

# BOSTON UNIVERSITY SCHOOL OF LAW

WORKING PAPER SERIES, PUBLIC LAW & LEGAL THEORY  
WORKING PAPER NO. 06-24



## **THE ROLE OF LAW IN CLOSE CASES: SOME EVIDENCE FROM THE FEDERAL COURTS OF APPEALS**

WARD FARNSWORTH

This paper can be downloaded without charge at:

The Boston University School of Law Working Paper Series Index:  
<http://www.bu.edu/law/faculty/scholarship/workingpapers/2006.html>

The Social Science Research Network Electronic Paper Collection:  
<http://ssrn.com/abstract=926316>

THE ROLE OF LAW IN CLOSE CASES:  
SOME EVIDENCE FROM THE FEDERAL COURTS OF APPEALS

Ward Farnsworth\*

What role does interpretive theory play when a judge decides a close case—or any case? What role is played by traditional legal materials, and what role by the judge’s policy preferences or private beliefs about the world? We can study these questions by considering how judges vote in close cases involving similar policy stakes but different sources of law; for we would expect those cases to produce different voting patterns if they boil down to interpretive disputes, but similar patterns if they depend on judgments of policy. This article presents evidence from one such inquiry. It examines the voting behavior of thirty federal appellate judges in two different types of criminal cases that were not decided unanimously: those involving disputes over constitutional questions and those involving disputes over statutes or rules. In prior work I used the same criteria to examine the behavior of Supreme Court justices,<sup>1</sup> and generally found little differences in their votes in the two types of cases; any justice who usually voted for the government in constitutional cases usually voted for the government in non-constitutional cases as well. This study produces similar results, but with some important underlying differences that arise from the different dockets of the federal supreme and appellate courts.

I. *Procedure and basic result.*

Prior work on the behavior of appellate judges has looked for correlations between judges’ ideologies and the votes they cast. Ideology is measured either by the party of the president who appointed the judge or by more elaborate measures—“common space” ideological scores—that also account for the politics of Senators who had a likely role in the selection process.<sup>2</sup> These studies have found impressive evidence that judges with different political ideologies tend to vote differently on their cases, and in predictable ways, though the evidence is stronger in some areas of law than in others.

---

\* Professor of Law and Nancy Barton Scholar, Boston University. Thanks to Jack Beermann, Andrew Kull, Gary Lawson, Richard Posner, and workshop participants at the Boston University School of Law for helpful comments, and to Nicholas Semanko and Josh Barlow for very helpful research assistance.

<sup>1</sup> Ward Farnsworth, *Signatures of Ideology: The Case of the Supreme Court’s Criminal Docket*, 104 Mich. L. Rev. 67 (2005).

<sup>2</sup> See Richard L. Revesz, *Environmental Regulations, Ideology, and the D.C. Circuit*, 83 Va. L. Rev. 1717 (1997); Frank B. Cross & Emerson H. Tiller, *Judicial Partisanship and Obedience to Legal Doctrine: Whistleblowing on the Federal Courts of Appeals*, 107 Yale L.J. 2155 (1998); Gregory C. Sisk and Michael Heise, *Judges and Ideology: Public and Academic Debates about Statistical Measures*, 99 Nw. L. Rev. 743 (2005).

