

Statutory Construction in Federal Appellate Tax Cases: The Effect of Judges' Social Backgrounds and of Other Aspects of Litigation

Daniel M. Schneider*

I. INTRODUCTION

The body of empirical legal research about how judges decide cases continues to expand, adding to the political science literature on the topic.¹ Specifically, much has been written about how judges should interpret the Internal Revenue Code.² However, the empirical research has failed to infuse the literature with observations about how judges *presently* interpret the Code. Indeed, lawyers have provided little empirical work addressing statutory interpretation.³ Yet different judges theoretically could justify the same result in a particular case either summarily, through reliance on precedent, or by deferral to the Internal Revenue Service. In fact, they actually do justify results for similar cases in different fashions. What leads judges down these different paths? Unlike the other literature addressing the Code's interpretation, this Article examines precisely that question.

* Professor of Law, Northern Illinois University College of Law. I would like to thank the participants at the Empirical Tax Workshop at Washington University School of Law for their comments, as well as Donald Songer, Jim Brudney, Fred Markowitz, Howard Erlanger, and Guadalupe Luna for data, advice, and comments. I would also like to thank Susan Boland and Terese Clarke, research librarians at the College of Law.

1. See, e.g., James J. Brudney et al., *Judicial Hostility Toward Labor Unions? Applying the Social Background Model to a Celebrated Concern*, 60 OHIO ST. L.J. 1675 (1999); Gregory C. Sisk et al., *Charting the Influences on the Judicial Mind: An Empirical Study of Judicial Reasoning*, 73 N.Y.U. L. REV. 1377 (1998). Although admittedly arbitrary, I choose to distinguish legal empirical research from more traditional social science research.

2. See *infra* Part II.B.

3. But see, e.g., Daniel M. Schneider, *Empirical Research on Judicial Reasoning: Statutory Interpretation in Federal Tax Cases*, 31 N.M. L. REV. 325 (2001); Brudney et al., *supra* note 1; Sisk et al., *supra* note 1.

In two earlier articles, I examined the effect of judges' social backgrounds on their federal tax trial court decisions.⁴ I analyzed cases decided by the Tax Court and by three federal district courts over a twenty-year period. In one of these studies, I examined the same issue presented here, except at the trial court level, whether background influenced the way in which judges justified their decisions. My earlier work concluded that some aspects of background—most notably education—were in fact closely associated with the judges' methods of justification.⁵

Later, I examined the same set of cases to see whether background factors influenced actual outcomes.⁶ This proved to be a richer ground for investigation because many variables, especially gender, were strongly correlated with the prevailing party. The non-traditional judges (e.g., judges who were not white, were women, or were educated at non-elite institutions) far more frequently and predictably found in the taxpayers' favor.

In my current research, I wanted to test the hypothesis that federal appellate judges' rationale in justifying their decisions is fairly *unaffected* by social backgrounds, both in patterns seen in descriptive statistics and in predictions suggested by regressions. I expected this research to lead to results similar to those obtained in my trial courts research, and my theory was borne out by my findings. Relatively few patterns among the "descriptive statistics," such as gender or race, were statistically significant. Those patterns that *were* statistically significant involved variables more related to "aspects" of the cases, such as taxpayer classification or legal representation, rather than those dealing with judges' personal backgrounds. Similarly, background factors were not predictive of the method of interpretation judges used to justify decisions. "Aspects" variables again were richer and more predictive than social backgrounds, but even these were not strongly suggestive of outcome. Thus, I believe that these results support my hypothesis that social background is a poor indicator of the methods of statutory construction judges use in

4. See Schneider, *supra* note 3; Daniel M. Schneider, *Assessing and Predicting Who Wins Federal Tax Trial Decisions*, 37 WAKE FOREST L. REV. 473 (2002).

5. Schneider, *supra* note 3.

6. Schneider, *supra* note 4.

justifying their decisions. I also believe that the current research reinforces my earlier work regarding statutory construction.

The remainder of this Article is divided into four parts. First, I summarize the existing literature, both on the impact of social background on decision-making, as well as on Internal Revenue Code interpretation. Next, I present my methodology, explaining both how I selected cases and variables for the study, and also my methods of statistical analyses. Third, I set forth the results of my empirical research. Finally, I conclude the Article by restating my belief that the social background of judges who decided the sampled cases neither explains nor predicts how they would have justified their decisions in these tax cases.

II. LITERATURE

A. Empirical Research

Law professors have published an increasing number of articles that use empirical research to examine judicial decisions. These articles ask whether a judge's social background influences her decision, a position that contrasts with the more traditional legal view that judges dispassionately examine the law before them, rendering wholly unbiased decisions.⁷

The legal research, in turn, flows from earlier political science literature looking at judges' decisions.⁸ One school of thought argues that judges' decisions are deeply influenced by their social backgrounds. Law professors appear to have applied this model more readily recently than political scientists. Regardless of the source of the model constructed, this empirical research concludes that background has some impact on judges' decisions.⁹

7. Compare Brudney et al., *supra* note 1 (using empirical research), and Sisk et al., *supra* note 1 (using empirical research), with Harry T. Edwards, *Public Misperceptions Concerning the "Politics" of Judging: Dispelling Some Myths About the D.C. Circuit*, 56 U. COLO. L. REV. 619 (1985) (applying traditional legal reasoning).

8. For a summary of earlier political science literature, see Schneider, *supra* note 4, at 477-83.

9. Compare Brudney et al., *supra* note 1, and Sisk et al., *supra* note 1 (law professors finding limited effect), with C. Neal Tate & Roger Handberg, *Time Binding and Theory Building in Personal Attribute Models of Supreme Court Voting Behavior*, 35 AMER. J. POL.

B. Interpreting the Internal Revenue Code

Statutes are necessarily subject to interpretation—they do not spring forth like the Greek goddess Athena, fully formed (and presumably fully understood) from the head of a parent. Compelling and competing theories have suggested all manners of interpretation. In turn, the tax literature reflects the broader debate about statutory construction; federal tax, after all, is based upon an extensive statutory system—the Internal Revenue Code.

Another article within this symposium volume proposes a profound, broad model for examining the Supreme Court's interpretation of the Code in federal tax cases that come before the Court.¹⁰ I have engaged in a more limited empirical analysis for interpretation of the Internal Revenue Code for several reasons. First, this narrower focus reflects the more traditional methods with which lawyers have interpreted the Code such as taking its words literally, relying upon regulations, or engaging in one of the other methods of interpretation. Literature about Code construction has used these traditional pigeonholes, and by following that literature, we may measure reasonable expectations against actual results. Another benefit of the more limited analysis is that it extends my prior

SCIENCE 460 (1991); S. Sidney Ulmer, *Are Social Background Models Time-Bound?*, 80 AMER. POL. SCIENCE REV. 957 (1986); Jilda M. Aliotta, *Combining Judges' Attributes and Case Characteristics: An Alternative Approach to Explaining Supreme Court Decisionmaking*, 71 JUDICATURE 277 (1988) (political scientists).

The social background model is not the sole political science model explaining judges' actions. Another view is the attitudinal model, the belief that "the justices base their decisions on the merits of the facts of the case juxtaposed against their personal policy preferences." Jeffrey A. Segal & Harold J. Spaeth, *THE SUPREME COURT AND THE ATTITUDINAL MODEL REVISITED* 312 (2d ed. 2002). This theory is drawn from "key concepts" derived "from legal realism, political science, psychology, and economics . . . [and] holds that the Supreme Court decides disputes in light of the facts of the case vis-à-vis the ideological attitudes and values of the justices." *Id.* at 86. A third, competing, view is that judges plan strategically, voting in certain ways in group settings in order to achieve their goals. For a readily available and succinct history of strategic planning, see Tracey E. George, *Developing a Positive Theory of Decisionmaking on U.S. Courts of Appeals*, 58 OHIO ST. L.J. 1635, 1655-65 (1998). My Article pursues the social background model in part because of its established ability to explain judicial decisions, and in part because of my past use of the model. By constructing a baseline using a consistent measurement—the social background model—the literature on federal tax decisions might more easily be advanced.

10. See Lee Epstein et al., *Judging Statutes: Thoughts on Statutory Interpretation and Notes for a Project on the Internal Revenue Code*, 13 WASH. U. J.L. & POL'Y 303 (2003).

research, and thus strengthens the baseline by which federal tax decisions can be assessed through empirical study.¹¹

The possible methods for interpreting the Internal Revenue Code considered and examined in this Article are:

- Strict construction of a statute;
- Deference to the Internal Revenue Service;
- Deference to the Code's structure;
- Deference to a Code section's legislative history;
- Reliance on precedent; and
- Total absence of interpretation revealed due to summary dispositions.

In its simplest form, Code interpretation devolves into an argument about whether it should be interpreted literally, a question which is also posed in the broader debate of statutory construction.¹²

Taking the Code's language literally seems fairly straightforward. This naturally defers to Congress, espousing a simple belief that Congress must have "meant what it said," and that it is not the proper function of a court to substitute its own judgment.¹³ It has been said that a judge should take the words in the statute within the range of "meanings that their text, taken in context, has in ordinary speech or in other provisions of the Code."¹⁴ By contrast, a nonliteral meaning might "provoke us to think, 'If that is what Congress meant, it should

11. Another significant aspect of this Article addresses the fact that, as has been suggested, *see id.*, most research regarding judicial decision-making focuses on the United States Supreme Court. Fewer law review articles have examined how judges decide at the appellate or trial level. *Cf.* Sisk et al., *supra* note 1 (examining trial decisions); Brudney et al., *supra* note 1 (examining appellate decisions).

