearity, and binned residual plots (Gelman *et al.* 2000) did not reveal problems with the fitted logistic regression model. (In contrast, had we stopped at the varying-intercept model shown in Figure 3, we would have found big problems with the model fit, most notably in the extreme income categories in the richest and poorest states.)

Ethnicity and Other Demographic Variables

Could the varying income effects we have shown be merely a proxy for race? This is a potentially plausible story. Perhaps the high slope in Mississippi reflects poor black Democrats and rich white Republicans, while Connecticut's flatter slope arises from its

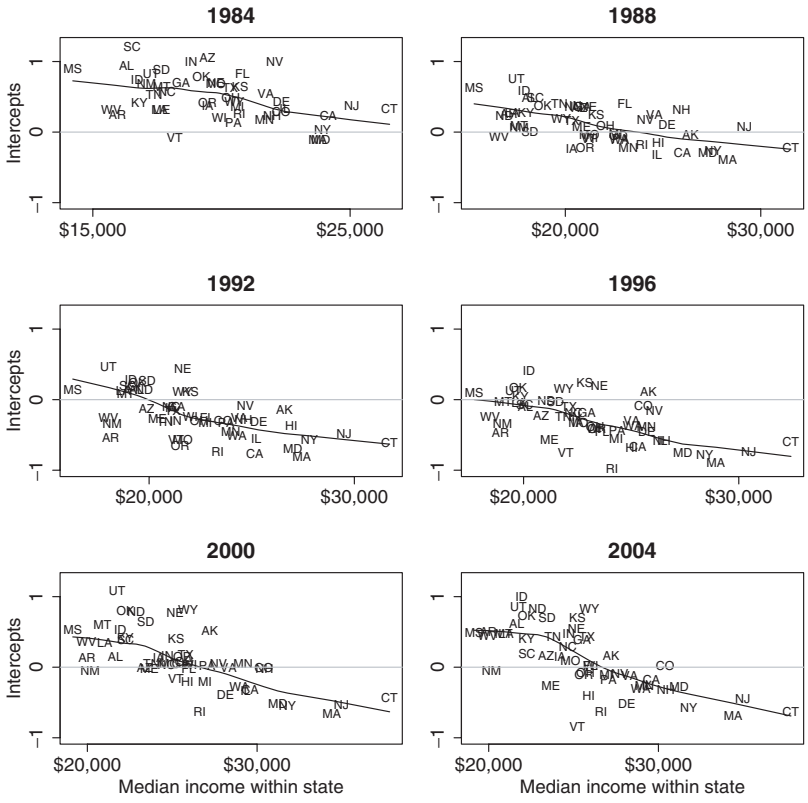

Figure 7. For the varying-intercept, varying-slope logistic regressions of Republican presidential vote preference on individual income: estimated state intercepts plotted vs. average state income. Models fit separately to exit poll data from each election year.

more racially homogeneous population. To test this, we replicate our analysis, dropping all African–American respondents. This reduces our key pattern by about half. For example, in a replication of Figure 5, the slopes for income remain higher in poor states than in rich states, but these slopes now go from about 0.2 to 0 rather than from 0.4 to 0.

To see if the income patterns could be explained by other demographic variables, we went back to the full dataset for the Annenberg surveys in 2000 and 2004 and added individual-level predictors for female, black, four age categories, and four education categories; and group-level predictors for percent black and average education in each state. After controlling for all these, the patterns for income remained: within states, the coefficient for individual income on probability of Republican vote was positive, with steeper slopes in poorer states; after controlling for the individual and group-level predictors, richer states supported the Democrats.

