

What's Law Got to Do with It?

Suzanna Sherry

The authors of this fascinating study modestly disclaim its significance, yet suggest that the results prove their model a success. As a legal expert, I have a rather different perspective on the results. I look at the numbers holistically, not statistically. And what I see tells a different story—if it tells any story at all.

The classification trees used by the model to generate predictions are a mass of tangled branches, because one justice's predicted vote often turns on another justice's predicted vote. By looking at the predominant variables for each justice, however, one can conclude that the predictions for Justices Ginsburg, Breyer, and Souter ultimately turn substantially on the predicted vote of Justice O'Connor, while the predicted votes of Chief Justice Rehnquist and Justices Kennedy and Thomas turn substantially on the predicted vote of Justice Scalia. By looking at the classification trees for Justices O'Connor and Scalia, then, we can learn a great deal about how the model works for eight of the nine justices. Moreover, the model beat the experts at predicting both O'Connor and Scalia; indeed, the largest gap between man and machine was in predicting how Justice O'Connor would vote.

So how did the model predict the votes of O'Connor and Scalia? For both justices, the first variable on the classification tree is whether the lower court decision was liberal or conservative. For Justice O'Connor,¹ the model predicts that she will always reverse a liberal lower-court decision. Whether she will reverse a conservative lower-court decision depends, first, on the circuit of origin (she will affirm cases from the Second, Third, DC, and Federal circuits), then on whether the respondent is the United States, and finally on the subject matter of the case. For Justice Scalia, the tree is slightly more complicated and slightly more tied to subject matter (see fig. 1). Again, however, besides the ideological classification of the lower-court decision, the circuit of origin plays a large role: for conser-

vative lower court decisions, Scalia will always affirm cases from the Third, Fourth, Tenth, DC, and Federal circuits, but will affirm only some types of cases from the other circuits; for liberal lower court decisions (in certain subject areas), he will affirm cases from the First, Third, Sixth, Seventh, Eighth, and DC circuits and reverse cases from the other circuits. Anyone who follows the politics in, quality of, or splits among the federal courts of appeals will recognize that these collections of circuits have little or nothing in common—they seem completely random.

While the study's authors suggest that the "circuit of origin" variable has to do with "the agenda process," one should still expect to find some commonalities among the circuits that particular justices favor or disfavor. I find none. While I do not expect the study's authors to be able to fully explain why their predictive model works, I am troubled when even with hindsight we cannot make sense of the variable.