12. *See, e.g.*, Michael Livingston, *Practical Reason, "Purposivism," and the Interpretation of Tax Statutes*, 51 TAX L. REV. 677 (1996).

13. *See, e.g.*, John F. Coverdale, *Text as Limit: A Plea for a Decent Respect for the Tax Code*, 71 TUL. L. REV. 1501 (1997); *cf.* Mary L. Heen, *Plain Meaning, the Tax Code, and Doctrinal Incoherence*, 48 HASTINGS L.J. 771 (1997) (noting the Supreme Court's increasing reliance on "plain meaning" of statutes when deciding tax cases). *See also* Joseph Isenbergh, *Review: Musings on Form and Substance in Taxation*, 49 U. CHI. L. REV. 859 (1982) (arguing the risks of nonliteral interpretations).

14. Coverdale, *supra* note 13, at 1503-04 (footnote omitted).

have said so rather than saying something very different.”¹⁵

Tax statutes, however, have long been interpreted nonliterally by some scholars. One approach is deference to the implementing agency or, in this case, to the Internal Revenue Service. A highwater mark of deference was etched by the Supreme Court in *Chevron, USA, Inc. v. Natural Resources Defense Council, Inc.*,¹⁶ which held that if Congress had not “directly spoken to the precise question at issue . . . [and if] the agency’s answer . . . [is] based on a permissible construction of the statute,” then the regulation controls.¹⁷ While rarely applied to tax cases,¹⁸ implicit in this environmental law decision is a theoretical deference first to Congress, and then to the agency.¹⁹

Another approach defers to the purpose of a statute, which in turn is part of a larger collection of statutes assembled into a framework. Arguably, there is a “theoretical construct that overarches the sum total of the entire Internal Revenue Code and is intended to be captured by it . . . and that statutory structure could come within the umbrella of statutory purpose.”²⁰ This sense of a section’s purpose or structure is “a joint effort between Congress and the courts . . . [in which] Congress’s law . . . includes that larger statutory structure.”²¹

Yet another approach is to rely upon a statute’s legislative history. As has been suggested, “[t]he argument in favor of using legislative history, simply put, is that context is important to the ascertainment

15. *Id.* at 1504.

16. 467 U.S. 837 (1984).

17. *Id.* at 842-43.

18. The *Chevron* case has garnered little attention in tax cases. See Heen, *supra* note 13, at 784-86; Linda Galler, *Emerging Standards for Judicial Review of IRS Revenue Rulings*, 72 B.U. L. REV. 841 (1992) (discussing the *Chevron* case); Ellen P. Aprill, *Muffled Chevron: Judicial Review of Tax Regulations*, 3 FLA. TAX. REV. 51 (1996) (discussing the *Chevron* case); Samuel B. Sterrett, *Suggested Approach for Judicial Interpretation of Regulations That Grant Discretion to Taxpayers*, 12 VA. TAX REV. 477 (1993) (discussing post-*Chevron* Supreme Court decisions).

19. See Heen, *supra* note 13, at 781.

20. Deborah A. Geier, *Interpreting Tax Legislation: The Role of Purpose*, 2 FLA. TAX REV. 492, 497, 502 (1995). See also Beverly I. Moran & Daniel M. Schneider, *The Elephant and the Four Blind Men: The Burger Court and Its Federal Tax Decisions*, 39 HOW. L. REV. 841, 928-42 (speaking about the “deep structure” of the Code).

21. See Geier, *supra* note 20, at 508 (citation omitted). Geier provides examples of courts respecting the purpose of a statute. *Id.* at 500-01.

of meaning, and legislative history is an important part of context.”²²

While two other approaches lack profound theoretical underpinnings, federal tax cases cannot be read without noting their prevalence among the tools judges use when making decisions. First, many judges rely strictly upon precedent when deciding cases. While there may be no strong, contemporary articulation of this view, it has certain and substantial roots in American legal traditions.²³ Second, courts frequently dispose of cases summarily. Some cases either do not require, or at least do not offer, any substantial guidance for interpreting the Code; sometimes a judge simply applies the law to the facts.

These normative approaches, of course, advance our understanding of the Internal Revenue Code. However, none offer empirical evidence of judges’ methods in tax cases. They neither suggest nor even comprehend, for example, that less elitely educated judges might use strict construction more often than their more elitely educated colleagues, or that we could similarly predict that the older judges would use that method more often than the more recent judges.

III. METHODOLOGY

The results drawn from my case review flow directly from my case selection and sorting methods. Criteria by which cases were selected and sampled are therefore set forth below. Some choices

22. Michael Livingston, *Congress, the Courts, and the Code: Legislative History and the Interpretation of Tax Statutes*, 69 TEX. L. REV. 819, 845 (1991) (footnote omitted).

23. See SPAETH & SEGAL, *supra* note 9, at 5-15 (discussing contemporary proponents of the precedent approach, such as Ronald Dworkin). Somewhat associated with the reliance on precedent is practical reasoning: the idea that, when interpreting the Code, a judge must consider statutory text, legislative history, and evolutive considerations—including judicial and administrative precedents and applicable current values—together with the consequences of alternate interpretations and the court’s own policy sense. While these sources of meaning typically would be considered in descending order of priority, the precise mix of sources depends upon the nature of the provision in question and the facts of the case at hand.

A practical reason approach provides both the correct method of deciding tax cases and an accurate description of what courts actually do in such cases. See Livingston, *supra* note 12, at 720. Practical reasoning is not the same as reliance on precedent because it uses other methods of interpretation and, more importantly, because it suggests a particular way to apply the methods of interpretation—strict construction, legislative history, and “evolutive” methods.

varied from those I made in previous studies, including my use of a shorter and different time period. Other choices did not vary, because I wanted to replicate earlier benchmarks as much as possible, in order to permit earlier results to be more easily contrasted with the current ones.

A. Data and Variables

1. Case Selection

The database includes all federal tax decisions rendered by all federal circuits. I wanted to sort the cases by circuit and by the appearance of tax matters, making traditional reporters like the Federal Reporter, and the specialty reporters, American Federal Tax Reports and United States Tax Cases, inadequate for the task. Instead, I engaged in online research,²⁴ sampling about ten percent of all listed circuit decisions regarding federal tax during a recent five-year period.

a. Courts selected

At the trial level, federal tax cases are decided by the Tax Court, district courts, and the Court of Federal Claims.²⁵ Appeals from decisions rendered by the former two courts appear before the appropriate United States Circuit Court of Appeals. If an action is filed in the Court of Federal Claims, however, then appeals are heard by the Federal Circuit.²⁶

One shortcoming of my earlier research on trial decisions was that I reviewed only selected district courts. Such an approach was necessitated by the large number of district courts (91), as well as my

24. Using the online WESTLAW database, I executed the following search:

PR([the circuit to be examined, such as "SIXTH"]) & CIRCUIT & DATE([a year from 1996 to 2000, inclusive])

25. Taxpayers who seek a refund file their claims either in the Court of Federal Claims or else in the appropriate district court. Taxpayers who do *not* seek a refund litigate in the Tax Court. See I.R.C. §§ 7441-42, 6213(a) (1999) (defining Tax Court jurisdiction); 28 U.S.C. § 1345(a)(1) (2000) (defining district court and Court of Federal Claims court jurisdiction).

26. See 28 U.S.C. § 1295.

desire to reach more discrete conclusions about particular district courts, rather than broad conclusions about all district courts. There are fewer federal appellate courts, making them easier than the trial courts to approach as a group. An anticipated and intended benefit of the current research was that results could be drawn about the breadth of federal appellate tax decisions included in the database. Thus, decisions by the District of Columbia Circuit and the First through Eleventh Circuits were examined.²⁷

b. The period reviewed

I reviewed the most recent five-year period for which cases were available when I began my research, from 1996 through 2000. Earlier research suggested the importance of gender and race in my study of the effect of social background factors on the prevailing party.²⁸ Because passage of time has only favored diversity on the bench,²⁹ a more recent period was more appropriate for this research.³⁰ Finally, a multi-year spectrum (five years) minimizes the risk of a temporary aberration than an even shorter, one-year base might present.

c. Published v. unpublished cases

Authors have varied in their approach as to whether to review *all* cases or only those that have been published.³¹ The wider universe of

27. Although I also collected data about the Federal Circuit, there was marginally less biographical information for the judges on this circuit. Furthermore, the Federal Circuit is a more specialized court, hearing several specialized appeals, including those of tax litigants from the Court of Federal Claims. In contrast, the circuits included in the database are more general, hearing appeals from any district court or Tax Court decision (assuming that jurisdictional prerequisites are met). *See supra* notes 25-26.

28. *See Schneider, supra* note 4.

29. *See* SHELDON GOLDMAN, PICKING FEDERAL JUDGES 346-65 (1997).