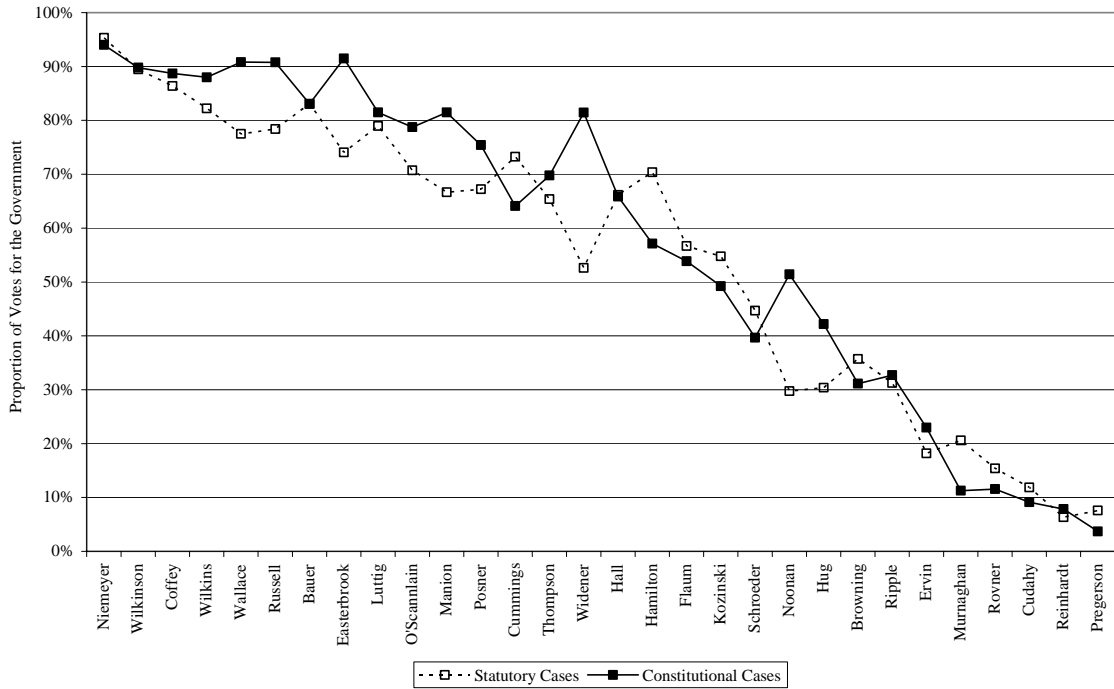
This study takes a different approach. Instead of trying to compare judges' votes with their ideologies, it compares their votes in cases that have similar stakes as a matter of policy but involve different sorts of legal materials. It also focuses on the votes that judges cast in cases that are not decided unanimously. The point is to learn what drives disagreement in close, disputed cases: disagreements about matters of policy, which might be considered ideological matters, or disagreements about matters of interpretation. Of course dissent is an imprecise way to find close cases; I will discuss some implications of this later.

This study examines the behavior of thirty judges of the federal courts of appeals. They all were selected from three circuits to enable cross-comparisons between judges hearing similar sets of cases. The three circuits from which they were taken were the Fourth (mid-Atlantic coastal states), Seventh (Midwest), and Ninth (West coast); these three circuits were selected to create geographical diversity between the judges. Within those circuits, I chose to study, first, the few judges who I considered best-known; then I randomly selected other, lesser-known judges—judges not well-known to me, in any event—who had a lot of service time and thus would generate a large sample size of cases. I believe the reader familiar with the three circuits will find the selections suitably representative, including judges who are prominent and judges who are not; some who are reputed to be ideological and some who are not; and some who are thought liberal, some thought moderate, and some thought conservative to the extent that they have political reputations. I then recorded the votes cast by the judges in every criminal matter that generated a dissent. “Criminal matter” was defined broadly to include direct appeals from convictions, petitions for habeas corpus from the state or federal courts, and civil rights suits brought by criminal defendants.

This method produced a total of just over 3,400 cases—an average of little more than one hundred cases per judge. The cases were then divided into those that raised constitutional issues and those that raised issues under other sources of law: usually statutes and rules. If a case involved more than one clear disagreement, the different disagreements were coded separately, as though they were different cases. Each judge's cases were selected and examined independently by two different coders. Where their codings disagreed, I refereed the dispute. The issues involved were divided almost evenly between the constitutional (52%) and not.

The basic result is shown in the graph below.

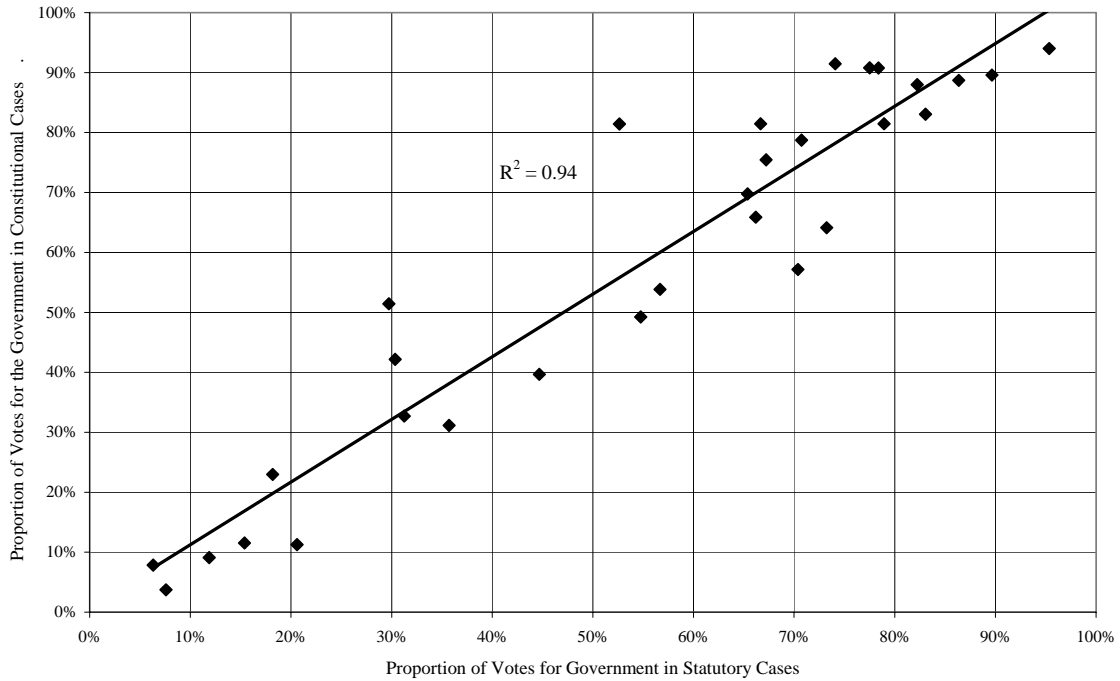
Chart 1: How Often Selected Federal Appellate Judges Have Voted for the Government in Two Types of Non-Unanimous Criminal Cases