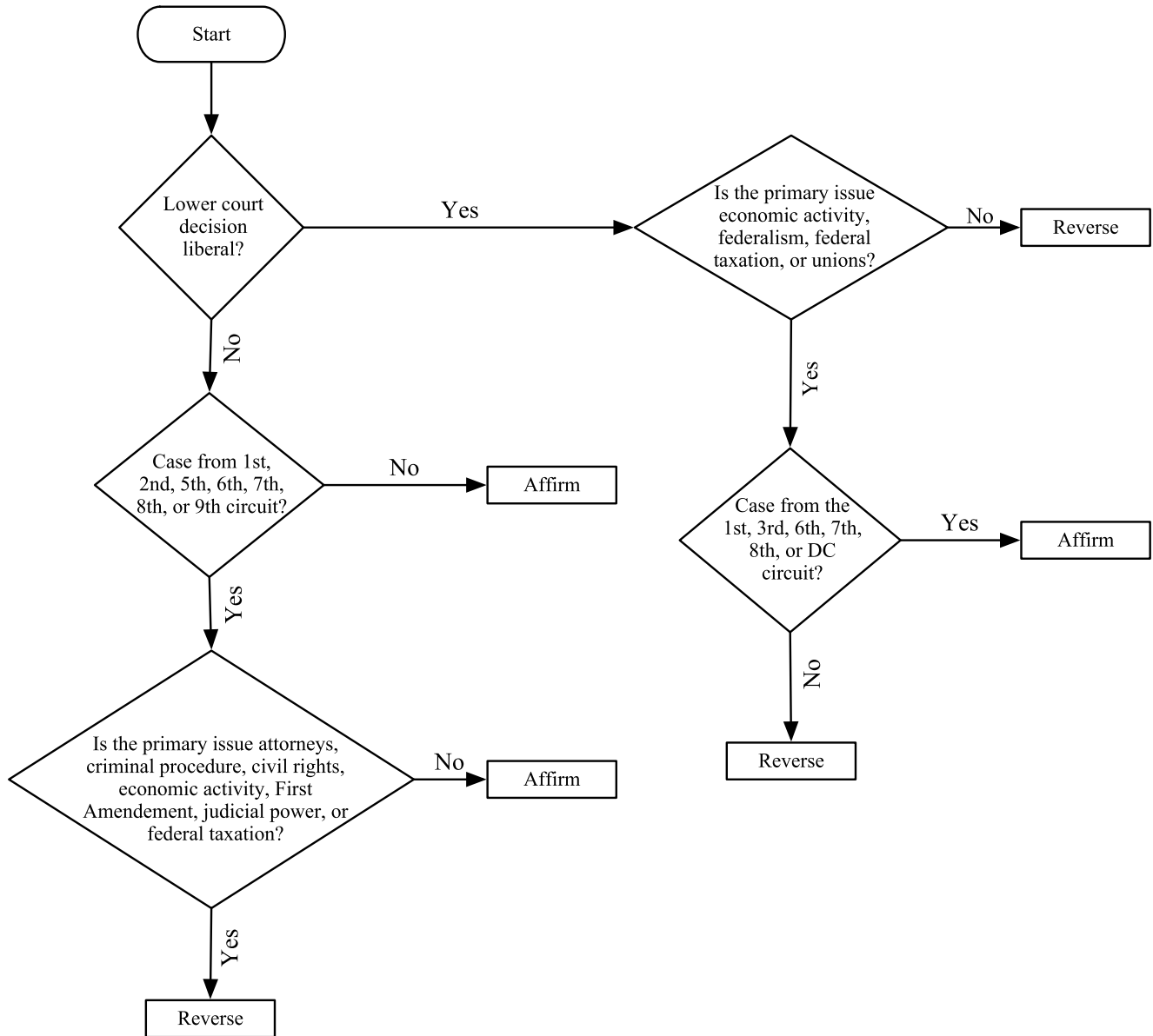
The classification trees for these two pivotal justices—on whose votes many of the other predictions rely—thus rest first on politics and second on a completely inexplicable factor. Leaving the latter aside for a moment, the emphasis on politics makes the model strongly attitudinal. So what are we to make of the fact that the experts also say they relied on an attitudinal model? However we explain it, we must also account for two striking results: the differences in which justices each method was better at predicting, and the differences in relative success rates of the two methods depending on the issue area of the case.

First, there is a significant difference between the model and the experts when it comes to predicting the votes of particular justices.² When the justices are arrayed from least to most conservative, the expert success rate forms a rough sideways V; they are least successful at predicting the centrist justices. The model's success rate on the same chart, however, is almost an evenly increasing percentage; the model is worst at predicting the liberal justices, better for the centrists, and best for the conservatives.

Second, comparing success rates for individual justices by subject area produces interesting outcomes, as table 1 shows. In criminal procedure, the experts bested the model for five of the nine justices, but for most of the justices the margins were small, less than five percentage points, and for all the justices it was less than ten. In civil rights, the model

Suzanna Sherry (suzanna.sherry@law.vanderbilt.edu) is the Cal Turner Professor of Law and Leadership at Vanderbilt University Law School. She thanks Rebecca Brown, Paul Edelman, and Erin O'Hara for comments on earlier drafts of this essay.

Figure 1
 Estimated classification tree for justice Scalia for forecasted non-unanimous cases



beat the experts on five of the nine justices, by larger margins—in four instances by more than ten percentage points. In the economic activity area, the model significantly outperformed the experts, beating them regularly by 20–30 percentage points. Even in those two fields, however, the experts managed to predict some justices better than did the model.³

The experts turned the tables in cases with less political salience. In federalism cases⁴ their predictions of individual justices' votes were more accurate than the model's in the aggregate by a large margin: 70.4 percent to 53.5 percent. And in cases involving judicial power, the experts swept the field, outpredicting the model on every justice, and racking

up success rates that were 7.9 to 82.2 percentage points above those of the model. As the authors point out, these statistics might be questioned because of the small number of cases and the flexibility in individual coding decisions, but they are intriguing nonetheless.