30. While a broader, twenty-year period was appropriate in the earlier articles, the passage of time turned out to be less critical to my analysis, whereas gender and race were more so. *See Schneider, supra* note 4, at 488 n.54, 506 nn.142, 515. Thus, I used a shorter period in order to obtain more decisions for which the judges had not been white or had been women. I was willing to risk the potential loss of trends defined by time, given the apparent unimportance of the passage of time in my earlier study.

31. For examples in the literature, compare Brudney et al., *supra* note 1 (reviewing all appellate decisions in database) with Orley Ashenfelter et al., *Politics and the Judiciary: The Influence of Judicial Background on Case Outcomes*, 24 J. LEGAL STUD. 257, 263-65 (1995)

cases, including those that were not published but were nevertheless set forth in slip opinions or online databases, are now easier to obtain than they once were. Reviewing all cases minimizes the risk of bias. On the other hand, reviewing only published cases might be justified if there is no actual bias or if limitations are imposed by resources. Obviously, reviewing only some cases is accomplished more easily than reviewing all cases.

This Article uses cases in the WESTLAW database, which includes both published and unpublished cases. Cases were excluded if they were so vague or brief that I could not characterize the approach used to justify the decision. These vague or sparse decisions, it turned out, tended to be the unpublished decisions.³²

d. Sampling

I read every tenth case. Because some of the cases in the database had to be excluded,³³ I actually read more than ten percent of the federal tax appellate decisions captured by the WESTLAW search.³⁴

(reviewing only relevant, filed decisions in database). *See generally* Deborah Jones Merritt & James J. Brudney, *Stalking Secret Law: What Predicts Publication in the United States Courts of Appeals*, 54 VAND. L. REV. 71, 72-73 (2001) (examining the effects of the selective publication of appellate decisions).

Case law also exists on this point. *Compare* Hart v. Massanari, 266 F.2d 1155, 1180 (9th Cir. 2001) (upholding Ninth Circuit rule that unpublished opinions may be cited only for factual purpose), *with* Anastasoff v. United States, 223 F.3d 898, 905 (8th Cir. 2000) (holding a rule prohibiting the use of unpublished opinions to be unconstitutional), *vacated for mootness*, Anastasoff v. United States, 235 F.3d 1054 (8th Cir. 2000). *See generally* Johanna S. Schiavoni, Comment, *Who's Afraid of Precedent?: The Debate over the Precedential Value of Unpublished Opinions*, 49 UCLA L. REV. 1859 (2002) (arguing in favor of greater reliance on unpublished opinions).

32. With very few exceptions, only published cases provided usable data. Most of the cases that appeared in the WESTLAW database (excluding AFTR2d and USTC cases) were unuseable. WESTLAW citations not appearing in either of these reporters usually lacked information, or else the area of the decision had nothing to do with federal tax, but had somehow been included among the cases captured by the WESTLAW search.

33. For example, in one search of Tenth Circuit cases, the database for some reason returned a Third Circuit decision. In another, the database returned a non-tax decision addressing drug possession. These cases, along with similar anomalies, were excluded from the present study.

34. After randomly selecting one of the first ten cases to come up in the search, I read every tenth case. This process was repeated for each circuit and each year. While it is unlikely that the first case decided by a circuit in a given year is somehow different than the second, third, or tenth, random selection of the first case for review minimizes the risk that differences

By following statistical protocols, I was able to minimize the inaccuracies that can result from sampling and not including all cases in a study.

Of the cases sampled, I then selected only those cases in which there was one or more signed opinion.³⁵ My rationale was that the judge who authored an opinion—whether in the majority, concurrence, or dissent—was the judge responsible for justifying the decision. This view was consistent with my prior research.³⁶ While judges on a panel might negotiate when deciding appellate cases, they are more likely to haggle over who should win the case than over the exact means of justification for the decision.³⁷ While others might take a different approach, such as assigning equal weight to all judges who consented to the opinion, I chose not to follow this course.³⁸

Per curiam cases were excluded from the sample because no one individual's background could be associated with those opinions. On the other hand, concurring and dissenting opinions *were* included, carrying the same weight as majority opinions; for each such opinion, an identifiable judge, with an assessable background, had authored that opinion.

will occur, avoiding bias in the results.

A case listed in the WESTLAW search might be excluded for a number of reasons: the case, as a practical matter, was entirely unrelated to tax; it was from a circuit other than the one under scrutiny; it could not be located; or it lacked enough information to be useful (i.e., the opinion was so summary that the subject matter was indiscernable).

To check the accumulated data, I compared a hard-copy source (AFTR2d more often than F.3d) with the same case in the WESTLAW database. The data was usually the same (e.g., the same judges, the same votes); the most notable inconsistency was the case's precedential value, as characterized by the deciding court. When only one source was available, I used that source; when the data differed between the two different sources, I used the data from the LEXIS database because it appeared to be more accurate.

35. I am indebted to the members of the Empirical Tax Workshop for this suggestion. There are other sampling options, most notably the assignment of equal weight to each vote on a judicial panel. This approach, however, accords the same weight to a judge who *merely voted* for the opinion as to the *authoring* judge.

36. See Schneider, *supra* note 4; Schneider, *supra* note 3.

37. *Id.* Note that highly dissatisfied judges always have the option to issue concurring or dissenting opinions, which were also reviewed for this study.

38. *Cf.* Brudney et al., *supra* note 1 (assigning equal weight to all three members of panel, but in assessing social background factors' effect on labor law cases).

2. Variables

I collected data about the cases and about the judges who decided the cases. The variables include:

*Data About the Judge*³⁹

a. Name.

b. Gender.⁴⁰

c. Race.

d. Educational background.

i) Did the judge attend an elite *college*?⁴¹

ii) Did the judge attend an elite *law school*?⁴²

39. The source of biographical information upon which I relied was the Federal Judicial Center, the website for which is found at <http://www.fjc.gov>. To minimize error, I checked my entries for accuracy of information—including educational institutions attended, gender, appointing president, and race—at least twice against information obtained from the website. I also re-checked the entries as I coded them for statistical analysis, *see infra* notes 40-41, at least one additional time.

40. The coding of data is important for multinomial logistic regression, which is the method used in this study to establish relationships between winning parties and social background variables. Coding is therefore set forth where appropriate (e.g., coding for gender, with female=0 and male=1).

41. Selectivity of college is revealed by the numerous guides to American colleges and universities. For this study, I used an older reference. ALEXANDER W. ASTIN, *WHO GOES WHERE TO COLLEGE?* 58-83 (1965). Astin establishes the relative selectivity of an undergraduate institution by dividing the “highly able students who [wanted] to enroll at the college . . . by the number of freshmen admitted.” *Id.* at 55. Compare Brudney et al., *supra* note 1, at 1703 (using Astin’s measurements), with Sisk et al., *supra* note 1 (not using undergraduate degree as a variable). *See also* GOLDMAN, *supra* note 29, at 346-65 (distinguishing between public and private institutions, as well as between the Ivy League and other schools).

Astin seemed to provide the most appropriate guide for several reasons. First, many of the judges were in college closer to the time when he was measuring selectivity of institutions (his treatise was published in 1965). The mean year of birth for the judges in the sampled cases was 1935. Second, perceptions of institutional eliteness change slowly. For example, Harvard has remained a prestigious institution over a long period of time. *See* Brudney et al., *supra* note 1, at 1704 n.105. Finally, I had used the Astin scale in my earlier articles, *see* Schneider, *supra* note 4; Schneider, *supra* note 3, and thus hoped to increase my ability to contrast the judgments made in each by remaining consistent.

42. The 1977 *Cartter Report* includes fifteen law schools in its “elite” category—Harvard, Yale, Stanford, Michigan, Chicago, Berkeley, Columbia, Pennsylvania, Virginia, UCLA, Cornell, Texas, Duke, Northwestern, and NYU. *See The Cartter Report on the Leading Schools of Education, Law, and Business*, CHANGE, Feb. 1977, at 44, 46. Another researcher substituted

e. Primary professional experience before appointment to the appellate bench.

i) Had his career been primarily in private practice, government (legal or nonlegal), other judicial appointment, law school teaching, or even some other, nonlegal capacity?⁴³

f. Political ideologies, as measured on a scale calculating the economic liberalism of the judge's appointing President.⁴⁴

three other schools of her choice in defining the fifteen elite law schools. See Diana Fossum, *Law Professor: A Profile of the Teaching Branch of the Profession*, 1980 AM. B. FOUND. RES. J. 501, 507 (1980) (substituting Georgetown, Iowa, and Wisconsin for Cornell, Duke, and UCLA). See also Sisk et al., *supra* note 1, at 1418-19 (providing the authors' own list of seven elite schools). For purposes of this Article, I treat all eighteen of the *Cartter Report's* and Fossum's schools as elite schools and all of the rest as non-elite. Again, this is consistent with my earlier articles, increasing my ability to contrast them. The judges' legal educations were coded for the study (nonelite=0 and elite=1).