As the two lines in the graph show, there are close relationships between the votes of each judge in the two types of case: the constitutional and the non-constitutional (for convenience the latter are referred to as “statutory” in the graph’s legend).

The next graph presents the same information differently; the votes in constitutional cases are arrayed along the left side, and the votes in other criminal matters on the bottom; each dot on the graph thus represents a judge.

Chart 2: Selected Federal Appellate Judges' Votes for the Government in Constitutional Cases by Nonconstitutional Cases (nonunanimous)



The graph shows that a tendency to vote for the government in the one type of case correlates closely with the same tendency in the other type. For reference purposes, I will include below the graphs showing the results when the same questions when asked about Supreme Court justices. The similarities between the results of that study and the results of the study of appellate judges will be obvious.

Chart 3: How often U.S. Supreme Court Justices have voted for the government in nonunanimous criminal cases since 1953.

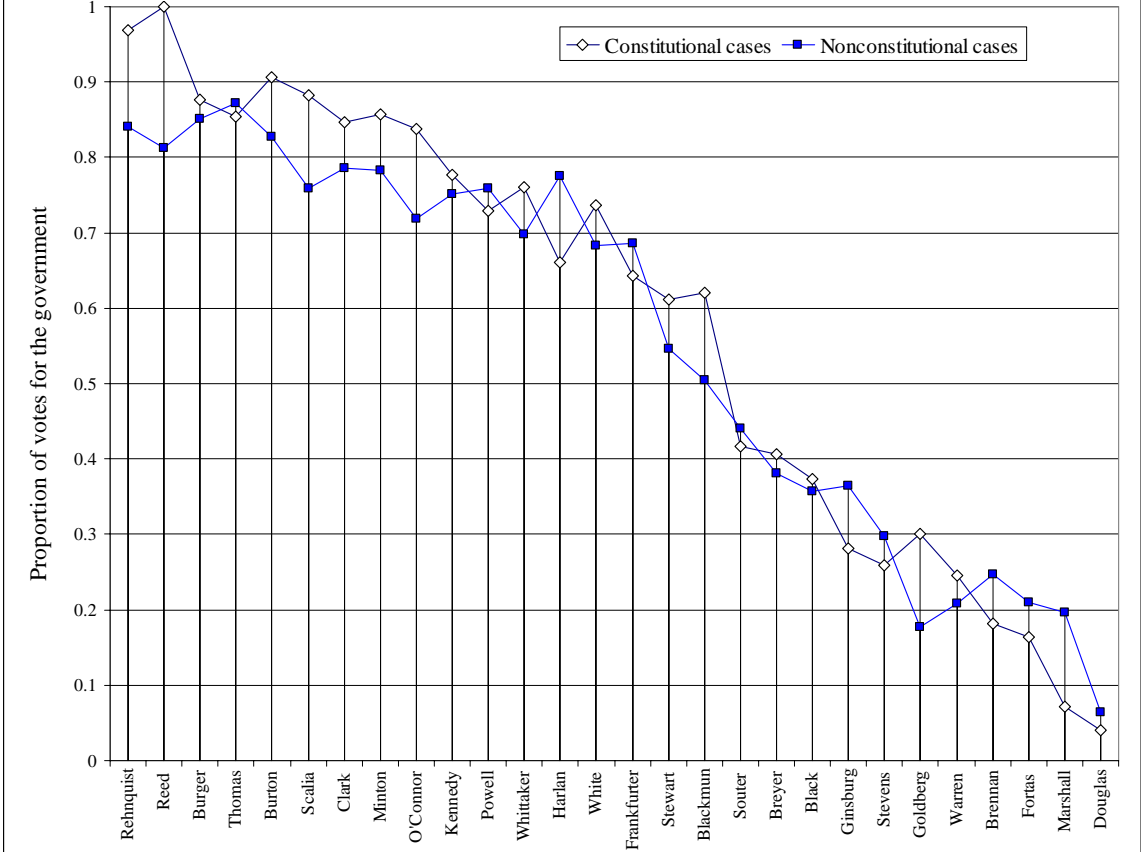
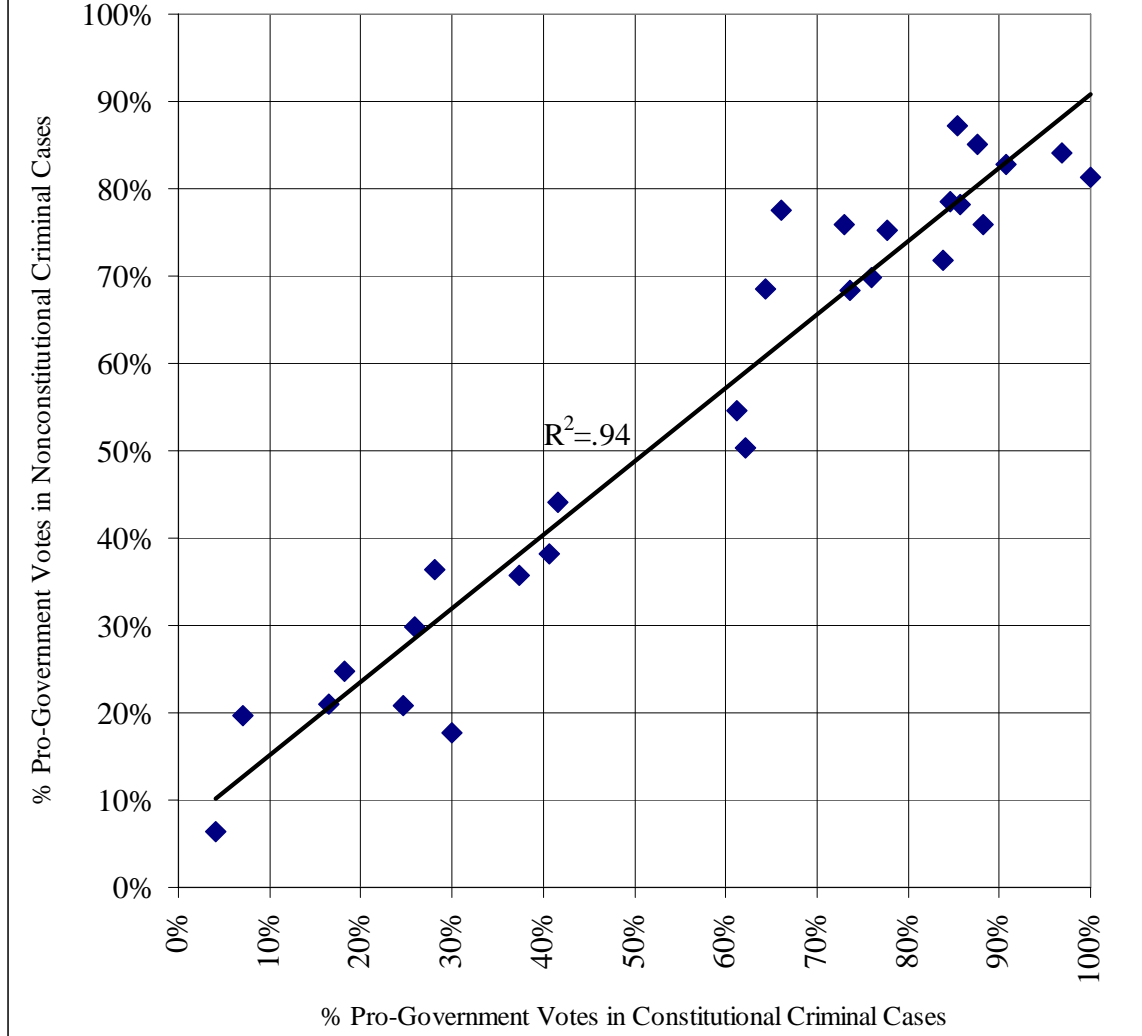


Chart 4: U.S. Supreme Court Justices' Votes for the Government in Constitutional Cases by Nonconstitutional Cases (nonunanimous), 1953-2002



If we compare the strength of the correlations found on the courts of appeals and on the Supreme Court, we find that they are the same: an r-squared of .94 in either situation. I do not wish to make too much of the exact identity between these figures. Remember that while the study of the Supreme Court included all the justices and all of the relevant cases during the study period, the study of appellate courts is more limited; it includes thirty judges, some of whom were picked randomly and some of whom were picked for their notoriety. If every federal appellate judge in the country were tested, it is possible that the strength of the overall correlation would be different; but it seems doubtful that the difference would be great.