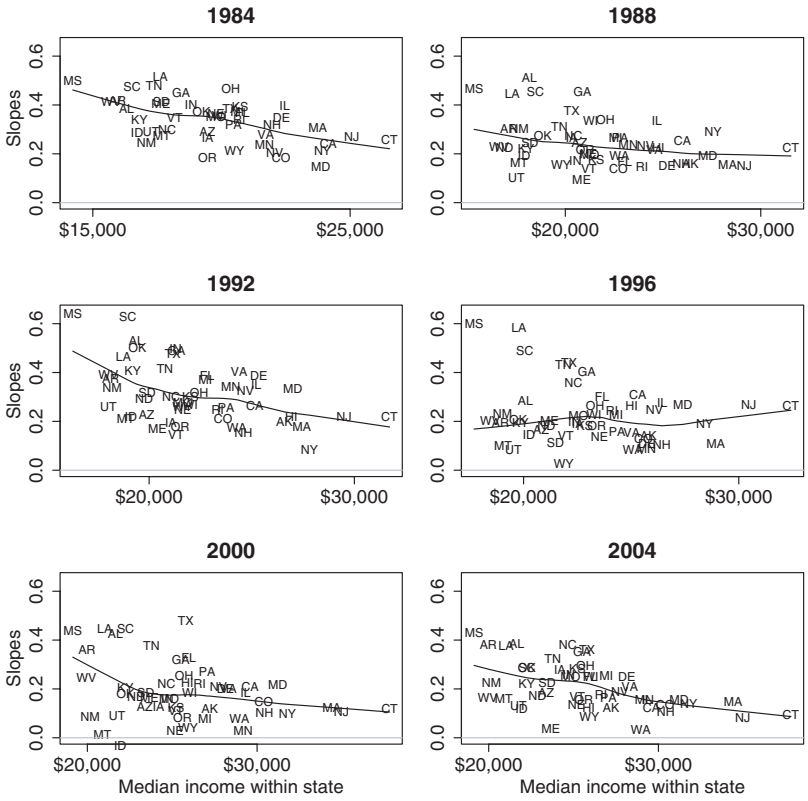

Figure 8. For the varying-intercept, varying-slope logistic regressions of Republican presidential vote preference on individual income: estimated state slopes plotted vs. average state income. Models fit separately to exit poll data from each election year.

Our varying-intercept, varying-slope model has thus redefined the puzzle: in asking why the patterns within states differ from those between states, we are specifically interested in why slopes have become so shallow in rich states — that is, what's the matter with Connecticut? We have found that the differences between rich and poor states have become much more prominent in the past 10 years, that they cannot simply be explained by race, and that they cannot be explained by the set of demographic variables that are typically used in adjusting survey respondents.

This is not to say that income is causing support for Republicans (or that such a causal relation is stronger in Mississippi than in Connecticut), but rather that richer voters within any state are more likely to support the Republicans, even after controlling for basic demographic variables — and this pattern is strong in poor states but weak in rich states.

DISCUSSION

Explaining the Differences Between States

As summarized in Figures 4–8, our multilevel analysis reveals three patterns that we would like to understand:

1. Voters in richer states support the Democrats — even though, within any given state, richer voters tend to support the Republicans.
2. The slope within a state — the pattern that richer voters support the Republicans — is strongest in poor, rural, Republican-leaning red states and weakest in rich, urban, Democrat-leaning blue states.
3. These patterns have increased in the past 10 or 15 years.

We have no conclusive explanations for these patterns — our contribution is to discover and highlight them — but we can consider some ideas. First, the positive slopes within states are no surprise — given both the history and the policies of the two parties, it makes sense that the Democrats would do better among the poor and the Republicans among the rich, a pattern that has persisted for decades. At the same time, votes are far from being determined economically — even in Mississippi, which is the state with strongest correlation between income and voting, over 30% of voters in the lowest income category support the Republicans. Income is one of the many factors contributing to voters' ideological and partisan worldviews, and one could, for example, use detailed survey data to try to understand individual-level positive correlation of income and Republican vote choice as coming from differential attitudes toward redistribution, as discussed by McCarty *et al.* (2006). Finally, about half of the pattern is explained by race: African–Americans mostly live in poorer states and themselves tend to be poorer and vote for Democrats.

Also interesting are the recent differences between rich and poor *states* that have gone in the other direction. Having ruled out the most obvious explanation — that rich and poor states represent the preferences of rich and poor voters — it makes sense to consider systematic differences between states, which are particularly interesting given the increasing mobility of Americans, the possibilities of self-stratification in exposure to news media and choosing where to live, and the increasing polarization of states and counties (Klinkner 2004). One direction is to separately analyze rural, suburban, and urban voters: replicating our analysis in this way revealed varying-slope patterns (as in Figure 8) within each group.

Another way of looking at this is to consider state average income as a proxy for secularism or some kind of cosmopolitanism. In other words, the cultural or social conservatism of states may be increasingly becoming negatively correlated with average income. At the same time, if these social issues are increasingly important to voters (perhaps made more salient by Clinton's scandals, as suggested by Fiorina *et al.* (2005)), this would induce changes in the relation between state income and individual vote. It would be interesting to study the relation between income and factors such as church attendance in different states.