43. I determined this variable by quantitatively examining years of service, choosing the profession with the longest tenure. This avoided a "last in time prevails" rule, which would be problematic. Using this measurement, most judges' primary professions were fairly clear; they rarely skipped from government work to private practice and back. Although I made separate entries for private and corporate practices, I collapsed the two into one category due to the small number of judges who had worked primarily for corporations. In both situations, the individual represented private interests, and so the combination was both apt and theoretically and methodologically justified.

In the event that a judge had devoted an equal amount of time to two separate lines of work, I treated the judge as having engaged primarily in the more recent of the two. If a judge appeared to engage in two overlapping jobs—most often, private practitioner and legislator, or private practitioner and teacher—I chose what appeared to be the more important career as the primary professional experience. Inevitably, I chose private practice.

In other instances, judges had been either public defenders or general counsel to nonprofit organizations. I treated those judges, respectively, as having been engaged in private practice and in corporate practice. This is because the predominant character of those jobs was the representation of individual clients or of corporate entities. There are other ways in which to characterize careers.

Other researchers have drawn more discrete distinctions when examining prior professional experience. See, e.g., Tracey E. George, *Court Fixing*, 43 ARIZ. L. REV. 1, 9 (2001); Brudney et al., *supra* note 1; James Edward Maule, *Instant Replay, Weak Teams, and Disputed Calls: An Empirical Study of Alleged Tax Court Judge Bias*, 66 TENN. L. REV. 351 (1999); Sisk et al., *supra* note 1. In part, I was again motivated by a desire to use the same measurements as I had used in my earlier articles.

44. In my earlier research, I used a standard proxy for the political views of the judge himself: the political affiliation of the President who appointed the judge. See, e.g., GOLDMAN, *supra* note 29, at 355-57 (suggesting that the Presidents from Eisenhower to Reagan appointed federal appellate judges from their own political parties over ninety percent of the time). President Carter appointed relatively fewer appellate judges from his own party, yet even his percentage was more than eighty-two percent. *Id.* at 355. See also ROBERT A. CARP & RONALD STIDHAM, *JUDICIAL PROCESS IN AMERICA* 237, 241 (4th ed. 1998) (reporting ninety percent of

g. Data about the judge's *judicial career*, including length of tenure on the bench at the time the opinion was rendered.⁴⁵

h. Religion.⁴⁶

j. Year of birth.

h. The *nature* of the judge's opinion (i.e., majority, concurrence, or dissent).

Data About the Decision

a. The *court* in which the case was decided.

b. The *year* of the decision.

c. The principal *Internal Revenue Code section* at issue in the case.

federal judges are of the same political party as his appointing President). To provide consistency, this was measured by the President who had appointed the judge to the bench from which he rendered the decision. Thus, if President Reagan had appointed Judge Smith to the district court, and President Clinton later appointed him to the court of appeals, Reagan was still treated as the appointing President for those opinions rendered while Smith was a district court judge. However, President Clinton was treated as the appointing president for Judge Smith's opinions rendered while he was an appellate judge.

Scholars have suggested a variety of ways to measure judges' and appointing Presidents' political views. See Robert A. Howard & David C. Nixon, *Local Control of the Bureaucracy: Federal Appeals Courts, Ideology and the Internal Revenue Service*, 13 WASH. U. J.L. & POL'Y 233 (2003). I have used an "economic liberalism measure." See generally Jeffrey A. Segal et al., *Buyer Beware? Presidential Success through Supreme Court Appointments*, 53 POL. RESEARCH Q. 557 (2000) (defining the economic liberalism approach). See also *infra* tbl. 1. These authors rank the Presidents' economic liberalism for all Presidents from Franklin Roosevelt through Bill Clinton. The scores ranged from zero (most conservative) to one hundred (most liberal). With a score of 17.6, Ronald Reagan was the most conservative, while Franklin Roosevelt, with a score of 82.5, was the most liberal. *Id.*

I used this scale for two reasons. First, it was readily understandable. Second, many of the alternative measures focused heavily on judges' views about civil liberties and similar topics, most of which are far afield from tax. While economic liberalism was not a clear match with tax law, I nevertheless felt that it came closer to accurately measuring judges' views about tax, via appointing presidents, than did the other standards.

45. In order to provide consistency, length of tenure between appointment and rendering of the decision was always measured from the time the judge assumed the particular bench from which he rendered the decision under scrutiny.

46. This information came from a database about federal appellate judges prepared by Professors Donald Songer and Gerry Gryski. Professor Songer graciously provided the data, which is available with the author of this Article. Data on religion was unavailable for judges not sitting on a Court of Appeals, judges whose religion was unknown, and appellate judges appointed after 1994.

d. The court's *action* on the case (e.g., affirmance or dismissal).

i) Courts were almost invariably clear about their actions. In those few cases in which the court was silent about its action, I engaged in an appropriate characterization (e.g., treating the court's summary rejection of all of the appellant's claims as an affirmance).

e. The *prevailing party*.⁴⁷

i) I determined who won a case by examining multiple factors. At the very least, I had no expectation about where my analysis would lead regarding this variable, which somewhat diminished bias.

f. Presence of *legal representation*.⁴⁸

g. The taxpayer's *legal status* (e.g., individual, business).⁴⁹

h. The *number of issues* decided by the court.⁵⁰

47. It was important to dichotomize the prevailing party into two groups in order to promote statistical analyses. While the most obvious division in tax is between the taxpayer and the government, such a split does not always literally describe the case. Sometimes an Internal Revenue Service agent was the named party in an action against a taxpayer. I treated that named party as if he were the government for purposes of coding, because the agent could clearly be aligned against someone aiming to protect her rights as a taxpayer.

As I read cases, I deemed decisions stated by the courts to be in favor of the government or against the taxpayer at face value (and vice versa), when coding, although I still read the cases in order to confirm independently the courts' conclusions. Even if the winner did not have all of its expectations met, for example if the taxpayer paid only some portion of the tax it argued it had never owed, that party was still treated as the winner.

48. One question often asked about my earlier research was whether the taxpayer's form of representation affected the measured outcome. While I did not think representation would matter, I was unable to prove the point empirically. I therefore included this variable in my current research.

When the taxpayer represented herself, but with legal counsel also listed, I treated that taxpayer as being represented by a lawyer. In such cases, the taxpayer had the benefit of counsel and should thus be treated as having representation, despite her choice to participate actively in her own action.

49. Another question raised regarding my earlier research was whether the taxpayer's status (i.e., corporation, individual, etc.) affected the outcome. For example, judges might have particular biases against corporations. While I did not think that this variable would correlate with judicial outcomes, I nevertheless included this variable in the current database.

I made judgments about taxpayer status based on information that usually was readily apparent: inclusion of "Inc." or "LLP" in the taxpayer's name and inclusion (or lack thereof) of any legal characterization of the taxpayer's name, description, or status within the opinion.

The benefit of coding for multiple issues is that it theoretically could strengthen the results of the current research. For example, if female judges were more likely to use precedent as a method for rationalizing their majority opinions, but not their concurring or dissenting opinions, then analyzing up to three methods for justifying the majority, or concurring or dissenting opinions, if there were any, could triple the chance of letting those patterns emerge.

The number of issues in a case were often signaled by the court—language such as “the sole issue presented in this case” or “the two questions to be answered here,” for example, revealed that one or two issues, respectively, were at play.⁵¹ In multiple-issue cases, I coded the first three issues, and also determined the winner on each of these issues, again using criteria already noted, such as the court’s stated action and my independent confirmation of the court’s conclusion.⁵² I tended to code the issues presented by a court in the same sequence in which they were presented by the court.⁵³

Few cases had more than three issues. In those few that did, my decision not to catalogue all the issues was not fatal, because I was interested in determining whether multiple issues might account for

50. See Schneider, *supra* note 3, at 341 n.91. One change in coding from my earlier research was the addition of a variable for whether one or more issues were decided by the court. I had been surprised that trial courts had mixed methods when interpreting the Internal Revenue Code in order to sustain their decisions. For example, they combined strict construction of a Code section with practical reasoning. Previously, I could not address whether they had done so simply because two issues were at play, as I had not coded for that possibility. While I believe that most cases in my earlier databases had been one-issue cases, I could not confirm the accuracy of that belief.

51. Courts frequently set forth preliminary matters—including jurisdiction, background material, the general legislative scheme, or the lower court’s decision—before discussing resolution of an issue. They usually used signals, such as “Discussion” or “Analysis,” to set out these preliminary matters, which I did not treat as “issues” for purposes of this Article.

52. One way to limit unconscious bias in the coding of cases would be through the use of multiple research assistants who lack knowledge of the expected outcome. Because I coded the cases by myself, however, I could not use that safeguard. I recoded the most subjective material—the number of issues, grounds for justifying the decision, and the prevailing party. I believe that I further minimized bias by recoding the subjective material without knowledge of the objective social background material (in some cases, coding the social background material only after recoding the subjective material). This prevented me from recoding the prevailing party, for example, in such a way as to promote a particular hypothesis. For more on bias, see *infra* note 59.

53. If I did not code in this manner, it was because I had coded the most important or most dispositive issue first, with a less important or more tangential issue coming later.

use of different rationales for justifying a decision. This could be accomplished by simply determining whether multiple issues existed instead of the number of issues.