## II. *Interpretation.*

In some respects, interpreting these new results is the same as interpreting the similar results produced when the identical questions are asked about Supreme Court justices. There may be occasional exceptions, but the implication over the long run of decisions seems the same for both types of courts: in close cases—cases close enough to provoke dissent—judges appear to seek guidance from the same place regardless of the source of law involved. Exactly what that same place *is* might be debated, but it seems likely to be some feature of the judges' own beliefs or values: how they weigh the interests of accuracy and finality when they compete; or how they weigh the risks of guilty people going free and innocent people being imprisoned; or how they assess various empirical probabilities: that a jury will faithfully follow its instructions, or that if some flaw had been removed from a jury trial, the result would have been different. (All these possibilities are discussed and illustrated in more detail below.) Those issues often arise a bit differently in appellate cases than they do at the Supreme Court, but in broad terms they come up in criminal cases of all sorts and at all levels. There may be theories of interpretation, such as originalism, that could explain why a judge would vote for the government much more or less often than for defendants in constitutional cases. But there is no comparable theory to explain such lopsided results in cases involving statutes and rules—or to explain why the results in constitutional and non-constitutional situations, even if each were based on a coherent interpretive theory, would end up being so similar.

There are a few differences to discuss between the new results here and the results of the similar, earlier study of the Supreme Court. Each study involves only non-unanimous cases—in other words, cases where at least one judge dissented. The significance of this limitation is different at the two levels. At the Supreme Court, where dissent is common, this criterion meant looking at roughly two-thirds of the criminal cases the justices had decided in a given term.<sup>3</sup> In the federal courts of appeals, however, only about 5% of all criminal matters provoke dissent. The number varies a bit; if one picks a very dissent-prone judge and looks only at the cases where he participated, the proportion of *those* cases ending in dissent can rise to a little more than 10%. Even then, though, the overwhelming majority of criminal matters decided by courts of appeals are decided unanimously, no matter what sorts of issues they raise, and thus aren't represented in the graphs above.

---

<sup>3</sup> Farnsworth, *supra* note 1, at n. 9.

This difference need not change the basic implication of the finding—in other words, the implications for how judges decide *close* cases. But it affects any efforts to move from those implications to larger claims about how judges decide cases generally. Simply put, most criminal cases in the federal courts of appeals aren't close. The legal materials speak to them clearly enough to produce agreement by judges with very different policy preferences. The reasons for such frequent agreement—and the reasons why agreement is so much more common in the courts of appeal than at the Supreme Court—are easy to understand. The Supreme Court usually declines to take cases unless they already have provoked judicial disagreement, either in the court below or between different courts that have considered the questions involved. But the courts of appeals have no discretion to duck appeals that obviously lack merit; a defendant has a right to one appeal from an adverse judgment. Nor are criminal defendants and convicts usually constrained by the costs of appealing, since it is common for them to be indigent and thus represented by volunteers or by lawyers in the pay of the government. So the circuit courts hear a lot of appeals based on very weak arguments, and they routinely decide them unanimously—and in favor of the government. The close-case effect thus seems the same in the Supreme Court, in the courts of appeals, and perhaps in most others—but not all courts hear anything like the same number of close cases.

Against this there are a couple of other interpretations to consider. The first is best imagined as a claim by a judge represented in the chart who believes it is misleading; he might offer the following argument: “In every case I adhere to the law without any reference to my own preferences or private beliefs. Unfortunately I often sit with an ideologue who constantly dissents in favor of criminal defendants even when their cases lack merit, and regardless of the source of law involved. Every one of those cases becomes nonunanimous, and so is included in the study; and taken together they make me look like a right-winger because I always vote for the government in them. But I vote for the government in those cases not out of obedience to any values of my own, but because I'm obeying the law while my colleague obeys values of *his* own that always carry him off to the left.” Or of course a comparable argument might be made by someone who frequently votes for defendants in nonunanimous cases, but says this is only because the nonunanimous cases tend to be the ones where his right-wing colleagues abandon the law in a zeal to find for the government.