Or, to put it another way, economic issues might well be more salient in poorer states such as Mississippi, and so one would expect voting to be more income-based. Conversely, in richer states like Connecticut, voters are more likely to follow non-economic cues. (These issues are raised by Ansolabehere *et al.* (2006), although without the focus on comparing rich to poor states.) In any case, a challenge for explanations of this sort is to understand why they become more relevant in the 1990s, given that the relative rankings of states by income have changed little in the past century. Journalists have also picked up on the 1990s as pivotal in voters' changing perceptions of the two parties (see, for example, Marlantes (2004), and Bishop (2004)), and these perceptions are increasingly important as the lens through which voters view political and economic issues (Bafumi 2004). As discussed by Fiorina *et al.* (2005), diverging ideological positions of the parties can lead to diverging attitudes about the parties among voters, even if the voters themselves remain largely centrist and do not show strong patterns of consistency in issue attitudes (Baldassarri and Gelman 2007).[12]

The Perils of Summarizing Categories by Typical Members; First- and Second-Order Availability Biases

As a result of the electoral college system and also, perhaps, because of the appeal of colorful maps, state-level election results are widely presented and studied. After seeing the pattern of richer coastal states supporting the Democrats and poorer states in the south and middle of the country supporting the Republicans — a pattern that has intensified in recent years (see Figure 1) — it is natural to personify the states and assume that the Democrats have the support of richer voters too. Psychologists have studied the human tendency to think of categories in terms of their typical members; for example, a robin and a penguin are both birds, but robins are perceived of as typical members of the bird category and penguins are not (Rosch 1975, Rosch and Mervis 1975). When looking at the electoral map, commentators are misled by the patterns in red and blue *states* into thinking of typical Republican and Democratic *voters* as having the characteristics of these states.[13]

If we had to pick a "typical Republican voter," he or she would be an upper-income resident of a poor state, and the "typical Democratic voter" would conversely be a lower-income resident of a rich state. But these are more subtle concepts, not directly readable off the red–blue map — and, in any case, we would argue that given the diversity among supporters of either party, choosing typical members is misleading.

---

[12] Similar patterns of varying slopes for individual income have been found in state-level analysis of Mexican presidential elections (Cortina *et al.* 2007) and in a cross-national analysis of legislative elections (Huber and Stanig 2006).

[13] Political scientists have also made the point that the division into red and blue is somewhat unnatural, considering that distributions of votes and issue preferences tend to be unimodal, with most voters, and most states, falling in the middle of the distribution (Ansolabehere *et al.* 2006, Fiorina *et al.* 2005). Here we are making a slightly different point, which is that a typical Republican (or Democratic) state does not look like an aggregation of typical Republican (or Democratic) voters.

In addition to the challenge of trying to summarize diverse groups by their typical members, journalists who compare Democrats and Republicans are subject to another cognitive illusion — the *availability heuristic*, which is the pattern of making judgments based on easily remembered experiences rather than population data (Tversky and Kahneman 1974).

In this case, we could speak of first-order and second-order availability biases. A national survey of journalists found that about twice as many are Democrats as Republicans (see Poyner Online (2003), summarizing the work of Weaver *et al.* (2003)). Presumably their friends and acquaintances are also more likely to support the Democrats, and a *first-order availability bias* would lead a journalist to overestimate the Democrats' support in the population, as in the notorious quote (mistakenly) attributed to the film critic Pauline Kael in 1972: "I can't believe Nixon won. I don't know anybody who voted for him" (see Rubio 2004). Political journalists are well aware of the latest polls and election forecasts and are unlikely to make such an elementary mistake.

However, even a well-informed journalist can make the *second-order* error of assuming that the correlations they see of income and voting are representative of the population.[14] Journalists are predominantly college graduates and have moderately high incomes (median salary in 2001 of $44,000, compared to a national average of $36,000; see Weaver *et al.* 2003) — so it is natural for them to think that higher-income voters such as themselves tend to be Democrats, and that lower-income voters whom they do not know are Republicans. In fact, a national survey of journalists finds a correlation between high income and support for the Democrats,[15] which is consistent with the latte Democrat, Nascar Republican storyline although not representative of the country as a whole, where Republicans are, on average, richer than Democrats.