What might explain all of these facts? I suggest that the liberal justices vote the law, the conservative justices vote their politics, and the centrist justices do neither.⁵ Consider the evidence regarding the different attitudinalist approaches of the model and the experts. Both the classification trees and the fact that the model did poorly in areas with low political salience—and best in cases involving civil rights and regulation of economic activity, which have perhaps

Table 1
Percent correct model and expert forecasts of votes, by justice and issue area

Justice	Criminal procedure		Civil rights		Economic activity		Federalism		Judicial power	
	Model	Experts	Model	Experts	Model	Experts	Model	Experts	Model	Experts
Rehnquist	69.2	77.8	85.7	91.4	81.2	51.3	80.0	33.3	75.0	84.2
Stevens	84.6	80.6	57.1	68.6	56.2	64.1	40.0	100.0	50.0	84.2
O'Connor	76.9	75.0	64.3	48.6	86.7	55.6	60.0	58.3	37.5	68.4
Scalia	84.6	75.0	78.6	97.1	75.0	48.7	60.0	50.0	37.5	68.4
Kennedy	69.2	77.8	85.7	82.4	75.0	51.3	100.0	75.0	50.0	57.9
Souter	84.6	61.1	85.7	57.1	56.2	61.5	20.0	91.7	25.0	73.7
Thomas	84.6	86.1	85.7	94.3	60.0	50.0	80.0	58.3	37.5	68.4
Ginsburg	76.9	77.8	71.4	60.0	62.5	59.0	0.0	100.0	12.5	94.7
Breyer	69.2	69.4	64.3	51.4	60.0	70.3	40.0	66.7	12.5	84.2

Note: For criminal procedure (n = 13); civil rights (n = 14); economic activity (n = 16), judicial power (n = 8), and federalism (n = 5) cases. Issue areas are coded according to Spaeth's protocol, and are mutually exclusive.

the highest political visibility—suggest that politics played a large role in the model's predictions. The experts, on the other hand, while confessing to using the justice's ideology in making their predictions, did worst in areas where one might expect an attitudinalist approach to succeed and best in more legalistic areas. Perhaps the experts did not actually focus on ideology, or perhaps they are just poor attitudinalists. Maybe they simply tempered their attitudinalist predictions with knowledge about the extent to which politics influences each particular justice, using an attitudinalist model for only some justices. (The questionnaires asked the experts what factors they considered for each case, not each justice, so they might have used ideology for only some justices.) In any case, despite the protests of the study's authors, it seems fair to suggest that the model adopted a more attitudinalist approach (with some additional nuances) and the experts a more legalist approach (again, with some nuances, and perhaps only for some justices).

It is then easy to explain the model's pattern of success. The more conservative the justice, the larger the role played by ideology and the more accurate the model's attitudinalist prediction. Moreover, this explanation is consistent with the facts surrounding the appointment of each justice. The most conservative justices—Scalia, Thomas, and Rehnquist—were nominated (or elevated) by conservative Republican presidents for whom ideology played an important role in the selection process, and confirmed by a Republican Senate or by a narrow margin that included only a few Democrats. Two of the four liberal justices, however, were nominated by Republicans—how ideological could they be?—and the other two by a moderate Democrat with only token opposition from Senate Republicans.⁶ This might

also explain the experts' ability to predict the three most conservative justices almost as well as the model does: their reliance on attitudinalism may have been strongest for the justices they perceived as most ideologically committed.

For the four most liberal justices, the model fared poorly using an attitudinalist approach to predict their votes—much worse than it did for the conservatives. The experts, however, predicted the liberal justices almost as well as they did

the conservatives. And their predictions were most accurate overall—ranging from a low of 58 percent for Justice Kennedy to a high of 95 percent for Justice Ginsburg—on cases raising questions of judicial power. The experts' success rate was higher than 70 percent for all four liberal justices in this area (see table 1). This pattern is exactly the converse of the one we find in the model's predictions of the liberals' votes, confirming that the experts were using a legalist, rather than attitudinalist, approach—at least for the liberal justices. And, indeed, the questionnaires confirm that the legal experts used legal precedent and legal analysis—again, at least for predicting some justices.