I do not think this explanation works well as a general matter. At first it might sound plausible; one imagines a moderate judge constantly paired with, say, Stephen Reinhardt, who enjoys a reputation as very liberal—a reputation supported by the data shown here—and whose

decisions are reversed by the Supreme Court more often than most.<sup>4</sup> The moderate judge would come to seem very conservative by comparison. But this mental picture of the situation does not fit the facts. Circuit panels are mixed regularly and systematically, so over the long run no individual judge will be very important to the patterns exhibited by any other individual judge. Consider Judge Reinhardt's impact on the numbers of Judge Clifford Wallace, both of whom are judges on the United States Court of Appeals for the Ninth Circuit. As Fig. 1 shows, Wallace usually votes for the government in criminal matters that are not unanimous. He has been on the court of appeals since 1972. Judge Reinhardt has been his colleague since 1980. The study considered 177 nonunanimous criminal cases in which Wallace participated. How many of those cases were *made* nonunanimous by Reinhardt's vote? Not many. Six of the 177 cases were indeed simple three-judge panels in which Reinhardt dissented, leaving Wallace in the majority. One of the 177 was a three-judge panel in which Wallace dissented, leaving Reinhardt in the majority. Then there were fifteen en banc cases in which Wallace and Reinhardt both participated. Reinhardt dissented in four of those, but was always joined by others. Of the other eleven en banc cases, Reinhardt and Wallace agreed in three; the remaining eight were cases where Reinhardt was in the majority and Wallace dissented—always with others. The net result is that Judge Reinhardt did not play a very large role in generating the nonunanimous cases from which Judge Wallace's numbers were derived.

The dilution of any single judge's influence is greatest in the Ninth Circuit because of its large size, but the basic point holds in the smaller Seventh Circuit as well. Thus one might wonder if, say, the reason Judge Cudahy usually votes for the defendant in split decisions is that he often sits with Judge Coffey, who votes for the government reflexively—or vice versa (maybe it's *Cudahy* who is constantly voting for defendants, and thus creating lots of split decisions; in those split decisions Coffey inevitably looks friendly to the government, but it's only because Cudahy's dissenting largely determined the content of the sample). Either way the probable answer is no. Cudahy and Coffey do sit together on three-judge panels more often than one would expect of any two judges of the Ninth Circuit; turning to the cases that interest us here, Cudahy and Coffey have participated together in 42 split decisions in criminal cases. But those cases still don't go far to explain either judge's location on the chart. The majority of those decisions—25—were matters heard en banc. The usual reasons for taking a case en banc are that the issue involved is one on which the judges of the circuit have substantial disagreement, or on which the federal appellate courts are divided (or on which the court

---

<sup>4</sup> See Eric Schippers, *Much Ado About the Ninth Circuit*, 50 Fed. Law. 20 (2003) (noting that opinions written by Reinhardt had been reversed by the Supreme Court five times in a single term).

threatens to create a division of authority with other appellate courts). So *en banc* attention to a case usually is a strong indication that the case is indeed close—i.e., that there are respectable arguments on both sides; the *en banc* cases don't suggest that either one of the two judges—Coffey or Cudahy—was *creating* the split decisions that then went into the data set by which the other was judged.<sup>5</sup>

If we therefore put the *en banc* cases to one side, we still *could* find that when Coffey and Cudahy sit together on three-judge panels, one of them constantly and eccentrically dissents, again causing all those cases to become part of the set by which the other judge is measured. But this does not appear to be the case. Coffey and Cudahy participated together in the decision of about 200 cases involving criminal matters during the study period, but only 17 of the cases produced split decisions. In six of those, Coffey dissented and Cudahy was in the majority. In nine, Cudahy dissented and Coffey was in the majority. In one case they both were in the majority against another dissenter; in still another they both dissented in part. In about 94% of all criminal matters where Coffey has participated in the decision, the result has been unanimous—and in about 92% of all such matters where Cudahy has voted. The picture that emerges is not one where any particular judge is constantly ruining the unanimity that otherwise would be present.

Yet another explanation might be tried along the same lines, only broader. This time our hypothetical judge reasons as follows: “True, it isn't any one judge who creates split decisions; rather, I'm *surrounded* by ideologues. The *usual* split decision in my court is a case where one of those ideologues departs from the law. Not me, of course; I simply vote every time whichever way the best reading of the law directs.” This interpretation cannot be falsified, at least using the data shown here. Any given judge, from anyplace on the chart, might say this and be right. I don't think the claim is plausible, because it seems to me that most close cases inevitably call on judges to make decisions based on their own prior beliefs about policy or about human behavior. That doesn't make them ideologues; it just makes them judges in a system where the legal materials sometimes leave some open space for decision. But even if the “surrounded by ideologues” explanation were correct, it wouldn't do much to undermine the general theory advanced here. It might save a few judges from the theory's reach, but not most of them.

---

<sup>5</sup> In those *en banc* cases, Cudahy was in dissent 18 times, but almost always with others, and often with three or four others. Judge Coffey dissented in only three of them. In four split *en banc* decisions, Cudahy and Coffey were on the same side.