Another form of availability bias is geographic. The centers of national journalistic activity are relatively rich states including New York, California, Maryland, and Virginia. Once again, the journalists — and, for that matter, academics — avoid the first-order availability bias: they are not surprised that the country as a whole votes differently from the residents of big cities. But they make the second-order error of too quickly generalizing from the *correlations* in their states. As we have discussed earlier, richer counties tend to support the Democrats within the "media center" states but not, in general, elsewhere. And as shown in Figure 5, richer voters support the Republicans just

---

[14] We use the term "second-order" because this bias does not involve inference about a frequency (that is what we refer to as first-order availability bias, for example thinking that muggings are more likely if you have been mugged, or thinking that cancer is rare because you do not know anyone with cancer), but rather inference about a correlation (for example, that richer people are more likely to vote for the Democrats). Correlation, or more precisely covariance, is a second moment in statistical terms $E((x - \mu_x)(y - \mu_y))$, as compared with simple frequencies $E(x)$ which are first moments. What we have termed the "second-order availability bias" is related to the systematic errors in estimation of covariation that have been found by cognitive psychologists (see, for example, Chapter 5 of Nisbett and Ross, 1980).

[15] For example, in the Weaver *et al.* (2003) survey, 37% of Democratic journalists reported incomes exceeding $50,000, compared to only 24% of Republican journalists. Much of this difference presumably arises because better-paid journalists tend to live in big cities which are politically liberal, but for our purposes here what is relevant is the correlation itself, not where it comes from.

about everywhere, but this pattern is much weaker — and thus easier to miss — within these states.

Much has been written in the national press about the perils of ignoring red America, but these second-order availability biases may have done just that, in a more subtle way. At this point, our hypotheses about journalistic biases are purely speculative; however, we hope these ideas can lead to a clearer picture, not only of the correlations between income, voting, and other variables, but of public understanding of these correlations. Future work in this area could include further analysis of journalists' beliefs and attitudes, along with studies of average citizens' perceptions of Democrats and Republicans, and how these perceptions differ by state.

Representation, Ideology, and Authenticity

> I come from Huntington, a small farming community in Indiana. I had an upbringing like many in my generation — a life built around family, public school, Little League, basketball and church on Sunday. My brother and I shared a room in our two-bedroom house.
>
> —Dan Quayle, 1992

> Clinton displays almost every trope of blackness: single-parent household, born poor, working-class, saxophone-playing, McDonald's-and-junk-food-loving boy from Arkansas.
>
> —Toni Morrison, 1998

Income is not the driving factor in politics in the United States. However, income is important in political perceptions and is also clearly relevant to a wide range of policies including minimum wage regulation, tax rates, Social Security, etc., and is also correlated with many measures of political participation (Verba *et al.* 1995). Similarly, geography is not an all-important factor in politics: red/blue maps of elections are appealing, but most of the states are not far from evenly divided. But, once again, geography is highly relevant to decisions on government spending, among other policies.

As the above quotations illustrate, both income and geography are relevant to politicians' claims of authenticity, just as the income and geography of a candidates' supporters are used to signify political legitimacy. In the 2000 presidential election, richer states voted for the Democrat. The recognition of this fact, and especially this long-term trend, was correctly noted by prominent journalists and pundits like David Brooks. But they went a step too far by attributing properties of red and blue states to red and blue voters and constructing inappropriate pictures of typical Republicans and Democrats. The psychological notion of typicality and the second-order availability bias discussed above give us insight as to how journalists could make this error, and the ongoing issues of authenticity and legitimacy explain why this error can have political consequences.

Sociologists and political scientists such as Brooks and Brady (1999), Stonecash (2000), McCarty *et al.* (2002), and Bartels (2006) have recognized that higher-income voters continue to support Republican candidates, and lower-income voters support Democrats

(in fact, this trend has been increasing since the 1950s). They have shown less interest in state-level differences in preferences (with notable exceptions being the Erikson *et al.* 1993, study of state opinions and state policies, and the comparison of party identification among rich and poor voters within states in McCarty *et al.* (2006)). As we have seen, state income is an important predictor of voting behavior in presidential elections, especially for people on the higher end of the income scale. Journalists' focus on red/blue maps has been somewhat misguided, but the differences between states are real, and indeed have changed in recent decades.

Geography matters politically. States are not merely organizational entities — mere folders that divide individuals for convenience. Nor are the differences cosmetic: a y'all here, a Hahvahd Yahd there. No–states have real, significant cultural and political differences. And despite the centripetal tendencies of a national media, drastically lower transportation costs, and a consumer economy frequently indistinguishable along regional lines (Starbucks everywhere) — regional political differences seem, if anything, to be getting more pronounced in the last decade or two.