If the model uses a predominantly attitudinalist approach throughout, and the experts use a legalist approach most of the time but an attitudinalist approach for the conservative justices, what can we conclude about the various justices from the pattern of success rates? Legalism, but not attitudinalism, is a better way to predict the votes of the four liberals because they place more emphasis on law than on politics. Attitudinalism, however, works just fine for the three conservatives (except in cases in which politics is unimportant), because their votes in fact depend more on politics than on law. In other words, the liberals vote the law and the conservatives vote their politics, as I contended earlier.

Liberal justices vote the law, the conservative justices vote their politics, and the centrist justices do neither.

What about the two centrists, Justices O'Connor and Kennedy? One possibility is that attitudinalism predicts their votes fairly accurately (although not as accurately as it does for the conservatives), but that the experts don't recognize that fact as well as they do with the more conservative justices. The experts' mistaken assumption that O'Connor and Kennedy behave like the liberals—looking at the law more than at politics—produces the distinctive V-shape, and the two justices' failure to conform entirely to an attitudinalist approach puts O'Connor and Kennedy about in the middle of the model's success rate. This suggests that neither legalism nor attitudinalism accurately captures the influences on these two justices. But such an explanation leaves us with a question: what *does* determine how O'Connor and Kennedy will vote?

Which brings us back to the additional factors on the classification trees. For Kennedy, the *only* nonpolitical variable is the circuit of origin; for O'Connor, the circuit of origin is the most important variable after the political slant of the lower-court decision, although she also favors plaintiffs with civil rights, economic liberties, or first amendment claims against the United States, and states' rights. (She is a libertarian "federalist," in other words—unsurprising considering that her background includes a stint in a sagebrush legislature.)⁷ But why does the circuit of origin explain what neither law nor politics can? I find that conclusion nothing short of bizarre. Something else must be going on, and the apparent influence of circuit of origin is simply coincidental.

What is the something else? I am afraid that it might be pure gut reaction, the "I know it when I see it" of Justice Stewart's definition of pornography.⁸ As the study authors note, "the legal experts put great weight on legal authority—primarily in the form of prior Supreme Court opinions—in making their predictions." And yet the experts were worst at predicting Justices O'Connor and Kennedy, suggesting perhaps that these two justices do not themselves place very much weight on legal authority.⁹

And, in fact, the votes of the two justices—especially O'Connor—in some key cases of the 2002 term illustrate a rather cavalier attitude toward precedent. Justice Kennedy, at least, was forthright in his rejection of precedent in *Lawrence v. Texas*, voting to overrule the 17-year-old *Bowers v. Hardwick*. Justice O'Connor in *Lawrence* was less candid on the inconsistency between her vote and her prior views: in 1986 she joined an opinion that essentially held that homosexuals had no rights which heterosexuals were bound to respect, but then ruled in *Lawrence* that legislatures could not rationally criminalize homosexual sodomy without criminalizing heterosexual sodomy as well. (I am not suggesting that *Bowers* should not have been overruled, only that *Lawrence*—and particularly Justice O'Connor's concurring opinion—illustrates the centrists' quirky decision making.)

Justice O'Connor was also a pivotal vote—and in one case the author of the opinion—in three other cases that paid only lip service to precedent. In *Nevada Department*

of Human Resources v. Hibbs, she joined a majority upholding Congress's abrogation of state sovereign immunity in the Family and Medical Leave Act, thus allowing states to be sued for damages by individuals who claimed that their state employer violated the FMLA. Prior cases had invalidated similar attempted abrogations under the Age Discrimination in Employment Act and the Americans with Disabilities Act, concluding that Congress did not have sufficient evidence of a pattern of state constitutional violations to enact remedial statutes. The evidence of state constitutional violations supporting the FMLA, however, was no stronger than the evidence supporting the ADA. Moreover, of the eight circuits that had previously confronted the question, seven had found the FMLA abrogation invalid under the precedents—further evidence that the Court's decision was inconsistent with its prior cases. Nevertheless, Chief Justice Rehnquist's majority opinion, joined by O'Connor, purported to apply the earlier precedent. O'Connor and Rehnquist were the only justices who were in the majority in both the earlier cases and in *Hibbs*, again suggesting that their view of the precedent was unusual.¹⁰