### III. *Examples.*

Why do judges vote so similarly in cases involving the Constitution and in cases that don't—and yet vote so differently from each other? The answer is that both kinds of cases often turn on the same types of questions, and the questions tend to be matters policy that cut different underlying sources of law. The point can be made clearer by considering some examples. Some were considered in my prior article pursuing the same questions with respect to the Supreme Court; here I will focus on disputes over the sorts of issues that more typically comprise the docket of the federal courts of appeals.

Consider two cases. In the first case, a group of appellate judges agree that a trial judge wrongly used the hearsay rule to exclude a piece of evidence the defendant wanted to use. The divisive question on appeal is whether the exclusion of the evidence was harmless.<sup>6</sup> In the second case, the defendant's lawyer unreasonably fails to pursue a possible alibi witness for his client, who is then found guilty. To obtain relief under the Sixth Amendment the defendant has to show not only that the lawyer made a judgment but that the misjudgment made a difference.<sup>7</sup> The first case just described doesn't involve the Constitution; the second one does. But in both cases the judge's job nevertheless is similar. It is to estimate the likelihood that the outcome of the trial would have been more favorable to the defendant if an error of some sort had not occurred. The standards for making those decisions are a bit different in the two cases, but disagreement over their application still is likely to involve the same variables: either empirical disagreements about the size of the chance that a jury would have made a different decision if it had seen the evidence, or disagreement about how large the likelihood of such a difference must be to warrant reversal. The former is an empirical disagreement. The latter is a disagreement about policy—a trade-off between ensuring a little more accuracy or gaining a little more finality, administrative savings, and so forth. It is hard to know which of these two sorts of disagreements is more important when judges disagree, because the legal tests on which cases like these depend are vague verbal formulations that don't require the judges to articulate their views of the underlying probabilities or their significance with any precision. Either way, however, the root of the matter isn't legal interpretation. The disagreement involves opinions about the world or about policy, which any given judge is likely to make consistently. The importance of those judgments to a case will depend on

---

<sup>6</sup> See, e.g., *United States v. Peak*, 856 F.2d 825 (7<sup>th</sup> Cir. 1988).

<sup>7</sup> See, e.g., *United States ex rel. Kleba v. McGinnis*, 796 F.2d 947 (7<sup>th</sup> Cir. 1986)

how clearly the legal materials speak to the case, but not on whether the materials happen to be statutory or constitutional.

Here is another illustrative pair of cases. In the first, a witness in a criminal case mentions a prior crime the defendant committed. The testimony was improper under the Federal Rules of Evidence, so the judge instructs the jury to disregard it. But is a mistrial in order?<sup>8</sup> In the second case the defendant is on trial for murder and raises an insanity defense. The prosecutor warns the jurors that if they believe the defendant he will “go laughing out the door of this courtroom.” The argument is misleading; the defendant will not go free if the jury accepts his defense. The judge instructs the jury not to treat the prosecutor’s argument as evidence.<sup>9</sup> Again, the first of these cases does not involve the Constitution, and the second one does; but both involve a similar judgment. It partly resembles the one mentioned in the previous paragraph, but this time there is a different wrinkle: how much a judge trusts juries to obey instructions to disregard what they shouldn’t have heard. Again there are two general ways to interpret this difference, one involving empirical views and the other involving view of policy. It might be that judges have different empirical ideas about the likelihood that jurors do what they are told, or it might be that judges agree about that but have different ideas about how likely disobedience by the jury must be to make a new trial worth the bother. As before, we can’t know which is the true source of disagreement because judges aren’t accustomed to explaining themselves at this level of detail. But whatever the deep source of these disagreements, they are not disputes about how to interpret a source of law. They are either disputes about what to expect from the behavior of jurors, or disputes about how much risk of error to tolerate in a criminal trial as a matter of policy.

Consider two last cases. In the first, the defendant is pulled over for a traffic offense and the police have trouble persuading him to get out of his car. It turns out that he is carrying a gun under his coat and is suffering from paranoid delusions. He doesn’t brandish the gun, however, and eventually is coaxed out of the car and arrested. The question is whether, for purposes of the federal sentencing guidelines, he committed a crime involving “a serious threat of violence.”<sup>10</sup> In the second case the defendant is arrested for dealing drugs; in his pocket are keys to an apartment where a woman with a different name lives, and the defendant also is driving a car registered in that woman’s name. The defendant will not explain why has the woman’s keys and car. The police enter the woman’s apartment without a warrant because they are worried that the defendant might have robbed her and done her other harm. That turns out

---

<sup>8</sup> See, e.g., *United States v. Boroni*, 758 F.2d 222 (7<sup>th</sup> Cir. 1985).