Once I coded for multiple issues in majority opinions and had decided to sample the concurring and dissenting opinions authored by a specific judge, I also decided to code these concurrences and dissents for multiple issues. I did so along the lines set forth above.

i. The *rationale used by the court* to justify its decision.

There were two aspects to the coding of rationale. First, what were the possible rationales or justifications? The methods listed above,⁵⁴ literal v. nonliteral, in turn became strict construction, and deference to the Internal Revenue Service through reliance on its regulations, rulings, etc., the Code's structure, a statute's legislative history, precedent, and mere summary disposition. The second question aimed to determine the rationales actually used by the court.

Judgments I made about the rationales used by a court depended, of course, upon the way in which the court drafted its opinion. As noted above, different judges might approach the same question in different ways.⁵⁵ For example, one court might discuss a regulation, revenue ruling, or procedure, which clearly suggests its deference to the IRS. Another court might turn to precedent to bolster its reliance on a regulation and, if it discussed other cases, these opinions were treated as relying on this deference and on precedent. A third decision in which cases were discussed, but *not* the regulation (or if the regulation was touched on, then only very briefly), was treated as a case for which the rationale was precedent only.

I listed up to three rationales for the first issue in the majority opinion, and only one rationale for each of the second and third issues. Again, because I wanted to ascertain whether the existence of multiple issues could account for use of different rationales, establishing multiple rationales for relatively less important issues was not critical and therefore not pursued. Because the first of three possible issues had been coded for the presentation of multiple rationales, I again increased the possibility that patterns might appear

54. See *supra* text following note 10.

55. See *supra* text following note 3.

in the results.⁵⁶ In other words, that Protestant judges could have been predicted to use structure to justify their decisions could be tested against three possible rationales for the first issue, rather than just one.

As with the number of issues, I also tried to follow a court's lead when coding rationales. If an opinion listed two grounds for deciding an issue, each of which it discussed, I was more likely than not to find two rationales for that issue.⁵⁷ As noted, I did not treat preliminary matters as issues, nor did I include the rationale used to justify that preliminary matter.⁵⁸

Concurrences and dissents were coded for rationales along the lines set forth above for coding rationales in majority opinion—one issue and up to three rationales.

With the methodology of data gathering set forth, statistical analysis may be briefly noted before turning to a discussion of my results.⁵⁹

B. Statistical Analysis

The explanatory (independent) variables in the database include the judge's gender, race, eliteness of college and law school, primary pre-judicial professional experience, the appointing President's economic liberalism, years on the bench when the sampled case was

56. See *supra* text accompanying notes 51-52.

57. See *supra* text accompanying note 52.

58. Just as courts often signal the analysis of issue with a signal, such as "Discussion," other textual signals were also used. For example, the court might say, "In the following two portions of this opinion, the court will discuss section *x*'s legislative history and the applicability of Revenue Ruling 98-xxx." I considered this kind of formal decision as one factor in determining the number of distinct issues covered by an opinion.

59. Empirical research is always at risk for bias. I took steps to minimize my biases, whether explicit or implicit. First, I did not code the majority of the independent variables, the social background factors surrounding the judges themselves, until after I had coded other data about the cases. Second, much of the data about the cases was "objective" and unalterable (e.g., the circuit in which the case was decided or the deciding judge). I also ensured accurate data entry, discovering only one mistake in each of these two objective variables when recoding ten percent of all of the cases for accuracy confirmation. Third, I recoded the "subjective" variables about the cases—the number of issues presented, the prevailing party on each issue, and the stated judicial rationale—without knowledge of the deciding judges' identities.

On this last point, I would like to make clear that the subjective matters, by their very nature, might be viewed differently by others, which is one of the reasons for my recoding.

decided, religious preference, whether the case was deemed by the court to have precedential value, whether the taxpayer was represented by a lawyer, and the taxpayer's legal status. The outcome (dependent) variable was the method the court used to interpret the Internal Revenue Code in justifying its decision.

Different statistical analyses sometimes dictated characterizations of outcomes, and so sometimes the outcome was expressed by noting all six methods of interpretation, especially in the regressions, and sometimes it was expressed as an either/or equation.⁶⁰ The phrasing of the outcome is obviously important for understanding my results, and so I have always clearly noted the outcome below.

Statistical significance for relationships was created through comparisons, such as between decisions by male and female judges, or between judges who had gone to more (or less) elite colleges, through appropriate tests.⁶¹ Multinomial logistic regression was used to establish relationships between the independent and the dependent variables.⁶²

As with the outcome, different statistical analyses sometimes dictated the use of different databases. The original database consisted of 419 decisions. Each of these decisions was parsed into its separate majority opinion and, because of concurrences and dissents, 25 opinions were added to the original database to create another database of 444 opinions. And, because I sometimes examined only authored decisions, I modified the 444 opinion database by excluding the 201 *per curiam* opinions, leaving a total of 243 opinions in a second modified database. Again, I have identified the particular database used in order to aid in understanding my

60. For example, either reliance on regulations or all other methods of interpretation; either reliance on precedent or all other methods of interpretation.

61. I used the chi-squared test for cross-tabulations of dichotomous variables (e.g., gender, political affiliation), means comparison to establish significance for continuous data (e.g., length of tenure when deciding the case), and the Wald test for logistic regression. Statistical significance is "designed to allow us to make statements about the probability that hypothetical relationships actually occur"; it permits the inference that two variables are interrelated, in both the sample and the general population, such that the correlation is not merely the result of random association. GEORGE W. BOHRNSTEDT & DAVID KNOKE, STATISTICS FOR SOCIAL DATA ANALYSIS 22-23, 158 (3d ed. 1994).

62. Logistic regression permits the isolation and measurement of each independent variable's effect, while controlling for the influence of other independent variables. See Schneider, *supra* note 3, at 490 n.61.

results. Generally, however, one of the two broader databases tended to be used more in the descriptive statistics while the third database was used exclusively for logistic regressions.⁶³

IV. RESULTS

A. Descriptive Statistics—Who Were the Judges and How Did They Justify Their Decisions?

The descriptive statistics drawn from the sampled cases tell several stories. Overall, some general observations can be made. We know something about who the judges were—white, male, and well-educated. They relied heavily on precedent, whether writing majority, concurring, or dissenting opinions. Beyond that, they engaged in few statistically significant patterns, and so we cannot negate the random nature of the patterns revealed by these statistics. In other words, for example, while Latino judges or judges appointed by more economically liberal Presidents might have used a certain method of interpretation to justify their decisions, we cannot negate the possibility that such patterns were random. Further, because there are few statistically significant patterns, I believe that methods of justifying tax decisions were not very important to the judges who decided the sampled cases.

Still, several observations may be made about the 419 cases in the original database and the judges in these cases.⁶⁴ The smallest

63. The statistical analysis does not use cluster sampling, a method of making observations about the larger population of cases based on the smaller units in the population presented by the circuits. Traditional reasons for using sample clusters, such as the difficulty or expense of sampling the larger population, were not at issue here; the risk of decreased precision rendered the model's use unadvisable. See SHARON L. LOHR, SAMPLING: DESIGN AND ANALYSIS 23-25 (1999) (discussing cluster sampling).

The use of a hierarchical model was also inappropriate. Hierarchical models use layered analyses, such as examining students who nest within a classroom, classrooms that nest within a school, and a school that nests within a community. See generally STEPHEN W. RAUDENBUSH & ANTHONY S. BRYK, HIERARCHICAL LINEAR MODELS (2002) (discussing this model). Other legal studies of judges' social backgrounds have similarly avoided hierarchical models. See, e.g., Sisk et al., *supra* note 1; Brudney et al., *supra* note 1.

64. As can be seen, I sometimes refer to the 419-case database and sometimes to the larger, 444-case database. While this may be confusing, I have done this because different inferences may be drawn from each database. The 444-case database includes all judges who authored opinions; it would be inappropriate to discuss the judges *without* using that base. See

percentage of opinions included in the database (three percent) were decided by the Court of Appeals for the District of Columbia, and the largest percentage, by the Ninth Circuit (twenty-four percent). Four hundred of the 419 cases were unanimous, where either one judge wrote an opinion with the other two judges joining, or else the opinion was simply an anonymously authored *per curiam* decision. In the remaining cases, five were cases in which there was only a concurrence, 13 contained only a dissent, one had two dissents, and five contained both a concurrence and a dissent. Solely unanimous decisions were rendered only by the First, Fourth, Seventh, Eighth and Tenth Circuits. The circuit with the lowest percentage of non-unanimous decisions was *not* the Ninth Circuit, the quality of whose opinions has been criticized,⁶⁵ but rather the Sixth Circuit, with fifteen percent non-unanimous decisions.⁶⁶ Of the *total* number of cases that were not unanimous, however, the Sixth and Ninth Circuits each decided 32 percent of them.

Some of the features of the 444-case database are set forth below in Table 1.

infra tbl. 1. The smaller database, however, examines only the *cases*; therefore, reference to that database is sometimes more appropriate. *See infra* tbl. 2.