To the extent political scientists want to understand political behavior in a federal system, we must recognize these differences. From a politician's perspective, given policies will be received differently in various states, even though those states are internally diverse. Therefore, an incentive to target policy geographically exists and has only gotten stronger. For policy analysts, then, increased attention to geography is also warranted.

The technique of multilevel modeling has allowed us to understand these patterns together. Individual income is positively correlated with Republican voting preference, but average state income is negatively correlated with aggregate state presidential voting for Republicans. The apparent paradox is *no paradox at all*, because Figure 4 clearly shows that these are not mutually exclusive relationships.

We can understand the state average income effect as one of context. The Mississippi electorate is more Republican than that of Connecticut; so much so that the richest segment of Connecticutians is only barely more likely to vote Republican than the poorest Mississippians. In poor states, rich people are very different from poor people in their political preferences. But in rich states, they are not.

REFERENCES

Alford, R. R. 1963. "The Role of Social Class in American Voting Behavior." *Western Political Quarterly* 16: 180–194.

Ansolabehere, S., J. Rodden, and J. M. Snyder. 2006. "Purple America." *Journal of Economic Perspectives* 20: 97–118.

Bafumi, J. 2004. "The Stubborn American Voter." Paper presented at the annual meeting of the American Political Science Association, Chicago.

Baker, S. G., and B. S. Kramer. 2001. "Good for Women, Good for Men, Bad for People: Simpson's Paradox and the Importance of Sex-Specific Analysis in Observational Studies." *Journal of Women's Health and Gender-Based Medicine* 10: 867–872.

Baldassarri, D., and A. Gelman. 2007. "Partisans Without Constraint: Political Polarization and Trends in American Public Opinion." Technical report, Department of Political Science, Columbia University.

Bartels, L. M. 2006. "What's the Matter with *What's the Matter with Kansas?*" *Quarterly Journal of Political Science* 1: 201–226.
Bertin, J. 1967. *Semiology of Graphics* (translation by W. J. Berg 1983), University of Wisconsin Press.
Bishop, B., ed. 2004. "The Great Divide." *Austin American-Statesman*, www.statesman.com/specialreports/content/special-reports/greatdivide.
Brooks, C., and D. Brady. 1999. "Income, Economic Voting, and Long-Term Political Change in the U.S., 1952–1996." *Social Forces* 77: 1339–1374.
Brooks, D. 2001. "One Nation, Slightly Divisible." *Atlantic Monthly* 288(5): 53–65.
Buckley, W. F., and L. B. Bozell. 1954. *McCarthy and His Enemies*. Chicago: Regnery.
Cortina, J., A. Gelman, and N. Lasala. 2007. "Income and Vote Choice in Mexican Presidential Elections." Technical report, Department of Political Science, Columbia University.
Datta, G. S., P. Lahiri, T. 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: 1074–1082.
Erikson, R. S., G. C. Wright, and J. P. McIver 1993. "" *Statehouse Democracy: Public Opinion and Policy in the American States*. Cambridge University Press.
Filer, J. E., L. W. Kenny and R. B. Morton. 1993. "Redistribution, Income, and Voting." *American Journal of Political Science* 37: 63–87.
Fiorina, M. P., S. J. Abrams, and J. C. Pope. (2005). *Culture War? The Myth of a Polarized America*. New York: Longman.
Frank, T. 2004. "Lie Down for America." *Harper's Magazine* 308 (April), 33–46.
Gelman, A., Y. Goegebeur, F. Tuerlinckx, and I. Van Mechelen. 2000. "Diagnostic Checks for Discrete-Data Regression Models Using Posterior Predictive Simulations." *Applied Statistics* 49: 247–268.
Gelman, A., and J. Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
Gelman, A., and T. C. Little. 1997. "Poststratification into Many Categories using Hierarchical Logistic Regression." *Survey Methodology* 23: 127–135.
Gelman, A., C. Pasarica, and R. Dodhia. 2002. "Let's Practice What We Preach: Using Graphs Instead of Tables." *American Statistician* 56: 121–130.
Greenberg, D. 2003. *Nixon's Shadow: The History of an Image*. New York: Norton.
Huber, J., and P. Stanig. 2006. "Voting Polarization on Redistribution Across Democracies." Technical report, Department of Political Science, Columbia University.
Issenberg, S. 2004. "Boo-Boos in Paradise." *Philadelphia Magazine*, April.
Klinkner, P. A. 2004. "Red and Blue Scare: The Continuing Diversity of the American Electoral Landscape." *The Forum* 2(2).
Kramer, G. H. 1983. "The Ecological Fallacy Revisited: Aggregate- Versus Individual-Level Findings on Economics and Elections, and Sociotropic Voting." *American Political Science Review* 77: 92–111.
Kreft, I., and J. De Leeuw. 1998. *Introducing Multilevel Modeling*. London: Sage.
Manza, J., and C. Brooks. 1999. *Social Cleavages and Political Change*. Oxford University Press.
Marlantes, L. 2004. "Inside red-and-blue America." *Christian Science Monitor*, July 14.
McCarty, N., K. T. Poole, and H. Rosenthal. 2002. "Political Polarization and Income Inequality." Unpublished manuscript, Department of Politics, Princeton University.
McCarty, N., K. T. Poole, and H. Rosenthal. 2006. *Polarized America: The Dance of Political Ideology and Unequal Riches*. Cambridge, MA: MIT Press.
McGirr, L. 2001. *Suburban Warriors: The Origins of the New American Right*. Princeton University Press.
Nisbett, R., and L. Ross. 1980. *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs, NJ: Prentice-Hall.
Park, D. K., A. Gelman, and J. Bafumi. 2004. "Bayesian Multilevel Estimation with Poststratification: State-Level Opinions from National Polls." *Political Analysis* 12: 375–385.
Perlstein, R. 2002. *Before the Storm: Barry Goldwater and the Unmaking of the American Consensus*. New York: Hill and Wang.
Poyner Online 2003. "The Face and Mind of the American Journalist." Quicklink A28235, 10 April.
R Project 2000. "The R Project for Statistical Computing," www.r-project.org.