In *Demore v. Kim*, Justice O'Connor provided the fifth vote upholding the constitutionality of detaining deportable aliens pending a deportation hearing, without any individualized determination of their dangerousness or risk of flight. Only two years earlier, she had provided the fifth vote in *Zadvydas v. Davis*, which held unconstitutional the detention of aliens who had been ordered deported (after a hearing) but who were required to remain in the United States because no other nation would accept them. Other than Justice O'Connor, there is no overlap in the majorities in the two cases, suggesting that perhaps they are inconsistent. While the cases may arguably be distinguishable on their facts, what is noteworthy is that the majority opinion in *Demore* cites Justice Kennedy's *Zadvydas* dissent more than it cites the *Zadvydas* majority opinion.

Finally, consider *Grutter v. Bollinger*, in which the Court upheld the University of Michigan Law School's affirmative action program. Justice O'Connor's majority opinion purported to apply strict scrutiny, which requires that the affirmative action program be "necessary to" or "narrowly tailored to achieve" the state's compelling interest in a racially diverse student body. She nevertheless deferred to the university's own determination of the benefits of a racially diverse student body, the lack of alternative methods of obtaining a racially diverse student body, and the temporary nature of the program. The Court has never before upheld a racially discriminatory state policy under strict scrutiny, nor has it ever suggested that the challenged program is due any deference from the Court. Instead, it has always demanded that such programs be subjected to the most searching scrutiny. (Three of the four justices who joined her opinion explicitly rejected the application of ordinary strict scrutiny to affirmative action.)¹¹ Moreover, on the same day that it

upheld the Law School's affirmative action program, the Court struck down the affirmative action program used by the University of Michigan in undergraduate admissions. Only Justices O'Connor and Breyer were in the majority in both cases. As the dissenters in *both* cases pointed out, the only difference between the undergraduate program and the law program was that the undergraduate program was more candid about the role played by race.

These cases provide anecdotal evidence that complements the project's results. Over the course of the term, Justices Kennedy and O'Connor (especially O'Connor) confounded experts who relied on legal precedent to predict their votes. In these cases, we can see how Justice O'Connor (and, to a lesser extent and more candidly, Justice Kennedy) treats precedent, which explains why legal experts were unable to predict her votes. We are now left only to explain why Justice O'Connor's erratic votes—based on neither precedent nor politics—seem to track such odd variables as circuit of origin.

If one steps back from individual cases in which a justice misapplied precedent, however, there is another possible explanation for the experts' poor performance in predicting the centrists' votes, although it does not explain the relative success of the computer model. If Justices Kennedy and O'Connor are taking an essentially pragmatist approach to legal questions—exploring every angle and relying heavily on fact-specific context—their decisions might be more difficult to predict, especially if the experts themselves were not pragmatists (or did not read Kennedy and O'Connor as pragmatists). Were it not for the lack of candor exhibited in the manipulation of precedent, this result might actually be encouraging, suggesting that the two justices engage in a typical common-law decision-making process.¹²

But perhaps the model's ability to predict Justice O'Connor better than the experts is a fluke. In other words, there might also be nothing going on. The authors admit that the study is of limited significance because of its necessarily narrow focus, the small number of cases, and the unscientific selection of experts.¹³ There is also a serious problem with relying on coding decisions, especially those created and applied by nonlawyers. The coding choices are inevitably ambiguous, and inevitably lead to internal inconsistencies. For example, the difference between "judicial power" and "federalism" is fuzzy, especially in any case raising the question of whether a federal court—as opposed to a state court—has jurisdiction. And, as noted above, several obvious (to lawyers) federalism cases were coded as economic activity cases—presumably by nonlawyers; a different coder might have made a different choice. Indeed, different coders *do* make different choices. Harold Spaeth originally coded about half the 68 cases in the sample. The project authors tentatively coded the other half, which Spaeth later recoded. Of the approximately 34 cases (half of 68) that Spaeth recoded, 16 were given different codes. Leaving aside the problem of "predictive" recoding *after* the Supreme Court has already issued its ruling, the fact

that Spaeth and the project authors disagreed about the appropriate coding in almost half the cases they both coded suggests the impossibility of consistent coding.¹⁴