65. Compare Richard A. Posner, *Is the Ninth Circuit Too Large? A Statistical Study of Judicial Quality*, 29 J. LEGAL STUD. 711 (2000) (arguing that the Ninth Circuit has too many judges and that the quality of its decisions suffers as a result), and Marybeth Herald, *Reversed, Vacated, and Split: The Supreme Court, the Ninth Circuit, and the Congress*, 77 OR. L. REV. 405 (1998) (noting the Ninth Circuit's high reversal rate), with Jerome Farris, *The Ninth Circuit—Most Maligned Circuit in the Country—Fact or Fiction?*, 58 OHIO ST. L.J. 1465 (1997) (arguing, as a Ninth Circuit judge, that the circuit is functioning well).

66. *Cf.* Posner, *supra* note 65, at 717 (arguing that, aside from the Federal Circuit, only the Sixth Circuit produced worse opinions than the Ninth Circuit).

Table 1
 Characteristics of the Judges (means and deviations, set forth in brackets, or percentages)⁶⁷

Characteristic	
gender (0=female, 1=male)	.84 [.37]
race	
black	4%
latino	5%
white	91%
Astin eliteness of undergraduate institution	63 [9]
eliteness of law school (0=nonelite, 1=elite)	.58 [.50]
prior work	
government	8%
law school teacher	12%
judge	17%
private practice other	62%
other	1%
religion (N=192)	
Jewish	17%
Catholic	26%
Protestant	57%
other	1%
seniority (tenure)	13 [8]
President's economic liberalism	39 [21]

Translated into more accessible language, several observations may be drawn from Table 1. Most of the judges—84 percent—were men.⁶⁸ The judges were overwhelmingly white (91 percent of the total). They had gone to colleges such as Fordham, Tufts, and West

67. Except for the Astin measurements of school selectivity (n=235) and religion (n=192), the sample size for all of these characteristics is 244. The number of authoring judges for whom many characteristics are known is smaller than the *total* number of cases in the database because many of the cases were issued *per curiam*. See *supra* text following note 62. Means and deviations are not set forth where they would be meaningless, such as for religion and other non-ordered, discrete, multinomial variables.

68. This may be deduced from the mean for the judges' gender, which is .84. This value tells us that the gender of sixteen percent of all judges were female (coded value=0), and eighty-four percent were male (coded value=1).

Point,⁶⁹ and were more elitely educated at law school than not, because 58 percent had gone to an elite law school. A fairly substantial majority of the judges had come from private practice, 62 percent, and others had come from other judicial positions, law school teaching or government work, in declining order. Most were Protestant, 57 percent, but some were Catholic and others, even fewer, Jewish. The average tenure of judges when they decided these tax cases was thirteen years, and they had been appointed by Presidents somewhat economically conservative (economically, more conservative than President Nixon, but more liberal than Presidents Ford and Eisenhower).

Table 2
Characteristics of the Cases (means and deviations, set forth in brackets, or percentages)

Characteristic	
Type of taxpayer (N=438)	
trust or estate	6%
business	16%
individual	76%
other	2%
Appellant (N=444) (0=taxpayer, 1=government)	.90 [.30]
Prevailing party (N=441) (0=taxpayer, 1=government)	.85 [.36]
Taxpayer represented by counsel (N=400) (0=no, 1=yes)	.67 [.47]
Reported decision? (N=442) (0=no, 1=yes)	.48 [.5]

Most of the taxpayers—76 percent—were individuals and then, in decreasing numbers, businesses, trusts or estates, or some other type

69. These schools were among those assigned a score of 63 (the mean score) by Alexander Astin. See ASTIN, *supra* note 41. Other schools of a similar caliber include Iowa, Marquette, and Texas (with Astin scores of 62), the University of Pennsylvania and Washington & Lee (with Astin scores of 64). *Id.*

of entity. Taxpayers appealed more frequently than the government (and thus must have been more dissatisfied with the decisions at the trial level), but the government almost always won on appeal, 86 percent of the time. Taxpayers usually were represented by counsel, two-thirds of the time, and the decisions were often unpublished, 55 percent of the time.

Tables 3–7 explain various aspects surrounding the numbers and types of issues presented in the cases, as well as the methods judges used to justify their decisions on these issues. As noted above, up to three issues were examined in majority opinions and one issue each was analyzed in concurring and dissenting opinions.

Table 3 sets forth the number of issues that appeared in each case.

Table 3
Number of Issues/Case

Issue	N
majority, first issue	419
majority, second issue	202
majority, third issue	139
concurrence	9
dissent	13

In other words, all 419 cases had at least had at least one issue and one rationale, but only 202 majority opinions went on to spot a second issue, and even fewer, 139, had a third issue. Thus, most cases were not terribly complex, tending to be only one-issue cases.

Table 4 sets forth the number of methods of interpretation used to justify each of the above issues.

Table 4
Number of Methods of Interpretation/Issue

Issue	N
majority, first issue, -first method	419
-second method	202
-third method	139
concurrence, first issue, -first method	9
-second method	3
-third method	3
dissent, first issue, -first method	13
-second method	8
-third method	6

As was true above, all 419 cases presented at least one issue in the majority opinion and had at least one method of construction in that issue. While not all cases presented multiple methods of construction, the drop-off from one to two (or three) methods of interpretation was not as conspicuous as the decline in the number of issues in Table 3.

The uses of the methods tabulated in Table 4 have been dissected further in Tables 5–7, into majority, concurring and dissenting opinions, so that use of precedent, in combination, for example, with reliance on legislative history can be compared.

Table 5
 Methods Used in Majority Opinions, by Percentage of all Methods
 Used in that Level of Method, with Absolute Number in Parentheses¹

Level of method/ method used	Legislative History	Regulations	Structure	Strict Construction	Summary	Precedent
First method (N=419)	1% (5)	3% (12)	5% (19)	6% (25)	30% (124)	56% (234)
Second method (N=202)	5% (9)	12% (24)	10% (21)	6% (13)	13% (26)	54% (109)
Third method (N=139)	7% (9)	7% (10)	11% (15)	2% (3)	15% (21)	58% (81)

1. These tables should be read horizontally. For example, six percent of judges (or twenty-five total) used strict construction as their first method of rationale. Similarly, ten percent of judges using at least two method (n=202) relied upon structure as their second method.

Table 6
 Methods Used in Concurring Opinions, by Percentage of all Methods
 Used in that Level of Method, with Absolute Number in Parentheses

Level of method/ Method used	Legislative History	Regulations	Structure	Strict Construction	Summary	Precedent
First method (N=9)	-	-	11% (1)	11% (1)	22% (2)	56% (5)
Second method (N=3)	-	-	-	-	-	100% (3)
Third method (N=3)	-	-	-	33% (1)	-	67% (2)

Table 7
 Methods Used in Dissenting Opinions, by Percentage of all Methods
 Used in that Level of Method, with Absolute Number in Parentheses

Level of method/ Method used	Legislative History	Regulations	Structure	Strict Construction	Summary	Precedent
First method (N=13)	-	8% (1)	8% (1)	23% (3)	8% (1)	54% (7)
Second method (N=8)	13% (1)	13% (1)	25% (2)	-	13% (1)	38% (3)
Third method (N=6)	-	17% (1)	-	17% (1)	17% (1)	50% (3)

Judges' reliance on precedent is unmistakable. Whether writing for the majority, concurring, or dissenting, and whether in the primary or some later method of justifying the decision, judges turned to what other judges had done in other cases a majority of the time (with one exception—dissenting judges, second method). Summary dispositions also were frequently made. Beyond that, however, there seems to be no discernable pattern and no predominate method upon which judges relied, nor were many of these patterns statistically significant.⁷⁰ While these results may not be surprising, they nevertheless sustain the conclusion that interpreting the Internal Revenue Code is not the product of a rational approach to statutory construction.

Attempts to discern other patterns, whether through other cross-tabulations or means comparison, also were unproductive. Either the results were not statistically significant or even approached statistical significance, which meant that a pattern about how judges acted could not be said to have been established, or else the results evidenced weak or largely unexplainable patterns.⁷¹

70. Statistical significance negates the possibility that the results were random. It may be inferred that patterns are discernible. I sought descriptive statistics that were either statistically significant (p-values $\leq .01$ or $.05$) or else that which approached statistical significance (p-value $\leq .10$). See generally Brudney et al., *supra* note 1, at 1709 n.121 (designating $p = .05$ as threshold for significance and $p = .10$ as threshold for approaching significance).

Statistical significance was examined through cross-tabulations, which were made in two ways: by cross-tabulating (1) all of the methods in one type of opinion against each other (i.e., first majority method by second majority method by third majority method), and (2) each method in one type of opinion against all other opinions of the same type (i.e., first majority method by second majority method, first majority method by third majority method, and second majority method by third majority method). Using the first way, only the patterns involving precedent in the majority opinions were statistically significant ($p \leq .01$). More patterns were statistically significant when using the latter cross-tabulations, generating the following p values: first majority method x second majority method, $p \leq .05$; first majority method x third majority method, $p \leq .01$; second majority method x third majority method, $p \leq .01$; first concurring method x third concurring method, $p \leq .10$; first dissenting method x third dissenting method, $p \leq .10$.