<sup>9</sup> See *Aliwoli v. Carter*, 225 F.3d 826 (7<sup>th</sup> Cir. 2000).

<sup>10</sup> See *United States v. Riggs*, 370 F.3d 382 (4<sup>th</sup> Cir. 2004).

not to be the case; the apartment turns out to belong to the defendant's mother. While the police are there, they see drugs, seize them, and use them against the defendant in court. The question is whether, under the Fourth Amendment, the apparent danger the police perceived was serious enough to justify their entry into the apartment.<sup>11</sup>

In this last pair of cases, as in the other pairs, the first case doesn't involve the Constitution and the second one does, but they involve similar underlying issues. In both cases there is a question about the amount of danger involved in a situation and how much liberty should give way as a result. As usual, there are two ways to look at the disagreements about this. The first is empirical; it involves estimates of the amount of danger that exists when a defendant is paranoid and has a gun under his coat, or when a drug dealer is driving the car owned by someone else and has keys to their apartment. The other way to look at the cases involves values. Maybe the judges agree on those likelihoods of danger, but disagree about how serious a danger must be to justify giving the defendant a longer prison sentence or sending the police into an apartment without a warrant. As usual, too, it is hard to sort out which of these accounts is more important because judges do not reveal their exact thinking on these points. But if either of those considerations is indeed driving the result, it doesn't have much to do with law or with a judge's distinctively legal philosophy.

All the cases used in these examples were split decisions that were part of the sample which produced the graphs shown earlier. They all were cases in which judges toward the pro-government side of the graph voted for the government, or where judges on the pro-defendant side of the graph voted for the defendant, or both. Countless other examples could be given. It is normal for disputes about criminal procedure to boil down to empirical judgments about the world or about policy. In some appellate cases, particularly at the Supreme Court, it is possible to view the close cases more than one way—either as debates about interpretation or as debates over policy; to figure out which consideration dominates the other requires a look at the patterns of a judges' decisions, and analysis of how they are more easily explained. But in the courts of appeal the source of dispute usually isn't hard to interpret; it is rare for an appellate case to even ostensibly involve a pure debate over the interpretation of a text, and this for several reasons. First, many criminal cases involve application of a test devised by the Supreme Court as a gloss on some legal rule, whether constitutional or otherwise; in those cases the appellate judges do not interpret a text directly, but try to fit the case in front of them into the categories the Court has created. Whether those categories are satisfied involves judgments about the world or judgments about how generously the categories should be applied (for they often contain vagueness at the

---

<sup>11</sup> *United States v. Brown*, 64 F.3d 1083 (7<sup>th</sup> Cir. 1995).

margins). If application of the test is close because the facts are novel or don't fit the categories in an obvious way, then the decision is more likely to be split—not because the judges disagree about interpretation, but because the decision has to be made on policy grounds over which judges disagree systematically. The examples recited a moment ago illustrate what those policy grounds sometimes are, and why they often are the same regardless of the underlying source of law at stake.

#### IV. Conclusion.

To understand the data presented here, I think, it helps to consider Judge Posner's recent account of how he decides cases:

The way I approach a case as a judge . . . is first to ask myself what would be a reasonable, sensible result, as a lay person would understand it, and then, having answered that question, to ask whether that result is blocked by clear constitutional or statutory text, governing precedent, or any other conventional limitation on judicial discretion.<sup>12</sup>

This is a complete explanation of why Posner's numbers on the graph are fairly close together; for in non-unanimous cases of any kind, constitutional or not, there usually will be no impediment to his wish to vote for the result he thinks sensible (the lack of unanimity suggests that the legal materials plausibly seem consistent with either side's position). And since his views about what result is sensible are matters of policy rather than law, they will tend not to vary between cases where the disputed source of law is different. The question is whether Posner's account also is a good description of what most other judges do, consciously or not. The data here suggest that it probably is.

The resulting conclusion, then, is not that judging is mostly political. Much of the time it is not, as in the usual, unanimous appellate cases where everyone can agree that the expression of whatever policy preferences the judges may have is, as Posner puts it, "blocked" by text, precedent, or other conventional limits. The view I mean to offer is slightly different. It is that when the legal materials bearing on a case are weak enough to comfortably admit of more than one reading, the choice between those readings tends to be made according to beliefs the judges bring to the case that don't owe much to law. The most important traits of most judges may well be their similarities—their ability to agree so often; but the most important *differences* between them tend to be their views about human behavior or about questions of policy.

---

<sup>12</sup> Richard A. Posner, *Tap Dancing* (debate with Philip Heymann), New Republic Online, <https://ssl.tnr.com/p/docsub.mhtml?i=w060130&s=heymanposner013106>.