Finally, the idea that every decision—whether by a lower court or by the Supreme Court—can be coherently coded as either "liberal" or "conservative" verges on nonsense. How, for example, should one code *Nguyen v. United States*, in which a panel of the Ninth Circuit upheld a criminal conviction against various criminal procedure challenges but the Supreme Court reversed on the ground that having a judge from the Mariana Islands (a federal judge but not an Article III judge) on the panel violated a federal statute? Unsurprisingly, the lineup in the Supreme Court was peculiar: Chief Justice Rehnquist wrote a dissenting opinion, which was joined by Justices Scalia, Ginsburg, and Breyer.

In addition to coding decisions, the results are also highly dependent on the particular cases in the sample, and there is evidence to suggest that this term's cases are unrepresentative of this Court in general. It was widely suggested that the Court was unusually and surprisingly liberal in the 2002 term.¹⁵ Some statistics bear this out. Of the 16 cases, or parts of cases, that were decided by a 5–4 majority, the five most conservative justices were in the majority in five cases and the four most liberal were in the majority in seven (in the remaining four cases, the majority consisted of a mix of conservatives and liberals). Of the 12 cases or parts of cases decided by a 6–3 vote, there were four in which the five most conservative justices were in the majority, and five in which the four most liberal justices were. (Again, the remaining cases were mixed.) This is the first time in this Court's nine years that the liberals won more close cases than the conservatives. The far right justices, Scalia and Thomas, were the most frequent dissenters; in past terms the far left Justice Stevens has been the most frequent dissenter. Moreover, the Fourth Circuit, the most conservative in the nation, was the least reversed circuit for the 2001 term (with a reversal rate of 64 percent) but had a reversal rate of 100 percent for the 2002 term. While none of these results are likely to be statistically significant, they do tend to confirm the impression of an unusually liberal term.

If this term was in fact aberrational, and especially if the pivotal votes of Justices O'Connor and Kennedy were unusual, then the fact that the project's results are statistically significant for this term is unlikely to be repeated. This term simply might have been one of the 2.5 percent of cases in which these results could be produced by random chance. Here's a challenge to the model from one expert: try again in a more representative term (and with experts who have less intellectual stake in the outcomes) and we'll beat you fair and square.

Conclusion

They say that no news is good news. This study is either no news or bad news. Either the results are an artifact of a

peculiar term, poor coding, and the wrong experts, or the results tell us that the votes of two pivotal Supreme Court justices depend on either politics or seemingly random factors. Those of us who have spent years arguing that constitutional law is truly law, not just politics, and that judges do the best they can, were disheartened in late 2000 by *Bush v. Gore*. Most of us recovered. Now comes the new millennium and a computer model to inform us that we regained our confidence too soon. Personally, I'd rather have no news.

Notes

A complete reference list for the entire symposium appears on pp. 791–93, below.