71. Cross-tabulations were calculated between each of the discrete independent variables and all of the methods that could be used as either the first, second, or third means of justifying decisions. For example, gender was compared among majority decisions, first method of justifying a decision in which use of all six methods was raised; majority decisions, second method of justification, all six methods being raised, etc. In contrast, means for all of the continuous independent variables were compared in each of the methods of interpretation used first, second, and last in the majority opinions, and so on. In other words, means for seniority was compared in majority opinions, first method of interpretation, legislative history (using

Generally, the strongest patterns are the most common. For example, occurrence of more frequent statistically significant cross-tabulations involving judges' gender would suggest that gender is a way in which to describe the patterns by which judges used certain methods of interpretation to justify their decisions.⁷² Similarly, repeated statistically significant differences of the means regarding judges' seniority would suggest that seniority could be employed to describe the methods judges had used to justify their decisions. As noted, however, the statistical significance of the results were less than overwhelming.⁷³

Whether the independent variable was the taxpayer's representation by a lawyer, the judge's religion, or the judge's race, the results were meager. Results that were statistically significant, or approached significance, seemed to have been generated more by continuous than discrete variables. Of these statistically significant results, one way to parse them is to distinguish between the less and more "sophisticated" or theoretical methods of interpreting the Internal Revenue Code—summary dispositions and mere reliance upon precedent, as opposed to any of the other four, more technical, methods of interpretation.⁷⁴

Using that approach, a general pattern might be discerned in the eliteness of the judges' college educations, with judges who had gone to less elite colleges tending to justify their decisions with the less sophisticated approaches and judges who had gone to more elite schools relying on the more technical approaches. This observation holds true across the breadth of opinion, with no apparent differences between majority, concurring, or dissenting opinions.⁷⁵ Similarly, but somewhat less convincingly, judges with more seniority seem to rely

dummy variables, legislative history v. all other methods), majority opinions, first method of interpretation, reliance on regulations (again involving dummy variables, relying on regulations v. all other methods), etc.

72. For example, gender could be tested against each of the six options for the first, second, and third methods of justification in the majority opinions, thereby producing eighteen instances in which gender could be statistically significant. The same could be done for the concurring and the dissenting opinions. Means could be tested for seniority in each of the three methods of justification in the majority, concurring, and dissenting opinions.

73. These results are set forth in the Appendix, in Tables A1 through A8.

74. See Sisk, *supra* note 1, at 1463-65.

75. See *infra* Appendix, at Table A1.

more on the less sophisticated approaches, while the more junior judges relied upon the more technical approaches.

The most intriguing aspect of the statistically significant cross-tabulations regarding judges' prior work experiences and the legal status of the taxpayer⁷⁶ is the use of the more technical approaches to interpreting the Code. While reliance on precedent and summary dispositions might be expected,⁷⁷ the judges in these cross-tabulations also relied more on other, technical, grounds to justify their decisions than proportionate numbers might have suggested.

B. Logistic Regressions—What Predicted How a Judge Would Decide?

As with the descriptive statistics, there were few statistically significant regressions, or even regressions approaching statistical significance, from which to infer that characteristics of judges' social backgrounds or aspects of the cases generally predict outcome.⁷⁸ Indeed, only the regressions for methods used in majority opinions are reproduced in the Appendix because none of the regressions for concurring or dissenting opinions led to statistically significant

76. See *infra* Appendix, at Tables A3, A8.

77. See *supra* Tables 5–7.

78. The regressions are set forth, *infra* Appendix, at Tables A9–A11. Multicollinearity was not problematic here. As has been suggested, what constitutes collinearity is not clear. See Schneider, *supra* note 3, at 345 n.101; Sisk et al., *supra* note 1, at 1432 n.233. Both adopt a model permitting collinearity where the bivariate correlation does not exceed .500; independent variables that would exceed .500 are excluded. With only one exception, the correlations between independent variables in regressions never went beyond +/- .240. The correlation between eliteness of college and law schools attended by the judges was .540. This was not surprising, as both variables have been viewed as indicators of social-economic status. Education, especially undergraduate education, is a strong socioeconomic marker. In other words, students who graduate from more elite colleges generally come from more privileged backgrounds than those who attend less elite colleges. See, e.g., James C. Hearn, *Pathways to Attendance at the Elite Colleges*, in THE HIGH-STATUS TRACK: STUDIES OF ELITE SCHOOLS AND STRATIFICATION 121, 122, 130-36 (Paul William Kingston & Lionel S. Lewis eds., 1990). The same assertion can be made for elite law schools, although the evidence is weaker. See, e.g., LINDA F. WIGHTMAN, LEGAL EDUCATION AT THE CLOSE OF THE TWENTIETH CENTURY 20-44 (1995). The risk presented by multicollinearity between these two variables notwithstanding, each has been used because of its importance to my other research.

results, or results that even approached statistical significance.⁷⁹ Again I conclude that the absence of many statistically significant regressions means that the method of justifying their decisions was not very important to the judges in the sampled cases.

One independent variable seems to appear more often: the legal status or type of taxpayer. This variable—namely, whether the taxpayer was an individual, business, etc.—appeared more frequently than others among the regressions that were statistically significant or approached statistical significance. The exponentiated β for this legal status also usually exceeded 1, thereby suggesting that a particular status increased the odds ratio of a particular method of statutory construction being used.⁸⁰ Further inspection also suggests that stronger associations existed between type of taxpayer and more sophisticated methods of interpretation. In other words, whether the taxpayer was a business or a trust or estate was more strongly correlated with the use of a sophisticated method of interpretation than with a summary disposition.

Seniority also correlated negatively. In other words, judges were less likely to use various methods of interpretation—regulations as a first method in majority opinions; legislative history, regulations, or structure as a third method in majority opinions—to justify their

79. Multinomial logistic regression was used to establish the connection between the independent and dependent variables. Logistic regression is well suited for the mix of continuous and discrete independent variables used here, permitting the effect of each independent variable to be isolated while controlling for the influence of other independent variables, and also measuring the magnitude of the influence. See Brudney et al., *supra* note 1, at 1680.

In turn, multinomial logistic regression predicts and measures associations by assessing the likelihood of an event's occurrence, and by doing so in relation to a reference category. In the logistic regression tables, Exp. β can be said to state the odds ratio of an event's occurrence. See SPSS REGRESSION MODELS 9.0 40-42 (1999); cf. Bohrnstedt & Knoke, *supra* note 61, at 178-81 (discussing odds and odds ratios). In Table A9, *infra* Appendix, for example, the odds that the judge would use the structure of the Code as the first method for justifying his majority decision was 5.56 times greater if the taxpayer was a business than if it was a member of the referent category, a taxpayer who was an individual. Exp. β for that type of taxpayer was 5.56 when structure was used.

Logistic regression also permits the ordering of multiple independent variables. For example, the odds were even greater (9.39) that the judge would use Code structure when the taxpayer was a trust or estate.

80. See *infra* Appendix, at Tables A9-A11.

decisions.⁸¹ That the judge was a woman, Latino, or Jewish occasionally led to positive correlations,⁸² and eliteness of education had both positive and negative correlations.⁸³

V. CONCLUSIONS

To summarize the results, and my interpretation of these results, there were relatively few descriptive statistics that were statistically significant. There were also relatively few regressions about judges' methods of justifying their decisions, such that it was difficult to predict how they might justify their decisions in this database. From these two sets of results, I conclude that the method of justifying a decision was not very important to the judges who decided the cases in the database.

It also appears that the results in this Article are consistent with my earlier article regarding methods of construction used to interpret the Internal Revenue Code. Social background factors are not highly predictive of how judges justify their tax decisions in the databases I assembled about recent appellate decisions, nor are aspects of the litigation themselves, such as whether the taxpayer was represented by a lawyer.

81. *See infra* Appendix, at Tables A9, A11.

82. *See infra* Appendix, at Tables A10-A11.

83. *See infra* Appendix, at Tables A10-A11.

APPENDIX

Table A.1. Statistically Significant Descriptive Statistics Regarding the Continuous Variables

Independent Variable	Type of Opinion/Method	P value	Method	All Other Methods
eliteness of college ¹ education	majority, first method	$p \leq .01$	72	63
	legislative history	$p \leq .10$	66	63
	strict construction	$p \leq .10$	62	64
	precedent	$p \leq .10$	67	63
	majority, second method	$p \leq .05$	62	65
	regulations	$p \leq .10$	51	68
	precedent	$p \leq .10$	81	64
	dissent, first method	$p \leq .10$	81	63
	summary disposition	$p \leq .10$	51	69
	dissent, second method			
	legislative history			
dissent, third method				
regulations				
summary disposition				

1. In other words, the mean Astin score of colleges for judges who used legislative history as the first method of justification in majority opinions was 72; it was only 63 for the combined judges who used all other methods. The mean Astin score of colleges for judges

Independent Variable	Type of Opinion/Method	P value	Method	All Other Methods
seniority ²	majority, first method regulations majority, third method legislative history regulations precedent dissent, first method summary disposition	p ≤ .10 p ≤ .05 p ≤ .01 p ≤ .10 p ≤ .05	10 9 9 14 1	13 13 13 12 11

who used summary dispositions as the third method of justifying their dissenting opinions was 51, while it was 69 for the combined judges who used all other methods.