Raudenbush, S. W., and A. S. Bryk. 2002. *Hierarchical Linear Models*, 2nd edition. Thousand Oaks, CA: Sage.
Robinson, W. S. 1950. "Ecological Correlations and the Behavior of Individuals." *American Sociological Review* 15: 351–357.
Rosch, E. 1975. "Cognitive Reference Points." *Cognitive Psychology* 7: 532–547.
Rosch, E., and C. B. Mervis. 1975. "Family Resemblances: Studies in the Internal Structure of Categories." *Cognitive Psychology* 7: 573–605.
Rovere, R. H. 1959. *Senator Joe McCarthy*. New York: Harper and Row.
Rubio, S. 2004. Kael/Nixon update, begonias.typepad.com/srubio/2004/12/kaelnixon_updat.html.
Snijders, T. A. B., and R. J. Bosker 1999. *Multilevel Analysis*. London: Sage.
Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde. 2004. "Bayesian Measures of Model Complexity and Fit (with discussion)." *Journal of the Royal Statistical Society B* 64: 583–639.
Spiegelhalter, D., A. Thomas, N. Best, W. Gilks, and D. Lunn. 1994. *BUGS: Bayesian Inference Using Gibbs Sampling*. Cambridge, England: MRC Biostatistics Unit, www.mrc-bsu.cam.ac.uk/bugs/.
Stonecash, J. M. 2000. *Class and Party in American Politics*. Boulder, CO: Westview Press.
Stonecash, J. M. 2005. "Scaring the Democrats: What's the Matter with Thomas Frank's Argument?" *The Forum* 3(3).
Sturtz, S., U. Ligges, and A. Gelman. 2005. "R2WinBUGS: A Package for Running WinBUGS from R." *Journal of Statistical Software* 12(3).
Tufte, E. R. 1990. *Envisioning Information*. Cheshire, CT: Graphics Press.
Tversky, A., and D. Kahneman. 1974. "Judgment Under Uncertainty: Heuristics and Biases." *Science* 185: 1124–1130.
Verba, S., K. L. Schlozman, and H. E. Brady. 1995. *Voice and Equality*. Harvard University Press.
Wainer, H. 2002. "The BK-Plot: Making Simpson's Paradox Clear to the Masses." *Chance* 15(3): 60–62.
Weaver, D., R. Beam, B. Brownlee, P. S. Voakes, and G. C. Wilhoit. 2003. *The American Journalist Survey*. Indiana University School of Journalism.
Wright, G. C. 1989. "Level-of-Analysis Effects on Explanations of Voting: The Case of the 1982 U.S. Senate Elections." *British Journal of Political Science* 19: 381–398.