- 1 See Martin et al. 2004, fig. 1, for O'Connor's classification tree.
- 2 Ibid., fig. 2.
- 3 Moreover, the results depend significantly on questionable coding decisions: 4 of the 16 cases coded as economic activity are, I would suggest, really federalism or judicial power cases. *Pierce County v. Guillen* raised the question whether Congress was entitled to make certain highway information collected by the state—as required by federal statute—undiscoverable and inadmissible in state civil litigation. *Cook County v. United States ex rel. Chandler* determined that counties are “persons” subject to qui tam actions; it was a follow-up to an earlier case holding that states are not persons in that context because otherwise the qui tam statute would raise grave constitutional (federalism) questions. *Pharmaceutical Research and Manufacturers of America v. Walsh* and *Hillside Dairy Inc. v. Lyons*; both involved the dormant Commerce Clause (and, in *Walsh*, preemption), which implicates federalism because it is part of the constitutional scheme of balancing national and local interests. For further discussion of coding difficulties, see below, p. 773.
- 4 Federalism may have political salience in some cases, but did not have much last term. Other than *Nevada Department of Human Resources v. Hibbs*, the federalism cases had neither high visibility nor significant political impact. The cases included one on ERISA, which no citizen understands (much less cares about); one on whether the federal Boat Safety Act preempts a state cause of action; and one on utilities regulation. The only federalism case (besides *Hibbs*) that was conceivably political was *American Insurance Association v. Garamendi*, which invalidated California's Holocaust Victim Insurance Relief Act; but any case in which Justice Ginsburg writes a dissent that is joined by Justices Stevens, Scalia, and Thomas confounds political classification. The addition of the cases in note 2 would not alter my conclusion that this term's federalism cases had low political salience. Whether it would change the success rates of the model cannot be determined without rerunning the program.
- 5 I am characterizing Justices Stevens, Souter, Ginsburg and Breyer as liberal, Chief Justice Rehnquist and Justices Scalia and Thomas as conservative, and Justices O'Connor and Kennedy as centrist. This accords with the project authors' array of justices in figure 2.
- 6 Chief Justice Rehnquist was originally nominated by Republican Richard Nixon and confirmed by a 68–26 vote in a Democrat-controlled Senate. He was elevated to chief justice by Republican Ronald Reagan and confirmed by a 65–33 vote in a Republican-controlled Senate. Justice Scalia was nominated by Reagan and confirmed by the same Republican-controlled Senate (on the same day). Justice Thomas was nominated by Republican George H. W. Bush and confirmed by a 52–48 vote in a Democrat-controlled Senate; only 11 Democrats voted in favor. Justice Stevens was nominated by Republican Gerald Ford and Justice Souter by Republican George H. W. Bush. Justices Ginsburg and Breyer were nominated by Democrat Bill Clinton and confirmed by a Democrat-controlled Senate by margins of 96–3 and 87–9, respectively.
- 7 “Sagebrush” states are often described as libertarian insofar as they favor less governmental regulation, especially of businesses. See Eskridge 1995.
- 8 *Jacobellis v. Ohio* (1964).
- 9 I once suggested that one might explain Justice O'Connor's votes at least partly on the basis of gender (Sherry 1986). Justice O'Connor has explicitly rejected my suggestion (O'Connor 2003).
- 10 For a more detailed examination of the use of precedent in this case (as well as those in the following paragraphs), see Sherry 2003.
- 11 In *Gratz v. Bollinger*, Justice Ginsburg's dissenting opinion, joined in relevant part by Justices Souter and Breyer, argued that the equal protection clause permits the government to “distinguish between policies of exclusion and inclusion,” and race should not automatically be a “‘suspect’ category” in the affirmative action context.
- 12 On this type of legal pragmatism, see Farber and Sherry 2002.
- 13 It is also possible that the experts' success rate was depressed by a “wishful-thinking” bias. Drawn from the ranks of those who were expert in the particular field, the experts came into the project with definite conclusions about how cases ought to be decided. In at least some instances, that may have influenced their predictions of how the case was going to be decided. The most glaring example is *Eldred v. Ashcroft*, in which all three experts predicted that the Supreme Court would reverse the lower court

and strike down the Sonny Bono Copyright Term Extension Act; the Court upheld it by a 7–2 vote. (The model correctly predicted the result, although not the exact vote.) I am no copyright expert, and as a constitutional scholar I would have thought reversal highly unlikely. But to those who are experts in the law of intellectual property, the CTEA was deeply flawed and *should* have been struck down. If only a few of the cases included experts whose own expertise convinced them that the Court simply had to agree with them, it would throw off the success rate.

In the interest of full disclosure, I should report that my own success rate was dismal. In the three cases I predicted, I was wrong on the outcome every time

(as were all the other experts in all three cases). The model, on the other hand, correctly predicted the outcome—although not the individual justices—in each of the three.

- 14 It is irrelevant that Spaeth's postdecision coding apparently shows high intercoder agreement. Predictive coding requires making decisions about the subject area of a case *before* the Supreme Court has focused the issues. The Court's opinion has the effect of highlighting some issues and suppressing others, making the coding more predictable. See Epstein and Knight 1998. Thus Spaeth's disagreement with the initial coding both undermines the validity of the predictions and illustrates the difficulties of predecision coding.
- 15 Greenhouse 2003j; Lane 2003.