2. The mean tenure or seniority of judges who used regulations to justify their decisions as the first method in majority opinions was ten years, while it was thirteen years for judges who used all other methods; when precedent was used as the third method in majority decisions, the mean seniority for judges was fourteen years, while it was only twelve for those who used all other methods.

Independent Variable	Type of Opinion/Method	P value	Method	All Other Methods
appointing president's economic liberalism ³	majority, first method	p<.05	32	40
	strict construction	p≤.05	41	33
	majority, third method	p≤.05	62	35
	precedent	p≤.05	18	33
	concurrence, first method	p≤.10	63	26
	summary disposition	p≤.10	63	26
	dissent, first method	p≤.05	18	53
	strict construction			
	summary disposition			
	dissent, second method			
legislative history				
dissent, third method				
precedent				

3. In other words, the mean score for economic liberalism for judges who used strict construction as the first method of justification in majority opinions was 32, while it was 40 for those judges who used all other methods.

Table A2. Statistically Significant Descriptive Statistics Regarding Representation by a Lawyer, by Absolute Number ($p \leq .01$)

Represented by Lawyer/Method for Majority, ⁴ First Method ⁴	Legislative History	Regulations	Structure	Strict Construction	Summary Disposition	Precedent
no	-	-	2	-	67	63
yes	5	12	17	24	35	151

4. Where only one result was statistically significant, as was true in this table, the p-value has been set forth after the caption for the table. The table should be read horizontally. In other words, taxpayers were not represented by counsel in two cases when the first method used to justify the decision in majority opinions was the Code's structure, in 67 cases when it was a summary disposition and in 63 cases when it was precedent. Tables A2-A8 also should be read horizontally. Therefore, for example, there were 132 cases involving the first method used in a majority opinion in which taxpayers were not represented by a lawyer (2+67+63).

Table A3. Statistically Significant Descriptive Statistics
Regarding Taxpayer Type

Taxpayer Type/Method ⁵	Legislative History	Regulations	Structure	Strict Construction	Summary Disposition	Precedent
for majority, first method ($p \leq .01$) other	-	1	-	-	-	6
trust or estate	2	1	2	3	3	10
business	-	5	7	14	13	29
individual	3	5	10	7	107	187
for majority, second method ($p \leq .10$) other	-	1	1	-	1	3

5. Where there is more than one statistically significant result, all of the results have been collapsed into a single table. There were four such instances: all three methods used to justify majority decisions, as well as the first method used to justify the dissenting opinions.

For example, taxpayer classification was trusts or estates in two cases where the first method used in majority opinions was legislative history, and in three cases where taxpayers were individuals. Similarly, of the first methods used to justify majority decisions when the taxpayer was a trust or an estate, regulations were used once; legislative history and structure each were used twice; strict construction and summary dispositions each were used three times; and precedent was used ten times.

Taxpayer Type/Method ⁵	Legislative History	Regulations	Structure	Strict Construction	Summary Disposition	Precedent
trust or estate	1	1	3	1	1	9
business	4	10	8	7	2	21
individual	4	11	9	5	22	76
for majority, third method (p≤.05) other	1	-	-	-	-	3
trust or estate	1	2	-	-	2	4
business	5	5	9	2	6	15
individual	2	3	6	1	13	58
dissent, first method (p≤.05) trust or estate	-	-	1	1	1	-
business	-	1	-	2	-	-
individual	-	-	-	-	-	6

Table A4. Statistically Significant Descriptive Statistics Regarding Judges' Gender, by Absolute Number

Gender/Method	Legislative History	Regulations	Structure	Strict Construction	Summary Disposition	Precedent
concurrency, first method ($p \leq .05$) female	-	-	1	-	2	-
male	-	-	-	1	-	4
dissent, first method ($p \leq .10$) female	-	1	1	-	-	1
male	-	-	-	3	1	6

Table A5. Statistically Significant Descriptive Statistics Regarding Judges' Race, by Absolute Number

Race/Method -for Concurrence, Third Method ($p \leq .10$)	Legislative History	Regulations	Structure	Strict Construction	Summary Disposition	Precedent
Black	-	-	-	-	-	2
white	-	-	-	1	-	-
dissent, second method ($p \leq .10$)	1	-	-	-	-	-
black	-	1	2	-	1	3
white	-	-	-	-	-	-

Table A6. Statistically Significant Descriptive Statistics Regarding Eliteness of Law Schools, by Absolute Number ($p \leq .10$)

Race/Method -for Concurrence, Third Method	Legislative History	Regulations	Structure	Strict Construction	Summary Disposition	Precedent
not elite	-	-	-	1	-	-
elite	-	-	-	-	-	2

Table A7. Statistically Significant Descriptive Statistics Regarding Judges' Religions, by Absolute Number ($p \leq .10$)

Religion/Method -for Dissent, Third Method	Legislative History	Regulations	Structure	Strict Construction	Summary Disposition	Precedent
Jewish	-	-	-	1	-	-
Protestant	-	1	-	-	-	3

Table A8. Statistically Significant Descriptive Statistics Regarding Prior Work, by Absolute Number

Prior Work/Method	Legislative History	Regulations	Structure	Strict Construction	Summary Disposition	Precedent
concurrency, first method ($p \leq .10$)	-	-	-	-	-	-
other	-	-	-	-	1	-
government	-	-	-	-	-	-
judge	-	-	1	-	-	-
private practice	-	-	-	1	1	4
dissent, first method ($p \leq .10$)	-	-	-	-	-	-
government	-	-	-	1	-	1
law school prof	-	1	-	-	-	1
judge	-	-	1	-	1	-
private practice	-	-	-	2	-	5

Table A9. Statistically Significant Logistic Regressions—Majority, First Method⁶

Method/Independent Variable	Variable	β	S.E.	Significance	Exp. β
legislative history	taxpayer=trust or estate	3.30	1.91	.08	27.09

6. The second column from the right sets forth the statistical significance of the relationships between the independent and dependent variables. The columns for β and Exp. β state the same fact differently. β indicates the effect of an independent variable on the logarithmic probability of the dependent variable's occurrence. An Exp. β greater than one indicates that an event is more likely to occur than not, and vice versa.

While I have conservatively chosen to examine only those independent variables that are statistically significant or approach statistical significance, a less cautious reader might choose to examine independent variables that are not statistically significant. When Exp. β is greater than one, the independent variable can be said to have a positive effect, although the way in which the variable is cast is important. Thus, for example, the dichotomous variable "represented by lawyer" is coded: 0 = no representation, 1 = lawyer. Thus, the taxpayer's lack of legal representation is associated with the judge's summary disposition of the issue as the first method in the majority decision, because Exp. β is positive (13.91). On the other hand, if Exp. β is less than one, then the effect is negative. Finally, the odds ratios are relative, such that a relatively greater odds ratio indicates a stronger correlative relationship.

The reference category or value for each of the independent variables is as follows:

- For method of interpretation, precedent;
- For gender, male;
- For race, white;
- For eliteness of law school, elite law schools;
- For prior work experience, private practice;
- For religion, Protestant;
- For legal counsel, the taxpayer was represented; and
- For type of taxpayer, individual.

Method/Independent Variable	Variable	β	S.E.	Significance	Exp. β
regulations	seniority	-14	.09	.09	.87
structure	prior work = government taxpayer=	1.59	.87	.07	4.88
	trust or estate taxpayer=	2.24	1.20	.06	9.39
	business	1.72	.83	.04	5.56
strict construction	taxpayer= business	1.94	.70	.01	6.99
summary disposition	no lawyer	2.67	.82	.01	14.41

Table A10. Statistically Significant Logistic Regressions—Majority, Second Method

Method/Independent Variable	Variable	β	S.E.	Significance	Exp. β
regulations	elite college taxpayer= business	.10	.06	.08	1.11
	gender	1.7	.86	.04	5.71
structure	gender	2.45	1.36	.07	11.62
strict construction	taxpayer= business	1.82	1.03	.08	6.16
summary disposition	elite college	.12	.07	.08	1.13

Table A11. Statistically Significant Logistic Regressions—Majority, Third Method

Method/Independent Variable	Variable	β	S.E.	Significance	Exp. β
legislative history	elite college	-.25	.14	.08	.78
	seniority	-.27	.14	.06	.77
	economic liberalism	.18	.11	.10	1.20
	judge=woman	4.92	2.30	.03	137
	judge=Latino	22.05	13.25	.10	3.77E+09
	judge=Jewish	6.55	2.77	.02	701
regulations	elite college	-.20	.11	.08	.82
	seniority	-.24	.14	.08	.78
	judge=Jewish	3.71	2.04	.07	41
	taxpayer=trust or estate	4.90	2.40	.04	135
structure	seniority	.69	.40	.09	2.0
	economic liberalism	-.40	.22	.07	.67

Method/Independent Variable	Variable	β	S.E.	Significance	Exp. β
summary disposition	elite college	-.27	.11	.02	.77
	judge=woman	5.00	2.1	.02	148
	judge=Latino	6.7	4.06	.10	815
	nonelite law school				
	prior work =government	-2.80	1.57	.07	6.129E-02
	religion=Jewish	-4.80	2.3	.04	8.253E-03
		5.50	2.02	.01	244